Introduction to Whole Genome Sequencing and Its Application to Enteric Outbreak Investigations

Jamie Yeadon
Supervisor, Enteric Microbiology, Parasitology, and Molecular Subtyping Laboratory

Nicole Stone, MPH
Enteric Epidemiologist
HISTORY OF SEROTYPE/MOLECULAR SUBTYPE
State Lab Process

- Enteric culture on selective media
- MALDI-TOF for Salmonella, Vibrio, and Listeria
- Biochemicals and Antisera for E. coli and Shigella
- PFGE for Salmonella
- WGS for Salmonella, STEC, Listeria, and Vibrio
History of Serotype/PFGE Pattern

• Commonality of serotype (if rare) or PFGE pattern
• Trend of PFGE pattern over time
• Geographical history of PFGE pattern (2x Baseline)
• Has this pattern been linked to outbreaks in the past? Are there non-human isolates with the same pattern?
Past Outbreak Clues

• **History of the PFGE pattern** or serotype
  – Some PFGE patterns are “repeat offenders”
  – Multiple outbreaks linked to the same source or region
  – Several places to look at past outbreaks

• **CDC or local/state outbreak databases**
  – Information on past outbreaks and suspected/confirmed vehicles
  – Documentation of the serotype/PFGE patterns

• **National Outbreak Reporting System (NORS)**
  – CDC surveillance system with data submitted by states and CDC
  – Good for localized clusters, but also captures multistate outbreaks

• **Literature review** can be helpful, especially for older outbreaks
Non-Human Isolate Clues

- Non-human isolates similar to the outbreak strain can provide clues about the outbreak source or reservoir
  - Indistinguishable by PFGE
  - Highly related by WGS
  - Same serotype (if rare)
- Non-human isolates are collected by:
  - State/local departments of health and agriculture
  - NARMS retail meat and food animal testing programs
  - USDA (FSIS, APHIS, NVSL) and FDA
- Use this information in early hypothesis generation, but keep an open mind!
BASICS OF WGS
Whole Genome Sequencing (WGS)

- The genome: made up of DNA.
- Each organism has a unique DNA sequence which is composed of bases (A, T, C, and G).
- Sequence of the bases in an organism → DNA fingerprint/pattern.
- **Sequencing:** Determining the order of bases
- **Whole genome sequencing:** A laboratory procedure that determines the order of bases in the genome of an organism in one process.
How does whole genome sequencing work?

- **DNA replication**: Copies of the DNA are made and using PCR reactions.
- **DNA bar-coding**: Adding small pieces of DNA tags (bar codes) to label DNA piece and associate with respective bacteria.
- **DNA Fragmented**: Molecular scissors are used to chop the DNA into small enough sections to be quickly examined (about 300bp)
- **Whole genome sequencing**: All specimens are combined and put in the whole genome sequencer. The sequencer identifies bases. Can be separated later using bar code system.
- **Data analysis**: Computer analysis tools are used to compare bacterial sequences nationally and locally and identify differences.
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 Library 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 Sequence 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.

Reconstructed Genome
PFGE and PulseNet

- **Pulsed-field gel electrophoresis (PFGE):** laboratory technique used by scientists to produce a DNA fingerprint for a bacterial isolate.
- **PulseNet** is a national laboratory network that investigates bacterial isolates from sick people, contaminated food, and the places where food is produced.

Image Source: ISDH
How will whole genome sequencing transform disease detection?

- Detailed and precise data
- One test will let you compare specimens but also will identify the serotype, antibiotic susceptibility genes, and mobile elements of the DNA.
How is ISDH notified of an outbreak?

1

**Subject:** Multiple Illness complaint

Attached is a complaint we received about 16 people who became ill after eating at [redacted] in [redacted] County has the complaint and is trying to get ahold of the complainant for more information.

Dear Colleagues,

PulseNet recently coded *Salmonella* Carrau cluster 1904OHJRG-1 (Xbal pattern JRGX01.0006) with 10 cases from IN(1), OH(5), KY(2), MI(1), and MN(1). Isolation dates ranging from 3/17/19 to 3/26/19, 50% female and ages range from 1-98 (median 7 years). *Salmonella* Carrau is an uncommon serotype and this PFGE pattern is very rare. Notably, there is one FDA cantaloupe isolate from 2005 with pattern JRGX01.0006. The last multistate *Salmonella* Carrau outbreak (PFGE patterns: JRGX01.0001, JRGX01.0022, JRGX01.0023) occurred in 2009 and the investigation team suspected melons (cantaloupe, honeydew and watermelon) as the likely vehicle. We have requested these clinical isolates for WGS.

Hello,

We have two E. coli O111 that appear to match by WGS. The NORS table and PFGE cluster have been updated.

<table>
<thead>
<tr>
<th>Key</th>
<th>SourceState</th>
<th>Serotype</th>
<th>Serotype_wgs</th>
<th>ReceivedDate</th>
<th>Outbreak</th>
<th>SourceCounty</th>
<th>SourceCity</th>
<th>SourceType</th>
<th>SourceSite</th>
<th>PatientAge</th>
<th>PatientSex</th>
<th>Toxin_wgs</th>
<th>WGS_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>C15031024</td>
<td>IN</td>
<td>O111</td>
<td>:H8</td>
<td>7/24/2019</td>
<td>Human</td>
<td>STOOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>stx1a;</td>
<td>PNUSAEO27690</td>
</tr>
<tr>
<td>C15030755</td>
<td>IN</td>
<td>O111</td>
<td>:H8</td>
<td>7/26/2019</td>
<td>Human</td>
<td>STOOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>stx1a;</td>
<td>PNUSAEO28300</td>
</tr>
</tbody>
</table>
WGS CLUSTERS

WGS Clusters are looked at using a tree and a heat map. The difference in alleles that are considered to be closely related depends on the pathogen. In general, anything less than 10 alleles difference would be considered suspect of an outbreak.
What does a WGS cluster match look like?
What does a WGS cluster match look like?

<table>
<thead>
<tr>
<th></th>
<th>C19033349</th>
<th>C19034702</th>
<th>C19034801</th>
<th>C19034886</th>
</tr>
</thead>
<tbody>
<tr>
<td>C19033349</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>C19034702</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C19034801</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>C19034886</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
Thank you!

Jamie Yeadon  
*Supervisor, Enteric Microbiology, Parasitology, and Molecular Subtyping Laboratory*  
Jhadley@isdh.in.gov

Nicole Stone, MPH  
*Enteric Epidemiologist*  
NStone2@isdh.in.gov
What Is Ebola Virus Disease?

- Ebola is a virus that causes Ebola Virus Disease (EVD).
- Discovered in 1976 near the Ebola River in present day Democratic Republic of the Congo.
- Rare and deadly disease most commonly affecting people and nonhuman primates.
- Case fatality rate is around 50% and has varied between 25% and 90% in the past.
What Is Ebola Virus Disease?

- Ebola virus (species *Zaire ebolavirus*)
- Bundibugyo virus (species *Bundibugyo ebolavirus*)
- Sudan virus (species *Sudan ebolavirus*)
- Taï Forest virus (species *Taï Forest ebolavirus*)
- Reston virus (species *Reston ebolavirus*)
- Bombali virus (species *Bombali ebolavirus*)

Known to cause disease in humans

Known to cause disease in primates and pigs

Unknown if causes disease in animals or humans; recently identified in bats
Symptoms

Symptoms may appear anywhere from 2 to 21 days following exposure but often appear between 8 and 10 days following exposure.

Common symptoms include:
- Fever
- Headache
- Joint and muscle pain
- Weakness
- Diarrhea
- Vomiting
- Stomach pain
- Lack of appetite

Some patients may experience:
- Rash
- Red eyes
- Hiccups
- Cough
- Sore throat
- Chest pain
- Difficulty breathing
- Difficulty swallowing
- Bleeding inside/outside the body
Case Definition

Person Under Investigation (PUI)
A person who has both consistent signs or symptoms and risk factors as follows should be considered a PUI:
• Elevated body temperature or subjective fever or symptoms, including severe headache, fatigue, muscle pain, vomiting, diarrhea, abdominal pain, or unexplained hemorrhage;
-AND-
• An exposure within 21 days before the onset of symptoms.

Confirmed Case
Laboratory-confirmed diagnostic evidence of Ebola virus infection
Diagnostics

Tests

• Antibody-capture enzyme-linked immunosorbent assay (ELISA)
• Antigen-capture detection tests
• Serum neutralization test
• Reverse transcriptase polymerase chain reaction (RT-PCR) assay
• Electron microscopy
• Virus isolation by cell culture

An Ebola patient is cared for at the ALIMA treatment centre in Beni, DRC. © WHO / Chris Black
Diagnostics (cont.)

Current WHO recommended tests include:
• Automated or semi-automated nucleic acid tests (NATs) for routine diagnostic management.
• Rapid antigen detection tests for use in remote settings where NATs are not readily available. These tests are recommended for screening purposes as part of surveillance activities; however, reactive tests should be confirmed with NATs.

The preferred specimens for diagnosis include:
• Whole blood collected in ethylenediaminetetraacetic acid (EDTA) from live patients exhibiting symptoms.
• Oral fluid specimen stored in universal transport medium collected from deceased patients or when blood collection is not possible.
Decision Guide for Initial Evaluation of Suspect Cases of Ebola Virus Disease (EVD)

Has the individual been in a country experiencing widespread transmission of Ebola in the last 21 days?

- NO
  - This patient is not at risk for having come into contact with Ebola virus.
  - You may evaluate the patient as you would normally.

- YES

Has the individual had ANY of the following exposures in the last 21 days?
- Direct contact with body fluids, from a person sick with Ebola who is showing symptoms, through a needle stick, splashes to eyes, nose or mouth, OR getting body fluids directly on skin
- Direct contact with a person with Ebola who has symptoms, or the person’s body fluids, while not wearing appropriate personal protective equipment (PPE)
- Worked in a lab processing blood or body fluids from a person sick with Ebola who has symptoms while not wearing appropriate PPE or without using standard biosafety precautions
- Providing direct care to a person showing symptoms of Ebola in a household setting
- Actively participated in a funeral or had any other contact with the remains of a known/suspect EVD patient

- NO
  - Patient is considered to be in low risk for Ebola.
  - Advise patient to monitor for symptoms and temperature closely for 21 days from date of departure from affected area.
  - Continue with your normal clinical evaluation.
  - Collect all patient information and notify SME via email.

- YES

Does the patient have ANY of the following symptoms?:
- Vomiting
- Diarrhea
- Headache
- Sore throat
- Fever
- Malaise
- Myalgias
- Abdominal pain

- NO

- YES

- Immediately place patient in isolation.
- Immediately notify SME.
When a Traveler Shows Symptoms

- LHD should not have direct contact.
- LHD contacts ISDH immediately to confirm symptoms.
- We will contact everyone to arrange for transport of the patient to the nearest assessment hospital:
  - Chief Medical Officer
  - State Epidemiologist
  - Deputy State Epidemiologist
  - DEP, Division Director
  - DEP, Director of Operations
- Once at an assessment hospital, blood tests are performed to confirm diagnosis.
- Patient is transported to the nearest treatment hospital.
rVSV-ZEBOV

• In a major trial in Guinea in 2015, the vaccination rVSV-ZEBOV proved to be highly protective.
• It is not yet licensed but is being used under expanded access or “compassionate use.”
• It is protective against the Zaire strain of the Ebola virus.
Second Vaccine

• Beginning in mid-October
• 2-dose course, 56 days apart
• Will complement the rVSV-ZEBOV vaccine
• Ring vaccination
Ebola Monitoring

• **Step 1:** ISDH is notified by the CDC that a traveler from an Ebola-affected area is entering the state.

• **Step 2:** ISDH notifies LHD, and they take over monitoring.

• **Step 3:** Traveler is monitored for 21 days.

• **Step 4:** Traveler takes follow-up survey.
Recent Outbreaks

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-2016</td>
<td>28,616</td>
<td>11,310</td>
</tr>
<tr>
<td>2018-2019</td>
<td>3,287</td>
<td>2,192</td>
</tr>
</tbody>
</table>

*Numbers current as of November 14, 2019*
Contact

Erin Swartz, MPH
Healthcare-Associated Infections Epidemiologist
Epidemiology Resource Center
Indiana State Department of Health
eswartz@isdh.in.gov
Thank You!