



COMPARATIVE GENOMICS RESULTS – TEAM 2

BY: KARA LEE, KRISTINE LACEK, COURTNEY ASTORE, UJANI HAZRA, JIANSHU ZHAO



PRESENTATION OUTLINE

- Overview & objective
- Summary of Epi Data
- Final comparative genomics pipeline
 - ANI results
 - MLST results
 - SNP-Based results
- AMR Profile
- Outbreak Source and CDC Recommendations

OBJECTIVES

- Determine the similarity and betweenness of our 50 isolates
- Construct phylogeny demonstrating which isolates are related and which differ
- Determine the source of the outbreak
- Map antibiotic resistance features of outbreak isolates
- Compile recommendations for outbreak response and treatment

FINAL PIPELINE

Data

- 50 Assembled Genomes
- Predicted and Annotated Genes
- Epidemiological data

Comparison Methods

- MLST
- SNP-Based
- ANI

Data Consolidation

- Phylogeny generation
- Virulence Profile
- AMR Features

CDC Recommendations

- Preventative measures
- Outbreak response
- Treatment strategy

OBJECTIVES FOR EACH APPROACH

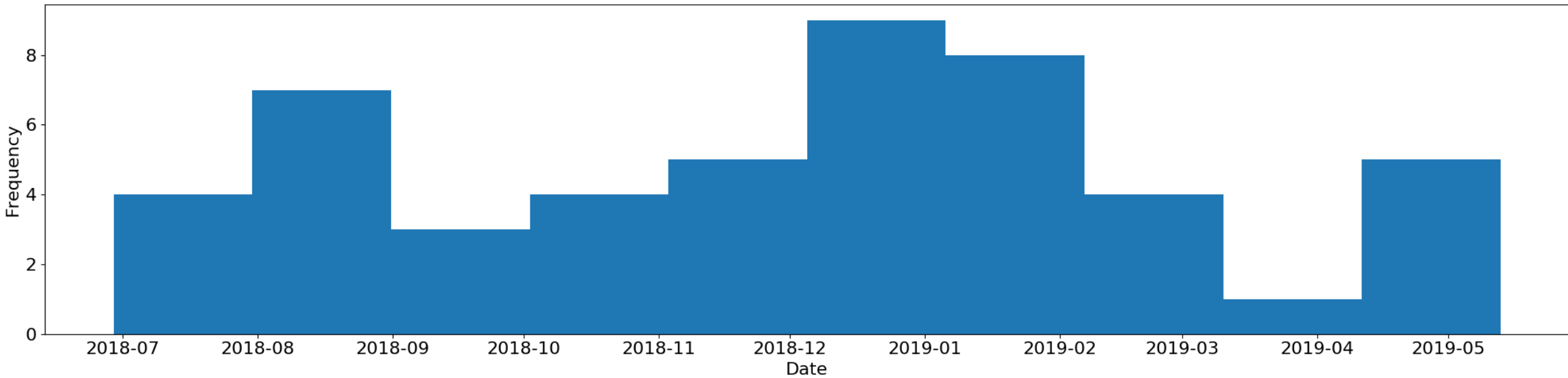
Approach	Isolate similarity assessment	Goal of Approach	Output
ANI	Species-specific	FastANI (Mashmap-based)	Species identity value
MLST	Based on allelic variations	Relatedness based on allelic variations	Sequence type
SNP	Based on SNPs	Determine relatedness from base-by-base analysis	SNP counts across isolates, phylogenetic trees



EPIDEMIOLOGICAL DATA

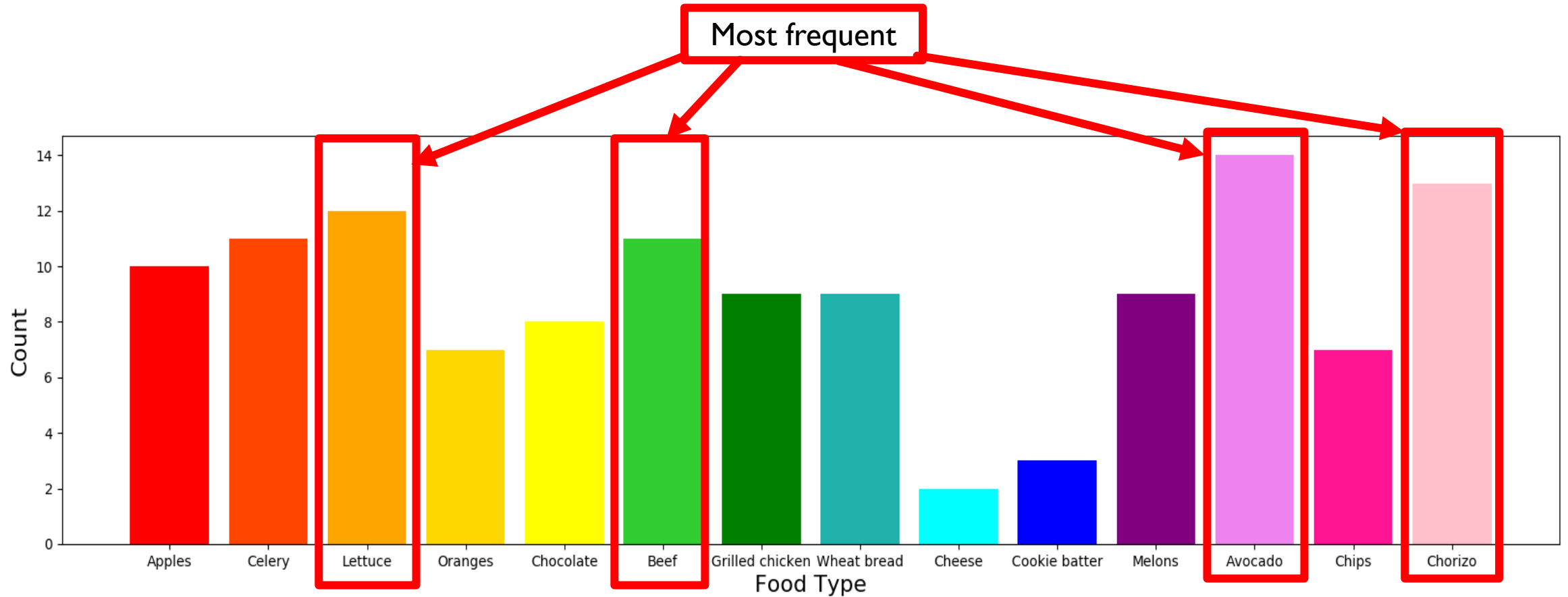
GEOGRAPHY, TIME, FOODS

NUMBER OF CASES REPORTED OVER TIME

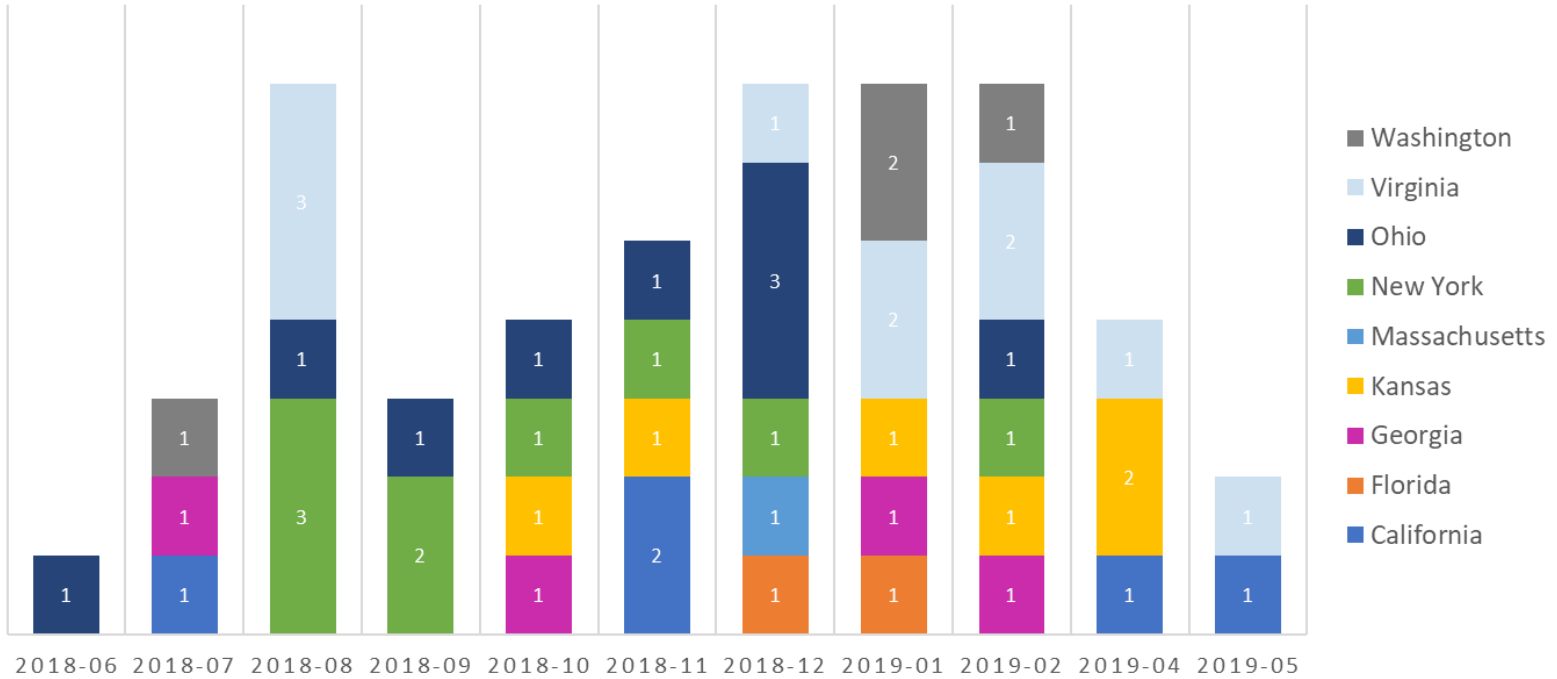
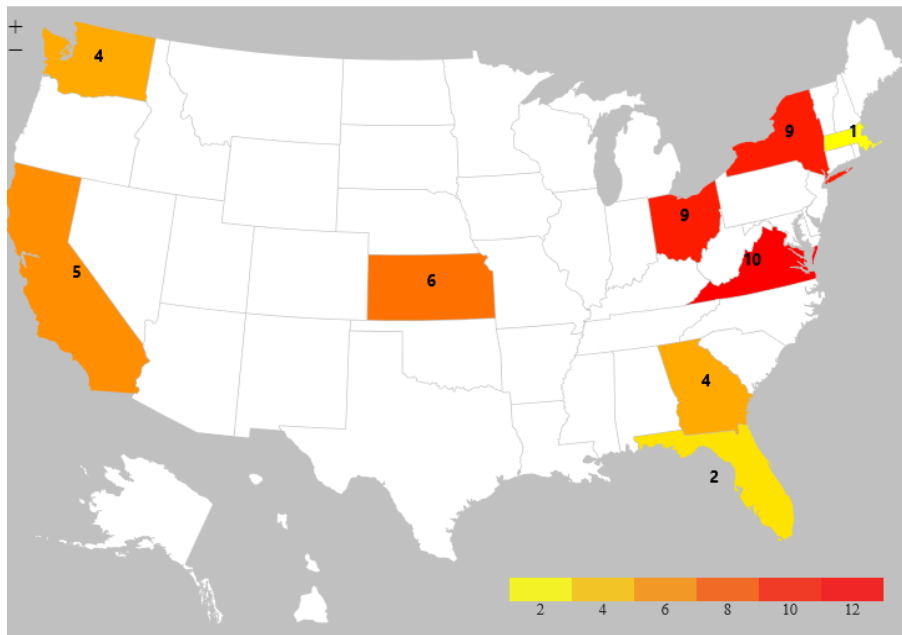


← Co-occurring *C. jejuni* outbreak linked to pet store puppies →

FREQUENCY OF FOOD TYPES ACROSS ISOLATES



GEOGRAPHIC VISUALIZATION OF CASES & TIME

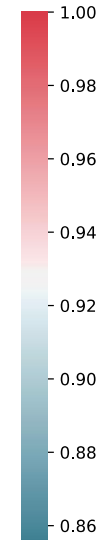
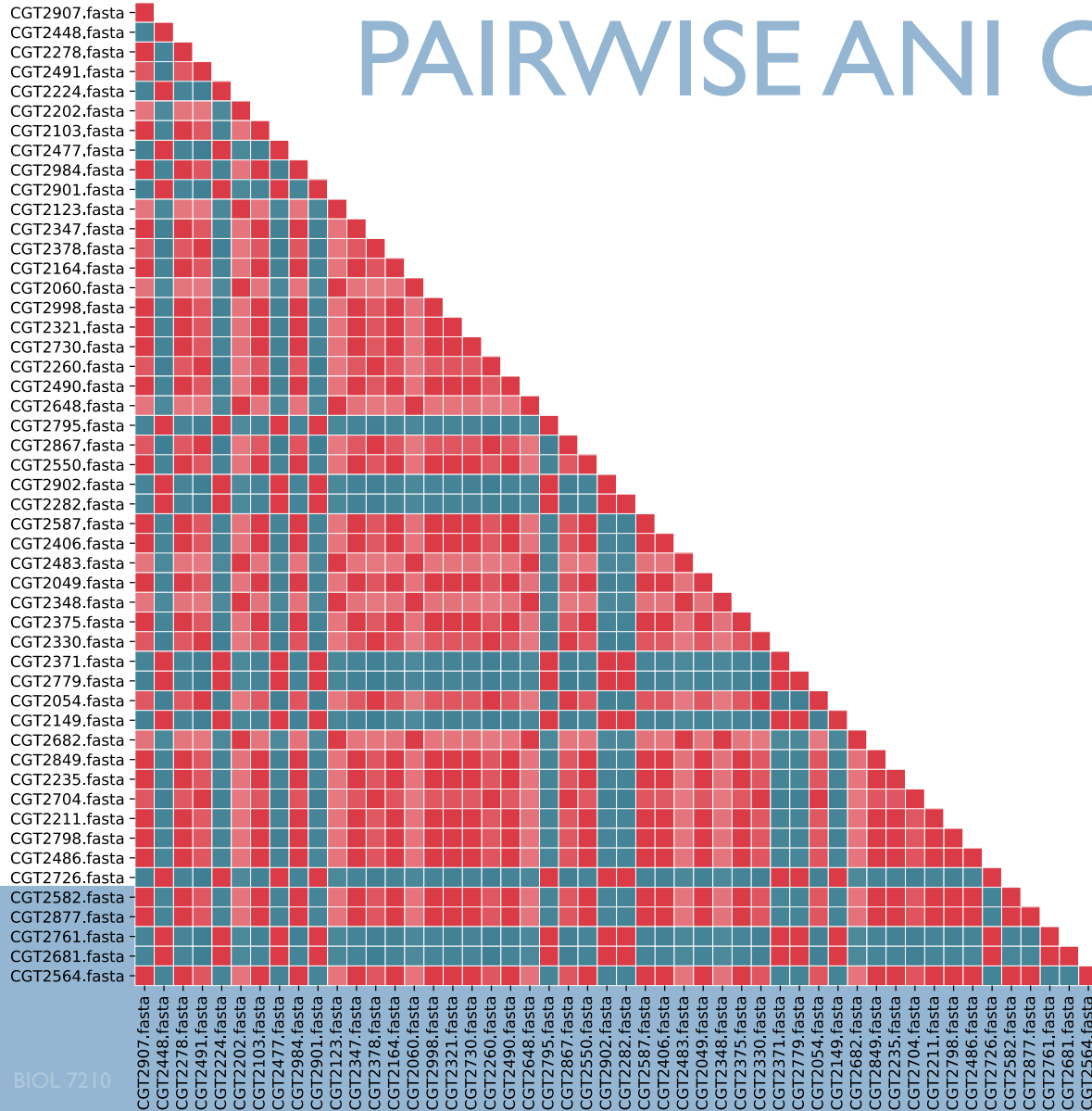




ANI APPROACH RESULTS

TOOLS: FASTANI

PAIRWISE ANI CALCULATION



- 80% ANI
- The same genus, Campylobacter

DATABASE FOR IDENTIFICATION OF SPECIES, A SUBSET OF CAMPYLOBACTER SPECIES GENOMES IN NCBI REFSEQ

<i>C. avium</i>	<i>C. insulaenigrae</i>
<i>C. butzleri</i>	<i>C. jejuni</i>
<i>C. canadensis</i>	<i>C. lanienae</i>
<i>C. cinaedi</i>	<i>C. lari</i>
<i>C. coli</i>	<i>C. mucosalis</i>
<i>C. concisus</i>	<i>C. mustelae</i>
<i>C. corcagiensis</i>	<i>C. nitrofigilis</i>
<i>C. cryaerophilus</i>	<i>C. peloridis</i>
<i>C. cuniculorum</i>	<i>C. pylori</i>
<i>C. curvus</i>	<i>C. rectus</i>
<i>C. fennelliae</i>	<i>C. showae</i>
<i>C. fetus</i>	<i>C. sputorum</i>
<i>C. gracilis</i>	<i>C. subantarcticus</i>
<i>C. helveticus</i>	<i>C. upsaliensis</i>
<i>C. hepaticus</i>	<i>C. ureolyticus</i>
<i>C. hominis</i>	<i>C. volucris</i>
<i>C. hyoilei</i>	
<i>C. hyointestinalis</i>	

Reference genome database (~2000 genomes, a representative subset of all **Campylobacter** genomes, included all the major species)

Find the highest ANI value in the genome to define the species information of query genome

Campylobacter jejuni
Campylobacter coli



MLST APPROACH RESULTS

MLST

CHEWBBACA

MLST TOOLS

1. **mlst**

- 7-gene mlst (aspA, glnA, gltA, glyA, pgm, tkt, uncA)
- 2 minutes 45 seconds for 50 samples (4:21 using stringMLST) using assembled genome
- Takes *de novo* assemblies as input on the command line and uses BLASTN
- Do not need to build db; bundle of all available databases in the software repository

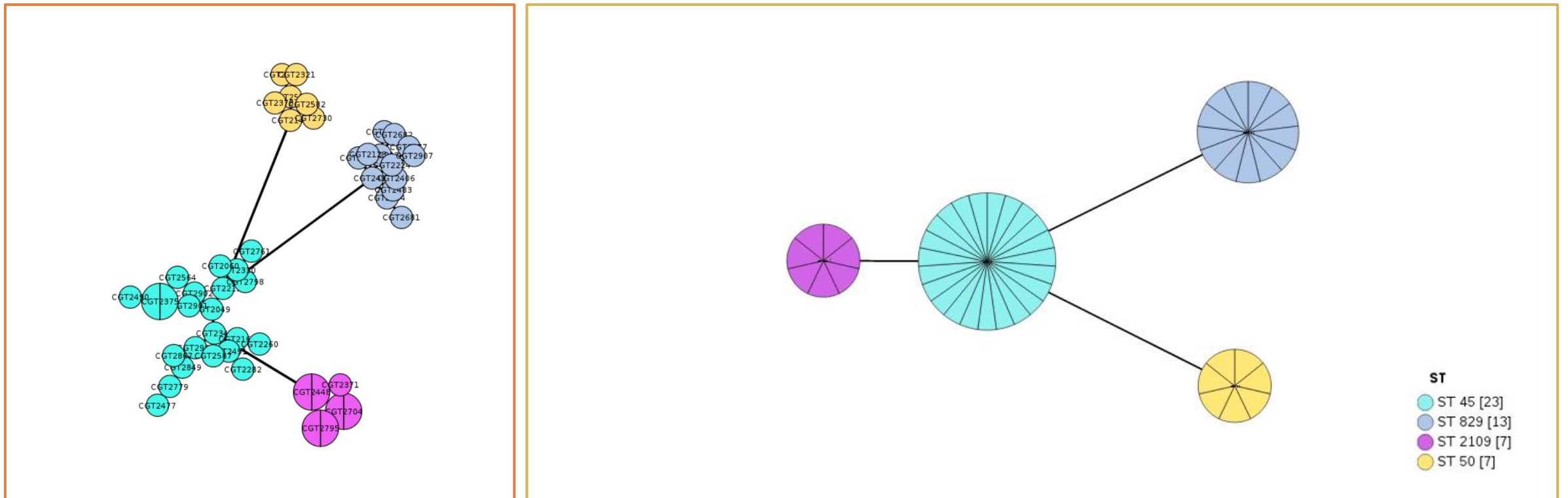
```
mlst --scheme campylobacter ./ * > /home/projects/group-b/comparative_genomics/test_mlst/mlst_50/output.txt
```

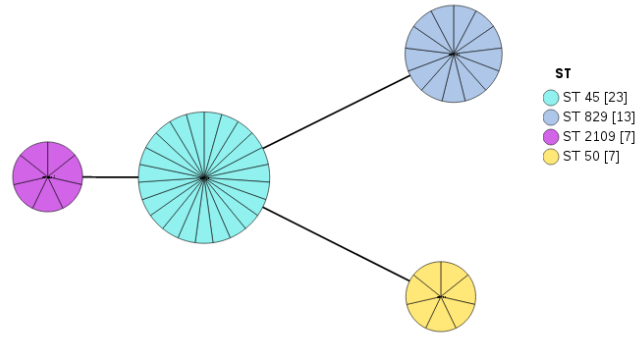
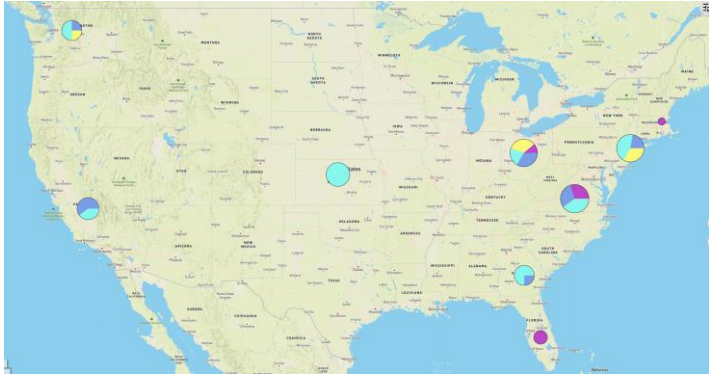
2. **ChewBBACA**

- wgMLST or cgMLST
- Schema creation and allele calls on complete or draft genomes resulting from *de novo* assemblies
- Allele calling algorithm based on Blast Score Ratio that can be run in multiprocessor settings

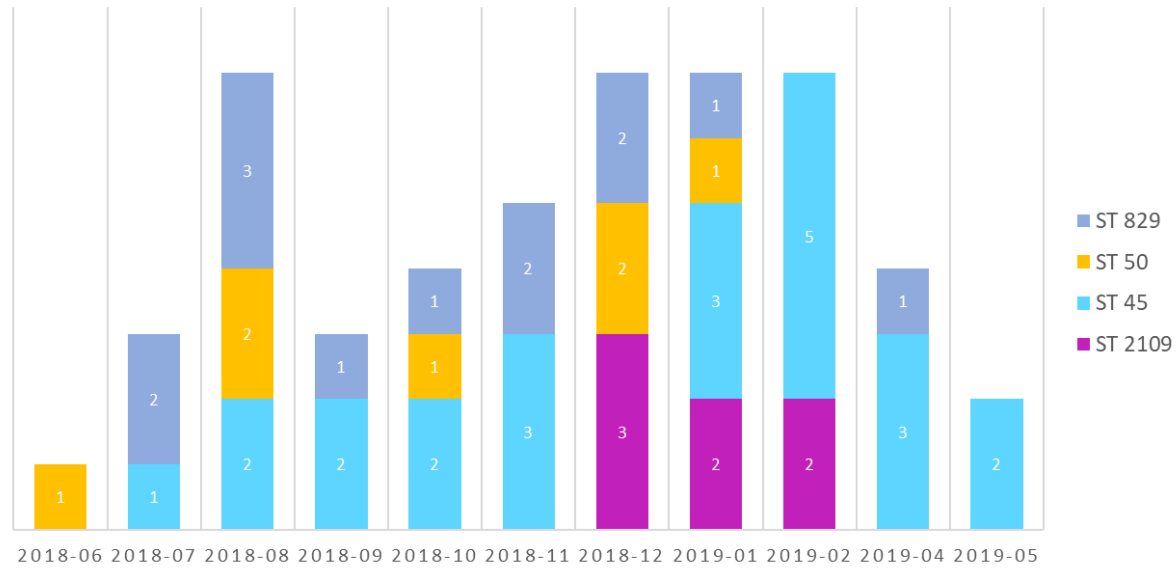
```
chewBBACA.py CreateSchema -i /home/projects/group-b/gene_prediction/contigs_file/ -o schema --cpu 4  
chewBBACA.py ExtractCgMLST -i results_alleles.tsv -o formatted
```

CHEWBBACA (CGMLST) VS. MLST RESULTS



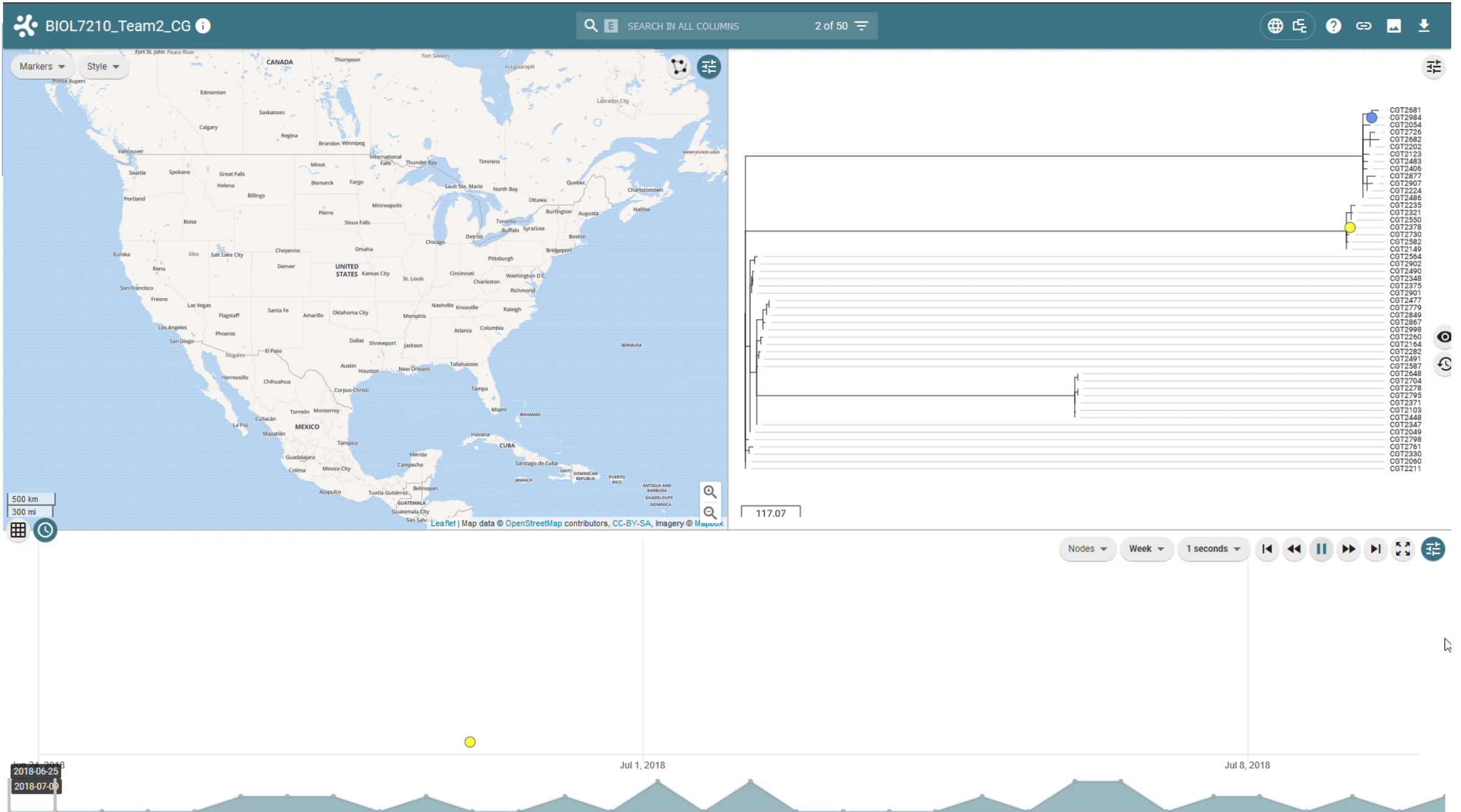


ST VS. LOCATION ST VS TIME



MICROREACT TIME-SERIES ANIMATION

<https://microreact.org/project/TYUxIN5L2>



ST VS. LOCATION

Virginia

New York

Massachusetts

Ohio

Georgia

Florida

Kansas

California

Washington



10 ITEMS

- 1 × CGT2049
- 1 × CGT2202
- 1 × CGT2260
- 1 × CGT2278
- 1 × CGT2371
- 1 × CGT2406
- 1 × CGT2486
- 1 × CGT2490
- 1 × CGT2564
- 1 × CGT2648

9 ITEMS

- 1 × CGT2149
- 1 × CGT2235
- 1 × CGT2348
- 1 × CGT2483
- 1 × CGT2491
- 1 × CGT2550
- 1 × CGT2587
- 1 × CGT2682
- 1 × CGT2867

1 ITEM

- 1 × CGT2448

9 ITEMS

- 1 × CGT2054
- 1 × CGT2060
- 1 × CGT2224
- 1 × CGT2378
- 1 × CGT2582
- 1 × CGT2681
- 1 × CGT2704
- 1 × CGT2730
- 1 × CGT2779

4 ITEMS

- 1 × CGT2123
- 1 × CGT2164
- 1 × CGT2375
- 1 × CGT2901

2 ITEMS

- 1 × CGT2103
- 1 × CGT2795

6 ITEMS

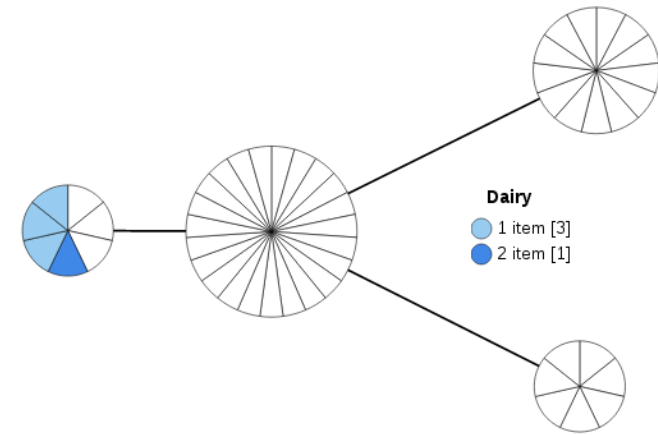
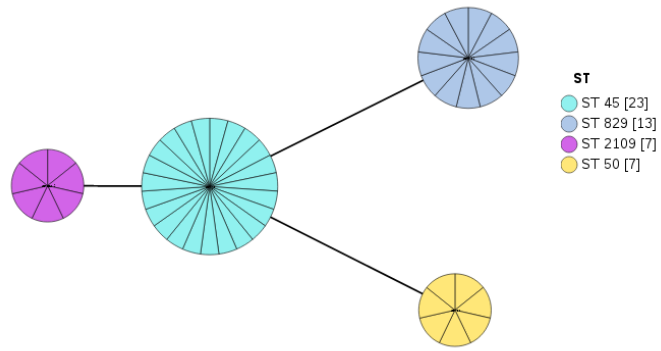
- 1 × CGT2282
- 1 × CGT2347
- 1 × CGT2761
- 1 × CGT2798
- 1 × CGT2849
- 1 × CGT2902

5 ITEMS

- 1 × CGT2211
- 1 × CGT2330
- 1 × CGT2726
- 1 × CGT2877
- 1 × CGT2907

4 ITEMS

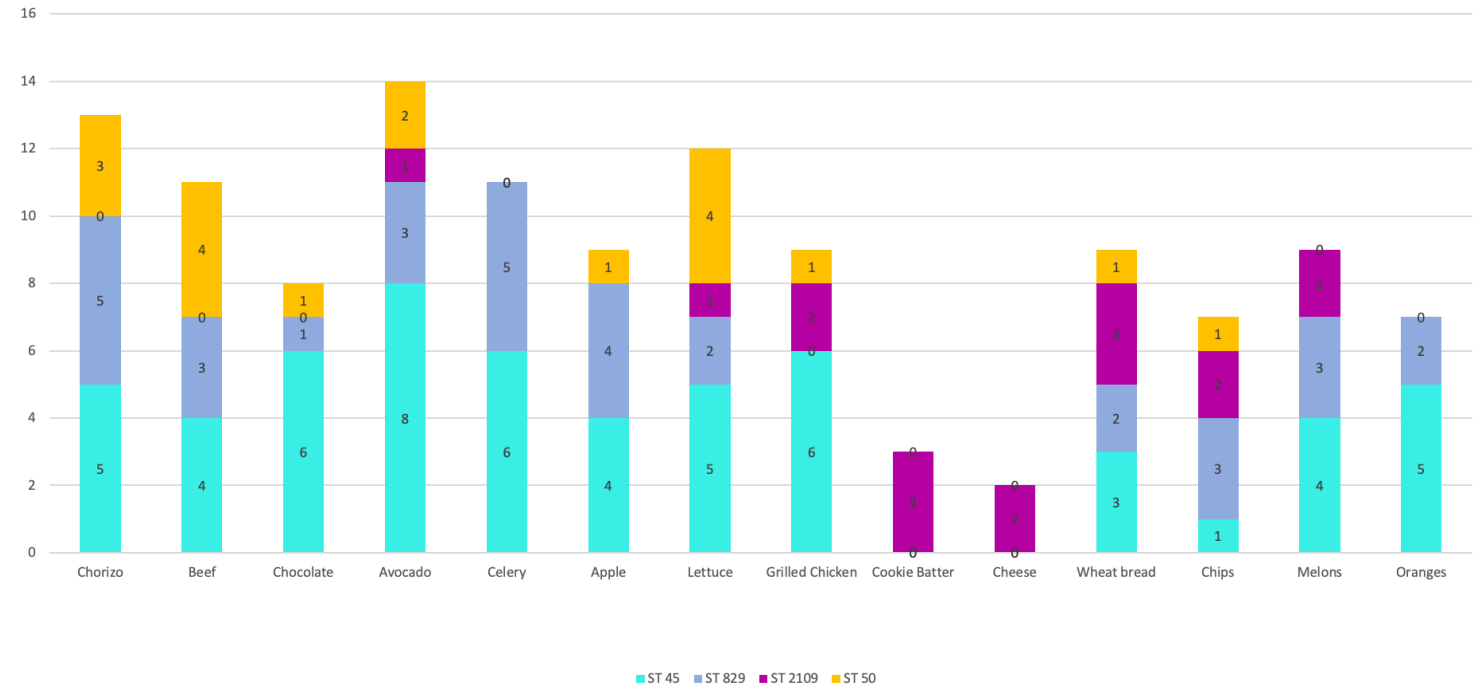
- 1 × CGT2321
- 1 × CGT2477
- 1 × CGT2984
- 1 × CGT2998



ST VS. FOOD

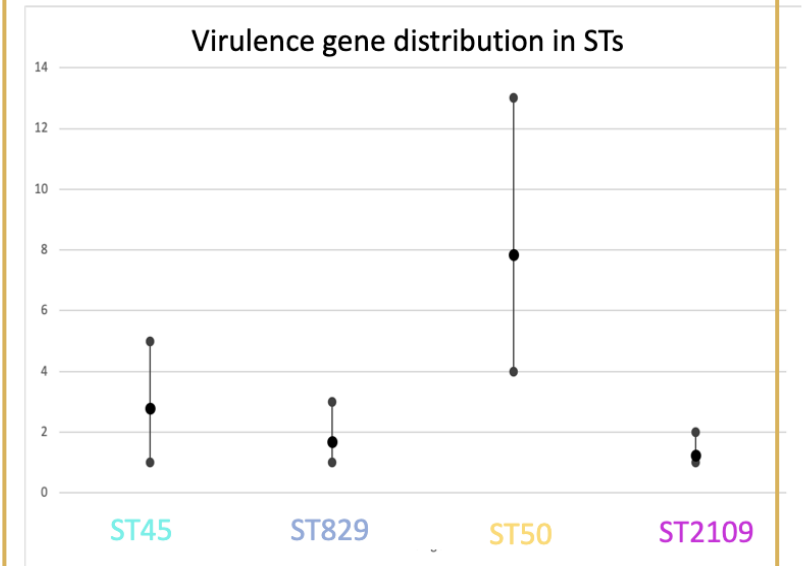
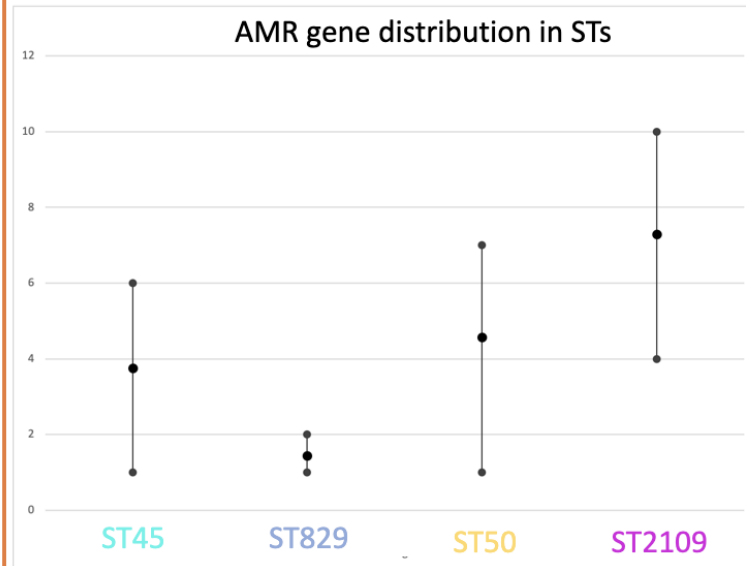
- ST2109 : Dairy products like cookie batter and cheese were only linked to this ST. The source might be linked to a dairy facility in the East coast
- ST45, ST829, ST50: The patients have consumed a combination of beef, chorizo, grilled chicken, lettuce, avocado. It might be linked to country-wide fast-food establishments,

Sequence Type vs Food



AMR AND VIRULENCE PROFILE OF ST

- ST2109 has high antibiotic resistance. This is also the only ST containing plasmids.
- ST50 is the most virulent strain with the second highest antibiotic resistance
- ST829 has low antibiotic resistance and virulence. All samples in this ST are *C.coli*





SNP-BASED APPROACH RESULTS

PARSNP AND KSNP

SNP OVERVIEW

- **What are SNPs?**
- A DNA sequence variation that occurs at a single position in the genome
- Prevalence of each variation > 1%

- **SNP Typing objective:**
 - Determine number of nucleotide differences between isolates
 - Measure genetic distance
 - Construct phylogenetic trees based on SNPs to study genetic and evolutionary forces of our isolates

KSNP3.1 : DETERMINING K

- Kchooser
 - Determines optimal k from 13 to 31 to maximize unique median length sequence
 - Results plateau at 95.8%
 - Reran with adjusted cutoff
- Maximal number of kmers are unique when k=25
- 406 of 963 unique kmers are core
- 96.6% uniqueness at k=25

```
initial value of k is 13.
When k is 13 0.807262916736116 of the kmers from the median length sequence are unique
When k is 15 0.925800944426724 of the kmers from the median length sequence are unique
When k is 17 0.952749944946047 of the kmers from the median length sequence are unique
When k is 19 0.957146407472129 of the kmers from the median length sequence are unique
When k is 21 0.957963860651176 of the kmers from the median length sequence are unique
When k is 23 0.958278716079754 of the kmers from the median length sequence are unique
When k is 25 0.958516575132611 of the kmers from the median length sequence are unique
When k is 27 0.958726842822602 of the kmers from the median length sequence are unique
When k is 29 0.958915168782195 of the kmers from the median length sequence are unique
(base) [klacek3@biogenome2020 kSNP3]$ less Kchooser.report
(base) [klacek3@biogenome2020 kSNP3]$ ./Kchooser ../CJ.fasta .958
Reading the input file...
There are 50 genomes.
Used 0 seconds so far.

Determining the optimum k...
The optimum value of K is 25.

Used 57 seconds so far.

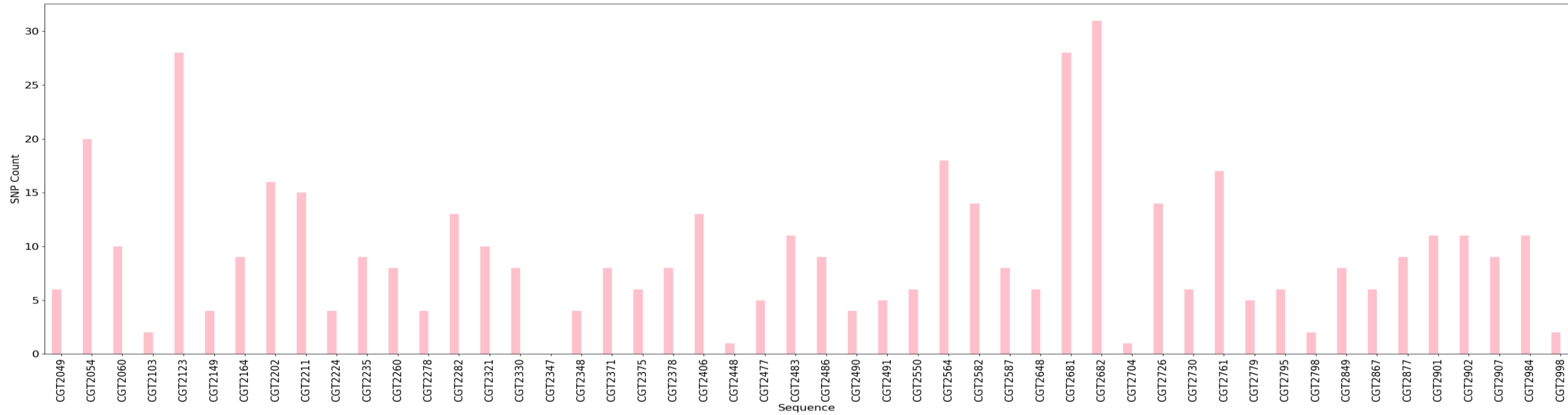
Calculating the fraction of kmers that are core...
From a sample of 963 unique kmers 494 are core kmers.
Used 406 seconds so far.

Done. Results are in the file named Kchooser.report.
```

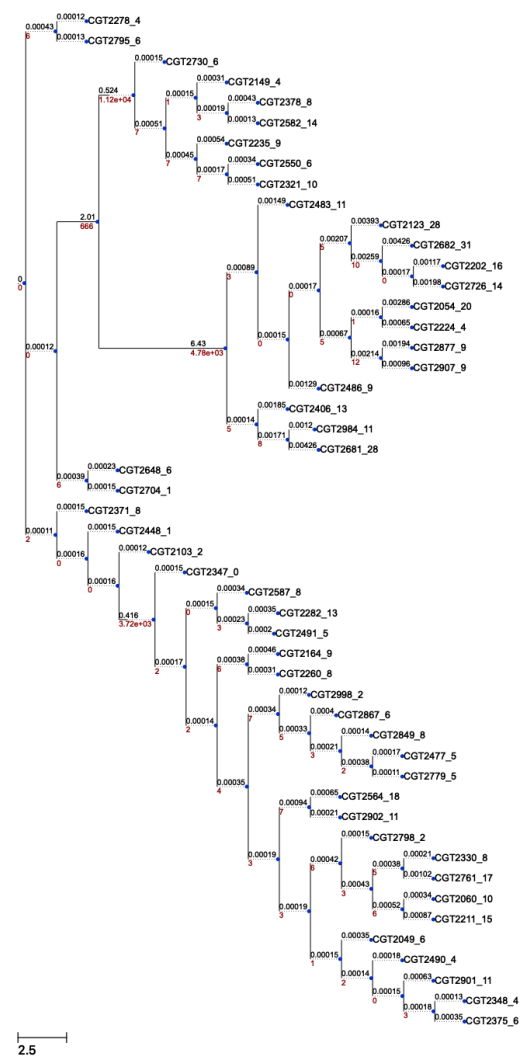
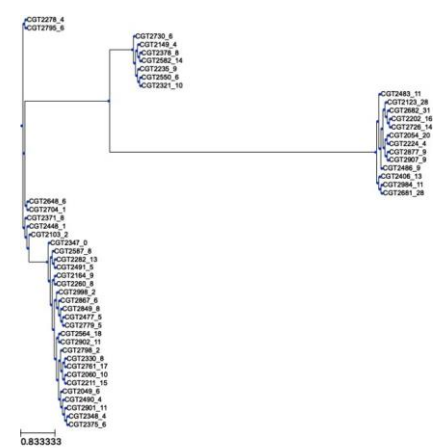
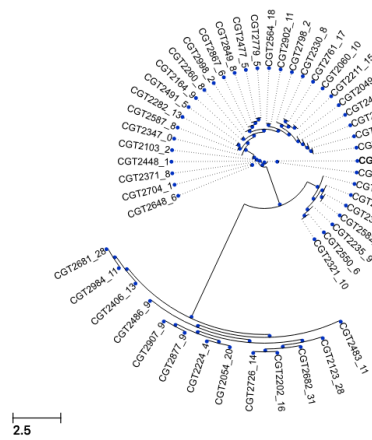
KSNP3.1 : CREATE SNP PHYLOGENIES

- KSNP3.1
 - The optimal k-value of 25 determined by KChooser was used
- Command:
 - `kSNP3 -in <input_file> -k 25outdir <output_directory> -ML`
 - Input file format: `<path/to/isolate_n.fasta> \t <isolate_ID>`
- Output files used:
 - tip_SNP_counts.ML :
 - Contains number of SNPs per isolate
 - Tree_AlleleCounts.ML.tre
 - Contains information for creating a phylogeny

KSNP RESULTS – K-VALUE = 25 (OPTIMAL K-VALUE)



Software	k-value	Number of SNPs	Min	Max	Median	Mean
BIOL 7210 kSNP	25	21,126	0	31	8	9.38 ²⁵

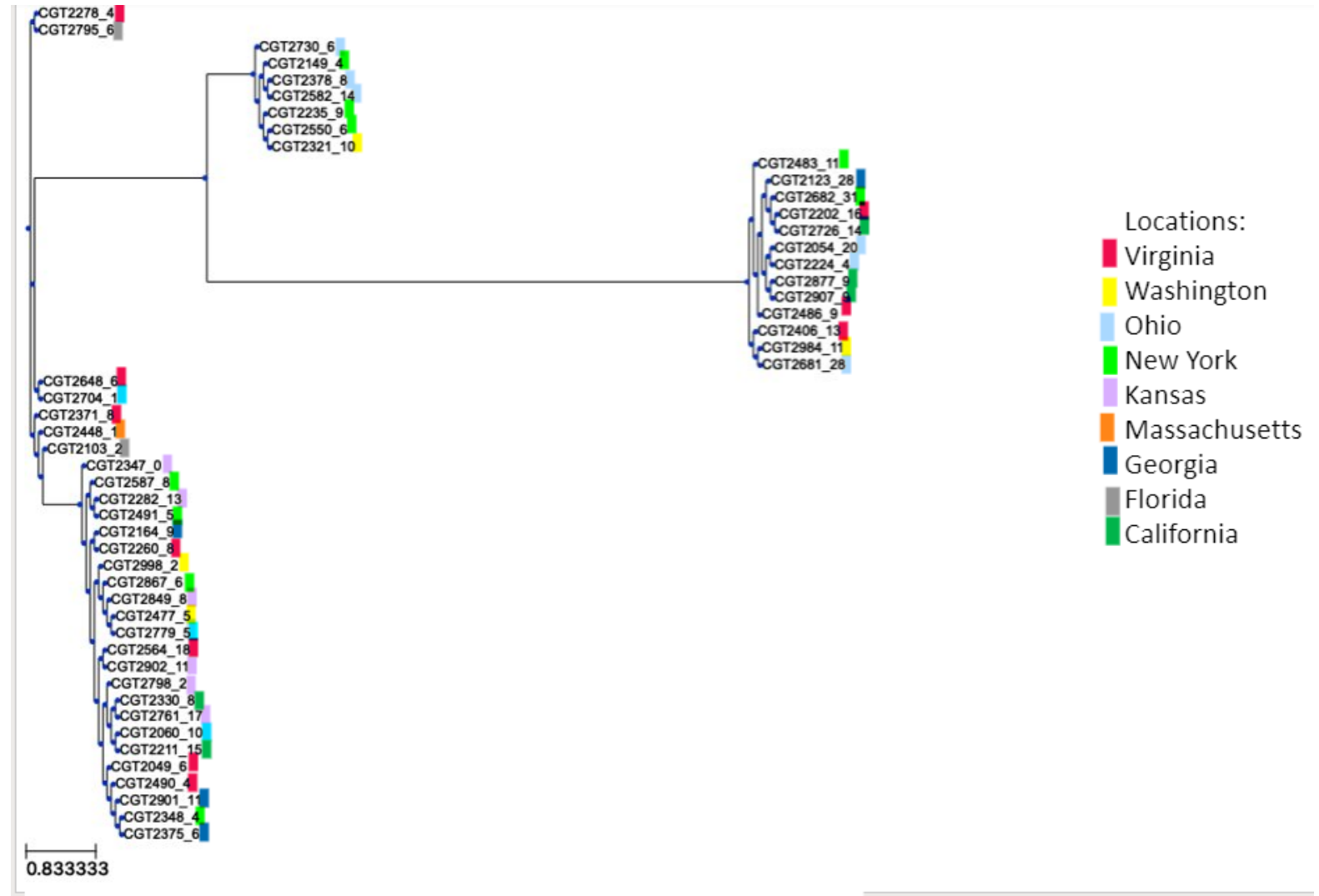


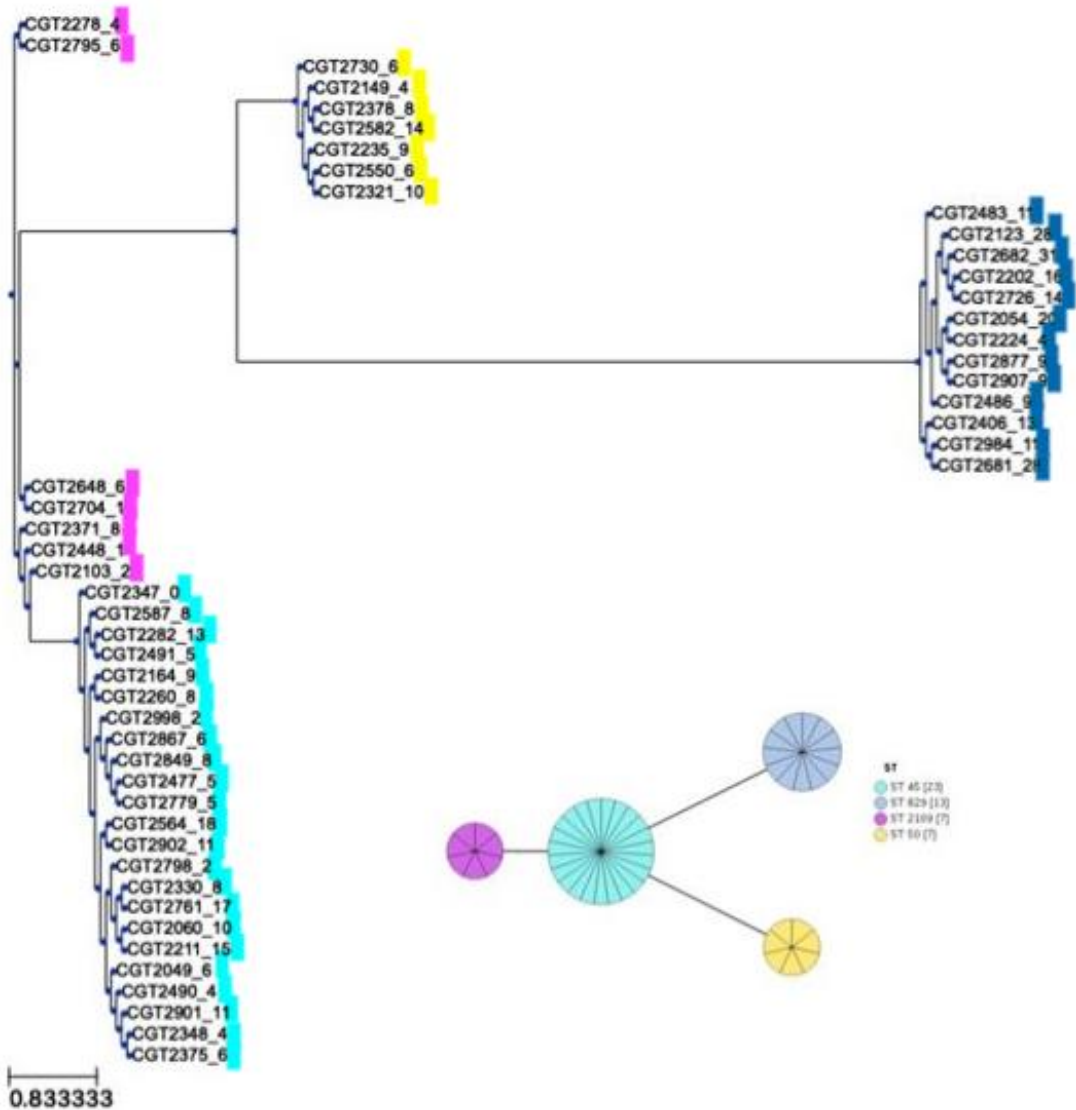
PHYLOGENY CREATION

- Ete3 python package
- Multiple types of phylogenetic trees can be generated
- Seamless usage in webservice

SNP RESULTS + LOCATION

- Little Convergence with Location
- Of 15 sister taxa (pairs of maximal relatedness), only 3 are from same state





SNP-BASED
APPROACH
CONVERGES
WITH MLST



PLASMID RESULTS AND INTERPRETING FUNCTIONAL ANNOTATION

AMR PROFILE

AMR Gene Detected	Resistant to Drugs/Class	Method of Antibiotic
APH 2/3	aminglycosides	Inhibition of 30s subunit
OXA184/451	Beta-lactamase inhibitors	Blocks cell wall synthesis
SAT-4	Streptothricin	Inhibition of 30s subunit
cmeA/B/C/R	ciprofloxacin, norfloxacin, cefotaxime, fusidic acid, erythromycin	**CmeABC Multidrug Efflux Pump
gryA	Flouroquinolone	DNA synthesis inhibitor
tetO	Tetracycline	30s Subunit Inhibitor

- Our recommendation:
 - 50s subunit inhibition
 - 5-7 days of a Macrolide
 - Azithromycin, Clarithromycin

OUTBREAK SOURCE AND CDC RECOMMENDATIONS

- **Outbreak Source**

- Dairy:

- Cheese, Cookie batter – Only in the east coast for a span of 3 months

- Vegetables, Meat:

- Chorizo, beef, lettuce, avocado

- **CDC Recommendations**

- Thoroughly wash and scrub fruits and vegetable under clean water

- Double-check expiration dates of meat/dairy products

- Do not consume undercooked meat

- Use clean cooking surfaces and utensils (i.e. cutting board, knives)

- Notify healthcare providers in 8 states of recommended antibiotics

- Drink extra fluids while diarrhea lasts, but for more serious illnesses, seek medical treatment and stay home from school or work



QUESTIONS?

WORKS CITED

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4870677/>
- <https://www.genome.jp/kegg/annotation/br01600.html>
- <https://card.mcmaster.ca/ontology/37153>
- <https://www.orthobullets.com/basic-science/9059/antibiotic-classification-and-mechanism>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3462042/>
- <https://www.statista.com/statistics/307383/largest-mexican-restaurant-chains-us/>
- <https://www.cdc.gov/campylobacter/outbreaks/puppies-12-19/index.html>
- 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