Integrating Whole Genome Sequencing of *Salmonella enterica* Serovar Enteritidis into the Public Health Laboratory for Surveillance and Outbreak Investigations

Levinson, K.J., Dickinson, M., Wirth, S.E., Anand, M., Baker, D.J., Bopp, D., Thompson, L., Musser, K.A., Lapierre, P., and Wolfgang , W.J.

NEW YORK STATE OF OPPORTUNITY. Department of Health | Wadsworth Center

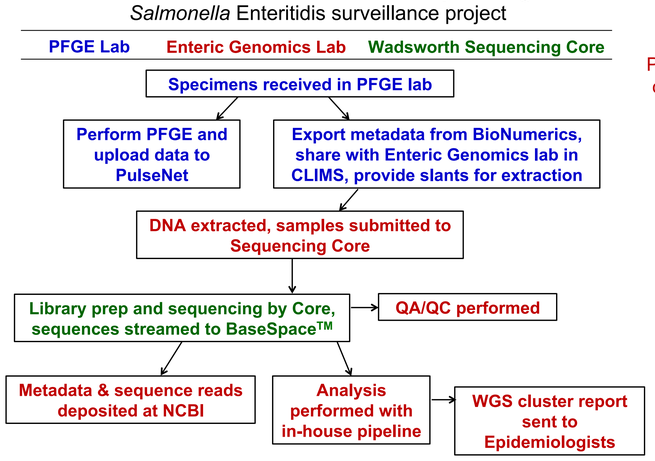
INTRODUCTION

Salmonella enterica serovar Enteritidis is a leading cause of foodborne illness in the United States. The low genetic diversity of *S.* Enteritidis limits how well pulse-field gel electrophoresis (PFGE) can detect outbreaks of enteric pathogens. To improve discrimination between sporadic and outbreak-associated-isolates, the Wadsworth Center performs whole genome sequencing (WGS) single nucleotide polymorphism (SNP) based phylogenetic typing on all *S.* Enteritidis isolates in addition to PFGE typing.

Study Objectives:

- Identify the benefits and challenges of incorporating WGS-based typing into routine surveillance.
- Establish an efficient means for reporting that is useful for both laboratorians and epidemiologists.

Workflow for Whole Genome Sequencing
Salmonella Enteritidis surveillance project



METHODS

- Nextera XT library preparation (250 paired end reads on MiSeq)
- Map over reference *Salmonella* str. P125109 with BWA-MEM
- SNP calling with Samtools
- Create full-length consensus sequence for each sample
- Extract positions with SNPs (called with 20x min depths and 95% of the reads in agreement), create alignment, calculate phylogenetic tree (PhyML using K2P model with no gamma)[1]

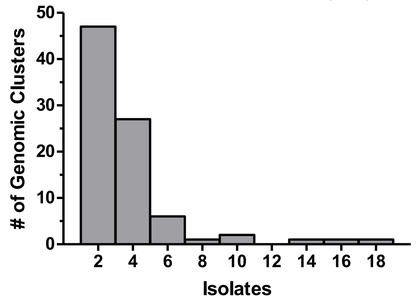
STUDY DESCRIPTION

502 isolates sequenced in real time
- Collected over 598 days (8/27/2013 to 4/16/2015)

32 PFGE patterns represented
- 438/502 (87%) isolates were part of endemic patterns

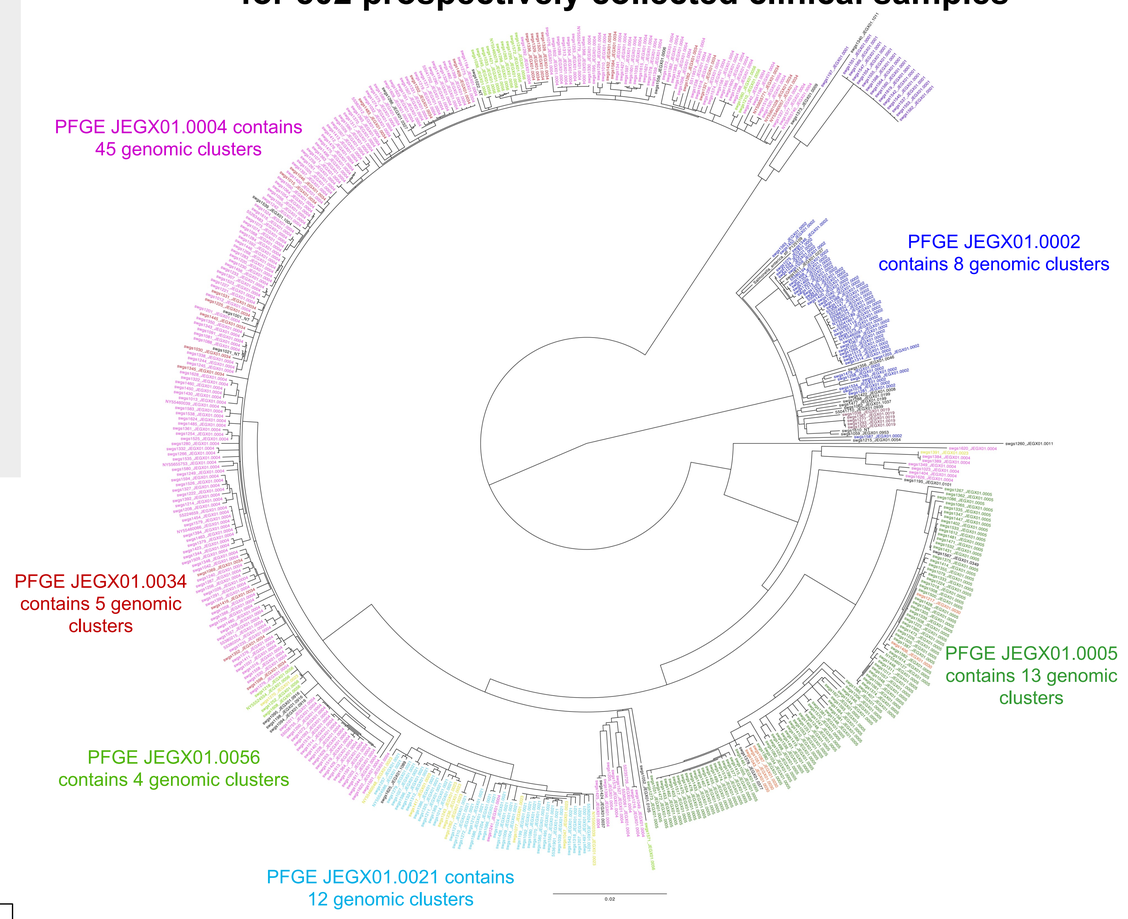
86 Genomic Clusters (GC) identified in dataset
- GC defined as ≤4 SNPs diversity among isolates
- 15/86 (17%) GC's contained 2 different PFGE patterns

Frequency Distribution of Isolates in Genomic Clusters (GC)



Most genomic clusters were comprised of 4 or fewer isolates (n=86 genomic clusters, mean=3.37, SD=2.87)

Cluster resolution is improved by WGS for 502 prospectively collected clinical samples



SNP-based phylogenetic tree of 502 *S.* Enteritidis isolates
- PFGE patterns are color-coded and interspersed throughout the phylogenetic tree
- PFGE patterns are not monophyletic, SNP-based phylogeny cannot predict PFGE type
- A single PFGE type can contain multiple genomic clusters (detailed below)

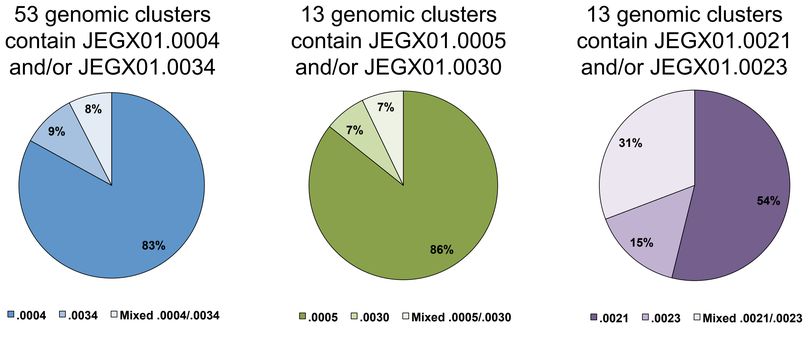
Whole genome sequencing subdivides PFGE types

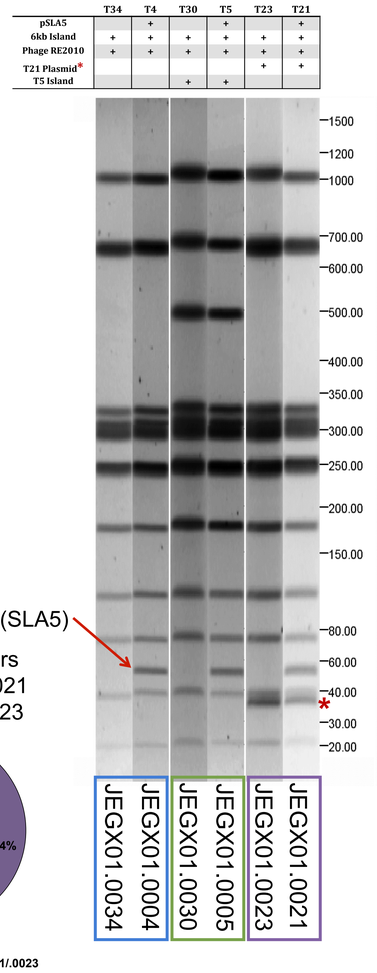
PFGE Type	Number of GCs Detected	Number of Isolates in GC (% of total)	Total number of Isolates in PFGE Type
JEGX01.0002	8	25 (52.1)	48
JEGX01.0004	45	121 (56.0)	216
JEGX01.0005	13	64 (66.7)	96
JEGX01.0021	12	20 (54.1)	37
JEGX01.0034	5	10 (40.0)	25
JEGX01.0056	5	10 (58.8)	17
JEGX01.0001	1	14 (93.3)	15
JEGX01.0019	1	4 (80.0)	5
JEGX01.0023	6	8 (80.0)	10
JEGX01.0030	2	5 (100)	5
JEGX01.0199	0	0 (0)	2
JEGX01.0916	1	3 (100)	3

- For endemic PFGE types (highlighted in gray) WGS analysis reveals many covert clusters
- For non-endemic rare PFGE types WGS confirms associations

17% of Genomic Clusters Contain Two PFGE Patterns

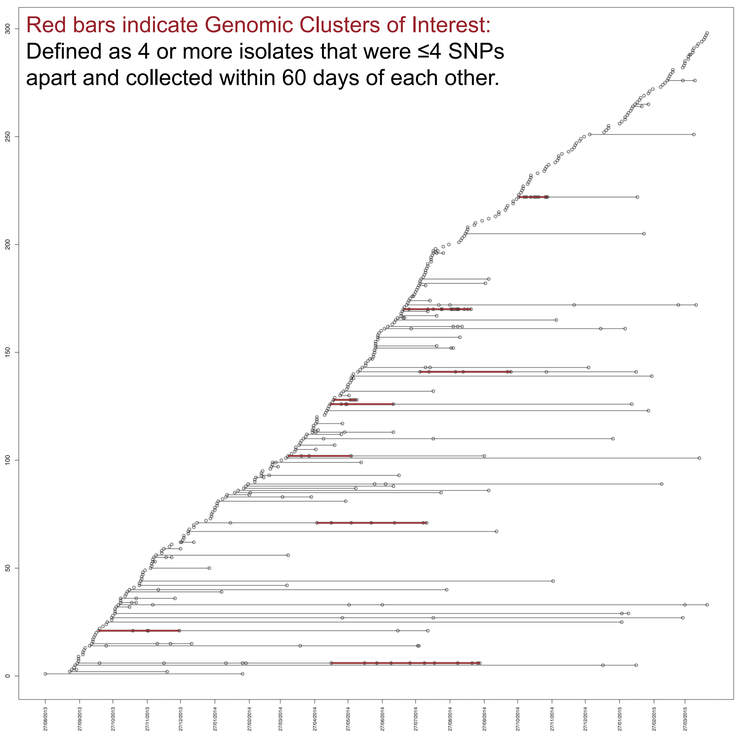
For the most common pairs shown below, this occurs from loss of a 59kb plasmid (SLA5)



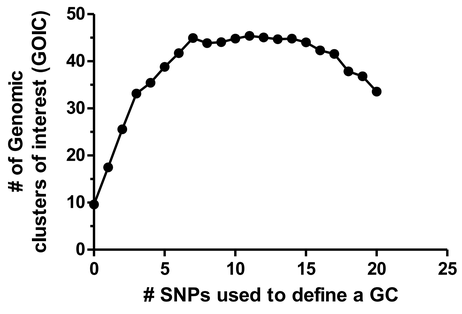


Prioritizing Genomic Clusters of Interest (GCOI) for rapid epidemiological follow-up: The LLWW Plot

Takes into account the frequency of isolate acquisition into a cluster:
- Variables can be changed depending on need



The affect of varying SNP diversity on number of GCOI



SUMMARY

- **WGS can subdivide endemic PFGE patterns into genomic clusters and can better discriminate between genomic clusters and sporadics.**
- **Created a tool to prioritize genomic clusters that can be utilized in real time by both epidemiologists and laboratorians.**

ACKNOWLEDGMENTS

We gratefully acknowledge support of this research by the Wadsworth Center Public Health Genomics Initiative, the Wadsworth Center Applied Genomic Technologies and Bioinformatics Cores, and the FDA Genome Tracker Initiative.

REFERENCES

1. Taylor AJ, Lappi V, Wolfgang WJ, Lapierre P, Palumbo MJ, Medus C, Boxrud D. Characterization of foodborne outbreaks of Salmonella enterica serovar Enteritidis with whole genome sequencing SNP-based analysis for surveillance and outbreak detection. *J Clin Microbiol*. 2015 PMID 26269623.
2. den Bakker HC, Allard MW, Bopp D, Brown EW, Fontana J, Igbal Z, Kinney A, Limberger R, Musser KA, Shudt M, Strain E, Wiedmann M, Wolfgang WJ. Rapid whole-genome sequencing for surveillance of Salmonella enterica serovar Enteritidis. *Emerg Infect Dis*. 2014. Aug;20(8):1306-14. PMID 25062035.