

Bioinformatics Approaches to Supporting Outbreak Investigations

CFSAN-SNP and the Lyve-SET Pipeline

Kevin G. Libuit, M.S.

Senior Informatics Scientist

Division of Consolidated Laboratory Services

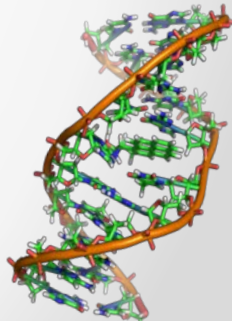
Inferring Genetic Relatedness from WGS

Clustering bacterial isolates to infer **epidemiological associations**

- Based on genetic relatedness inferred from whole-genome sequencing (WGS) data

Predominant approaches in public health bioinformatics:

- Multiple sequence alignment (MSA)
- Core/whole genome multilocus sequence typing (c/wgMLST)
- Single nucleotide polymorphism (SNP)



Single Nucleotide Polymorphism (SNP)

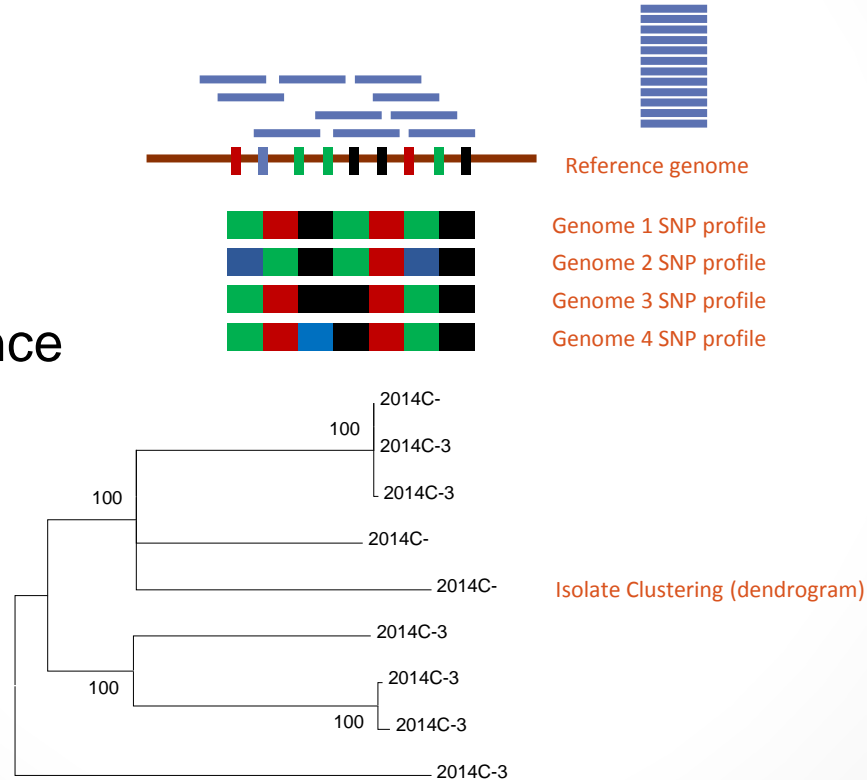
SNP:

- *Significant* changes in single nucleotide positions, with respect to a reference genome
 - Isolates clustered through a pairwise comparison of SNPs identified

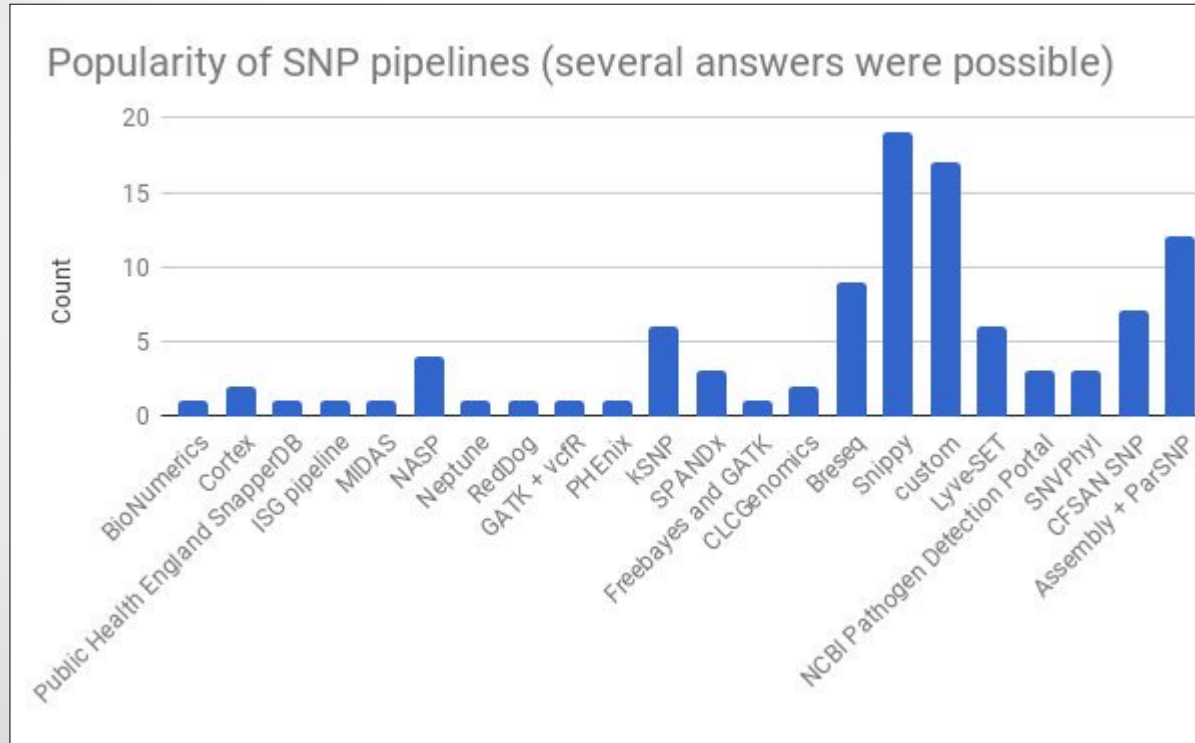
Common processes of a SNP Pipeline

Single Nucleotide Polymorphism (SNP)

1. Read Mapping
2. SNP-Calling
3. Phylogenetic Inference



Twitter Poll from Anita Schürch (UMC) on SNP Pipeline Popularity (conducted October 2017)



66 responses
22 different pipelines



Identifying the Appropriate SNP Pipeline

Literature review

- Clustering bacterial isolates to infer **epidemiological associations**
- Microbial foodborne pathogens

Communicating with collaborators

- Other state and federal public health laboratories

CFSAN-SNP¹ & Lyve-SET Pipeline²





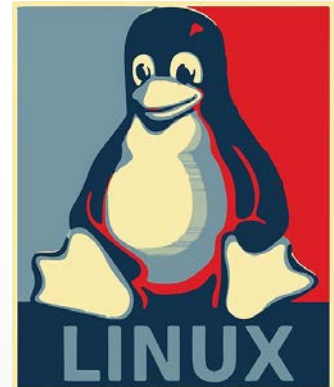
CFSAN-SNP and the Lyve-SET Pipeline

	CFSAN-SNP (FDA)	Lyve-SET (CDC)
Phage Masking	FALSE	TRUE
Read Mapping	BowTie2	Smalt
SNP Caller	VarScan	VarScan
Coverage	8x	20x
Consensus	60%	95%
Density Filtering	333bp	5bp
Phylogenetic Inference	FastTreev2.1*	RAxMLv8

Access and Operation of the SNP Pipeline

Download, installation, and usage:

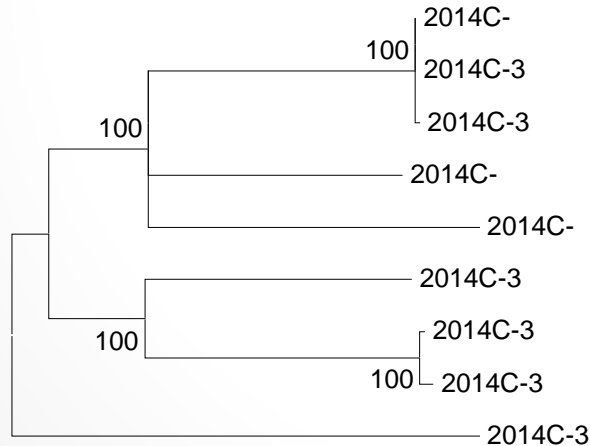
- Operating system and compute specifications
- Graphic user or command line interface
- Bioinformatics experience and background of personnel



Validating SNP Pipelines

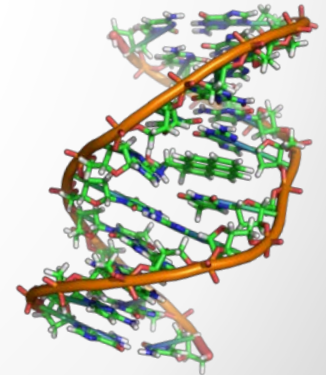
FDA and CDC curated benchmark dataset expectations

- Comparison of a group of isolates with known cluster profile
 - Outbreak isolates should cluster separately from non-outbreak strains



Virginia's SNP Analysis Workflow

- Active surveillance by PFGE
- Requests for SNP analysis when above baseline or temporal/geographic clustering observed
- CFSAN-SNP and Lyve-SET
- Trees and matrices assessed internally
 - Line lists shared with state epidemiologists



CFSAN-SNP and Lyve-SET Pipeline

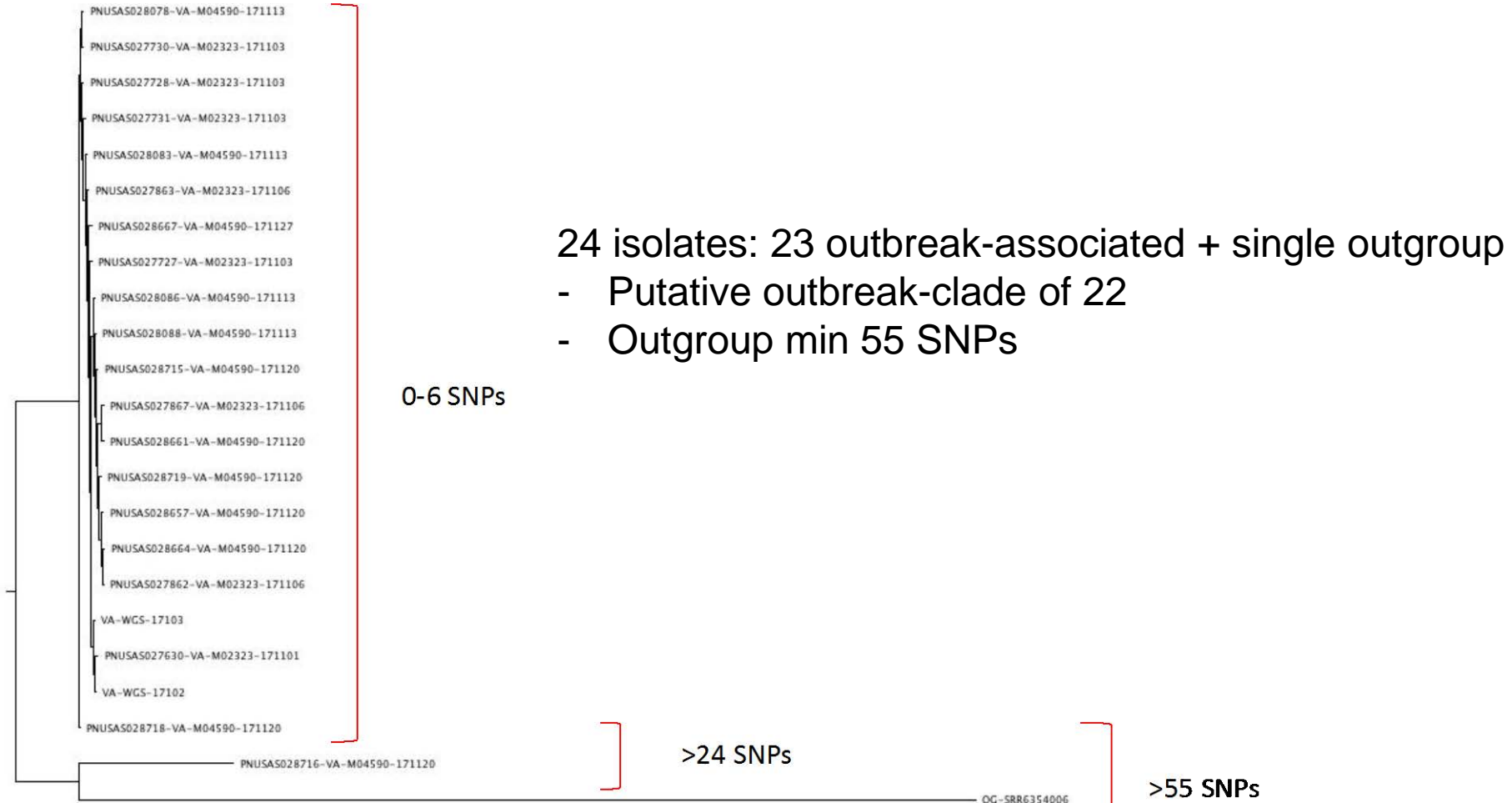
In general:

- Topological agreement
- Minor discrepancies in SNP-distances

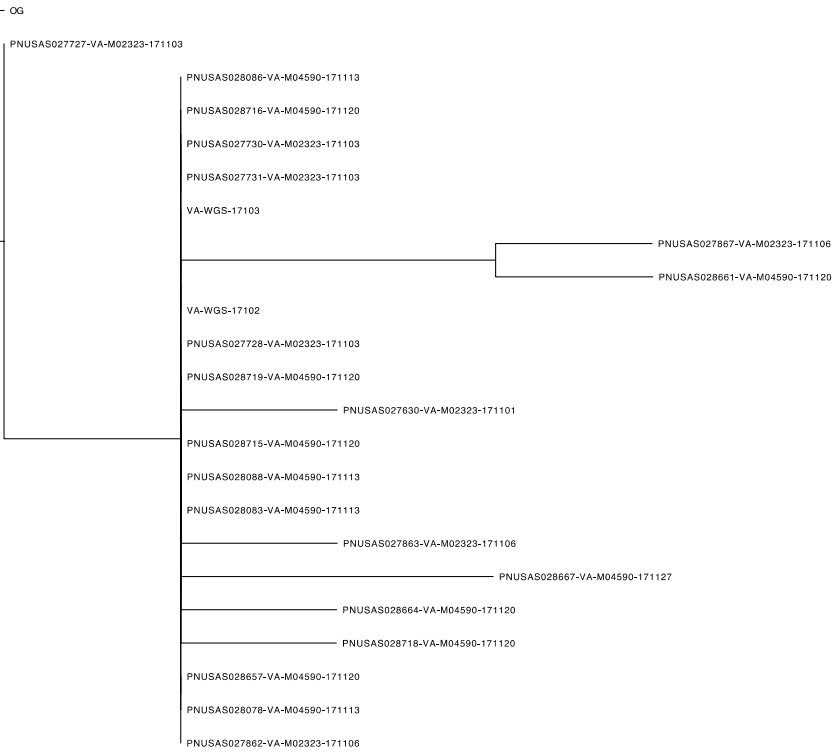
Occasional topological discrepancies with vast SNP discrepancies

- Multiple approaches allows for further investigation and troubleshooting, if necessary

Salmonella enterica subsp. Heidelberg outbreak investigation: CFSAN-SNP Output



Salmonella enterica subsp. Heidelberg outbreak investigation: LYVE-SET Output



0-10 SNPs

- 24 isolates: 23 outbreak-associated + single outgroup
- Putative outbreak-clade of... 24 isolates (?)
 - All isolates with SNP-distance <8



Integrating Bioinformatics Solutions

Factors to consider:

- How bioinformatics and WGS is going to inform public health decisions
- Multiple approaches extant in the field
- Accessibility and operation of the bioinformatics tool
- Validate local functionality
- If possible, employ more than one approach





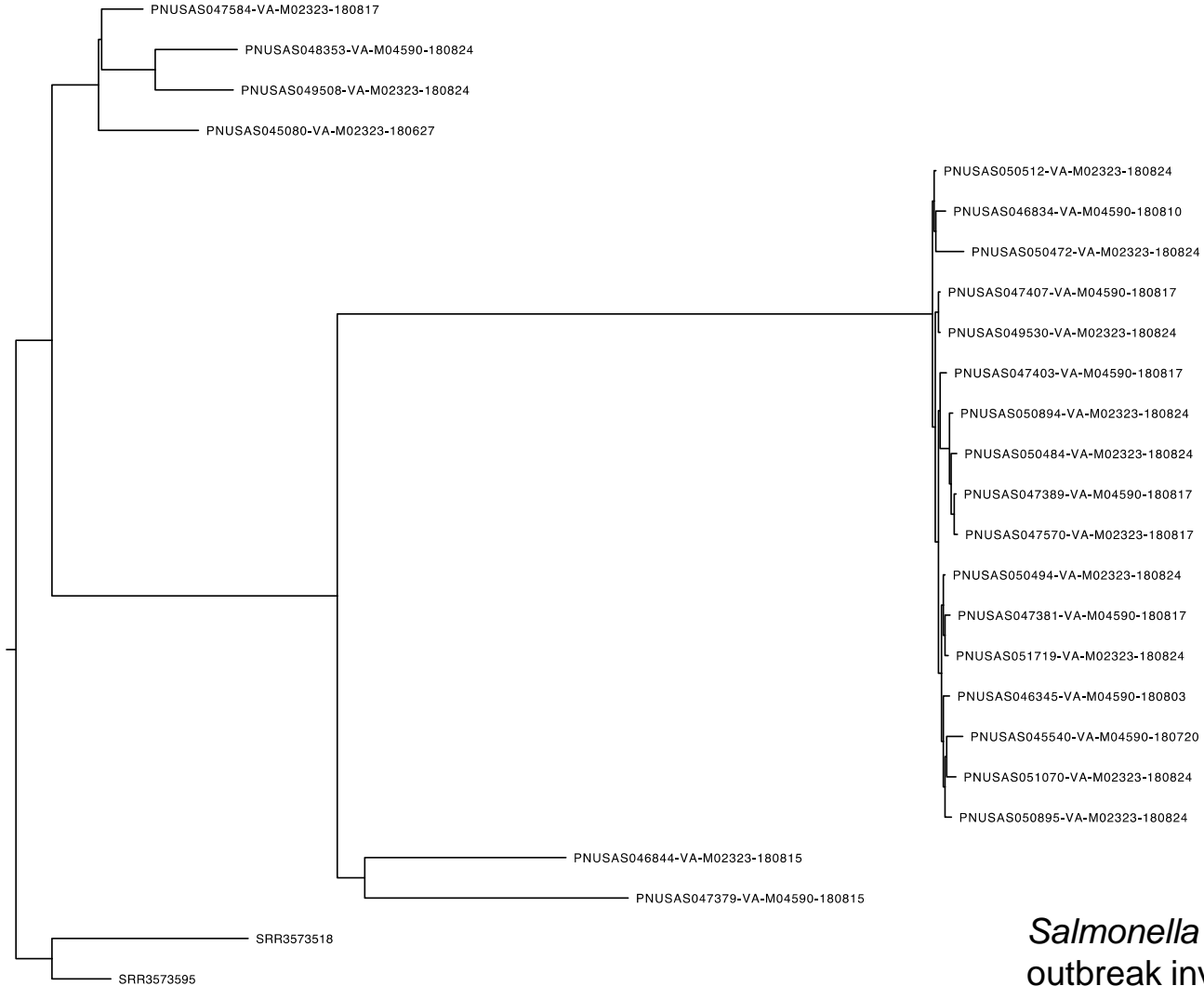
Kevin G. Libuit, M.S.

Senior Informatics Scientist

Division of Consolidated Laboratory Services

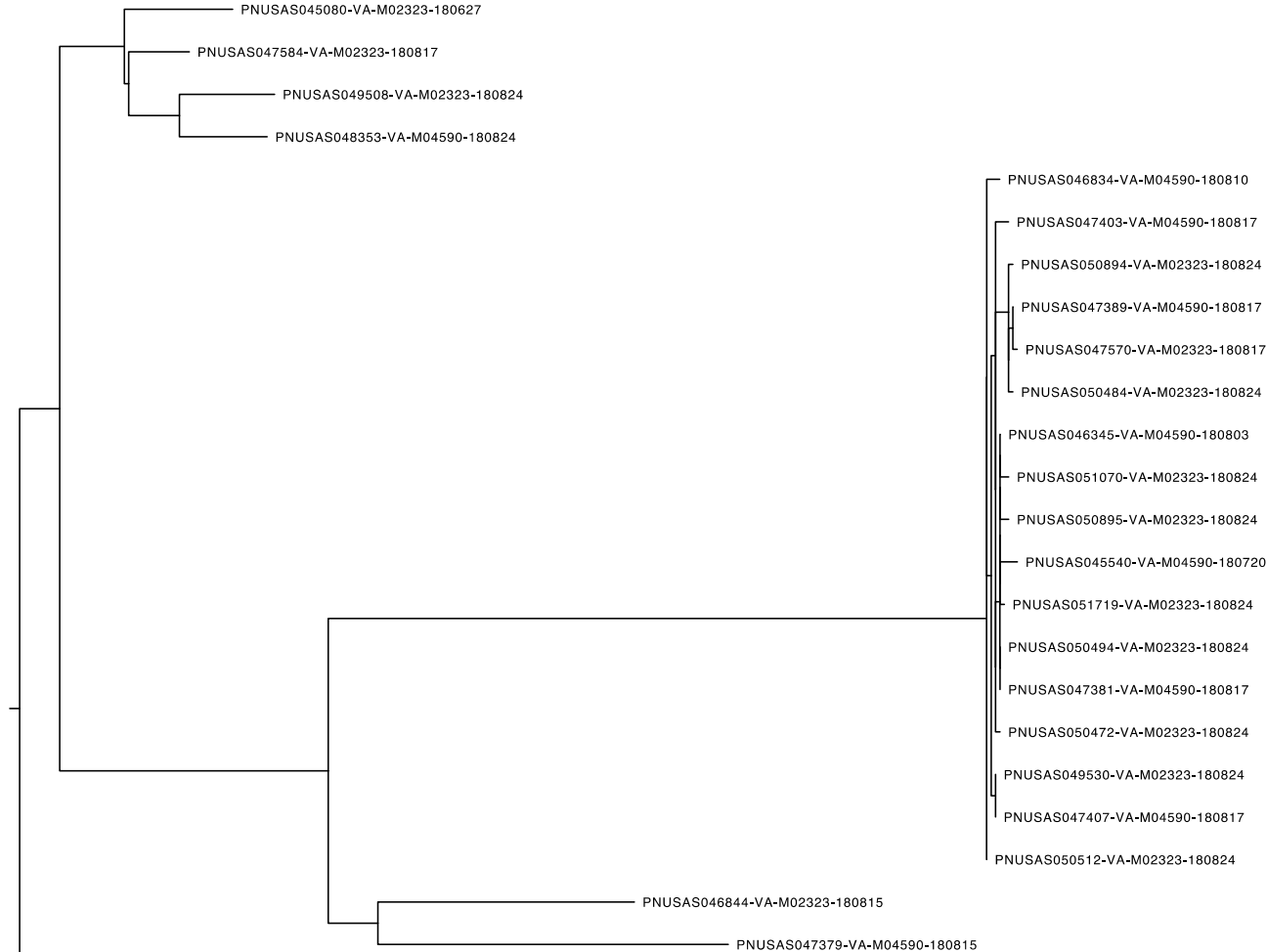
Email: kevin.libuit@dgs.virginia.gov

1. Davis S, Pettengill JB, Luo Y, Payne J, Shpuntoff A, Rand H, Strain E. (2015) CFSAN SNP Pipeline: an automated method for constructing SNP matrices from next-generation sequence data. PeerJ Computer Science 1:e20.
2. Katz LS, Griswold T, Williams-Newkirk AJ, Wagner D, Petkau A, et al. (2017) A Comparative Analysis of the Lyve-SET Phylogenomics Pipeline for Genomic Epidemiology of Foodborne Pathogens. Frontiers in Microbiology 8:375.



0-16 SNPs

Salmonella enterica subsp. Enteritidis
 outbreak investigation: CFSAN-SNP Output



0-10 SNPs

Salmonella enterica subsp. Enteritidis
outbreak investigation: Lyve-SET Output

SRR3573595

SRR3573518