NEW YORK STATE | Department of Health | Wadsworth Center

Whole Genome Sequencing for National Surveillance of Enteric Pathogens

William Wolfgang Ph.D.
Wadsworth Center / NYSDOH
Division of Infectious Diseases
william.wolfgang@health.ny.gov

NEBASM Symposium 11/7/19

WGS for Enteric pathogen surveillance

Wadsworth: 2012 acquires its first bench top sequencer
- pilot studies with *Salmonella* Enteritidis

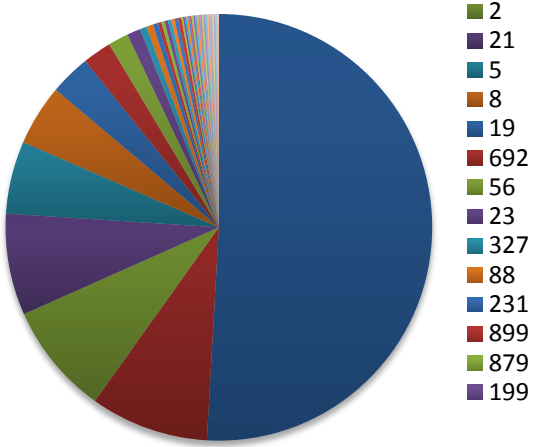
FDA: 2013 GenomeTrackr initiative.
- Surveillance of **Environmental** pathogens

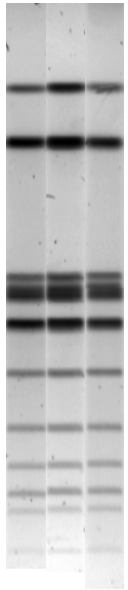
CDC: 2014 Advanced Molecular Detection initiative.
- Surveillance of **Clinical** pathogens

NCBI: creates public databases to hold NGS data.
- Pathogen Detection Portal

For *Salmonella* Enteritidis (SE) Outbreak clusters are hard to detect using PFGE

- 50% of the isolates we receive have the same PFGE DNA fingerprint.
- And 2/3 have a very common PFGE DNA fingerprint.
- These **Endemic** types are of limited use to our epidemiologists.





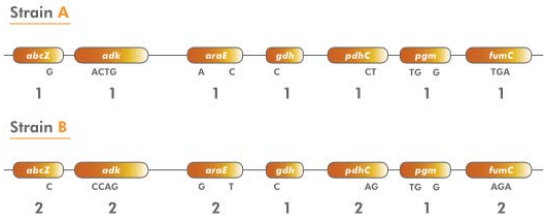
52 PFGE types

NEW YORK STATE | Department of Health | Wadsworth Center

SNP and wgMLST analysis

SNP

wgMLST



Applied Maths

Compare single nucleotides (SNPs)

- Highest resolution
- Typing nomenclature is not possible

Compare gene by gene (alleles)

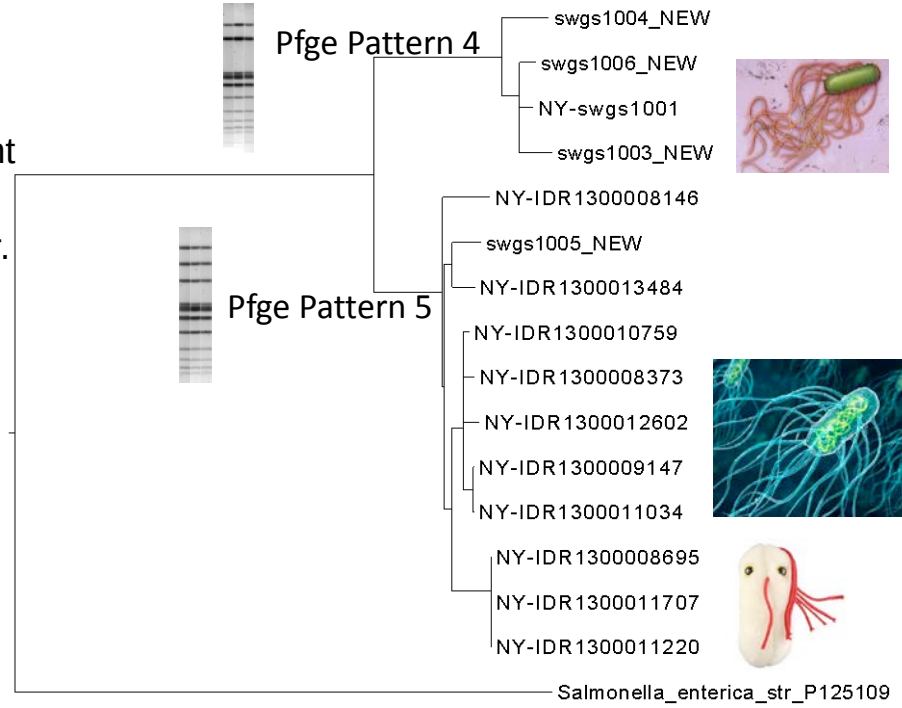
- Lower resolution
- But typing nomenclature is possible

NEW YORK STATE Department of Health | Wadsworth Center

Salmonella phylogenetic tree

- Patient isolates are on right
- Branch points indicate putative common ancestor.
- Sum of Horizontal lines measure genetic closeness.

- Tree reveals
  - Genetic closeness
  - Ancestral relationships



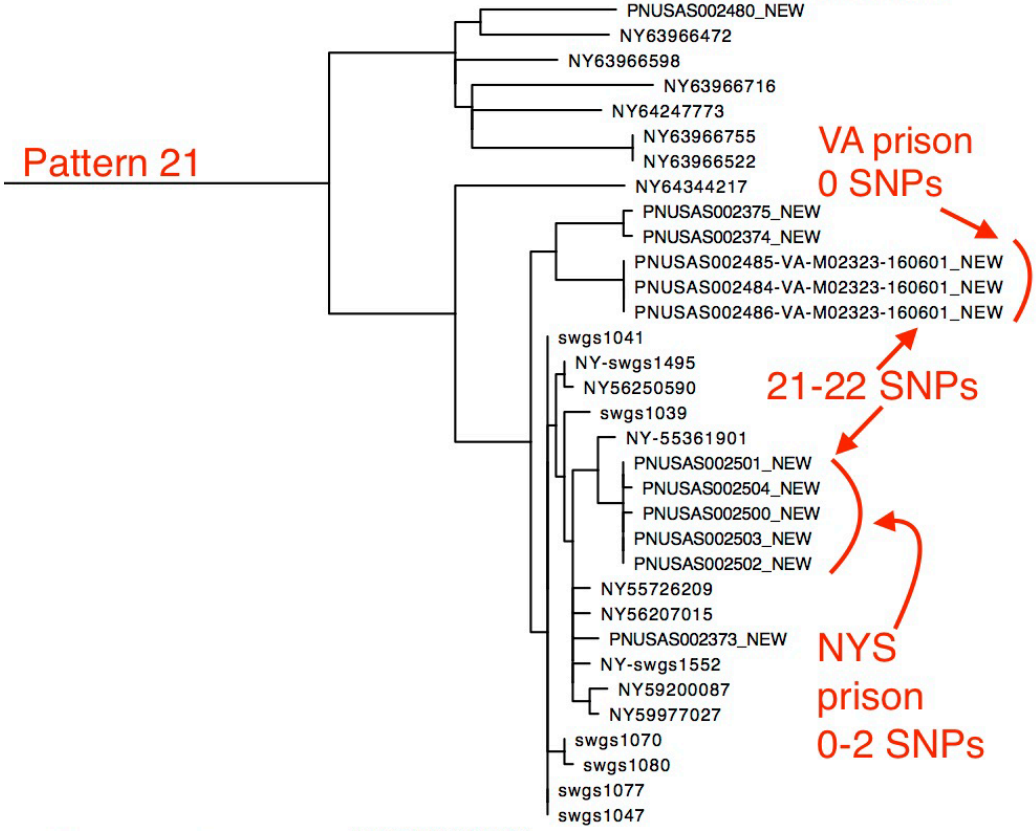
Prison outbreaks in Virginia and New York

- **5/16**-**Virginia** reports an SE outbreak associated with a correctional facility.
- At the same time **NYS** is investigating a SE outbreak also associated with a correctional facility.
- Both have the same PFGE pattern: JEGX01.0021

Q. Could they be from a common source?

NEW YORK STATE | Department of Health | Wadsworth Center

Prison associated cluster



- Clusters were distinct.
- Suggest different source for each outbreak.

NEW YORK STATE Department of Health | Wadsworth Center

Cluster Thresholds that Trigger a Report to NYS Epidemiologists

Organism	Min # isolates	Timeframe	Alleles or SNPs
Listeria monocytogenes	2	Indefinite	20
E. coli	2	1 year	10
***Salmonella* other serovars**	2	60 days	10
***S.* Enteritidis,Typhimurium & Newport**	3	60 days	5

Thresholds as of Oct 2019

We have made 6 revisions since May 2019

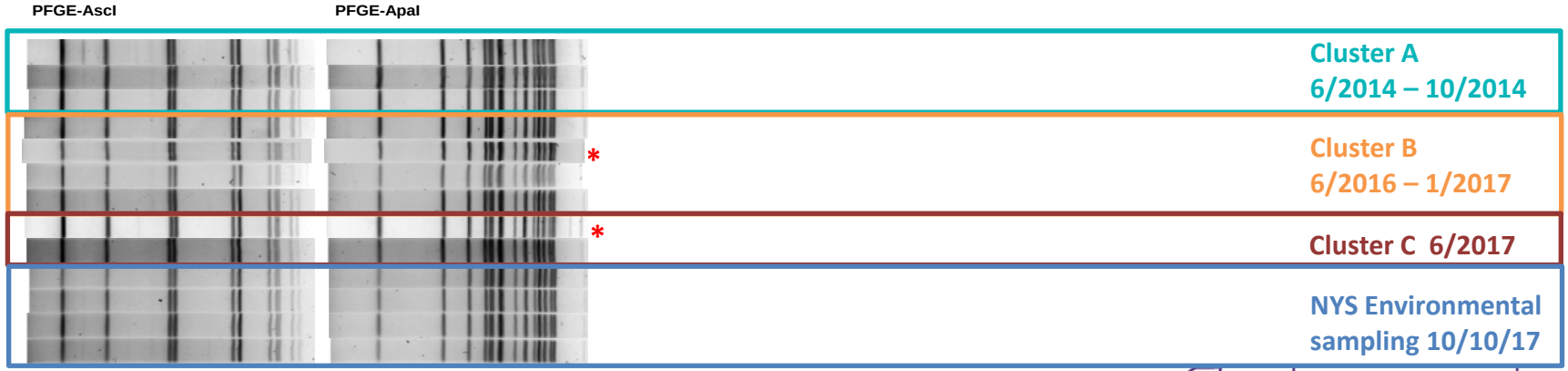
3 case studies

- *Listeria monocytogenes* – food preparation facility
- *E. coli* O103 – NYS geographical cluster
- *Salmonella* Enteritidis – The cluster that never ends

Intermittent *Listeria* outbreak

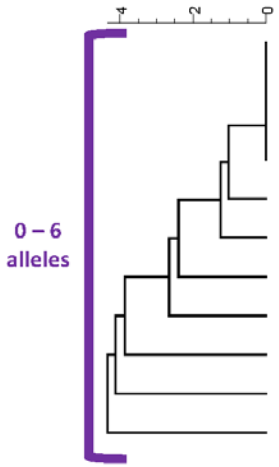
From June. 2014 to Oct. 2017

- 13 isolates with matching PFGE were detected
- NY residents were from the same counties
- Based on 120 day time frame- treated as 3 separate clusters
- But should they really be considered as one cluster?



NEW YORK STATE Department of Health | Wadsworth Center

wgMLST analysis shows all are 0 to 6 alleles apart

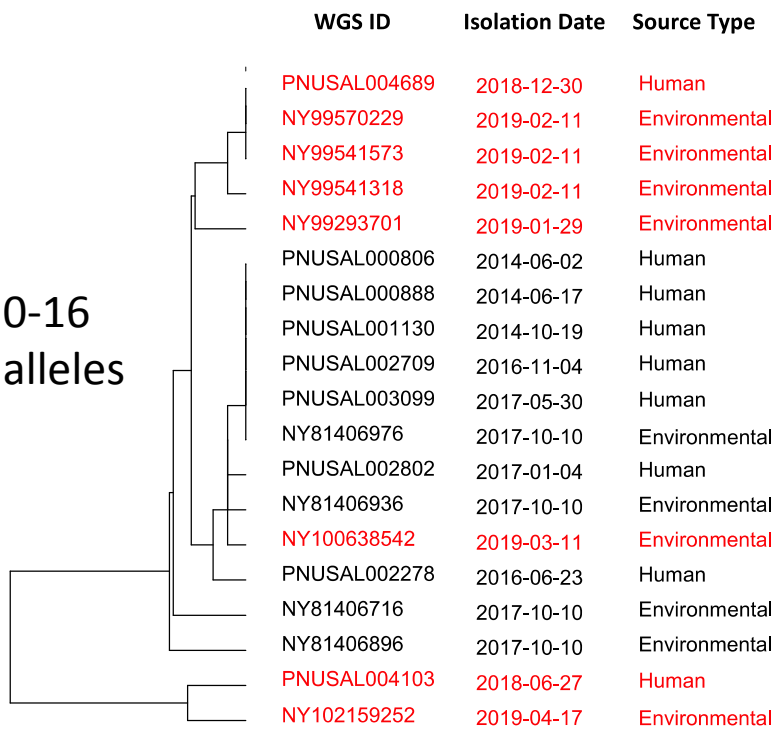


WGS ID	Iso Date	Outbreak	Ascl Pattern	Apal Pattern	SourceType
PNUSAL002709	2016-11-04		GX6A16.0677	GX6A12.0770	Human
PNUSAL001130	2014-10-19	1410NYGX6-1	GX6A16.0677	GX6A12.0770	Human
PNUSAL000888	2014-06-17	1410NYGX6-1	GX6A16.0677	GX6A12.0770	Human
PNUSAL000806	2014-06-02	1410NYGX6-1	GX6A16.0677	GX6A12.0770	Human
PNUSAL002802	2017-01-04		GX6A16.0677	GX6A12.0770	Human
NY81406936	2017-10-10	OMS201715450	GX6A16.0677	GX6A12.0770	Environmental
PNUSAL003099	2017-05-30		GX6A16.0677	GX6A12.0770	Human
PNUSAL002278	2016-06-23		GX6A16.0677	GX6A12.0770	Human
NY81406896	2017-10-10	OMS201715450	GX6A16.0677	GX6A12.0770	Environmental
NY81406716	2017-10-10	OMS201715450	GX6A16.0677	GX6A12.0770	Environmental
NY81406976	2017-10-10	OMS201715450	GX6A16.0677	GX6A12.0770	Environmental

- The 2 out-of-state isolates did **not** matched cluster.
- The 4 environmental samples were also closely related

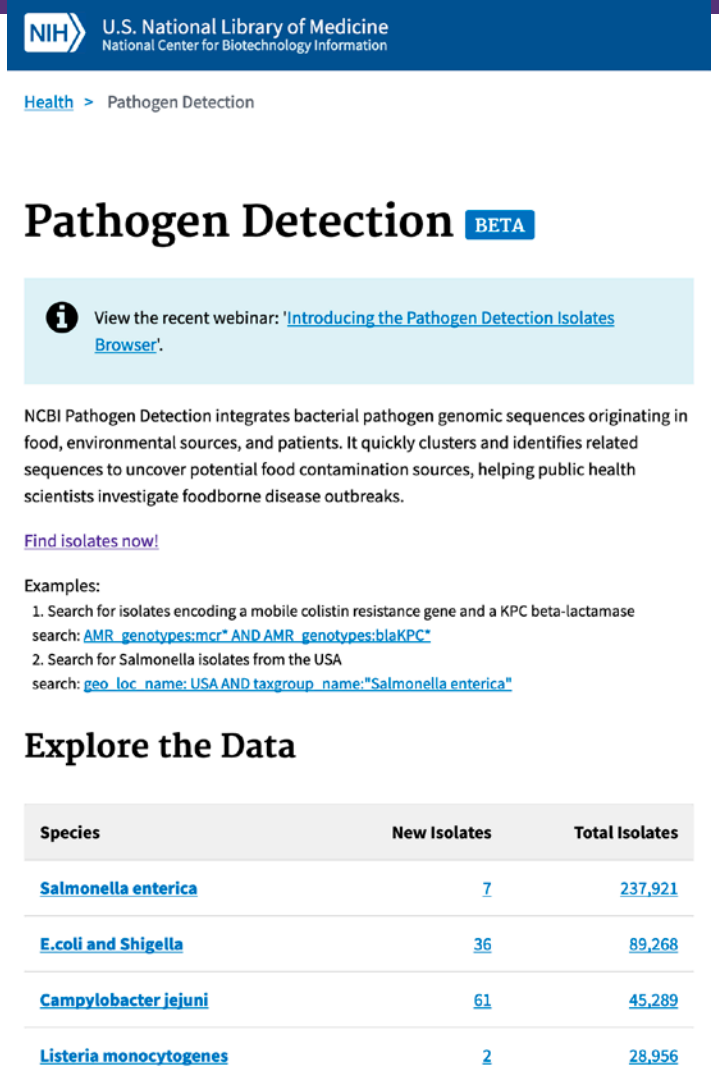
NEW YORK STATE | Department of Health | Wadsworth Center

Surveillance and environmental sampling continued



- In 2018 and 2019
  - 2 patient
  - 6 environmental
- Remediation of the facility was undertaken
- After March 2019 no more positive environmental or clinical specimens
- But were there out of state isolates?

NEW YORK STATE | Department of Health | Wadsworth Center

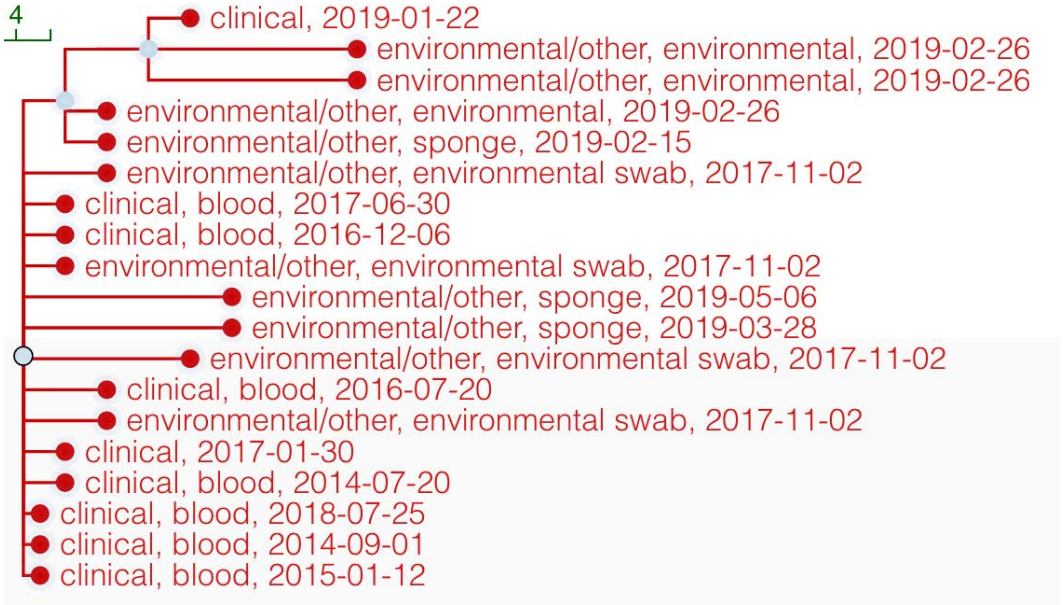


NCBI Pathogen Detection

- All Genometrakr and PulseNet samples end up here.
- NCBI builds trees daily for Enteric pathogens.
  - 29,000 *Lm* samples
  - 2,548 *Lm* trees
- Easily accessed through a web portal.
- Data is publicly available.

NEW YORK STATE | Department of Health | Wadsworth Center

NCBI Pathogen Detection shows no closely related out of state isolates



- All isolates fell into a single NCBI tree.
- No other isolates on the tree.
- 0 to 13 SNPs

And so were not closely related to 29,000 other *Lm*

NEW YORK STATE Department of Health | Wadsworth Center

So how is this working with *Listeria monocytogenes*?

Pretty Good

Improved resolution of WGS allowed:

- Identification of a long term source of *Lm* contamination
- Exclusion of out of state samples
- Able to accurately track remediation
- No additional isolates were detected nationally

- Supports allele range of 0-20
- Supports time frame of forever

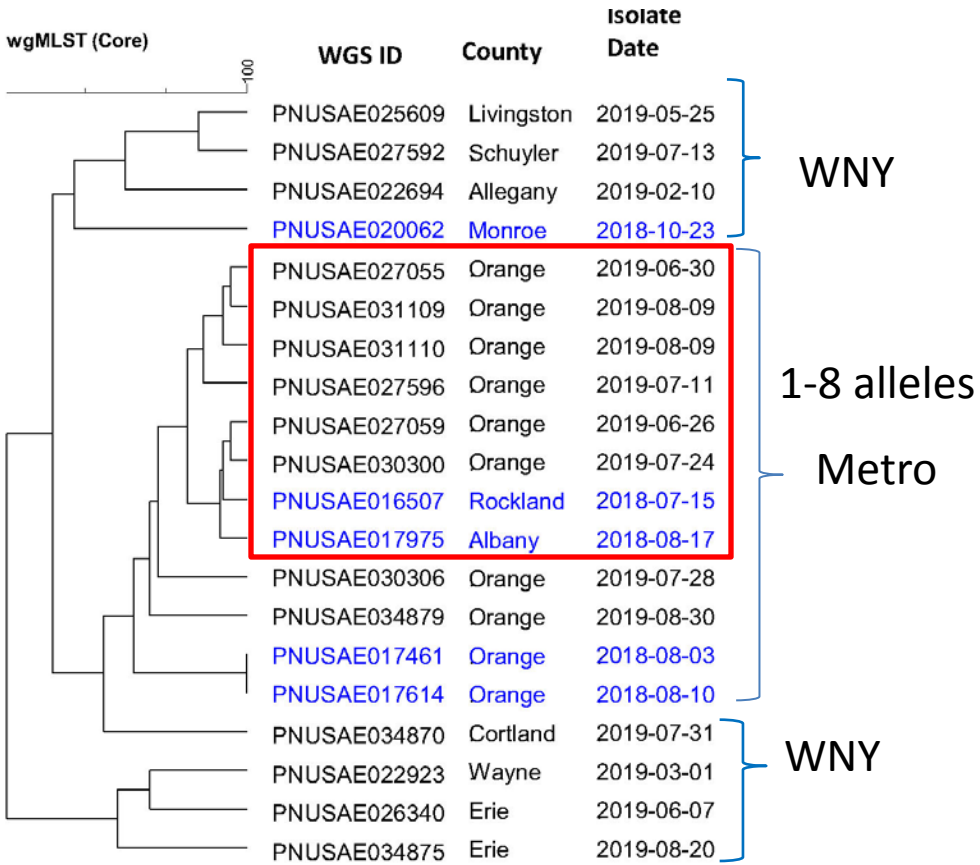
NYS *E. Coli* O103 cluster

From July,11 to Oct 7, 2019

- 20 isolates within a single genomic cluster (0-20 alleles) were detected.
- No PFGE was done

- Epidemiology supported two origins- Western NY and the Metro area.
- Some Metro isolates were associated with kosher food consumption.

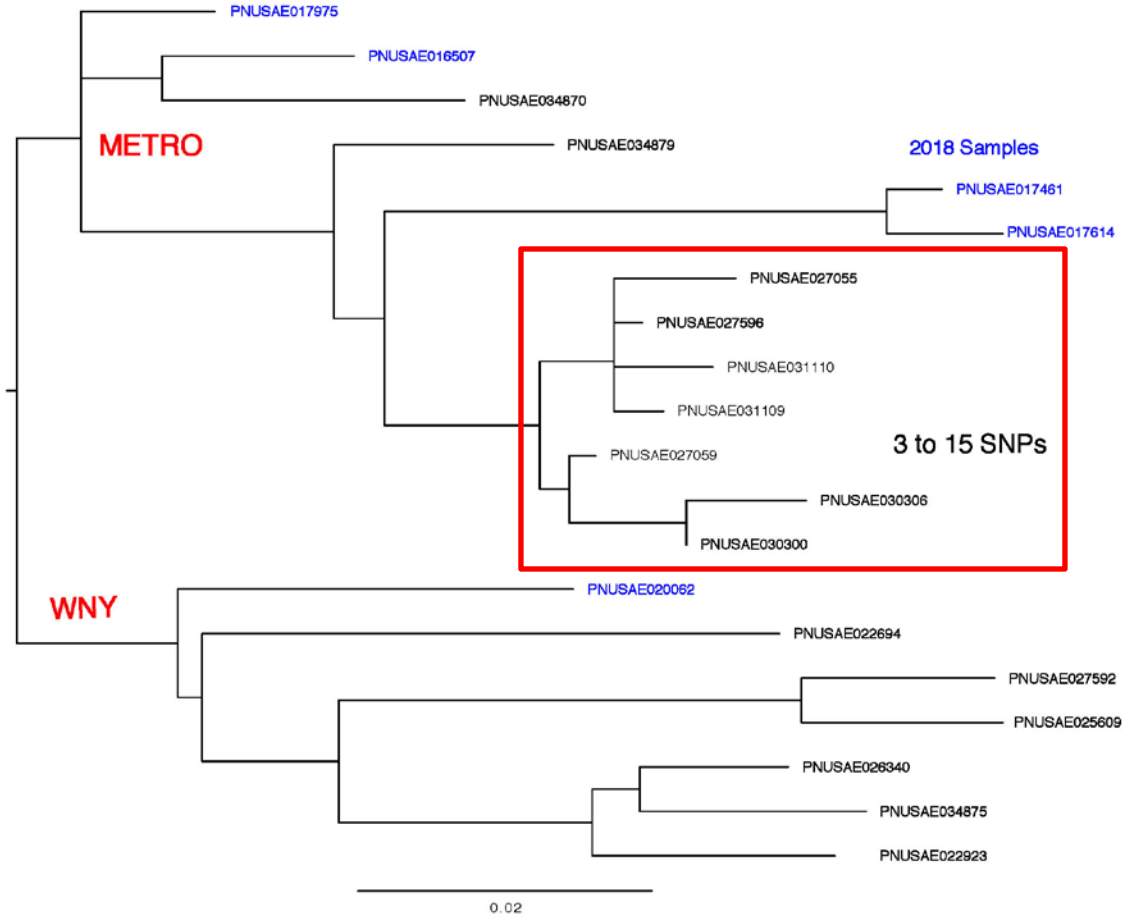
cgMLST allele analysis



- 0 – 20 alleles
- One sub-cluster with 1-8 alleles; associated with kosher food consumption
- cgMLST did not support geographic clustering

NEW YORK STATE | Department of Health | Wadsworth Center

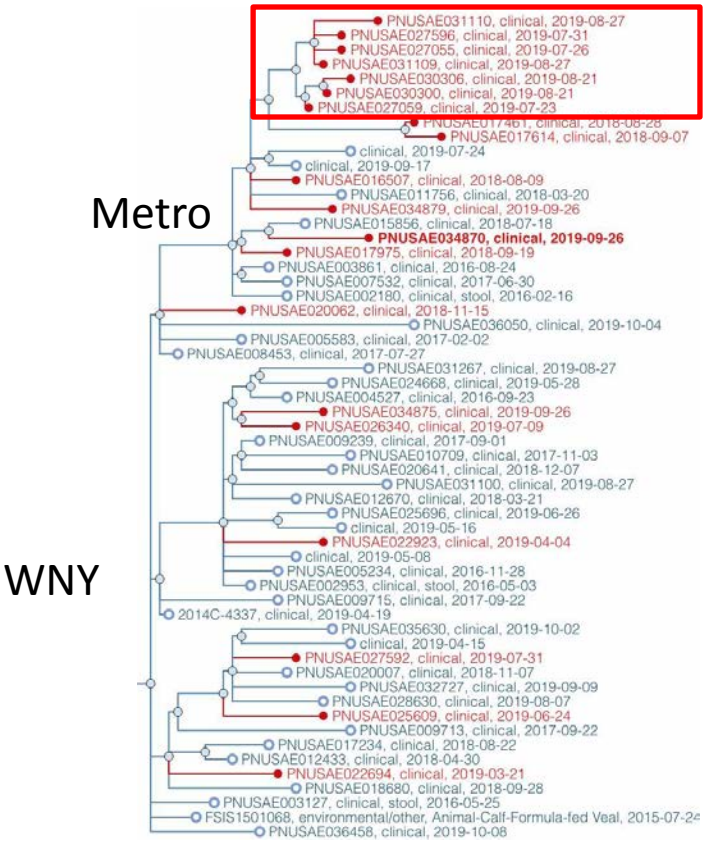
SNP Analysis is similar but different



- 2-61 SNP diversity
- One sub-cluster with 3-15 SNP diversity
- SNP analysis supported geographic clustering

NEW YORK STATE Department of Health | Wadsworth Center

NCBI Pathogen detection



- 2 - 51 SNPs overall
- One sub-cluster 2-15 SNPs
- SNP analysis supported geographic clustering
- WNY interspersed among many other isolates

NEW YORK STATE | Department of Health | Wadsworth Center

Comparing Allele vs SNP trees

Type of tree	Full diversity	Sub-cluster
CGMLST	0-20 alleles	1-8 alleles
FDA SNP	2-61 SNPs	3-15 SNPs
NCBI SNP	2-51 SNPs	2-15 SNPs

- Alleles underestimate full diversity.
- Structure of the allele tree less concordant with epi. data.
- When resolution is needed SNP trees should be built.

So how is this working with STEC?

Pretty good but

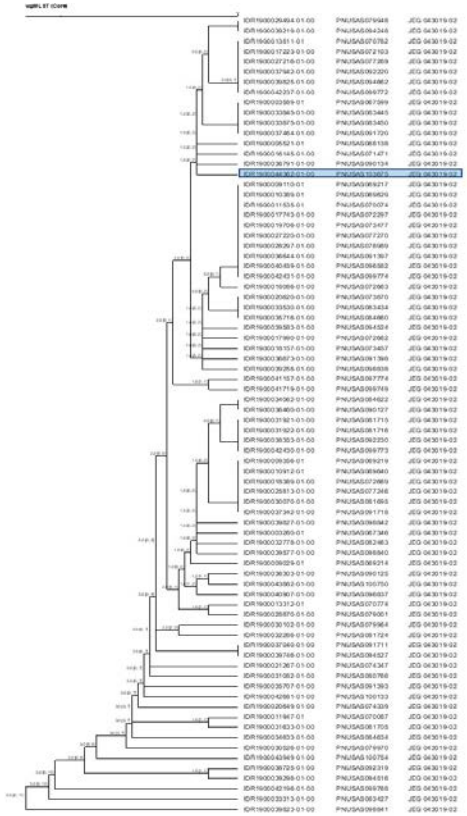
In this case cgMLST was discordant with geography

- SNP tree was concordant and had higher resolution

Demonstrated need to rethink cluster reporting thresholds for STECs

- reduce allele or SNP diversity
- shorten timeframe

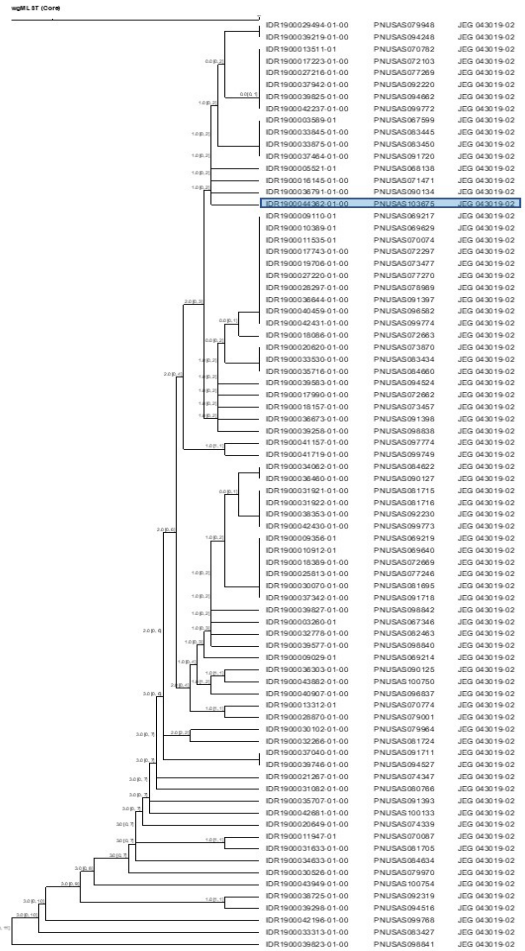
The *Salmonella* Enteritidis cluster that won't stop



From 4/30/19 to 10/08/19 we sent **8** seperate reports tracking a single cluster to our epidemiologists.

- **The cluster** eventually contained **84** patient samples.
- With no strong epidemiological links.
- 0-11 allele diversity.
  - Why is the cluster allele diversity greater than 5?

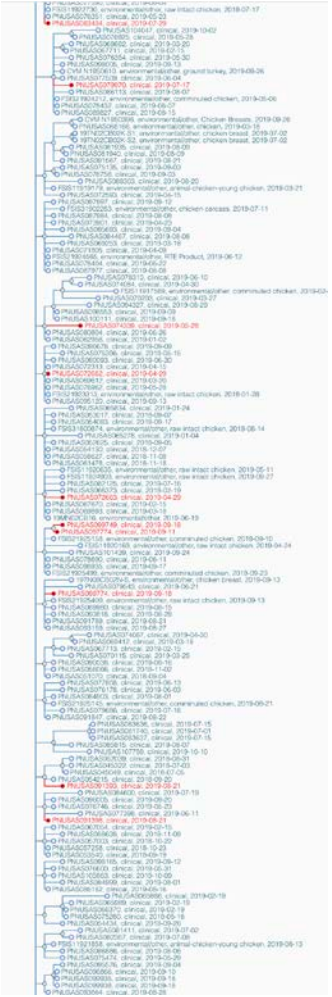
NEW YORK STATE | Department of Health | Wadsworth Center



The problem of chaining

- New samples are received within 60 days that are within 5 alleles to at least one other sample.
- Clustering criteria are inadequate.
- Overlaying epi. data was not helpful.

NEW YORK STATE | Department of Health | Wadsworth Center

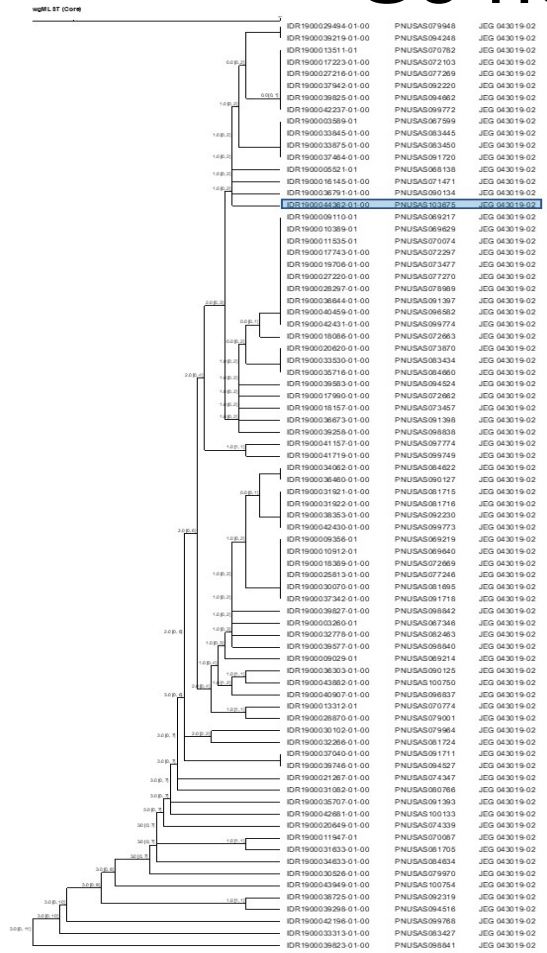


And worse at NCBI

- 353 samples within 0 SNPs of this cluster.
- 1645 within 5 SNPs.
- 2375 within 15 SNPs.
- And these numbers are constantly increasing

NEW YORK STATE Department of Health | Wadsworth Center

So how is this working with *SE*?



Not so well for common *S*. Enteritidis types

Improved resolution was not helpful

- Lab results can not inform epi.
- Yet a huge amount of work.

Solutions for *Salmonella*

- Consider other parts of the genome.
- Try to identify these common types prior to analysis.
    - then do not analyze unless requested

NEW YORK STATE | Department of Health | Wadsworth Center

Where do we stand now?

WGS does improve surveillance.

- More clusters will be identified than with PFGE.
- More sources will be identified.
- And cases of foodborne illness should decline.

But challenges remain

- Too many clusters.
  - Need to prioritize, but how?
  - Refine cluster definitions.
- Identification of **endemic clusters**.

Acknowledgements

Wadsworth Center

- Kim Musser
- Samantha Wirth
- Lisa Thompson
- Lisa Mingle
- Nellie Dumas
- Michelle Dickinson
- Deborah Baker
- Sai Gubbala

- Sequencing core
- Bioinformatics core

BCDC

- Amy Robbins
- Kate Purcell
- Alexandra Newman
- Paula Huth
- Madhu Anand
- Kadri Ajileye

BCEHFP

- David Nicholas

FDA

- Eric Brown
- Marc Allard
- Errol Strain
- Ruth Timme

CDC

- Eija Trees
- Kelley Hise
- Steven Stroika

NEW YORK STATE | Department of Health | Wadsworth Center