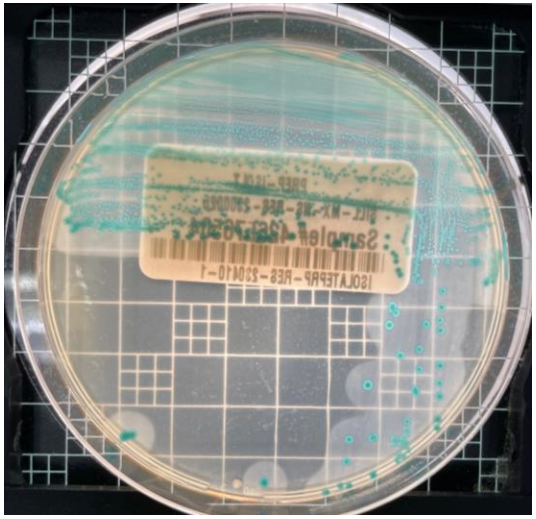
Whole Genome Sequencing and Outbreak Tracking

Catharine Carlin, PhD
Director, Microbiology Innovation
Mérieux NutriSciences

NORTH AMERICAN MEAT INSTITUTE

FOUNDATION FOR MEAT & POULTRY RESEARCH & EDUCATION

Agenda



1. *L. monocytogenes* – brief overview
2. PulseNet and outbreak detection
3. Subtyping: PFGE vs. WGS
4. You need the Epi data!
5. Industry & WGS

L. monocytogenes

A brief overview

NORTH AMERICAN MEAT INSTITUTE

FOUNDATION FOR MEAT & POULTRY RESEARCH & EDUCATION

Listeria

- There are currently 28 *Listeria* spp.
- Only *L. monocytogenes* (LM) is a public health concern
  - Identified in 1926, classified as a foodborne pathogen in 1981
  - *L. ivanovii* is pathogenic, but rarely infects humans
- US has a zero-tolerance policy for LM
  - Other regions (EU and Canada) tolerate certain levels
  - Reluctance to speciate in the US
- Whole-genome sequencing (WGS)
  - If not already determine, WGS will ID the species

LM & Foodborne Illness (US)

- **Outbreaks** - Not as many compared to other bacteria
  - **From 1998-2017:**
    - *Salmonella* NTS– 811
    - *E. coli* O157 – 242
    - *Campylobacter* – 236
    - ***Listeria monocytogenes* – 40**
- **Hospitalizations** – Illnesses are more severe
  - % of cases that lead to hospitalization/death
    - *Salmonella* NTS– 28%/0.5%
    - *E. coli* O157 – 46%/0.5%
    - *Campylobacter* – 17%/0.1%
    - ***Listeria monocytogenes* – 94%/16%**

LM contamination

- *Listeria* is ubiquitous (everywhere)
- Source of contamination is often a post-processing from the plant environment
  - *Listeria* does not survive heat treatment
- RTE foods stored for long periods at refrigeration temperatures
  - *Listeria* grows at refrigeration temperatures

PulseNet & Outbreak detection

How PulseNet revolutionized outbreak surveillance

NORTH AMERICAN MEAT INSTITUTE

FOUNDATION FOR MEAT & POULTRY RESEARCH & EDUCATION

Some Terms

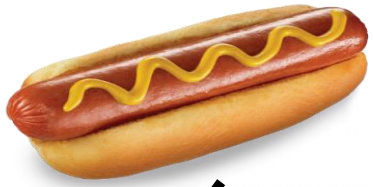
- Strain – A unique isolate of bacteria within a species
  - There are many genotypically different strains of LM
- Subtype – Also referred to as strain typing. The process of differentiating strains of the same genus and species.
  - Are 2 strains of *L. monocytogenes* the same or different?
- Cluster – A group of strains that are highly similar
  - These strains are the same, highly related, or very different

Strains are uploaded to PulseNet for subtyping and detection of clusters

PulseNet

- The goal is to link related isolates or clusters

Patient isolate



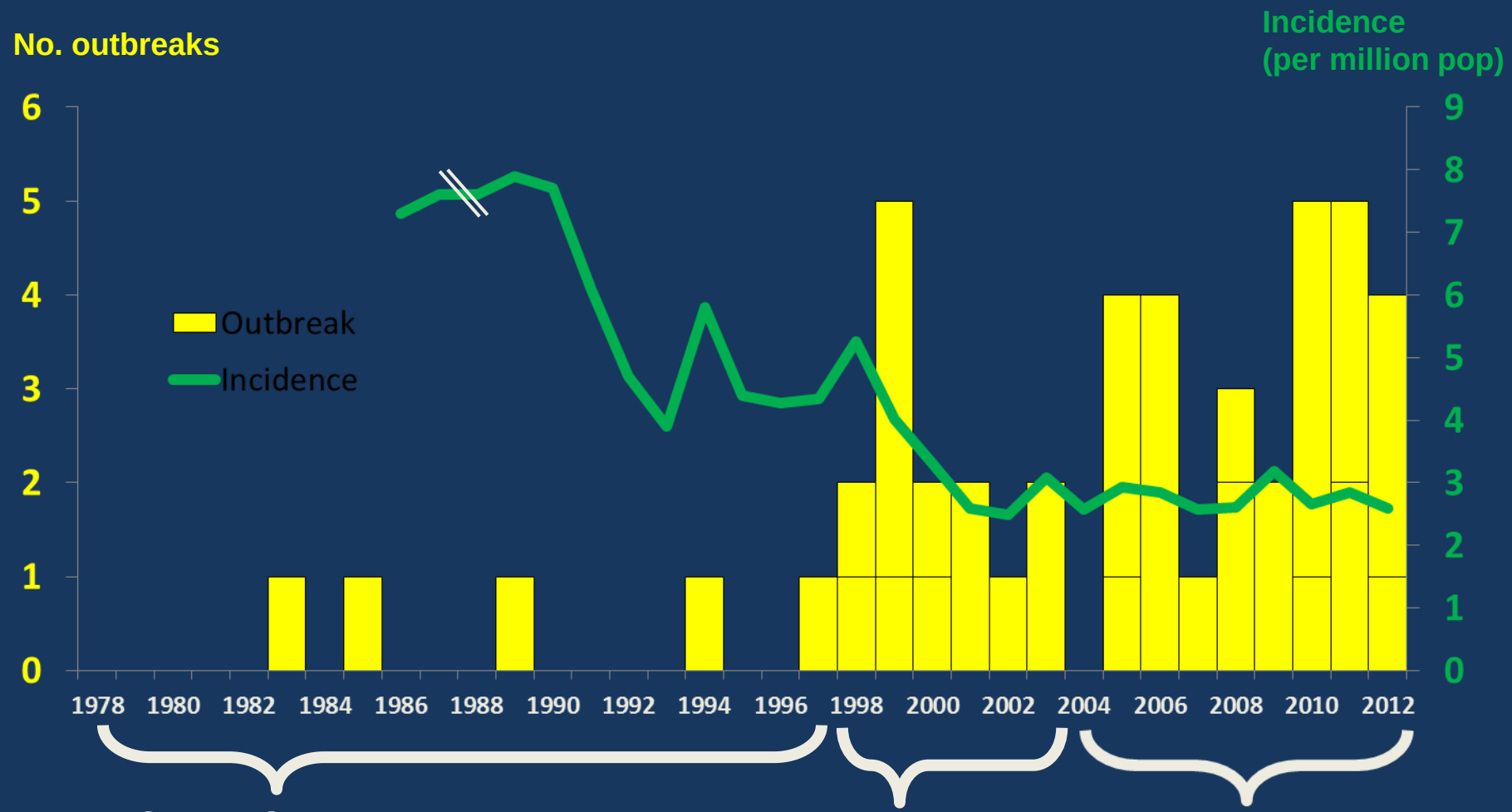
Food and/or environmental isolates

Bacterial DNA data loaded into a database: Are the *L. monocytogenes* isolates genetically related?

PulseNet

- **1996**: CDC launched PulseNet, a national network for outbreak detection
  - FDA & USDA collaborate
- **1996 – 2003**: Pulse Field Gel Electrophoresis (PFGE) for subtyping and cluster identification of LM
- **2004**: Pilot project using WGS for LM outbreaks
- **2005**: Next Generation Sequencing (NGS)
- **2013**: Transition to WGS
- **2019**: WGS is the gold standard for source tracking
  - CDC, FDA, USDA
  - 83 participating laboratories

L. monocytogenes - Outbreaks and Incidence, 1978-2012



Before PulseNet
(20 years)
1978-1997
5 outbreaks
Median **69** cases/outbreak

PulseNet's first years
(6 years)
1998-2003
14 outbreaks
Median **11** cases/outbreak

***Listeria* Initiative & PulseNet** (9 years)
2004-2012
28 outbreaks
Median **5.5** cases/outbreak

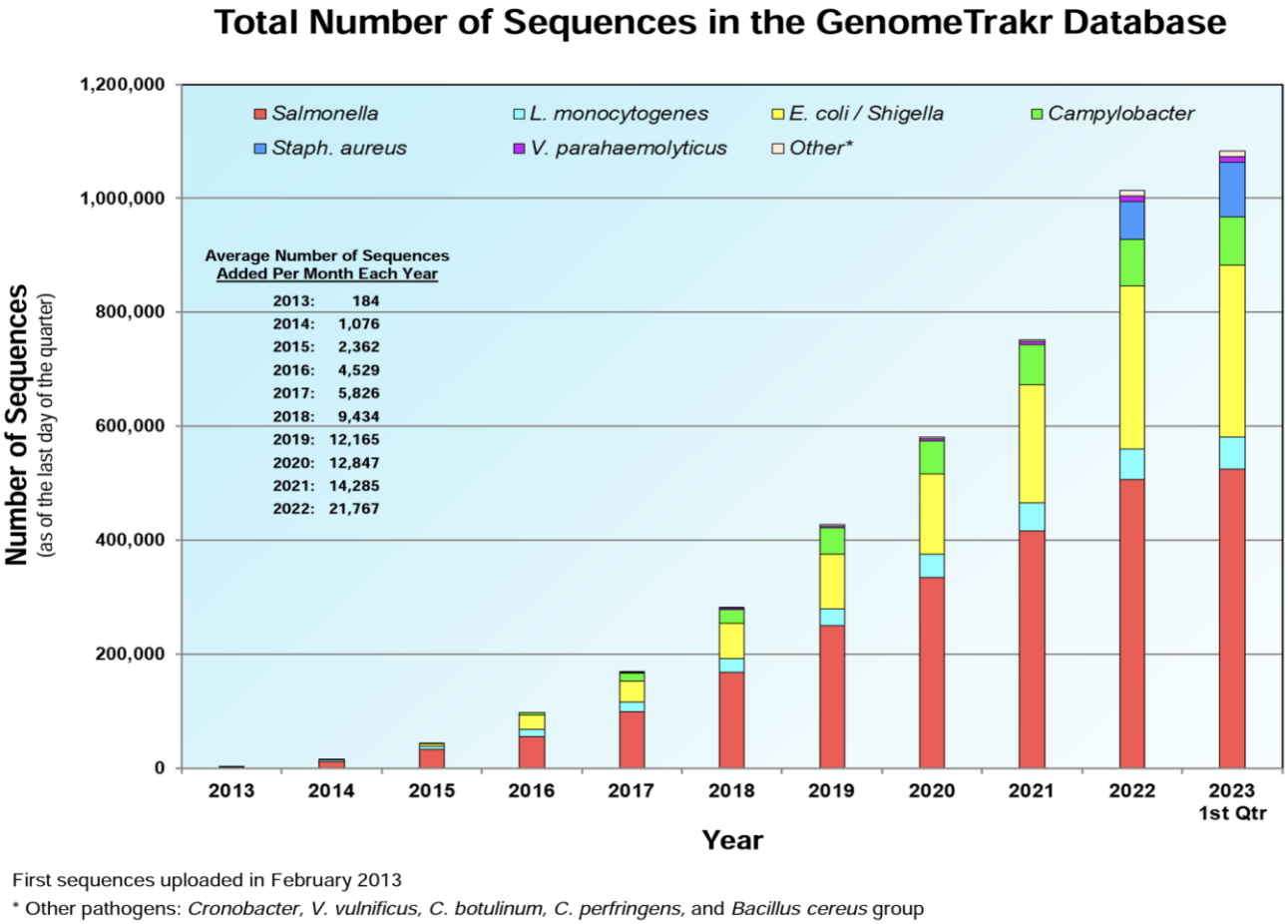
SOURCE: John Besser (CDC)

PulseNet

- **Revolutionized foodborne disease surveillance**
  - Outbreaks are smaller, detected faster
- **Outbreaks identified faster**
  - *E. coli* outbreak before PulseNet: 38d, >700 case
  - *E. coli* outbreak after PulseNet: 18d, ~40 cases
- **Each year, PulseNet identifies:**
  - About 1,500 clusters of foodborne disease at local or state levels
  - About 280 foodborne disease clusters that span multiple states
  - About 30 multistate or national outbreaks
- **PulseNet prevents an estimated 270,000 illnesses in the US annually**

Current Status – WGS & PulseNet

- Updated every week
  - April 2022: 750K
  - April 2023: >1 million



Source: https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genometrakr-fast-facts

PFGE vs. WGS

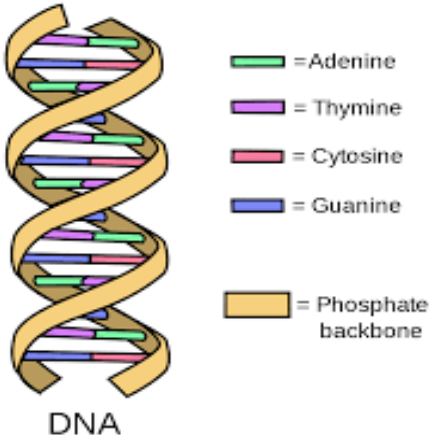
Why WGS is superior for subtyping

NORTH AMERICAN MEAT INSTITUTE

FOUNDATION FOR MEAT & POULTRY RESEARCH & EDUCATION

PFGE vs. WGS Overview

- PFGE and WGS are both whole-genome-based subtyping methods
- With WGS: Millions of DNA base pairs compared vs. 15 to 30 large DNA fragments with PFGE



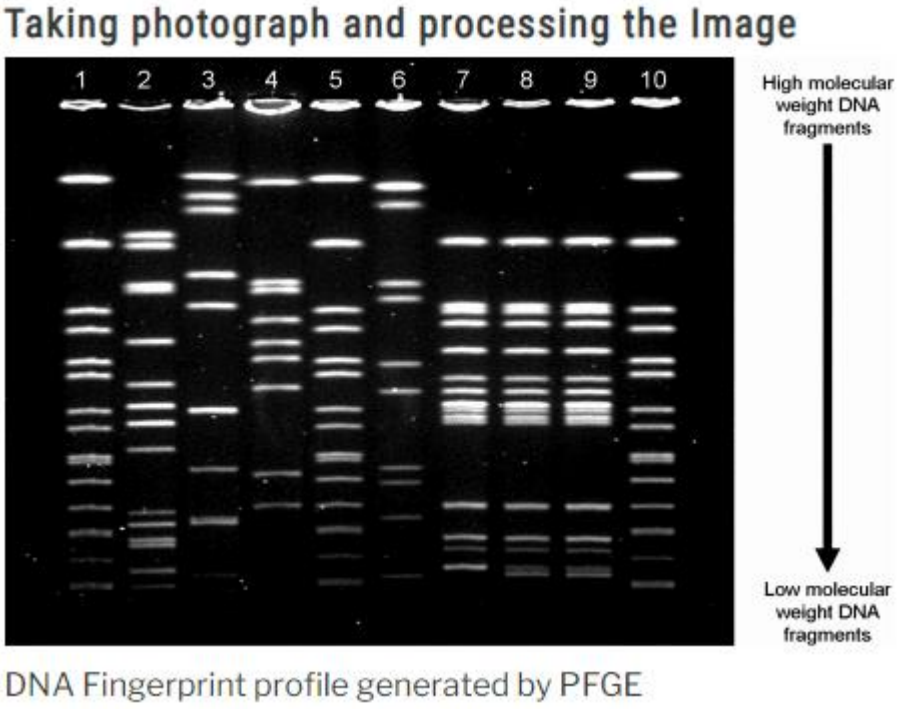
The genome is made up of **4 nucleotides (A,T, G, C)** and the sequence is unique to each strain

```
>NODE_1_length_1401924_cov_73.350607
AAAGTCTCCTCACGCAAACCCTTCTTGGTTGGGTCACGGAAGATAACCTTGGCCAAAGGA
GCACCTCTACCAGGGTCGTGAATGATCTGCTTAACATATCCACGGATCTGTCCTTCACGT
TCGGAAGAGTCCAAAGCTCTAAGCTTGGCAGCTCCCTTTCTCAGACGAGTGTGAGCGGTG
AAGATGGAACCAGCACCCTTTCTTTGAGCACGAATAACTCTACCCATGTTGTGTTAATGT
TTCTTGCTGTAAGAGGACTTGAAATTTTTTATTGGTTTTTTTTTTTGGGGAGTATGAGGG
TTCTTGTGTTCGCGGGTTAACCCTAGTGCTGGTCACGTGCCTTATTGGGCAAGCTGTGTG
AGGTATCATAAGGTGGTAGTTGAAAGGTACCTTATGGAAGACTTCGTTAGGAAGGTGTCT
GTATGATTAGAGTGGCGTAGGGTGAATGATTTAATTCTTCTTCG.....
>NODE_2_length_1392447_cov_73.757244
TGTACCTACTAGCTTGAATAACAAGTTTATCTTTGAGGAAACTTGGTTTCAGAGACAAAA
GTTAAGTACTTGACATTGGGAGCTAAGCTTCTGCATTGCTCCTCTCTGAAAGACTTCAAG
ACTTACCATTTGGAAACAAGTGAGTTTGCATTAGTATCAAAAGGTTGGTATGATATAGGT
```

- *Listeria* contains ~ 3 million base pairs

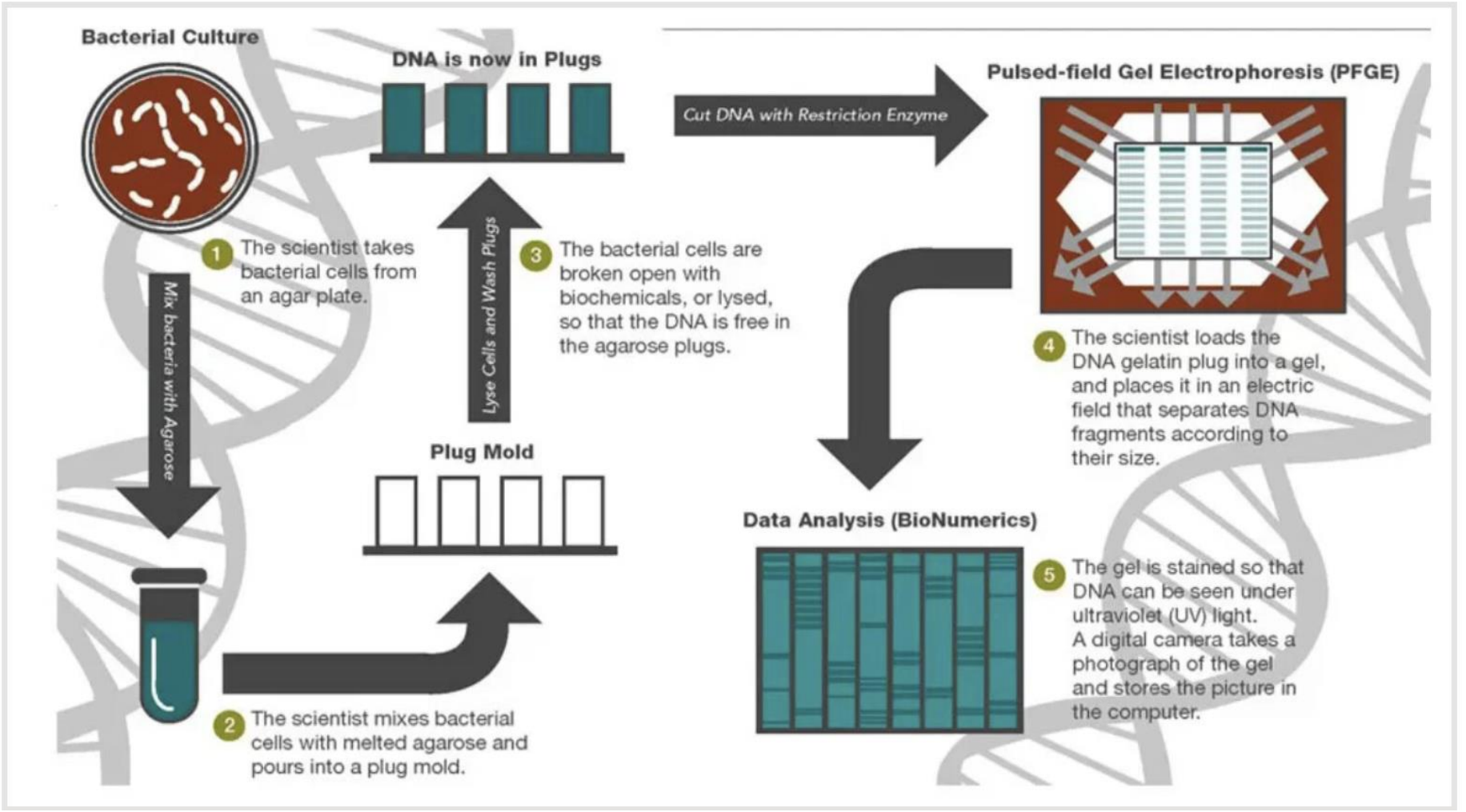
PFGE Basics

- How does PFGE work?
  1. The bacterial genome is cut into multiple large DNA fragments
     - Genome = all DNA
  2. The DNA fragments separated based on size
  3. 15-30 bands are generated to create a DNA fingerprint
  4. The DNA fingerprint patterns are compared
     - Different strains yield different band patterns



Source: A. Tankeshwar, 2022

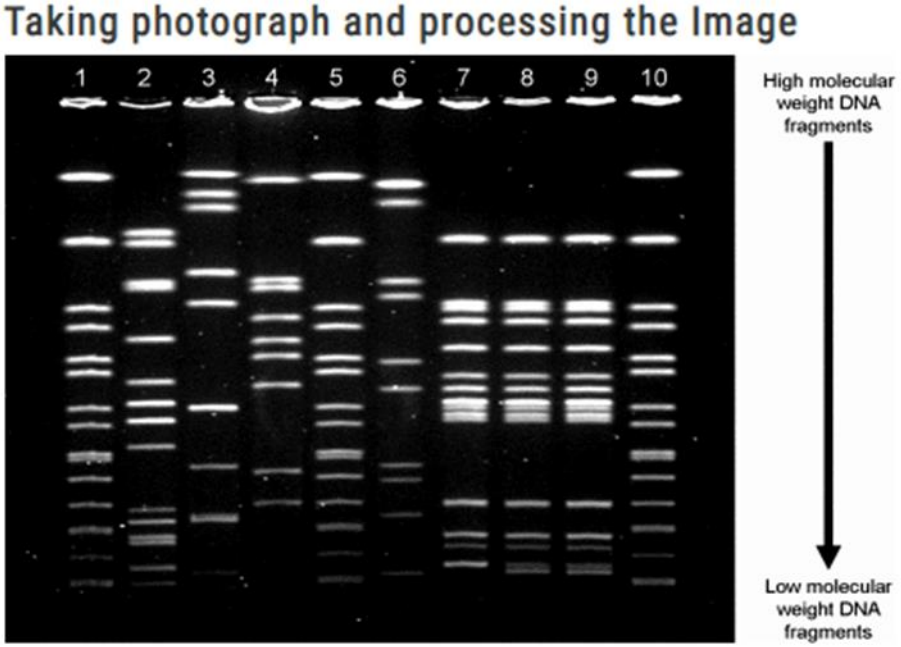
PFGE Basics



Source: S. Aryal Microbe Notes

Limitations of PFGE

- 15 – 30 “bands” are compared
- Isolates that are not genetically related may appear the same
  - Complicates an investigation
- Isolates that are genetically related by PFGE may not be related
  - Leads to false associations



DNA Fingerprint profile generated by PFGE

Source: A. Tankeshwar, 2022

WGS Basics

- **Whole Genome Sequencing (WGS)**
  - A method for determining the DNA sequence of an organism's genome
- **Next Generation Sequencing (NGS)**
  - NGS = massively parallel sequencing (or high throughput sequencing)
  - NGS is the technology, WGS is the application
- **Sequence data analyzed**
  - Accomplished using bioinformatics software
- **Differences can be detected to a single base pair**
  - SNP = Single Nucleotide Polymorphism

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

1 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

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

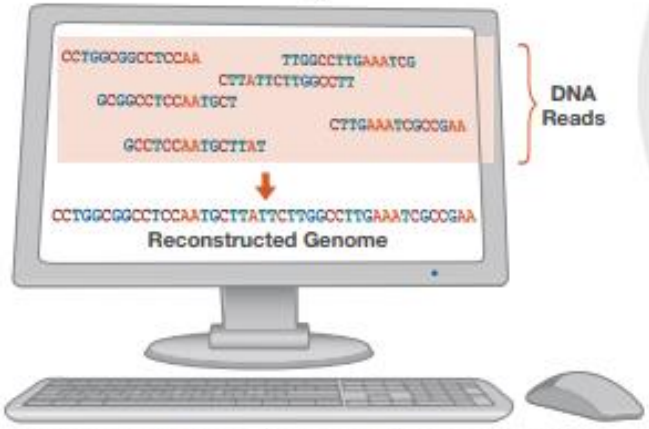
3. DNA Library Preparation

3 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

4 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



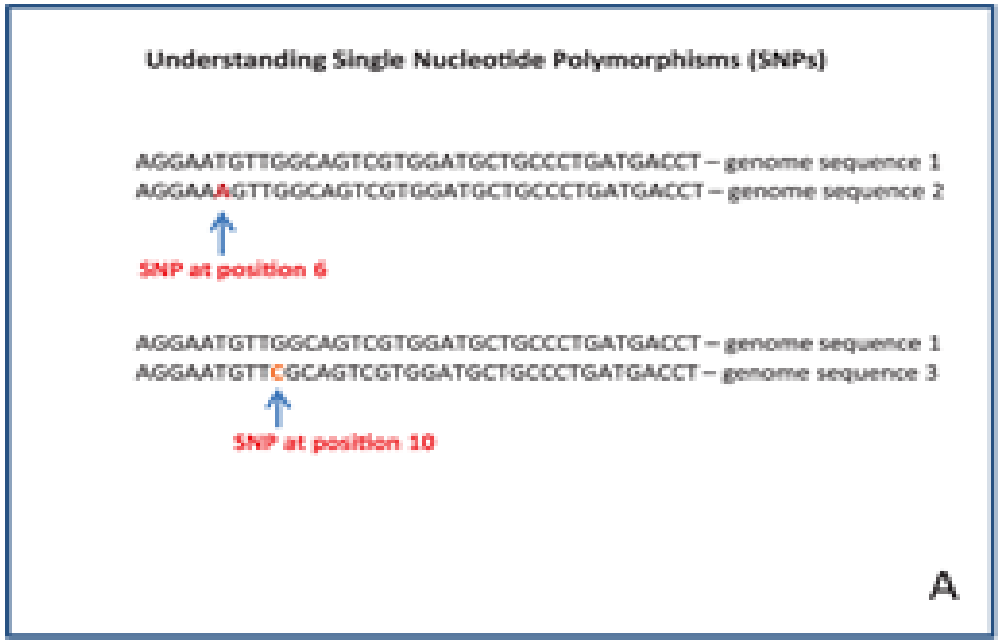
5 The sequencer produces millions of DNA reads and specialized computer programs are used to put them together in the correct order like pieces of a jigsaw puzzle. When completed, the genome sequence containing millions of nucleotides (in one or a few large pieces) is ready for further analysis.

https://www.cdc.gov/pulsenet/pdf/Genome-Sequencing-508c.pdf

Comparing WGS Sequence Data

- Single Nucleotide Polymorphism (SNP)
- Compares single nucleotide differences
- The number of SNPs is used to assess relatedness
- How many SNPs = different strains?
  - These is no definite number
  - ***General*** rule of thumb ≥ 20

*Ice Cream outbreak strains had up to 29 SNP differences, 9 PFGE profiles



Source: E.L. Stevens, 2022

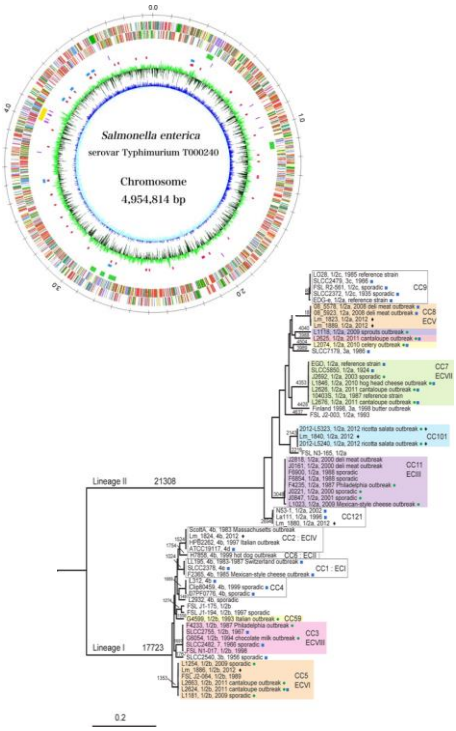
WGS – Beyond Strain Differentiation

Compared to PFGE which only provides a qualitative comparison, WGS can identify

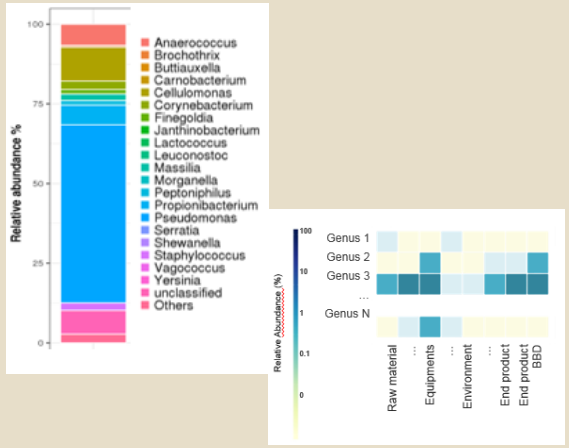
- ✓ Strain relatedness with high resolution
  - Fewer strains needed to identify a cluster
- ✓ Serotype
- ✓ Virulence
- ✓ Antimicrobial Resistance (AMR)

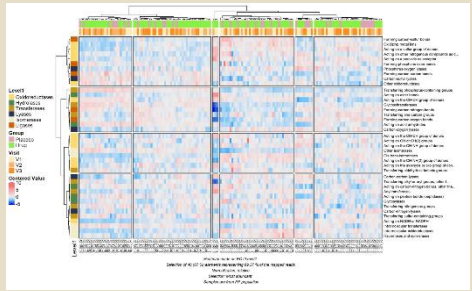
WGS – Beyond Strain Differentiation

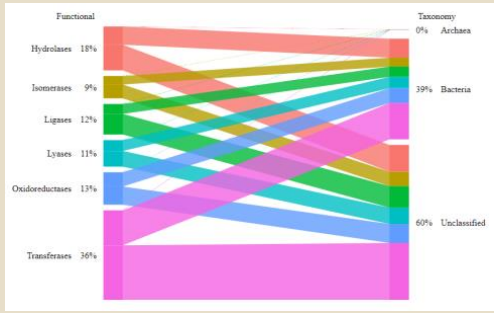
WGS	Metabarcoding	Shotgun metagenomics
Identification and properties	*Microbial composition*	*Detailed composition & functionality*



16S for bacteria
ITS for fungi





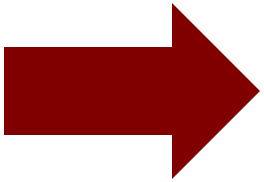


Why replace PFGE with WGS?

- **PFGE served practical public health function but data are qualitative**
- **WGS provides provides more than just strain differentiation**
- **WGS provides much higher resolution for relatedness than PFGE**
- **Outbreaks are solved faster with WGS compared to PFGE**
  - Fewer cases with WGS compared to PFGE investigated outbreaks
- **WGS shotgun metagenomics applications**
  - Sequencing without a bacterial isolate

WGS – The future

- **The use of Culture Independent Diagnostic Tests (CIDTs) among public health labs continues to increase**
  - No bacterial isolate available for sequencing
- **Shotgun Metagenomics is showing promise**
  - Regulatory agencies are conducting studies comparing isolate vs. enrichment-based sequencing



Quasimetagenomic source tracking of *Listeria monocytogenes* from naturally contaminated ice cream

Andrea Ottesen ✉, Padmini Ramachandran, Yi Chen, Eric Brown, Elizabeth Reed & Errol Strain

BMC Infectious Diseases **20**, Article number: 83 (2020) | Cite this article

You still need the Epi data!

The importance of epidemiological evidence in source tracking investigations

NORTH AMERICAN MEAT INSTITUTE

FOUNDATION FOR MEAT & POULTRY RESEARCH & EDUCATION

Considering All the Evidence

- Epidemiological “epi” data
  - Similar WGS strains may no be epidemiologically linked
  - Different WGS strains may be identified from a common source
    - Caramel apple outbreak!
- Case and non-case exposures – determine if there is statistically significant association
  - Food product, brand, region
- Traceback investigation data
  - Evaluate all steps of production (e.g., distribution pattern)

Basic Epi

- The study of disease distribution and determinants
  - What exposure factors are associated with a disease
- Patients with listeriosis are asked to report what they ate over a 4-week period

- Quantify and statistically analyze the findings

Basic Epi

1. Cluster detected by PulseNet
2. Conduct patient interviews, include control samples
3. Form a hypothesis around the potential source
4. Collect traceback samples (food, environmental)
5. Calculate a measure of association

	Listeriosis cases	No illness
Ate Company A hot dog	a 250	b 19
Did not eat	c 9	d 90

Odd Ratio (OR) = (a*d)/(b*c) = (250*90)/(19*9) = 13.0

OR >1 → People that ate the hot dog have a greater odd than those that did not of getting listeriosis

Basic Epi

- Real-world example: *Salmonella* and flour
  - On-going as of May 1, 2023
- From the CDC website: https://www.cdc.gov/foodsafety/outbreaks/lists/active-investigations.html

Epidemiologic Data

State and local public health officials are interviewing people about the foods they ate in the week before they got sick. Of the eight people interviewed, seven (88%) reported eating raw dough or batter. Of six sick people with brand information, all six (100%) reported buying Gold Medal brand flour. The only brand reported was Gold Medal.

Traceback and Laboratory Data

FDA conducted a traceback investigation and identified a single production facility of the flour consumed by sick people. FDA initiated an inspection at the General Mills Kansas City, Missouri facility and collected samples from retain flour. The outbreak strain was identified in one of the samples of flour.

Industry & WGS

NORTH AMERICAN MEAT INSTITUTE

FOUNDATION FOR MEAT & POULTRY RESEARCH & EDUCATION

When Could Industry Use?

- **Find the root cause of a product or facility pathogen contamination (strain tracking)**
  - Pinpoint a growth niche in equipment or facility
  - Identify a problem ingredient or supplier
- **Determine if isolates from product or facility match with illness databases (much less common)**
- **Find the root cause of spoilage issues**
  - Would require a metagenomics approach to address meat spoilage

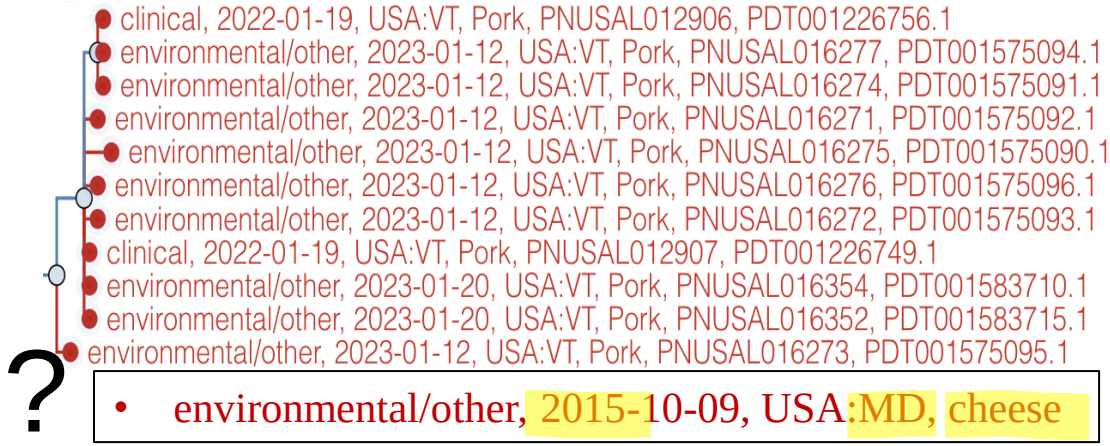
When It Might Help

- **Extended Sporadic Findings in Product or Environment**
  - Subtyping can tell if it's a "house bug" or multiple strains coming in
  - May be able to narrow down the source to a specific ingredient or supplier

- **Outbreak Situations**
  - Vindicate – not our bug
  - Pull the trigger – it's our bug

Industry Adoption of WGS – The Reality

- **Very few (even large companies) doing this**
  - Primary reason is the potential regulatory implications
    - WGS data could potentially lead to legal action
    - Concerns with lack of epi data when regulatory determines WGS sequences are highly-related
  - What will happen if a historical isolate matches a current outbreak strain?



- **WGS may provide too much information**
  - *Listeria* spp. level identification

Industry Adoption of WGS – The Reality

- **Time and cost**
  - **Still relatively expensive and requires high level of technical skill**
    - Costs continue to come down, but still more expensive than less discriminatory subtyping methods
  - **Turn-around times relatively long: 1 – 3 weeks**
    - Faster options are being introduced
- **PFGE continues to be utilized because it is not subjected to regulatory scrutiny**
  - PFGE technical limitations (e.g., PFGE "matches" may not be the same strain) also limit regulatory scrutiny
- **Ribotyping is also utilized (faster, easier cheaper, but lower resolution than PFGE)**

Conclusions

- **PulseNet and foodborne disease surveillance**
  - Fewer cases, faster resolution
- **The gold standard for subtyping is WGS (replaced PFGE)**
  - WGS has identified more outbreaks, allowed for faster response
- **WGS sequence data provides more information than strain-to-strain comparison**
  - Metagenomics
- **Proactive applications – root cause analyses to better inform facilities where to focus money and time**
  - Justify the costs of better hygienic design
- **Protection from legal scrutiny is needed**
  - "Safe Harbor"