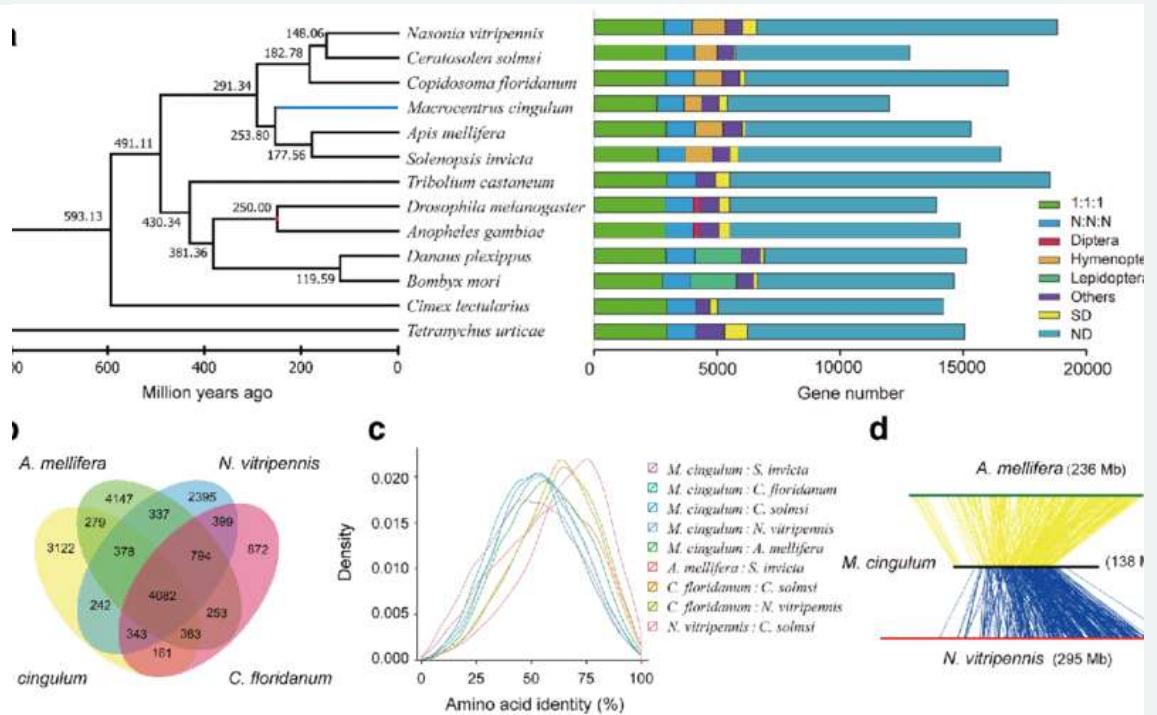
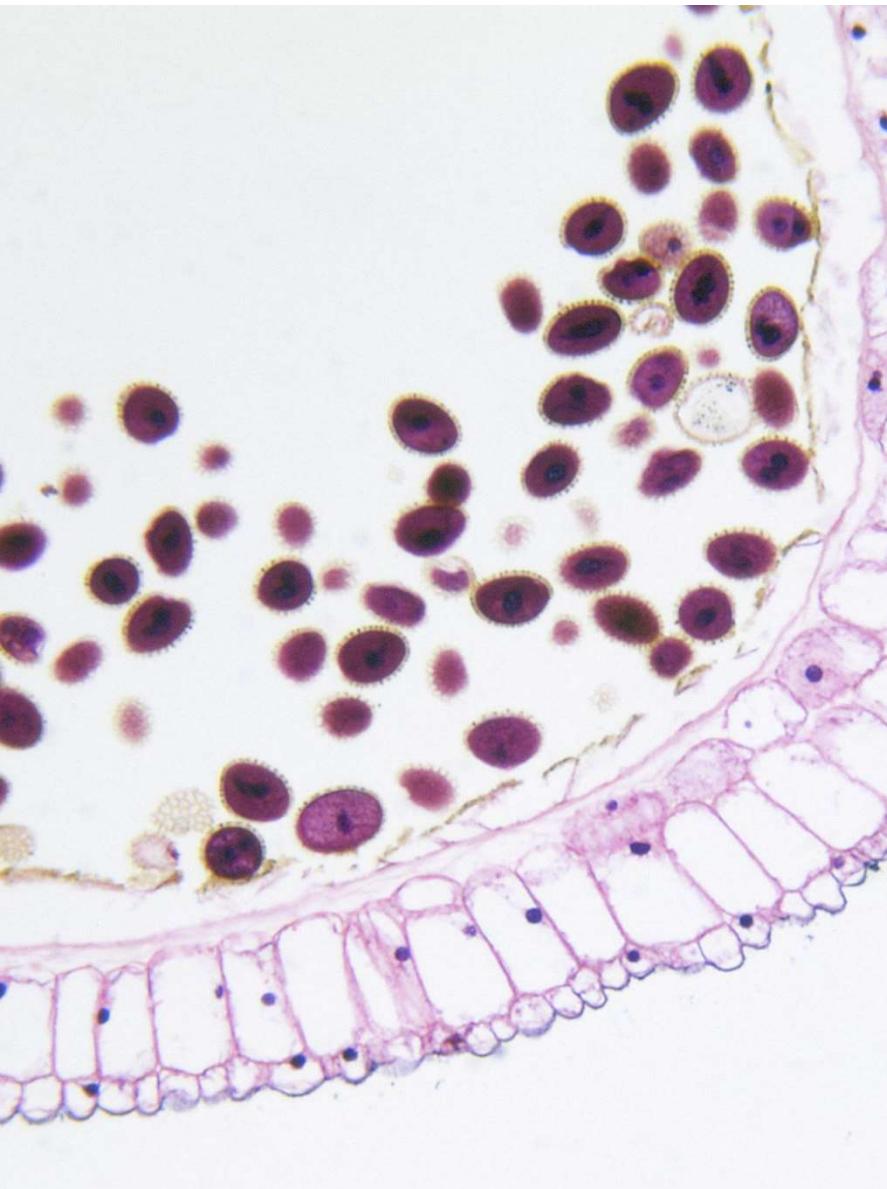


Comparative Genomics

TEAM: F (GROUP-4)

TEAM MEMBERS: ANNA, AYUSHI,
KHUSHI, PAVITHRA, RAMYA





Introduction

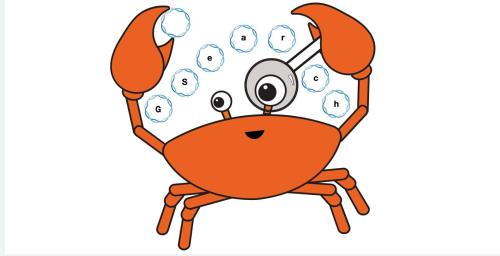
- *Listeria monocytogenes*, a pathogenic species within *Listeria* genus, is responsible for causing listeriosis, a serious foodborne illness.
- It is significant in public health due to their widespread distribution in natural environments and ability to adapt to various habitats.
- This presentation outlines the investigation of an outbreak using comparative genomics approaches.
- The analysis aims to provide insights into the causative *L. monocytogenes* strain, its genetic characteristics, virulence factors, and antimicrobial resistance profiles.

Approach

- Comparative genomics is a robust scientific approach that involves comparing the genetic content of multiple organisms to understand their evolutionary relationships, identify genetic variations, and elucidate functional capabilities.
- In the context of pathogen outbreak investigation, comparative genomics can reveal how specific strains differ from each other and from known reference genomes, which is crucial for pinpointing the outbreak source and understanding the pathogen's evolution.

Pipeline

- In this pipeline, we will utilize three key tools for comparative genomics:
- gsearch: This tool is used to perform advanced genomic searches and comparisons using MinHash-like signatures and Hnsw (Hierarchical Navigable Small World) graphs.
- parsnp: This tool is used for phylogenetic reconstruction and identifying core genome single nucleotide polymorphisms (SNPs).
- roary: This tool is used for pangenome analysis, identifying the core and accessory genes across multiple genomes.
- These tools work together to provide a comprehensive analysis of the genomic data, from pangenome analysis to phylogenetic reconstruction.

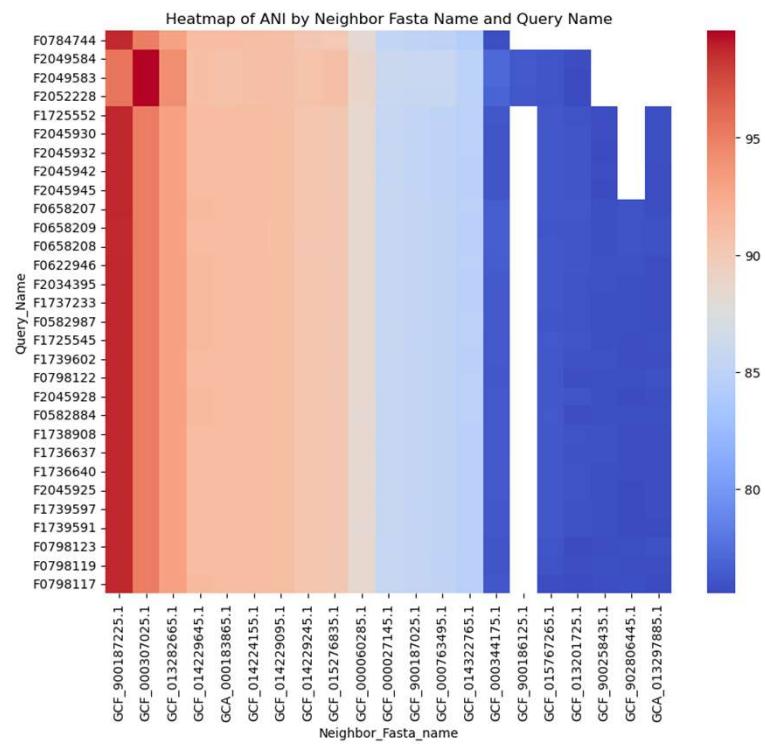


gsearch

- Performs fast search for microbial genomes similar to target genomes using algorithm that employs a MinHash-like measurement to estimate similarities
- Builds hierarchical navigable small world (HNSW) graph database that uses the Jaccard index as a proxy for average nucleotide identity
- Determines nearest neighbors of query genomes against HSNW graph database
- Approximates nearest neighbor embedding using UMAP-like algorithm

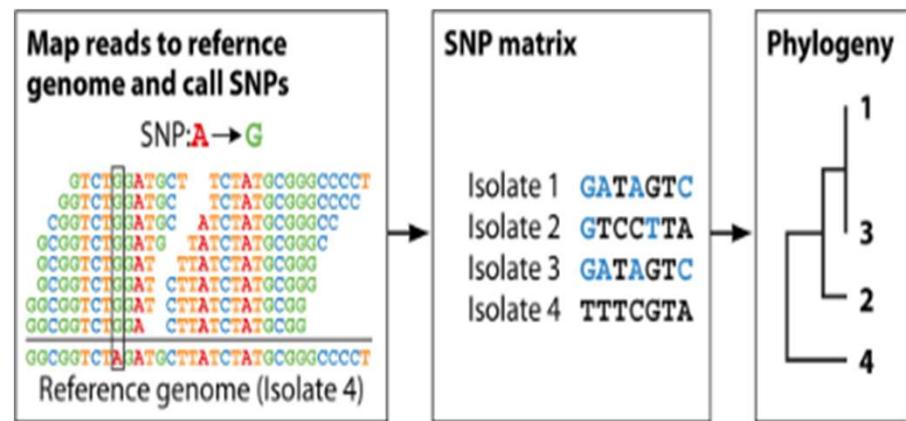
gsearch ANI

- Heatmap: Red indicates greater similarity between genomes, blue indicates less similarity between genomes
- Most similar sequence for outbreak group is the reference genome with serotype 1/2a; same as determined by group 3

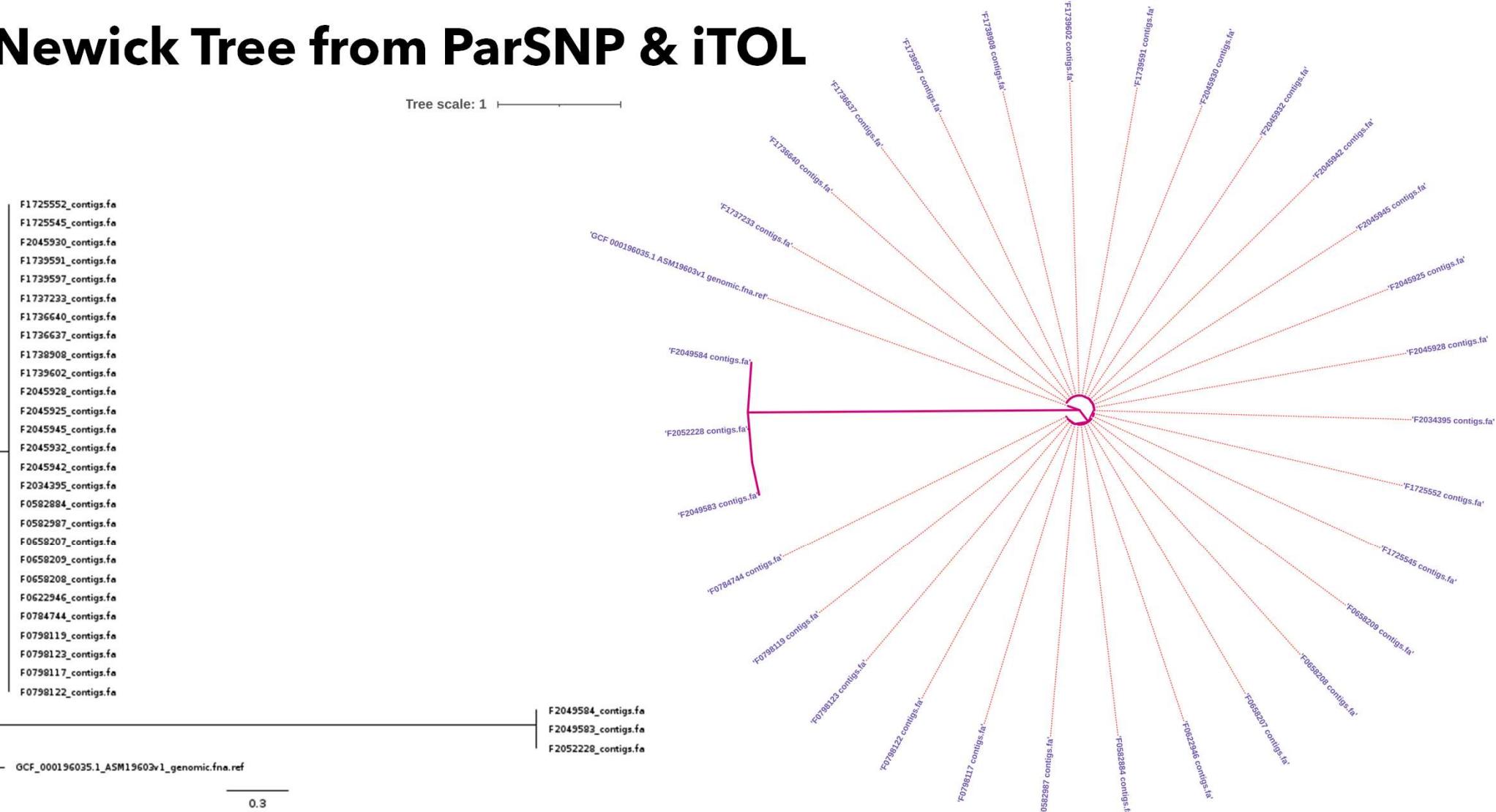


Phylogenetic analysis with ParSNP

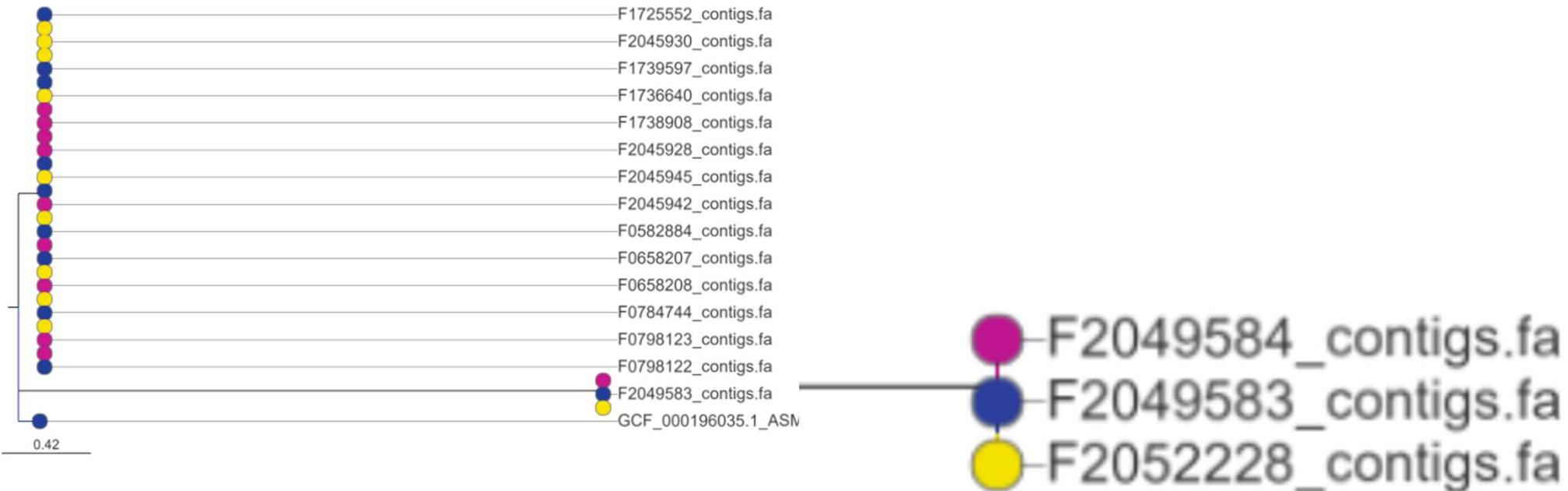
- Performs a rapid core genome multi-alignment and phylogenetic tree construction.
- ParSNP identifies the core genome regions shared across the input genomes and generates a multiple sequence alignment and a maximum-likelihood phylogenetic tree.
- The core genome represents the majority sequence of the outbreak, and should be used to identify potential causative sources of the outbreak



Newick Tree from ParSNP & iTOL



Using Microreact

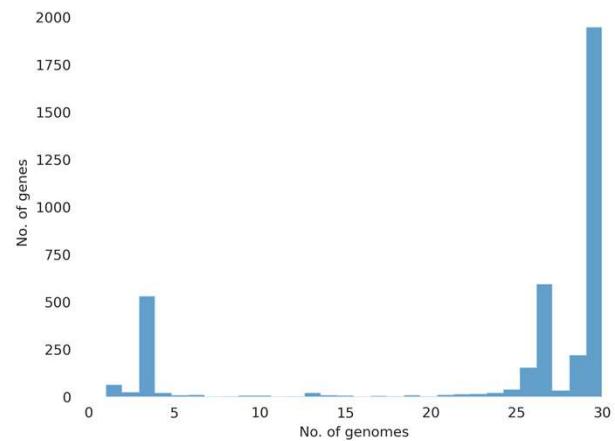


Pangenome analysis with Roary

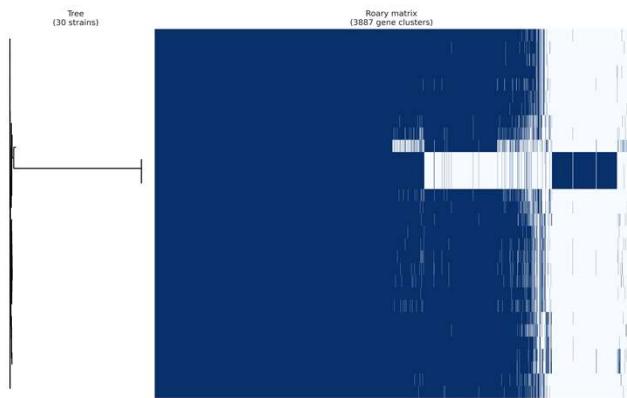


- **Pangenome Construction:** Roary takes the annotated genome files (in GFF3 format) as input and constructs the pangenome, which is the complete set of genes found across all the input genomes.
- **Core and Accessory Genes Identification:** Roary identifies the core genes (present in all genomes) and the accessory genes (present in some but not all genomes) within the pangenome. This information is crucial for understanding the shared and unique genetic features among the *Listeria monocytogenes* isolates.
- **Phylogenetic Analysis:** Roary can generate a core gene alignment and use it to construct a phylogenetic tree, providing insights into the evolutionary relationships between the *Listeria monocytogenes* strains.
- **Visualization:** Roary produces several output files, including a pangenome matrix and a presence/absence visualization, which can be used to explore the genetic diversity and similarities among the *Listeria monocytogenes* isolates.

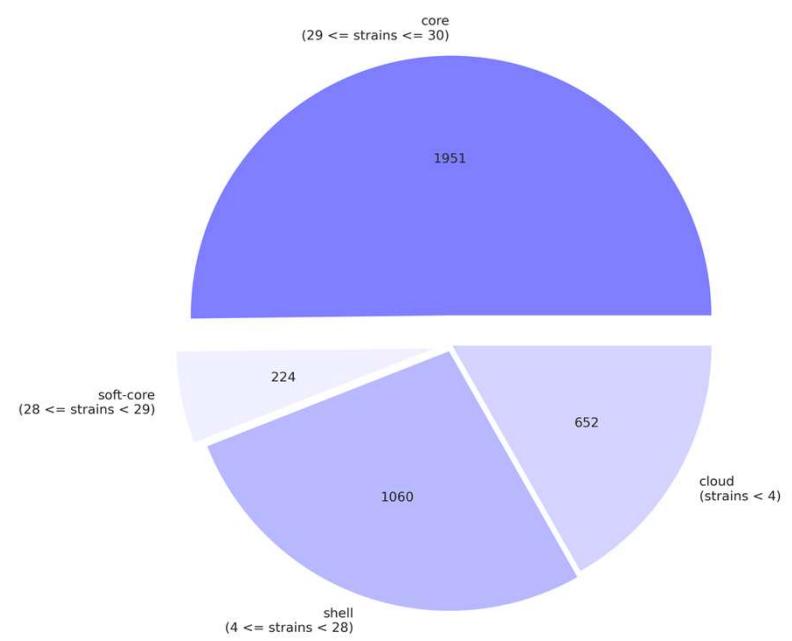
1.

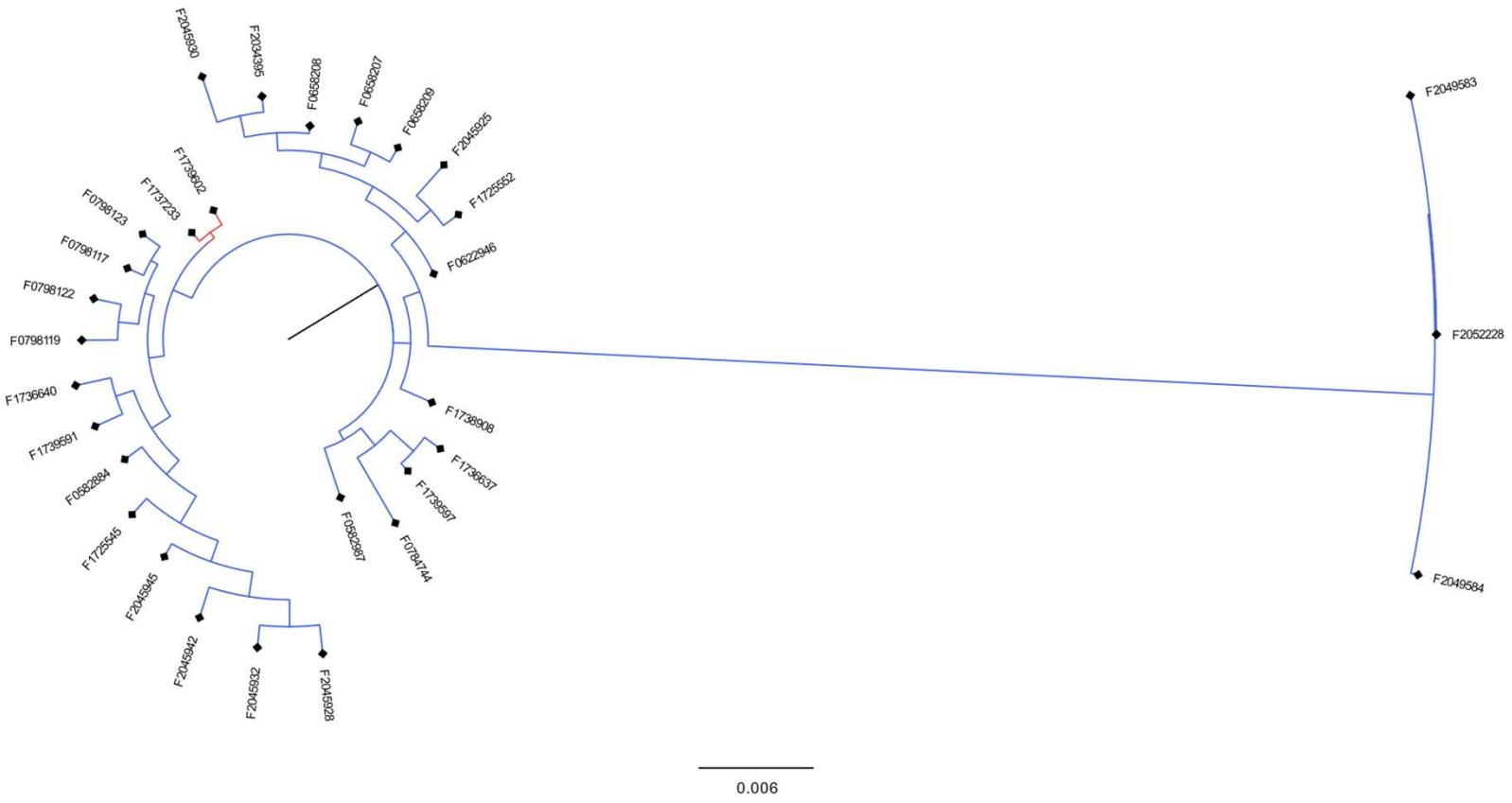


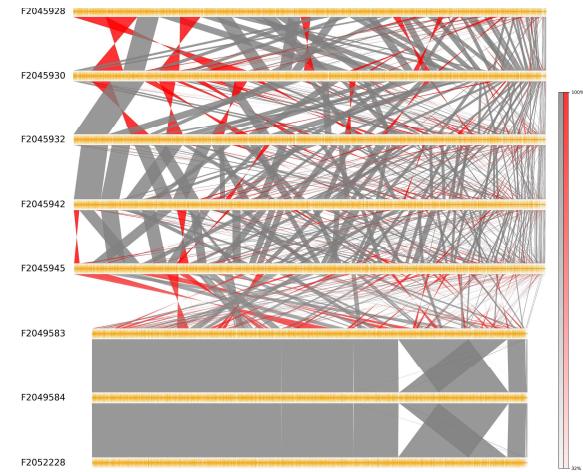
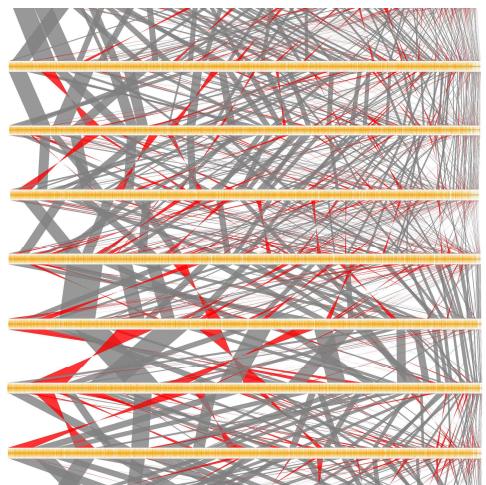
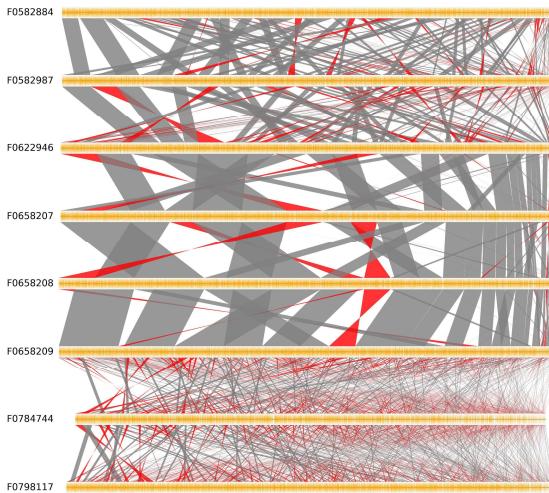
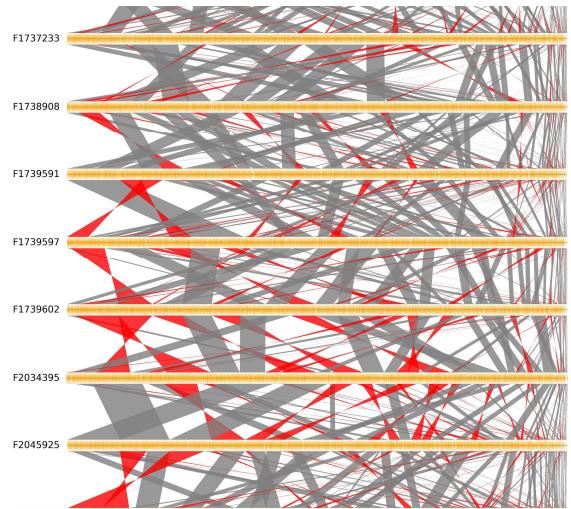
2.

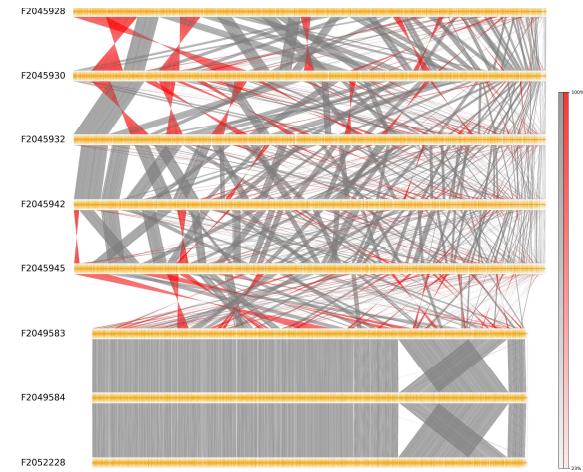
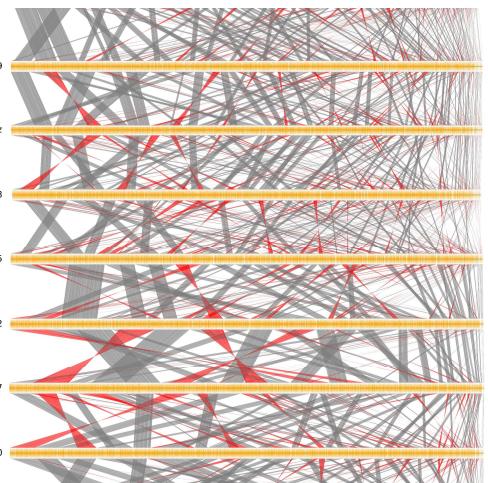
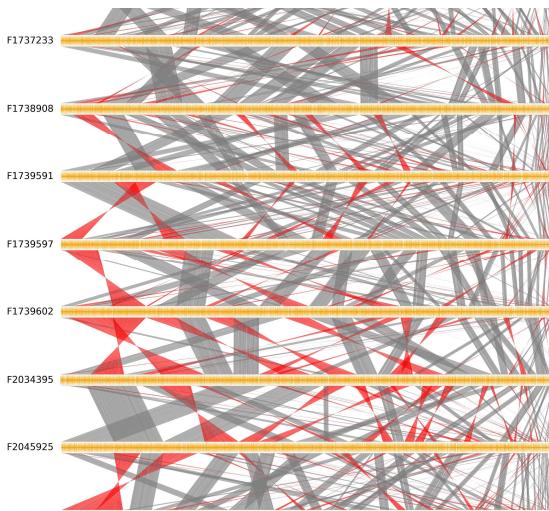
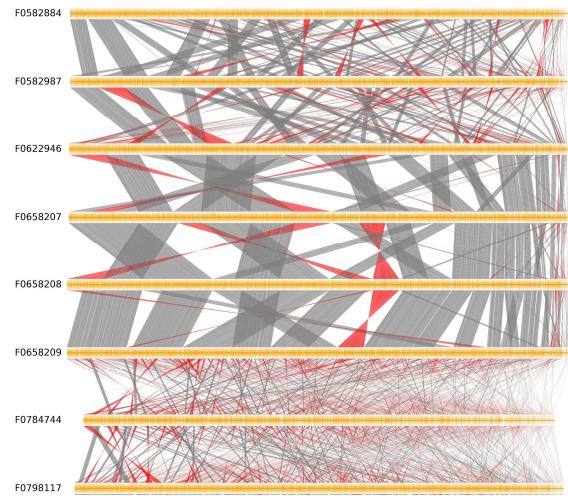


3.





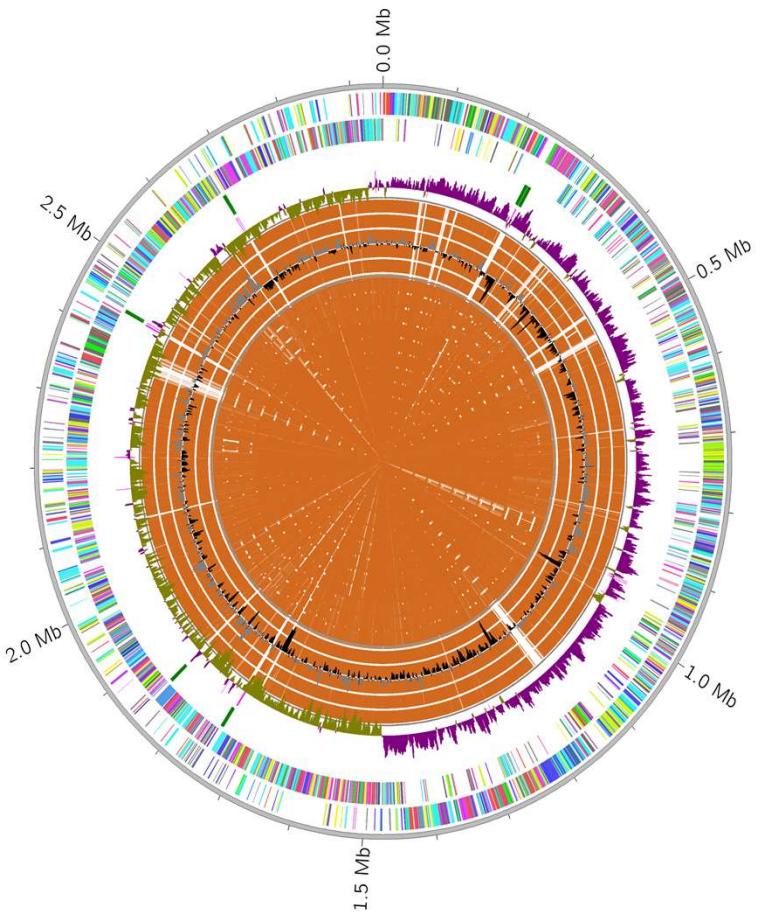
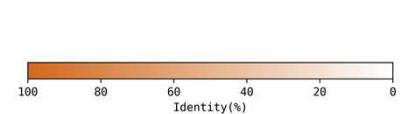




■ J : Translation, ribosomal structure and biogenesis
 ■ A : RNA processing and modification
 ■ K : Transcription
 ■ L : Replication, recombination and repair
 ■ B : Chromatin structure and dynamics
 ■ D : Cell cycle control, cell division, chromosome partitioning
 ■ Y : Nuclear structure
 ■ V : Defense mechanisms
 ■ T : Signal transduction mechanisms
 ■ M : Cell wall/membrane/envelope biogenesis
 ■ N : Cell motility
 ■ Z : Cytoskeleton
 ■ W : Extracellular structures
 ■ U : Intracellular trafficking, secretion, and vesicular transport
 ■ O : Posttranslational modification, protein turnover, chaperones
 ■ X : Mobilome: prophages, transposons
 ■ C : Energy production and conversion
 ■ G : Carbohydrate transport and metabolism
 ■ E : Amino acid transport and metabolism
 ■ F : Nucleotide transport and metabolism
 ■ H : Coenzyme transport and metabolism
 ■ I : Lipid transport and metabolism
 ■ P : Inorganic ion transport and metabolism
 ■ Q : Secondary metabolites biosynthesis, transport and catabolism
 ■ R : General function prediction only
 ■ S : Function unknown
 ■ - : No COG classified

Track Contents

- Forward CDS
- Reverse CDS
- rRNA
- tRNA
- Conserved CDS
- ▲ GC Content (+)
- ▼ GC Content (-)
- ▲ GC Skew (+)
- ▼ GC Skew (-)



- All *Listeria Monocytogenes*

- o Two different sequence types (ST's)

1 abcZ(3) bglA(1) cat(1) dapE(1) dat(3) ldh(1) lhkA(3)

ST 1 (3 Samples)

F2049583, F2049584, F2052228

31 abcZ(7) bglA(14) cat(10) dapE(19) dat(9) ldh(8) lhkA(1)

ST 31

(27 samples)



Food Safety News

<https://www.foodsafetynews.com> › 2024/02 › maryland... :

Maryland company expands recall of soybean sprouts ...

Feb 1, 2024 — Nam & Son of MD of Jessup, MD, has expanded a recall of one-pound bags of soybean sprouts, with the sell-by date of Dec.



Centers for Disease Control and Prevention (.gov)

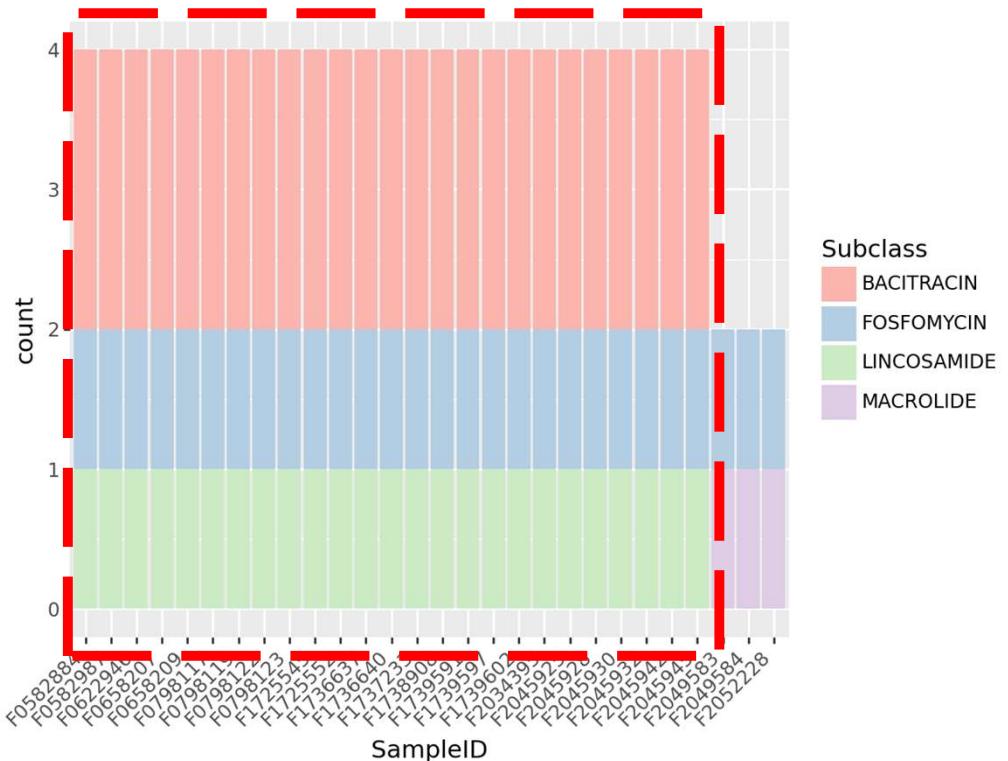
<https://www.cdc.gov> › outbreaks › bean-sprouts-11-14... :

Wholesome Soy Products, Inc. Sprouts

Jan 27, 2015 — On August 28, 2014, Wholesome Soy Products, Inc. conducted a voluntary recall of mung bean sprouts due to possible *Listeria monocytogenes* ...

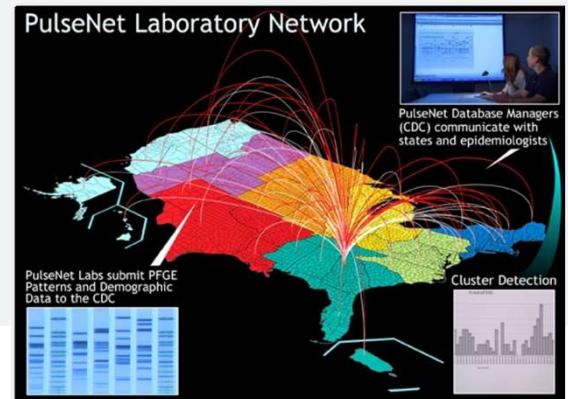
USA: TX, FL, NY, MA
Canada

AMRFinder



- Compared subclasses from AMRFinder between samples that were part of the outbreak and outlier samples
- Most *L. monocytogenes* inherent resistant to fosfomycin and lincosamide antibiotics
- Outbreak samples do not have macrolide resistant gene
- Clinical case studies suggest the combination of amoxicillin and gentamicin is the best option.

Recommendations to CDC



- Use PulseNet to compare core outbreak genome sequence(s) with national database
- Identify other individuals with the same strain
- Interviews about food consumed prior to illness
- Check for State Department of Health Food and Drug Branch records of *L. monocytogenes* detections at food processing facilities.
- Determine causative product and recall

Timeline

References

- Andrew J. Page, Carla A. Cummins, Martin Hunt, Vanessa K. Wong, Sandra Reuter, Matthew T.G. Holden, Maria Fookes, Daniel Falush, Jacqueline A. Keane, Julian Parkhill, Roary: rapid large-scale prokaryote pan genome analysis, *Bioinformatics*, Volume 31, Issue 22, November 2015, Pages 3691-3693, <https://doi.org/10.1093/bioinformatics/btv421>
- Herbert Hof, Therapeutic options, FEMS Immunology & Medical Microbiology, Volume 35, Issue 3, April 2003, Pages 203-205, [https://doi.org/10.1016/S0928-8244\(02\)00466-2](https://doi.org/10.1016/S0928-8244(02)00466-2)
- Radoshevich, L., Cossart, P. Listeria monocytogenes: towards a complete picture of its physiology and pathogenesis. *Nat Rev Microbiol* 16, 32-46 (2018). <https://doi.org/10.1038/nrmicro.2017.126>
- Treangen, T. J., Ondov, B. D., Koren, S., & Phillippy, A. M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome biology*, 15, 1-15.
- Zhao, J., Both, J. P., Rodriguez-R, L. M., & Konstantinidis, K. T. (2022). GSearch: Ultra-Fast and Scalable Microbial Genome Search by Combining K-mer Hashing with Hierarchical Navigable Small World Graphs. *bioRxiv*, 2022-10.

Recent Listeriosis Outbreak

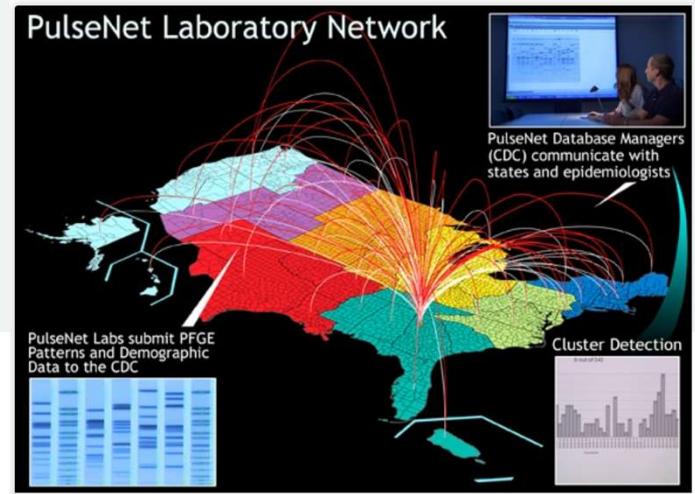
- 26 people infected with the outbreak strain of Listeria were reported from 11 states
- Illnesses started between Jun 2014 to Dec 2023
- 23/26 people with information available were hospitalized
- Health officials interviewed people about the foods they ate in the month before they got sick
- 16/22 people interviewed reported eating queso fresco, cotija, or other similar cheeses, four people reported eating brands made by Rio Lopez



<https://www.fda.gov/>

PulseNet

- Started in 1996
- Network of national labs that process samples associated with foodborne, waterborne, and other zoonotic illnesses to detect outbreaks
- CDC manages a national database of DNA fingerprints of bacteria that cause foodborne illnesses
- Compares the WGS DNA fingerprints of bacteria from patients to find clusters of disease that represent unrecognized outbreaks



<https://www.cdc.gov/>