One Team, One Purpose

Food Safety and Inspection Service
Protecting Public Health and Preventing Foodborne Illness
FSIS: Analytical Tools for WGS

October 26, 2017

Glenn E. Tillman, PhD
Chief, Microbiology Characterization Branch
Eastern Laboratory (Athens, GA)
Office of Public Health Science
Food Safety Inspection Service, USDA
Food Safety and Inspection Service:

FSIS: Update on WGS

- Background
- What are we doing (or planning to do) with our data?
- Concluding remarks
Food Safety and Inspection Service:

Why WGS?

- **Improved resolution for foodborne illness investigations**
  - Improved strain discrimination, illness cluster detection, and case classification

- **Supports FSIS mission goals**
  - Effectively use science to understand foodborne illness and emerging microbiological trends
  - Identification of environmental harborage or recurrences of pathogens in FSIS-regulated establishments/products to further support the inspection and verification process

- **Alignment of pathogen surveillance with our domestic public health and regulatory partners**
  - Collaborative efforts with US Food and Drug Administration Center for Food Safety and Applied Nutrition (FDA-CFSAN), the US Centers for Disease Control and Prevention (CDC), the US National Institutes of Health National Center for Biotechnology Information (NCBI), and also state/local health partners/laboratories
Food Safety and Inspection Service:

**WGS at FSIS: Current Status**

- FSIS has built capacity for conducting WGS on all isolates obtained from FSIS sampling programs
  - Currently 12 sequencers in FSIS Field Service Laboratories, and expect all to be operational in early FY18
  - In FY17, FSIS sequenced 7,282 isolates

- In collaboration with our public health and regulatory partners, FSIS currently considers available WGS analyses in addition to PFGE, epidemiological and traceback information to further understand the relationship between clinical and food isolates

- FSIS works with National Antimicrobial Resistance Monitoring System (NARMS) partners (FDA, CDC) to understand the occurrence or introduction of antimicrobial resistance genes in pathogens of interest
Food Safety and Inspection Service:
FSIS Sequencing Efforts and Data Sharing

FSIS Submissions to NCBI Bioprojects

- **PRJNA242847**
  - GenomeTrakr Project: USDA-FSIS (*Salmonella*)
- **PRJNA215355**
  - GenomeTrakr Project: FDA (*Listeria monocytogenes*)
- **PRJNA287430**
  - USDA-FSIS: *Campylobacter*
- **PRJNA268206**
  - GenomeTrakr Project: USDA-FSIS (STEC)
- **PRJNA292666**
  - FSIS NARMS *Salmonella*
- **PRJNA292667**
  - FSIS NARMS *E. coli*
- **PRJNA292668**
  - FSIS NARMS *Campylobacter*
- **PRJNA292669**
  - FSIS NARMS *Enterococcus*
Food Safety and Inspection Service:
WGS Data Sharing – Metadata

- Description about an isolate that has an experiment assigned
- Metadata, such as source type, organism, serotype etc.
- BioSample number is specific to a bacterial isolate for FSIS
Food Safety and Inspection Service: FSIS: WGS Analytical Tools Overview

- Quality Control Assessment
- Antibiotic Resistance gene detection
- Salmonella and STEC serotype determination
- STEC Virulence Gene characterization
- Phylogenetic Comparisons
Further Characterization of Isolates Using WGS

- Campylobacter speciation
- Serotyping/serogrouping
- Antimicrobial Resistance (Phenotype prediction)
  - Salmonella
  - Campylobacter
  - E. coli
  - Enterococcus
- Adulterant STEC
  - Resistance to environmental factors (heat, acid, etc)
  - Virulence factors
- stx/eae sub-types (STEC)
- Identify characterized genes of interest
- Alternative to PFGE for comparison of genotypes
  - wgMLST analyses
  - SNP analyses

A single workflow for many characterization approaches via informatics
Food Safety and Inspection Service: FSIS: WGS Data Analyses Work Flow Overview

De novo Assembly

Input: FASTQ
- QC Pipeline
  - Coverage
  - Average Quality
  - Nucleotide balance

Output: FASTA
- wgMLST BioNumerics 7.6
- Lyve-SET, SNP Pipeline
- NCBI Pathogen Isolate Browser

Input: FASTA
- MLST Sequence Type
- Antibiotic Resistance genes
- Virulence Profile
- Salmonella and STEC serotype
- MASH Tree comparison

De novo Assembly
- QC Pipeline
  - File Size
  - N50 & No. contigs
  - Correct organism
Food Safety and Inspection Service: Genotypic Screening for Antimicrobial Resistance Using WGS Data

- Ability to rapidly identify new genes of concern in sequenced isolates
- FSIS works with NARMS and other public health and regulatory partners to identify presence of genes or genetic markers of interest
  - When possible, FSIS currently compares genotype prediction to phenotype using the NARMS panel
- Examples:
  - Extended spectrum beta-lactamases ($bla_{CTX-M-65}$)
  - Colistin
  - Quinolone
  - Linezolid
  - Daptomycin

\[\text{2015: } \text{Salmonella and Campylobacter from cecal with R > 1 antimicrobial} \]
\[\text{2016: } \text{All Salmonella and Campylobacter from cecal} \]
\[\text{2017: } \text{All Salmonella and Campylobacter from both cecal and HACCP AMR E. coli and Enterococcus} \]
Food Safety and Inspection Service: WGS usage: Antibiotic Resistance Gene Detection

FASTQ files → QC & Assembly Pipeline → Resfinder (BLAST) → Resistance gene profile
Food Safety and Inspection Service:
WGS usage: Antibiotic Resistance Gene Detection

https://bitbucket.org/genomicepidemiology/resfinder
Antimicrobial Resistance (Phenotype Prediction) Using WGS Analyses

- Analyzed ~1190 *Salmonella* isolates from FY2016 NARMS cecal sampling
  - Includes isolates from all animals tested
- Compared genotypic prediction for resistance phenotype to reported phenotype using NARMS panel
  - Overall concordance for 1130 isolates (95.0%)

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Number of Isolates with Phenotype Using NARMS Panel</th>
<th>Percent Concordance of Genotype with Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pan-susceptible</td>
<td>769</td>
<td>99.1%</td>
</tr>
<tr>
<td>Beta-lactam resistance</td>
<td>110</td>
<td>99.1%</td>
</tr>
<tr>
<td>Aminoglycoside resistance</td>
<td>167</td>
<td>98.8%</td>
</tr>
<tr>
<td>Phenicol resistance</td>
<td>42</td>
<td>97.6%</td>
</tr>
<tr>
<td>Tetracycline resistance</td>
<td>330</td>
<td>99.7%</td>
</tr>
</tbody>
</table>
Food Safety and Inspection Service: WGS usage: Salmonella Serotype Determination

FASTQ files → Assembly Pipeline → SeqSero (BLASTdb) → Serotype Determination
Food Safety and Inspection Service: Using WGS Analyses for Serotype Determination

- FSIS sequenced 4205 *Salmonella* isolates from various sampling programs from Jan 2015 through Jun 2017
- Compared serotype reported by routine methods (molecular serotyping or traditional serology) with serotype determined using a custom query to SeqSero database
  - For 4045/4205 (96.2%) of isolates, WGS that matched reported serology result
  - For 160/4205 (3.8%) of isolates, WGS did not match reported serology result
    - Includes isolates with incomplete genetic factor set (cannot call/identify serotype)
Food Safety and Inspection Service: WGS usage: MLST and virulence typing in Shiga toxin-producing *Escherichia coli*

FASTQ files → Assembly Pipeline → Virulence genes and serotype → VirulenceFinder SerotypeFinder (BLAST)
Food Safety and Inspection Service:
WGS usage: MLST and virulence typing in Shiga toxin-producing Escherichia coli

<table>
<thead>
<tr>
<th>FSIS_Number</th>
<th>Serotype</th>
<th>Sequence_type</th>
<th>stx_type</th>
<th>eae_allele</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSIS1608854</td>
<td>O26:H11</td>
<td>ST-21</td>
<td>stx1a</td>
<td>Beta1</td>
</tr>
</tbody>
</table>
Food Safety and Inspection Service:

Using WGS Analyses for MLST and virulence typing in Shiga toxin-producing Escherichia coli

<table>
<thead>
<tr>
<th>Serogroup (no. sequenced)</th>
<th>Top 7 gene MLST sequence types</th>
<th>Top stx types</th>
<th>eae types</th>
<th>Top Serotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>O26 (50)</td>
<td>ST-21 (82%)</td>
<td>stx1a (80%)</td>
<td>Beta1 (100%)</td>
<td>O26:H11 (100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>stx2a (16%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O45 (20)</td>
<td>ST-17 (95%)</td>
<td>stx1a (100%)</td>
<td>Epsilon (95%)</td>
<td>O111:H8 (100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O103 (147)</td>
<td>ST-17 (67.8%)</td>
<td>stx1a (98.0%)</td>
<td>Epsilon (84.9%)</td>
<td>O103:H2 (84.9%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O111 (44)</td>
<td>ST-16 (93.0%)</td>
<td>stx1a (86.0%)</td>
<td>Theta (100%)</td>
<td>O111:H8 (100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O121 (15)</td>
<td>ST-655 (93.3%)</td>
<td>stx2a (100%)</td>
<td>Epsilon (100%)</td>
<td>O121:H19 (100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O145 (18)</td>
<td>ST-32 (100%)</td>
<td>stx1a stx2d (38.9%)</td>
<td>Gamma-1 (100%)</td>
<td>O145:H28 (100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>stx2a (27.7%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>stx2c (11.1%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O157 (149)</td>
<td>ST-11 (96.00%)</td>
<td>stx1a stx2a (28.2%)</td>
<td>Gamma-1 (100%)</td>
<td>O157:H7 (100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>stx2c (22.8%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>stx2a (20.1%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- *E. coli* serotypes, O157:H11 & O157:H29 identified to be lacking eae/stx through WGS
- *stx*-negative strains can be identified as serotype *E. coli* O157:H7 based on WGS
FSIS uses pipelines developed by public health partners
Lyve-SET
https://github.com/lskatz/lyve-SET/blob/master/

NCBI Pathogen Detection Isolates Browser

FDA SNP Pipeline
https://github.com/CFSAN-Biostatistics/snp-pipeline

wgMLST
BioNumerics 7.6 CDC-PulseNet
WGS usage: SNP tree from NCBI Pathogen Pipeline


<table>
<thead>
<tr>
<th>Product</th>
<th>Serotype</th>
<th>Total #</th>
<th>% in SNP cluster (#)</th>
<th>% w/in 20 SNPs of clinical</th>
<th>% w/in 10 SNPs of clinical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Serotypes</td>
<td>Enteritidis</td>
<td>447</td>
<td>99.8% (446)</td>
<td>98.9% (442)</td>
<td>90.6% (405)</td>
</tr>
<tr>
<td>Kentucky</td>
<td>607</td>
<td>99.8% (606)</td>
<td>0% (0)</td>
<td>86.4% (64)</td>
<td>7.3% (9)</td>
</tr>
<tr>
<td>Heidelberg</td>
<td>74</td>
<td>98.6% (73)</td>
<td>98.6% (73)</td>
<td>86.4% (64)</td>
<td>7.3% (9)</td>
</tr>
<tr>
<td>Schwarzengrund</td>
<td>122</td>
<td>100% (122)</td>
<td>47.5% (58)</td>
<td>50.9% (111)</td>
<td>10.3% (16)</td>
</tr>
<tr>
<td>Infantis</td>
<td>218</td>
<td>94.5% (206)</td>
<td>92.2% (201)</td>
<td>50.9% (111)</td>
<td>10.3% (16)</td>
</tr>
<tr>
<td>Typhimurium</td>
<td>155</td>
<td>96.8% (150)</td>
<td>30.3% (47)</td>
<td>10.3% (16)</td>
<td></td>
</tr>
</tbody>
</table>
Large range of SNPs between PFGE pattern JEGX01.0004 isolates, isolates with different PFGE patterns clustering with pattern JEGX01.0004 isolates.
Food Safety and Inspection Service

WGS usage: hqSNP analyses and wgMLST with *Listeria monocytogenes*
<table>
<thead>
<tr>
<th>Lab Number</th>
<th>Primary_PN_Pattern</th>
<th>Sec_PN_Pattern</th>
<th>Lyve-SET SNP Clade</th>
<th>CFSAN Snp Pipeline Clade</th>
<th>wgMLST Clade</th>
</tr>
</thead>
<tbody>
<tr>
<td>FSIS1400040</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1400020</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1400249</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1400248</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1400246</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500681</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500682</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500683</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500684</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500685</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500686</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500687</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500688</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500689</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500690</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500691</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1500692</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FSIS1502844</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>FSIS1502845</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>FSIS1502847</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>FSIS1502848</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>FSIS1502849</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>FSIS1502850</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>FSIS1502851</td>
<td>GX6A16.0018</td>
<td>GX6A12.0022</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Food Safety and Inspection Service

WGS usage: hqSNP analyses and wgMLST with *Listeria monocytogenes*
Food Safety and Inspection Service:
WGS usage: Usage of genotypic data to predict phenotypic characteristics

- FSIS is beginning to explore using genotypic data to predict phenotypic characteristics
  - Previous examples of antimicrobial resistance markers
  - Resistance to other environmental factors (i.e. heat, acid, certain chemicals, etc)
- Example: Locus of Heat Resistance (LHR) - 14-kb genomic island
  - Present in a diverse group of Enterobacteriaceae, including *Cronobacter sakazakii*, *Klebsiella pneumoniae*, *Enterobacter cloacae*, *E. coli*, and *Salmonella*
  - A BLAST database including genes contained in the LHR was built based on the published sequence of *E. coli* AW1.7
  - *Salmonella* isolates from FSIS-regulated products were sequenced, assembled and queried against the LHR BLAST database
    - 11 FSIS *Salmonella* isolates contained LHR genes of interest
    - Additional phenotypic testing required to determine if isolates exhibit resistance to heat
- Heat resistance can be linked to different biological pathways/genes
  - Advantage of WGS: As new pathways/genes are identified, data can be queried again
FSIS’ new strategic plan is focused on the use of new technology to prevent foodborne illnesses and protect public health
  – Includes two objectives specific to incorporating WGS into FSIS surveillance and regulatory efforts
FSIS has built sufficient capacity for conducting WGS on all FSIS pathogen isolates
  – Sequenced 7282 isolates in FY17 with the future target of ~10,000 isolates/year
FSIS is exploring how we can use WGS data beyond outbreak investigations, including understanding the link between genotypes and phenotypes of interest
FSIS continues to engage with national and international partners
FSIS continues to use WGS analyses in conjunction with other metadata, including epidemiological and traceback information, to further understand the relationship between clinical, food and environmental isolates
Food Safety and Inspection Service:

**Acknowledgements**

- USDA FSIS Offices
- USDA ARS
- CDC PulseNet and NARMS
- FDA CFSAN
- FDA CVM
- NCBI
- State Laboratories