Food Safety Uses of Genomics Tools: What they can and can’t do

Martin Wiedmann
Gellert Family Professor of Food Safety
Department of Food Science
Cornell University, Ithaca, NY
E-mail: mw16@cornell.edu
Phone: 607-254-2838
Take home messages

• WGS and other genomics tools will be and already are “game changers” in many areas of microbial food safety and quality, including
  – Detection of more and smaller foodborne disease outbreaks
  – More rapid detection of new pathogens
  – Improved characterization of (potential) foodborne pathogen isolates
  – *Metagenomics* approaches for QA, fraud detection, identification of spoilage root causes etc.

• WGS and other genomics tools are not magic bullets
  – There will be a continued (and maybe increased) need for (good) epidemiology to identify outbreak sources
    • Also tremendous need for understanding of potentially complex contemporary supply chains
  – These tools cannot rewrite biological paradigms
What they can do

• Detect human disease clusters better
  – More clusters (particularly for organisms where PFGE has limited discriminatory power)
  – Smaller clusters
  – Clusters detected more rapidly

• Identify and characterize microbial isolates more rapidly and more accurately

• Identify root causes of spoilage issues more rapidly

• Identify deviations from “normal” (if we can get the algorithms right...)
Den Bakker et al. 2011. AEM.

Includes isolates form Salmonella outbreak linked to sausages (Rhode Island) and isolates from pistachios.
Tip-dated maximum clade credibility tree based on SNP data for 47 Montevideo isolates
Food Safety News
Breaking news for everyone's consumption

CDC/FDA Partnership Targets Whole Genome Sequencing of Listeria Monocytogenes

By Brian Saunders | November 27, 2013

In a prior APHLTech blog post (NGS in Action: FDA’s Genome TRAKR Network), Victor Waddell of the Arizona State Public Health Laboratory described the newly formed network of laboratories formed by the U.S. Food and Drug Administration (FDA). Known collectively as Genome TRAKR, the member laboratories perform whole genome sequencing (WGS) on bacterial foodborne pathogens isolated primarily from food and environmental sources.

On Sept. 1, 2013, the Centers for Disease Control and Prevention (CDC) began a partnership with the FDA Genome TRAKR network to utilize the network to conduct WGS of all Listeria monocytogenes collected from reported human illness cases in the United States. This effort leverages public health resources to evaluate and
Listeria Outbreaks and Incidence, 1983-2015

- **Era**
  - Outbreaks per year:
    - Pre-PulseNet: 0.3
    - Early PulseNet: 2.3
    - Listeria Initiative: 2.6
    - WGS: 7.5
  - Median cases per outbreak:
    - Pre-PulseNet: 69
    - Early PulseNet: 11
    - Listeria Initiative: 5.5
    - WGS: 3.5

Incidence (per million pop)

- 1983: 9
- 1985: 8
- 1987: 7
- 1989: 6
- 1991: 5
- 1993: 4
- 1995: 3
- 1997: 2
- 1999: 1
- 2001: 0
- 2003: 1
- 2005: 2
- 2007: 3
- 2009: 4
- 2011: 5
- 2013: 6
- 2015: 7

Courtesy: Amanda Conrad, Outbreak Preparedness and Response Branch
Improved characterization of foodborne pathogens

• WGS data will replace a suite of previous tests as WGS can be used to predict:
  – Serotype
  – Antimicrobial resistance
  – Presence of gene that will allow organisms to cause disease

• Impact includes more rapid identification and detection, including of “new pathogens”
  – Salmonella that produce DNA damaging toxins
Metagenomics is the characterization of all genetic material in a sample (for example meat sample). Allows for comprehensive characterization that can be used for QA and troubleshooting:
  - Already regularly used to identify causes of quality issues
Increasingly used and applied by US FDA, regulatory agencies worldwide, and industry.
50 shades of gray: *Pseudomonas* causes gray discoloration in HTST milk
Key gene unique to Pseudomonas causing gray (and blue) color defects

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>trpD</td>
<td>Anthranilate phosphoribosyltransferase</td>
</tr>
<tr>
<td>trpF</td>
<td>N-(5'-phosphoribosyl)anthranilate isomerase</td>
</tr>
<tr>
<td>trpA</td>
<td>Tryptophan synthase alpha chain</td>
</tr>
<tr>
<td>trpB</td>
<td>Tryptophan synthase beta chain</td>
</tr>
<tr>
<td>iolG</td>
<td>Inositol 2-dehydrogenase/D-chiro-inositol 3-dehydrogenase</td>
</tr>
<tr>
<td>mdh</td>
<td>Malate dehydrogenase</td>
</tr>
<tr>
<td></td>
<td>Peptidase M</td>
</tr>
<tr>
<td>degT</td>
<td>Xylose Isomerase Domain-Containing Protein</td>
</tr>
<tr>
<td></td>
<td>Pleiotropic regulatory protein</td>
</tr>
<tr>
<td></td>
<td>Oxidoreductase Domain Protein</td>
</tr>
<tr>
<td>trpC</td>
<td>Indole-3-glycerol phosphate synthase</td>
</tr>
<tr>
<td>solR</td>
<td>Transcriptional activator protein solR</td>
</tr>
<tr>
<td>rhtB</td>
<td>Homoserine/homoserine lactone efflux protein</td>
</tr>
</tbody>
</table>

This reaction requires 3 oxygens!!
Consequences of unreliable differentiation between pathogens and non-pathogens can be costly

- Suspected botulism-causing bacteria identified in whey products
- ~1,000 tones of products recalled in 7 countries
- No disease cases
- Detailed strain characterization confirmed species misclassification
Clostridium tepidum sp. nov., a close relative of Clostridium sporogenes and Clostridium botulinum Group I

Anatoly P. Dobritsa, Kirthi K. Kutumbaka, Kirsten Werner, Martin Wiedmann, Aaron Asmus and Mansour Samadpour

Abstract

Obligately anaerobic, Gram-stain-positive, spore-forming bacteria indistinguishable by pulsed-field gel electrophoresis were isolated from non-dairy protein shakes in bloated bottles. One of the isolates, strain IEH 97212\textsuperscript{T}, was selected for further study. The strain was closely related to Clostridium sporogenes and Clostridium botulinum Group I based on 16S rRNA gene sequence similarities. Phylogenetic analysis also showed that strain IEH 97212\textsuperscript{T} and strain PE (=DSM 18688),
16S rRNA-based differentiation among *B. cereus* group species

Miller et al., 2016
What they can’t do
A hypothetical cases study

- July 2010: lettuce from packing house X found positive for *L. monocytogenes* when a sample collected at a supermarket was tested by government lab
  - Isolates was subsequently characterized by WGS
- June 2017: Human isolates “matches” July 2010 isolate
  - Both isolates are DNI (Darn Near Identical); i.e., 3 SNP difference
- What if the person reports eating lettuce labeled ”company X”? What if the company is “huge”? What if the person did not report eating lettuce? etc.
What WGS REALLY can’t do

• Replace good epidemiology
  – Detect and trace-back outbreaks of “1”
    • Maybe there are rare exceptions....
• Replace an in-depth understanding of food supply chains
• Rewrite the textbook on microbial evolution:
  – WGS will not give us a magic SNP cut-off that defines “same” or “not same” (“in” versus “out”)
• Fix microbial taxonomy
  – Unless you make that the only focus of your life
• *Cure dyslexia or attention deficit disorder*
The theoretical background

- Bacteria divide asexually: Bacterial populations can be seen as large populations of “identical twins”
- Mutation rate during replication is low: extremes of the suggested mutation rates range from $2.25 \times 10^{-11}$ to $4.50 \times 10^{-10}$ per bp per generation
  - With a genome size of around 5 Million bp per bacterial genome ($5 \times 10^6$) between approx. 450 and 9,000 generations are needed for a single SNP difference
  - Eyre et al. estimated evolutionary rate of 0.74 SNVs per successfully sequenced genome per year for C. difficile (N. Engl. J. Med. 2013)
    - “Whole-genome sequencing ... identified 13% of cases that were genetically related ($\leq 2$ SNVs) but without any evidence of plausible previous contact through a hospital, residential area, or family doctor.”
  - Unknown bacterial generation time in different environments complicates interpretation
    - Salmonella growth will differ considerably in poultry house versus a dry processing plant
- Ignoring all the complexities, a current rule of thumb is 0.5 SNP differences per year (but uncertainty is huge)
Practical observations – 1

BMC Genomics

Research article

Short-term genome evolution of *Listeria monocytogenes* in a non-controlled environment

Renato H Orsi¹, Mark L Borowsky²,⁷, Peter Lauer³, Sarah K Young², Chad Nusbaum², James E Galagan²,⁴, Bruce W Birren², Reid A Ivy¹, Qi Sun⁵, Lewis M Graves⁶, Bala Swaminathan⁶ and Martin Wiedmann*¹

bacterial genome evolution in natural environments is limited. We thus performed full genome analyses on four *Listeria monocytogenes*, including human and food isolates from both a 1988 case of sporadic listeriosis and a 2000 listeriosis outbreak, which had been linked to contaminated food from a single processing facility. All four isolates had been shown to have identical

**Results:** The two *L. monocytogenes* isolates from 1988 and the two isolates from 2000 had highly similar genome backbone sequences with very few single nucleotide (nt) polymorphisms (1 – 8 SNPs/isolate; confirmed by re-sequencing). While no
Figure 3
Schematic of the putative evolutionary history of the *L. monocytogenes* strain in the food facility between 1988 and 2000. Numbers on the arrows represent new mutations. Ancestor A is the ancestor of F6854 and F6900 (the food and human isolate, respectively, from the sporadic case in 1988) and Ancestor B is the ancestor of J0161 and J2818 (the food and human isolate, respectively, from the outbreak in 2000).
Practical observations –2

Whole genome sequencing allows for improved identification of persistent Listeria monocytogenes in food associated environments.

Stasiewicz MJ¹, Oliver HF², Wiedmann M³, den Bakker HC⁴.

In one case, isolates with < 3 SNP differences were found in retail delis in their different states.
In January 2017, CDC identified a cluster of _Salmonella enterica_ serotype Newport infections with isolates sharing an indistinguishable pulsed-field gel electrophoresis (PFGE) pattern, JJPX01.0010 (pattern 10).

**Epidemiologic Investigation:** 106 cases were identified in 21 states. Most illnesses (72%) were reported from southwestern states, including AZ (30), CA (25), NM (14), and TX (7). Illness onset from 4 Oct 2016, through 19 Jul 2017. Among 65 interviewed patients, 52 (80%) reported eating ground beef at home in the week before illness began. This percentage was significantly higher than the 2006-2007 FoodNet Population Survey. Among the 52 patients who ate ground beef at home, 31 (60%) reported that they bought it or maybe bought it from multiple locations of 2 national grocery chains, and 21 (40%) reported that they bought ground beef from locations of 15 other grocery chains.

**Traceback Investigation:** USDA-FSIS conducted traceback on ground beef for 11 patients who provided shopper card records or receipts. Approx. 20 ground beef suppliers belonging to at least 10 corporations were identified; 10 of the 11 records traced back to 5 company A slaughter/processing establishments, 7 of 11 traced back to five company B slaughter/processing establishments, and 4 of 11 traced back to 2 company C slaughter/processing establishments.

**Product and Animal Testing:** Opened samples of ground beef from 3 patients’ homes were collected. All were purchased from 1 of 2 national grocery chains that had been identified by a majority of patients. One sample, collected from ground beef removed from its original packaging, yielded the outbreak strain. The outbreak strain was also isolated from four NM dairy cattle. One was collected from a spontaneously aborted fetus in July 2016, and one was isolated from feces from a young calf in November 2016. The 3rd isolate was identified by searching the USDA-APHIS NVSL database for Salmonella Newport isolates collected from cattle in AZ, CA, TX, NM, and WI; the only Newport pattern 10 isolate identified was from a fecal sample from a NM dairy cow collected during November 2016. The 4th isolate was from a USDA-FSIS routine cattle fecal sample collected at a TX slaughter establishment in 2016; FSIS determined the sample was from a dairy cow and identified the NM farm of origin. Officials were not able to identify the farm or farms of origin for the dairy cows associated with the other 3 samples or whether the 4 dairy cows were associated with a single farm.
Laboratory Investigation: SNP analysis showed that 106 clinical isolates were closely related to each other genetically, to the 4 dairy cattle isolates, and to the leftover ground beef isolate (range = 0-12 SNP differences), suggesting that the Salmonella bacteria found in patients, ground beef, and dairy cattle all shared a common source.

Public Health Response: Because the USDA-FSIS traceback investigation did not converge on a common production lot of ground beef or a single slaughter/processing establishment, and no ground beef in the original packaging yielded the outbreak strain, a recall of specific product was not requested. A public warning was not issued to consumers because specific, actionable information was not available (e.g., a specific brand or type of ground beef).
What genomics tools have hard time doing – a few examples
Report: Human DNA found in hot dogs

After analyzing hot dogs from 75 different brands, Clear Foods discovered human DNA in 2% of the samples studied. Rachel Holt (@itsRachelHolt) dishes what else the study found. Buzz60
DNA blunder creates phantom serial killer

Police admit they wasted 15 years hunting for the 'Woman Without a Face'

She was a mysterious serial killer known as the "The Woman Without a Face" and detectives across Europe spent more than 15 years doing their utmost to bring her to justice for at least six brutal murders and a string of break-ins. Yesterday, however, they were forced to admit that she probably didn't exist.

The only clues that "The Woman Without a Face" left behind at 40 different crime scenes were DNA traces. These were collected on cotton swabs, supplied to the police in a number of European countries. Now police investigators have established that in all probability the DNA had not been left by their quarry but by a woman working for the German medical company supplying the swabs, who had inadvertently contaminated them.

German police who had been leading the hunt said they had probably been involved in one of the longest and most perplexing wild goose chases in criminal history. "This is a very embarrassing story," admitted police spokesman Josef Schneider.
What they have hard time doing

• Metagenomics
  – Overcome the limitations of 16S rDNA-based characterization methods (16S metagenomics)
  – Detect organisms at a sensitivity of 1 cfu/25g (without enrichment)
    • Detection of rare microbial DNA particularly challenging in food samples
      – Thins 10,000 bacteria per ml of milk versus 100,000 cow cells per ml (at 1,000 the DNA per cell relative to bacterial cells)
• Correctly assign and annotate 100% (?) of sequences, while also eliminating any (?) contaminants
• Transcriptomics/RNA-based metagenomics
  – Differentiate live & dead: not all RNA is highly unstable (e.g., rRNA is pretty stable)
Take home messages

• WGS and other genomics tools will be and already are “game changers” in many areas of microbial food safety and quality, including
  – Detection of more and smaller foodborne disease outbreaks
  – More rapid detection of new pathogens
  – Improved characterization of (potential) foodborne pathogen isolates
  – *Metagenomics* approaches for QA, fraud detection, identification of spoilage root causes etc.

• WGS and other genomics tools are not magic bullets
  – There will be a continued need for (good) epidemiology to identify outbreak sources
    • Also tremendous need for understanding of potentially complex contemporary supply chains
  – These tools cannot rewrite biological paradigms
Concluding thoughts on (industry) needs

• Have access to expertise in WGS and metagenomics
  – Need to have experts that understand the biology
• Assess your vulnerabilities
• Be prepared to see more outbreaks traced back to source
  – raw meat
  – Salmonella Enteritidis
• Make sure you have the data to keep recalls small
• We need some sort of safe harbor that will encourage industry use of WGS