Comparative Genomics - Team 1
Final Results

Heather Patrick, Lawrence McKinney, Laura Mora, Manasa Vegesna, Kenji Gerhardt, Hira Anis

April 14, 2020

OUTLINE

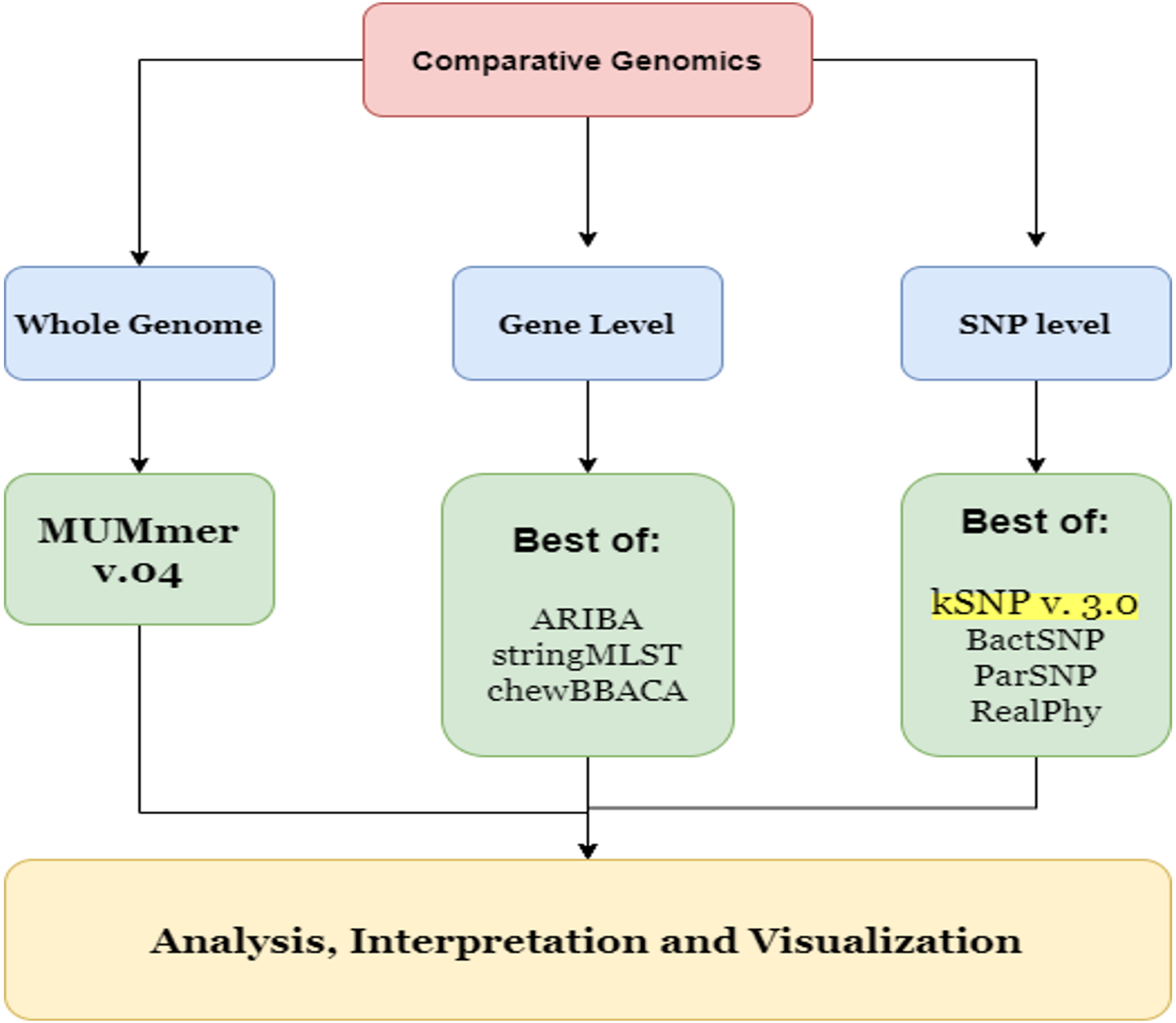
- Objectives
- Proposed Pipeline
- Comparative Genomics Approach
- Final pipeline
- References

OBJECTIVES

Compare genome sequences using various bioinformatic tools to gather knowledge that will answer the following questions:

1. What is the identity of the species/strains that cause the outbreak?
2. How are the isolates related to each other? how do they differ?
3. Which isolates correspond to outbreak versus sporadic strains?
4. What are the virulence and antibiotic resistance profiles of the outbreak isolates?
5. What is the recommended outbreak response and treatment?

The Proposed Preliminary Pipeline



Comparative Genomics

Whole Genome

Gene Level

SNP level

MUMmer v.04

Best of:

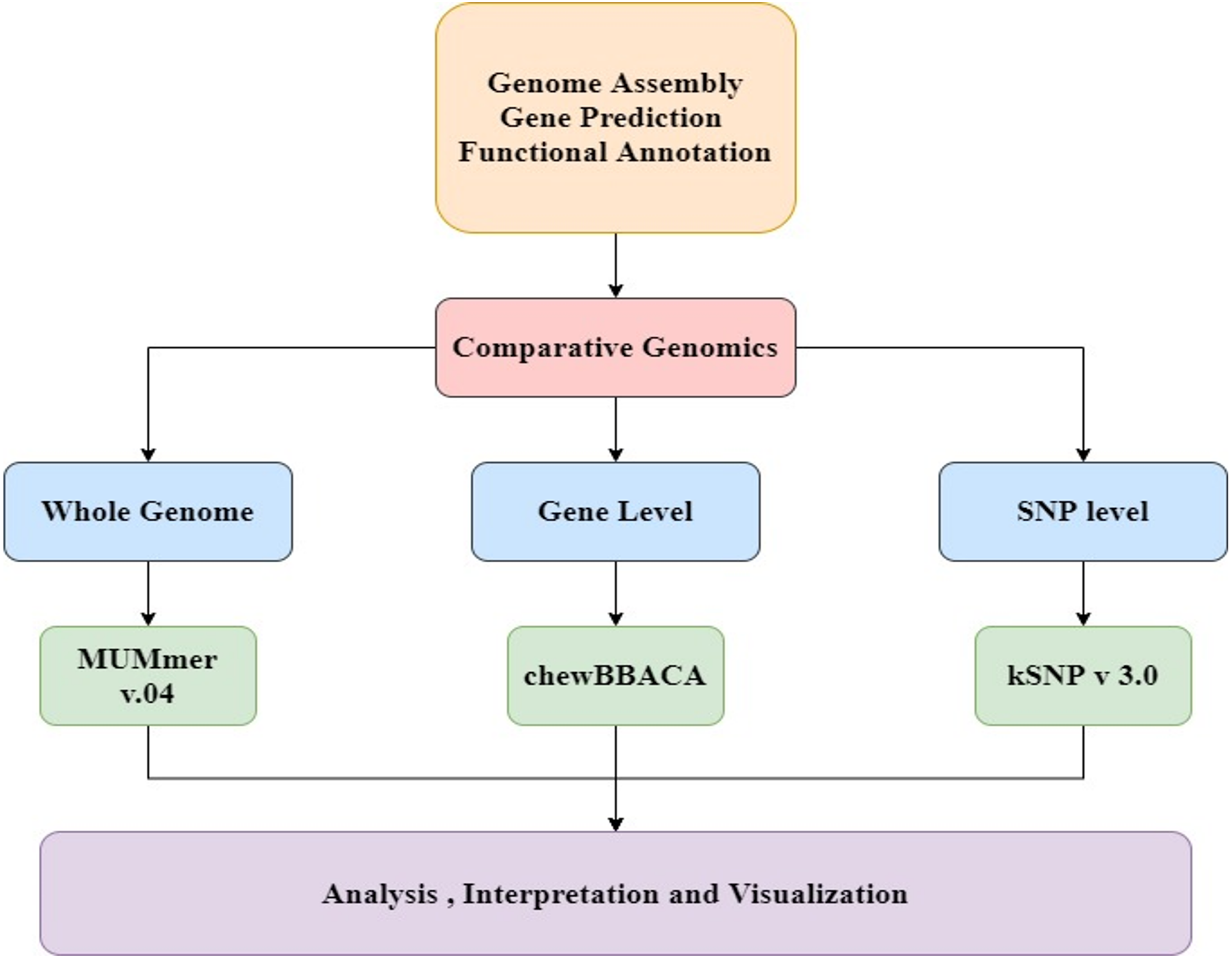
ARIBA
stringMLST
chewBBACA

Best of:

kSNP v. 3.0
BactSNP
ParSNP
RealPhy

Analysis, Interpretation and Visualization

The Final Pipeline



Genome Assembly
Gene Prediction
Functional Annotation

Comparative Genomics

Whole Genome

Gene Level

SNP level

MUMmer v.04

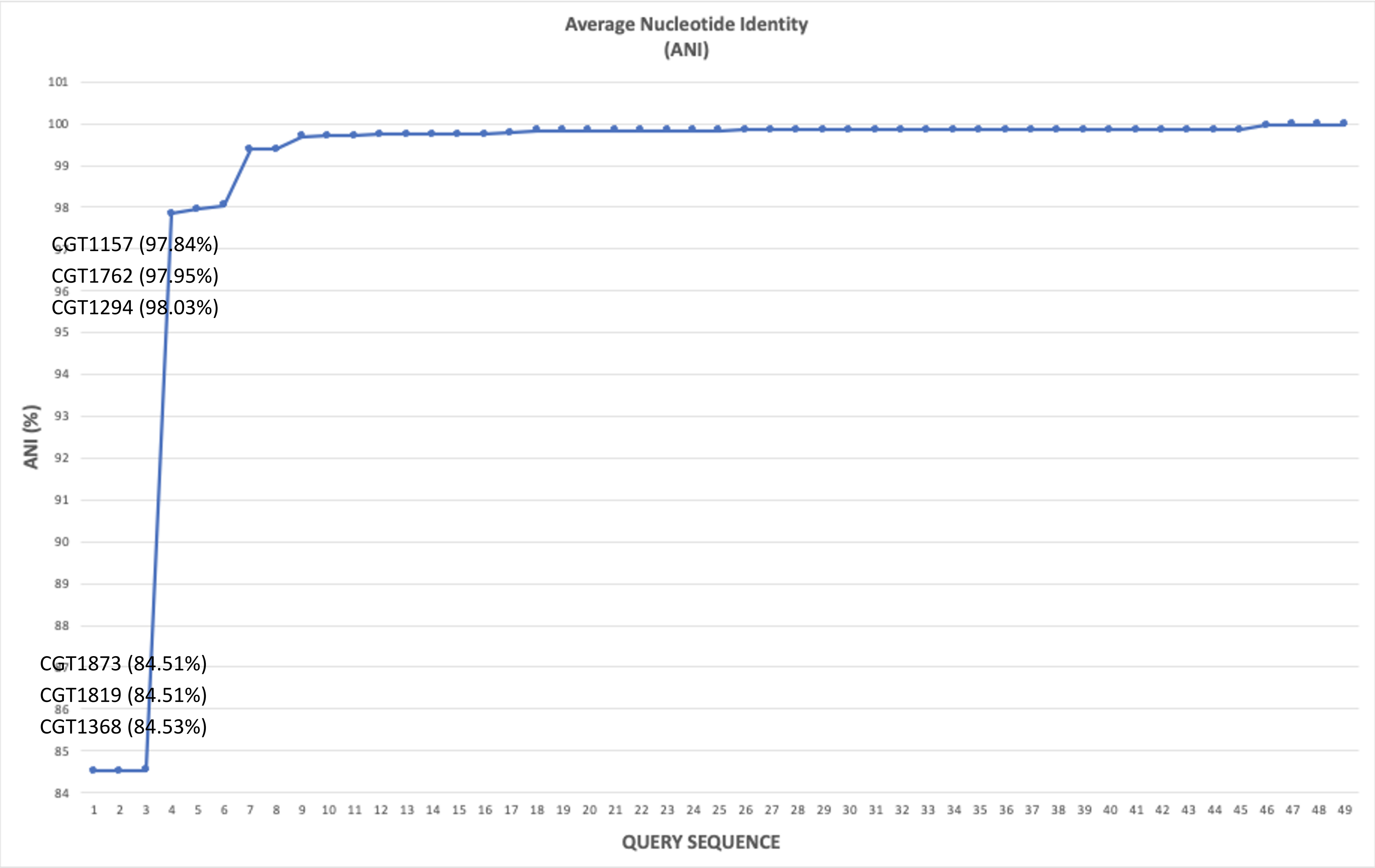
chewBBACA

kSNP v 3.0

Analysis , Interpretation and Visualization

What is the identity of the strains that cause the outbreak?

MUMMER-4.0 Result



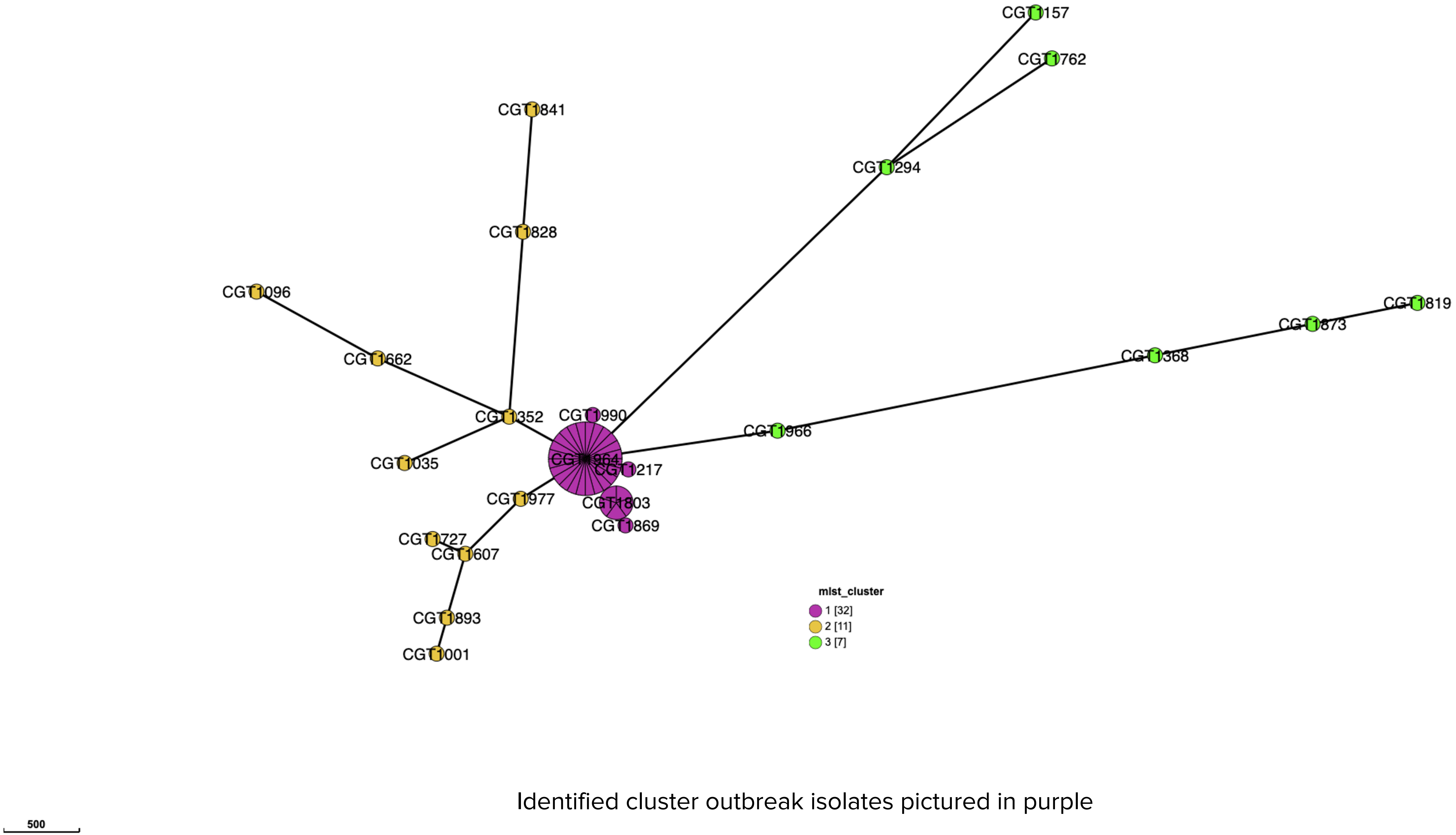
CGT1157 (97.84%)
CGT1762 (97.95%)
CGT1294 (98.03%)

CGT1873 (84.51%)
CGT1819 (84.51%)
CGT1368 (84.53%)

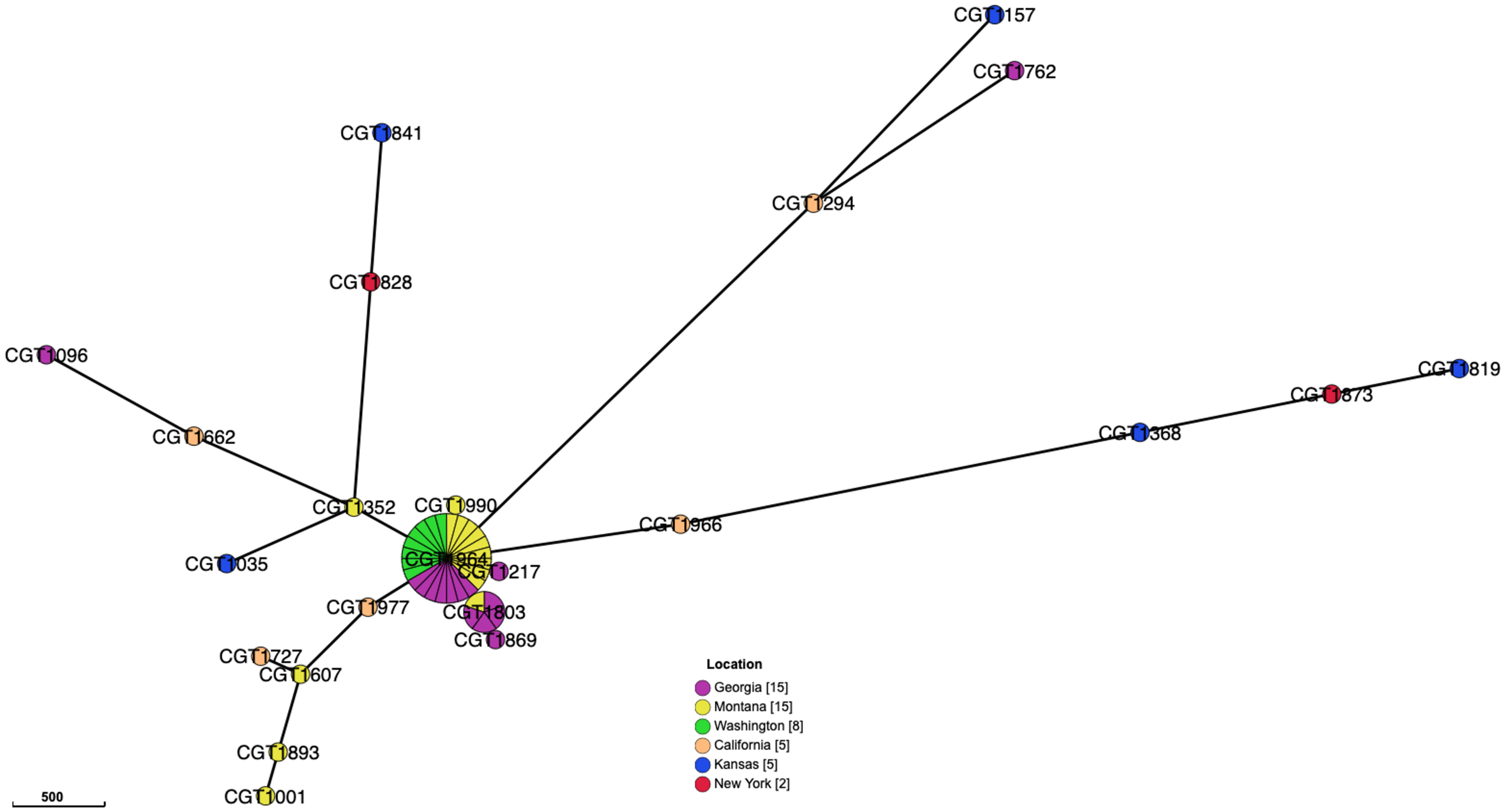
How are the isolates related to each other? How do they differ?

MLST Analysis

- Used **chewBBACA** to create a schema and do allele calling on the assembled genomes of the 50 isolates
- Initial results were visualized using Grapetree before doing deeper epidemiological analysis

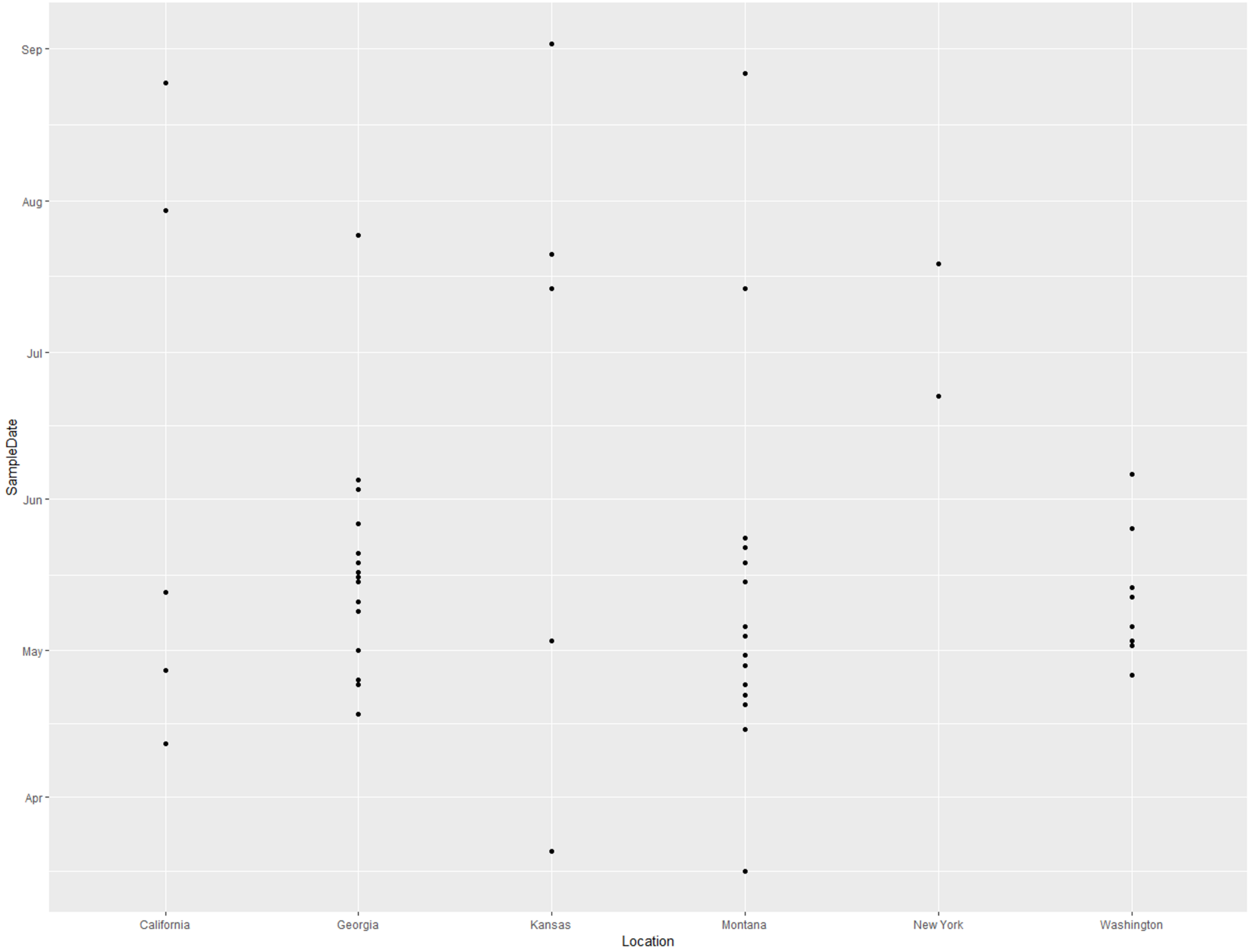


Identified cluster outbreak isolates pictured in purple


CGT1157
CGT1762
CGT1841
CGT1294
CGT1828
CGT1096
CGT1819
CGT1873
CGT1368
CGT1662
CGT1352
CGT1990
CGT1966
CGT1035
CGT1964
CGT1217
CGT1977
CGT1803
CGT1869
CGT1727
CGT1607
CGT1893
CGT1001
500
Location
Georgia [15]
Montana [15]
Washington [8]
California [5]
Kansas [5]
New York [2]

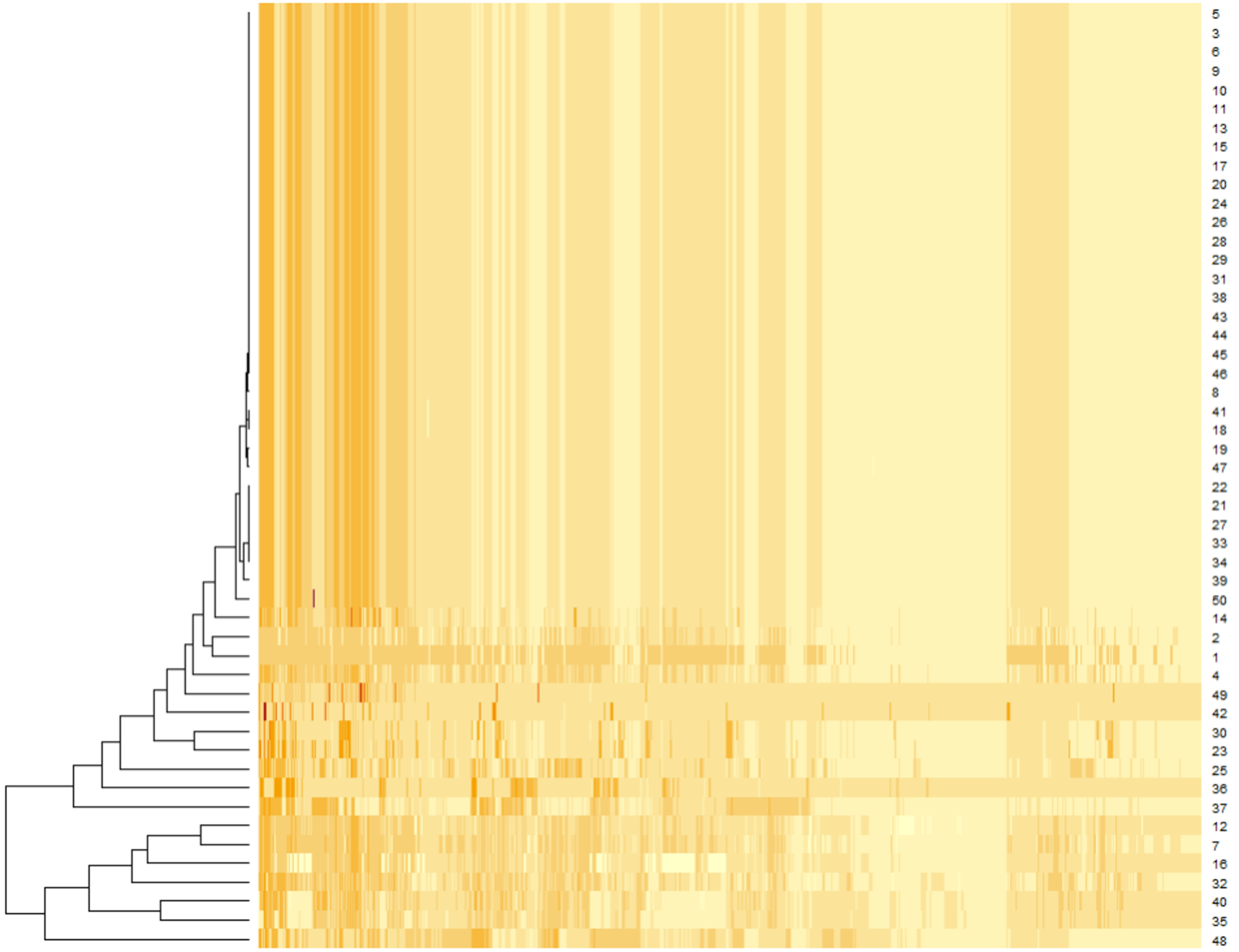
Initial Impressions

- We decided to contextualize our other results with the epidata
- To this end, we tried a few visualizations, but ended up keeping one:
- To the right:
  - X-axis: State of sample
  - Y-axis: Date of sample
- We seem to have 3 outbreaks
  - similar chronologically
  - GA, MT, WA



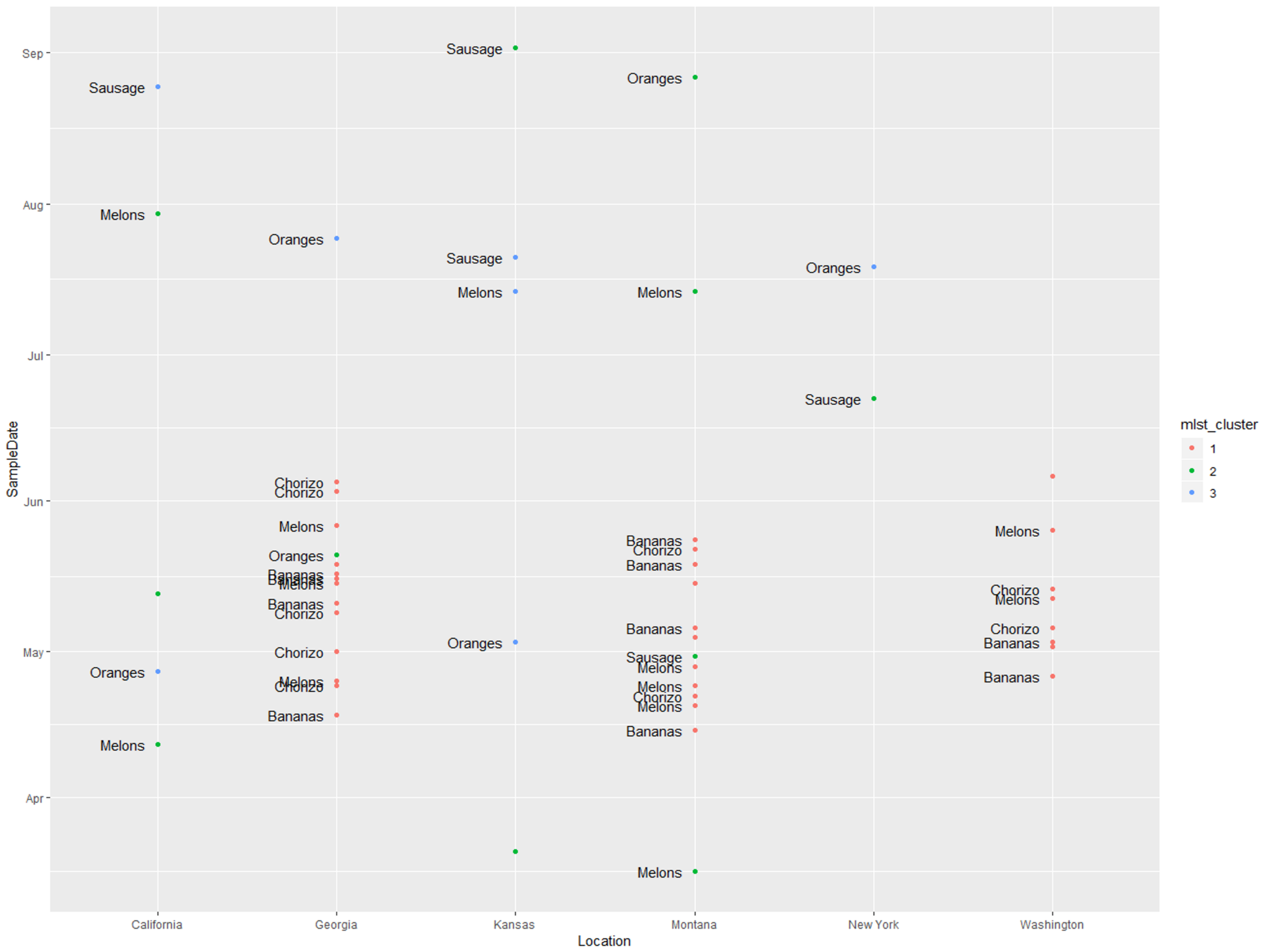
MLST results

- Several tools were tried; MLST produced clear results early on
- To the right:
  - X axis: MLST loci
  - Y axis: Samples
- Our interpretation:
  - 3 clusters:
    - Outbreak
    - Sporadic 1
    - Sporadic 2



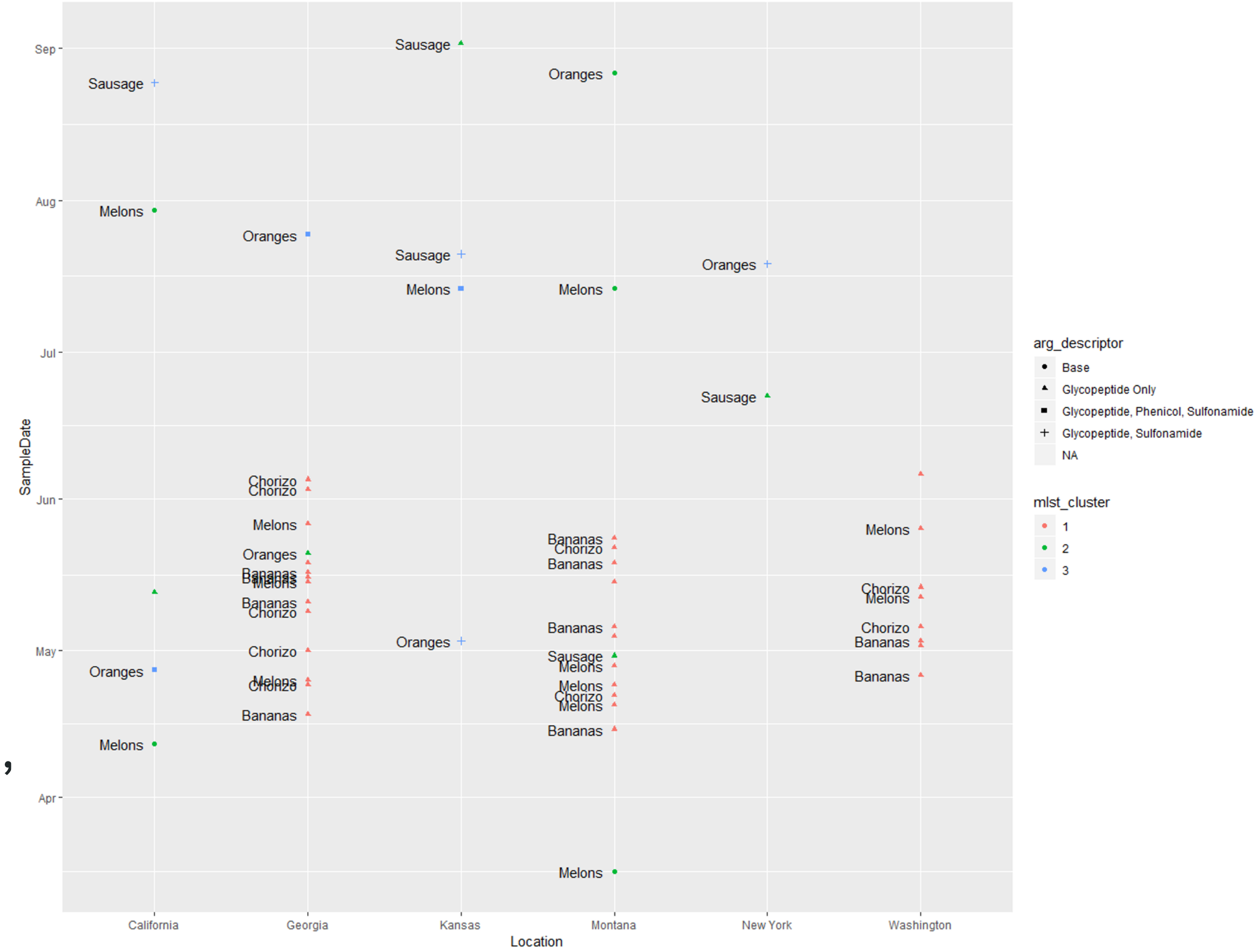
A deeper look

- We ran the appropriate data through strain analysis and incorporated functional annotation results
- MLST results perfectly supported what appeared from the epidata - an outbreak strain and perhaps a few sporadic strains
- United on 3 foods: Melons, chorizo, and bananas



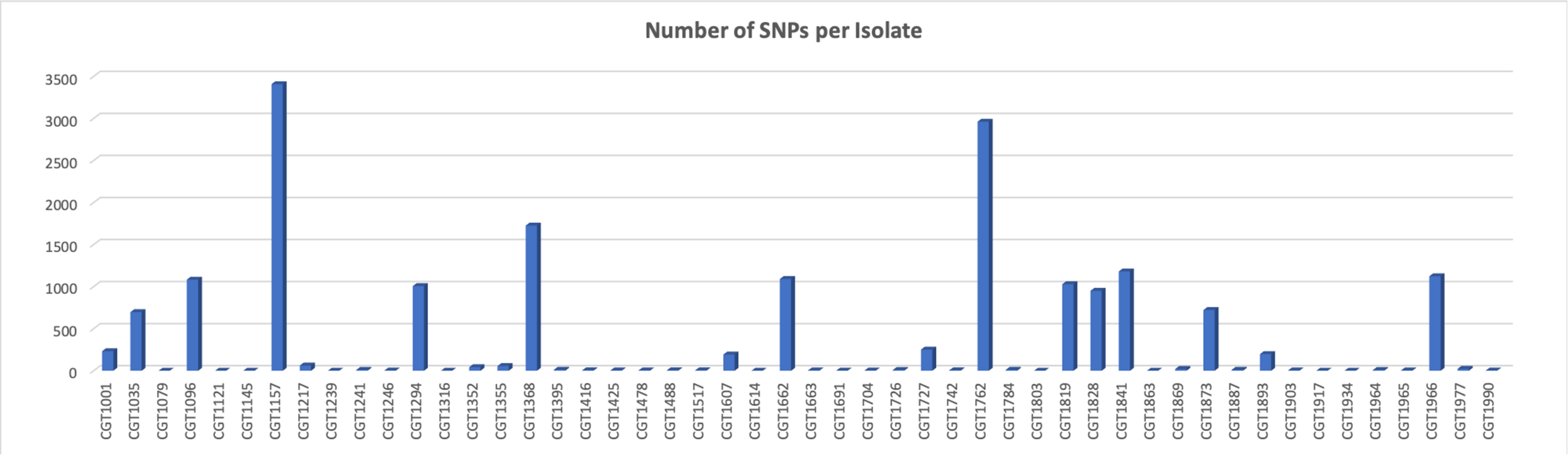
Epidata finale

- With clear strains, possessed of clear genetic relatedness, the question was whether they were treatable in a similar fashion
- Yes.
- All strains shared a base ARG set, according to deepARG
- The outbreak strain was (fortunately) identical on this basis, and was quite vulnerable.
- Phenicol and sulfonamides both work on the outbreak strain

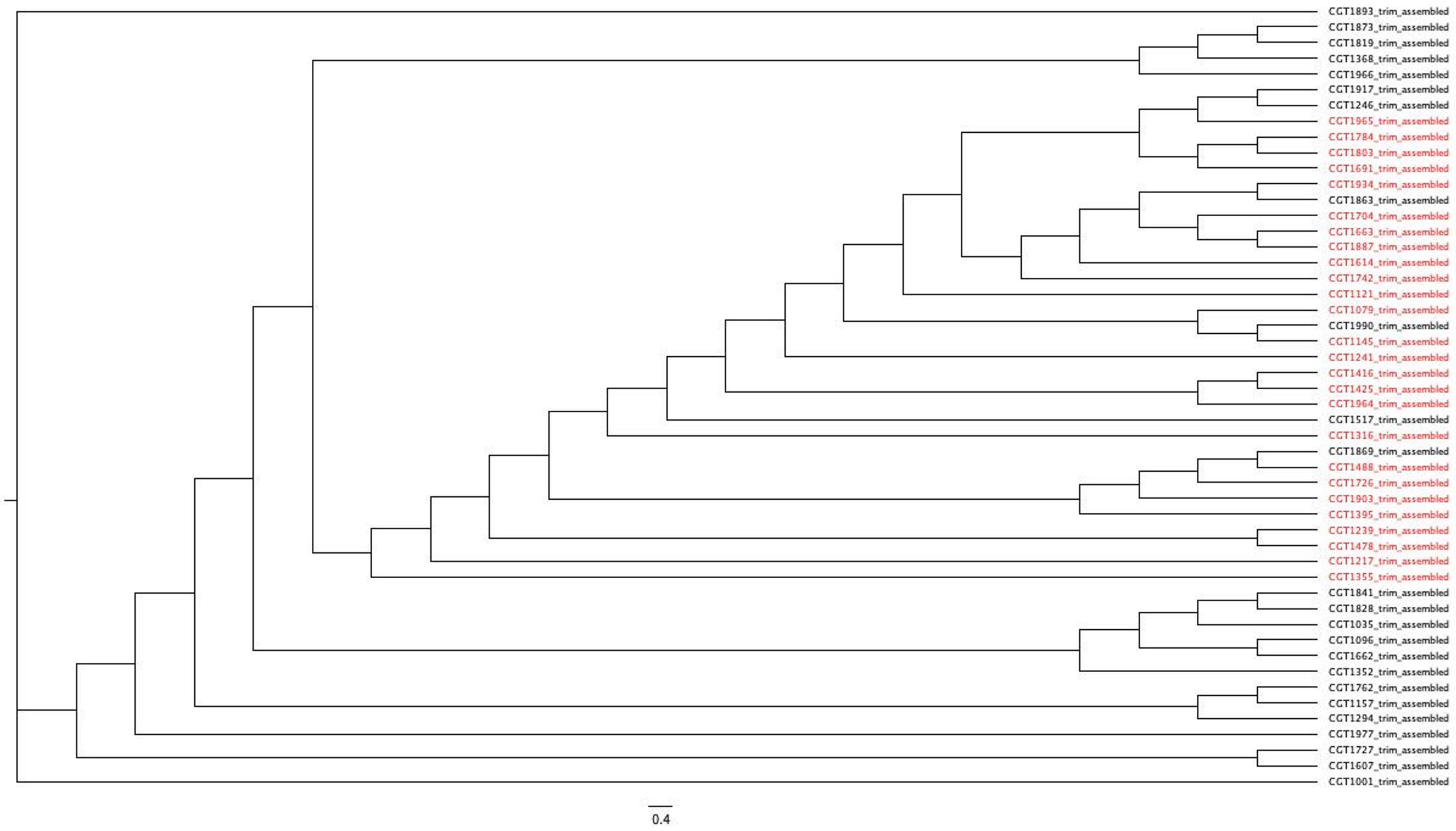


SNP Analysis

- Used **kSNP 3.0** to analyze and determine SNPs across the 50 isolates
- kSNP uses k-mer analysis and the appropriate k-mer size for our dataset was 19.
- FCK: 0.422 (measure of sequence diversity)
- built phylogenetic trees to understand the diversity among the isolates



Phylogenetic Tree estimated using Parsimony Method



Outbreak Isolates are highlighted in red

What are the virulence of the outbreak isolates?

Virulence Profile

Gene*	Allele	Length (bp)	Description
b0557 (iss)	8	294	Increased Serum Survival (ISS) Protein
ECO26_RS04705 (cif)	4	830	Effector Protein (Type III)
efa1	7	9672	Adhesin Protein
nleA	1	1221	Effector Protein

What is the recommended outbreak response and treatment?

Outbreak Response

- The outbreak strain has a relatively limited ARG profile
- Although some drugs may be able to treat all strains, inhibiting the selection of ARG response to new drugs is wise
- Our results recommend the use of an antibiotic of either the phenicol or sulfonamide class
    - resistances exist to these in the sporadic, but not the outbreak strains
- Preemptively suggest recalls of chorizo, banana, and melon from stores

Outbreak resistances

- aminoglycoside
- bacitracin
- beta-lactam
- diaminopyrimidine
- fluoroquinolone
- fosmidomycin
- macrolide
- peptide
- tetracycline

References

1. Maiden MC, Jansen van Rensburg MJ, Bray JE, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol*. 2013;11(10):728-36.
2. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, et al. (2018) MUMmer4: A fast and versatile genome alignment system. PLOS Computational Biology 14(1): e1005944. https://doi.org/10.1371/journal.pcbi.1005944
3. Perez-Losada M, Arenas M, Castro-Nallar E. Microbial sequence typing in the genomic era. Infection, Genetics and Evolution. 2018;63:346-359. http://dx.doi.org/10.1016/j.meegid.2017.09.022
4. Strockbine N, Bopp C, Fields P, Kaper J, Nataro J. 2015. *Escherichia*, *Shigella*, **and** *Salmonella*, p 685-713. *In* Jorgensen J, Pfaller M, Carroll K, Funke G, Landry M, Richter S, Warnock D (ed), *Manual of Clinical Microbiology, Eleventh Edition*. ASM Press, Washington, DC. doi: 10.1128/9781555817381.ch37
5. Sultan, I., Rahman, S., Jan, A. T., Siddiqui, M. T., Mondal, A. H., & Haq, Q. M. R. (2018). Antibiotics, Resistome and Resistance Mechanisms: A Bacterial Perspective. *Frontiers in Microbiology, 9*(2066). doi:10.3389/fmicb.2018.02066
6. Trees E**,** Rota P, Maccannell D, Gerner-smidt P**.**. Molecular Epidemiology, p 131-159. *In* Jorgensen J, Pfaller M, Carroll K, Funke G, Landry M, Richter S, Warnock D (ed), *Manual of Clinical Microbiology, Eleventh Edition*. ASM Press, Washington, DC. 2015. doi: 10.1128/9781555817381.ch10
7. Silva M, Machado M, Silva D, Rossi M, Moran-Gilad J, Santos S, Ramirez M, Carriço J. 15/03/2018. M Gen 4(3): doi:10.1099/mgen.0.000166
8. Z Zhou, NF Alikhan, MJ Sergeant, N Luhmann, C Vaz, AP Francisco, JA Carrico, M Achtman (2018) "GrapeTree: Visualization of core genomic relationships among 100,000 bacterial pathogens", Genome Res; doi: https://doi.org/10.1101/gr.232397.117