#020

Assessment of Plasmids for Relating the 2020 *Salmonella enterica* Serovar Newport Onion Outbreak to Farms Implicated by the Outbreak Investigation

Seth Commichaux[1], Hugh Rand[1], Kiran Javkar[2,3], Erin K. Molloy[2,3], James B. Pettengill[1], Arthur Pightling[1], Maria Hoffmann[1], Mihai Pop[2,3], Victor Jayeola[1], Steven Foley[4], Yan Luo[1]; 1. Center for Food Safety and Nutrition, Food and Drug Administration, MD, USA; 2. Center for Bioinformatics and Computational Biology, University of Maryland, MD, USA; 3. Department of Computer Science, University of Maryland, MD, USA; 4. National Center for Toxicological Research, Food and DrugAdministration, AR, USA

Background: The *Salmonella enterica* serovar Newport red onion outbreak of 2020 was the largest foodborne outbreak of *Salmonella* in over a decade. The epidemiological investigation suggested two farms as the likely source of contamination. SNP analysis showed that none of the Salmonella isolates collected from the farm regions were linked to the clinical isolates. We explored an alternative method for analyzing the whole genome sequencing data driven by the hypothesis that if the outbreak strain had come from the farm regions, then the clinical isolates would disproportionately contain plasmids found in isolates from the farm regions due to horizontal transfer.

Conclusions: Phylogenetic analysis of the plasmids provided limited information about their geographic origin, isolation source, or time of transfer, seemingly due to promiscuous horizontal transfer. However, our resampling analysis suggested that observing a similar number and combination of highly similar plasmids in random samples of environmental *Salmonella enterica* was unlikely. A limitation of our study was that the environmental *Salmonella* isolates were the only representatives of the farm microbiome. It would have been informative to survey the farm microbiome, the gut microbiomes of local animals, and the gut microbiomes of patients for plasmids to increase the chance of detecting horizontally transferred plasmids and to better identify their source. Nonetheless, horizontally transferred plasmids provided evidence for a connection between clinical isolates and the farms implicated as the source of the outbreak. Our case study suggests that such analyses might add a new dimension to source tracking investigations, but highlights the need for detailed and accurate metadata, more extensive environmental sampling, and a better understanding of plasmid molecular evolution.

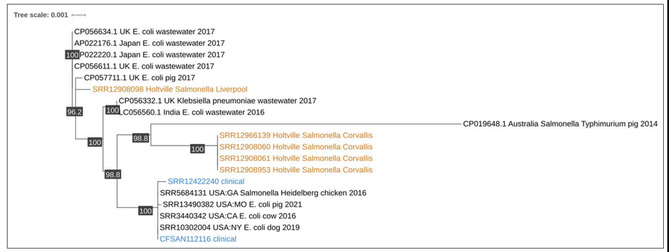
FDA Mission Statement: This study provides an additional method for using whole genomic sequence data to trace the source of foodborne illness outbreaks. Our results indicate that plasmids might be an important, and often overlooked, source of information for outbreak investigations.

Plasmid type	Number of clinical isolates with plasmid	Number of Holtville isolates with plasmid	Number of Bakersfield isolates with plasmid	Median contig length (bp)	Median number of genes
IncFII(S)	1728	3	7	72766	76
Unclassified	373	235	42	6170	7
Col440I	3	60	8	4322	5
IncI1-I(Gamma)	27	13	3	85161	93
IncQ1	0	20	6	8259	10
IncFIB(pB171)	0	25	0	28614	33
Col(pHAD28)	1	12	4	4751	7
IncFII	1	15	0	70562	85
ColpVC	12	4	0	2223	1
IncX3(pEC14)	0	14	0	38955	51
IncFII(pCRY)	0	3	10	49311	62
Col(BS512)	12	0	0	2216	2
ColRNAI	0	1	10	9712	11
IncI2(Delta)	6	0	0	58474	75
IncI(Gamma)	4	0	0	84482	95
Col(MG828)	4	0	0	1672	1
IncFII(Yp)	0	4	0	171753	190
IncR	0	4	0	13431	15
pXuzhou21	3	0	0	40119	54
IncL	3	0	0	59108	81
Col156	2	0	0	4937	5
IncX4	2	0	0	30161	44
Col8282	1	0	0	4218	3
IncB/O/K/Z	1	0	0	21481	25
IncN	1	0	0	38347	49
IncX1	1	0	0	40944	51
pSL483	1	0	0	38560	53
IncM1	1	0	0	58540	76
IncFII(SARC14)	0	1	0	31141	24
Total	2187	413	90		

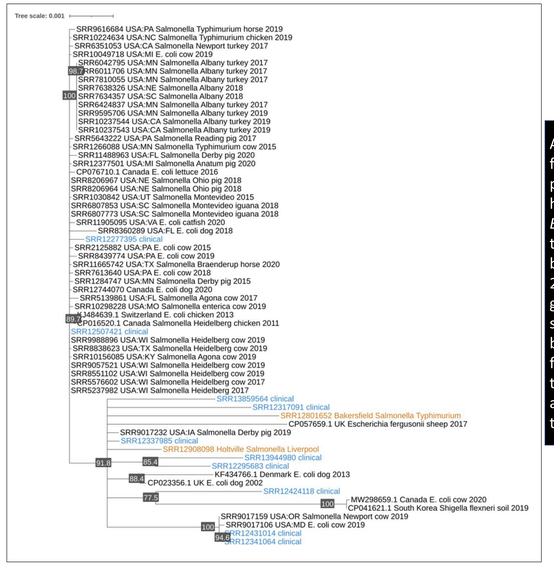
For the clinical clade, ≥64% of the accessory genes were carried on plasmids. More plasmid types (n = 20) were observed in the clinical clade than in the literature for *Salmonella* Newport. The IncFII(S) plasmid type was the only plasmid present in all clinical isolates.

The 512 farm isolates, consisting of 30 different *Salmonella* serovars, carried 14 known plasmid types. No plasmid type was common to all farm isolates and 346 isolates had no plasmid contigs.

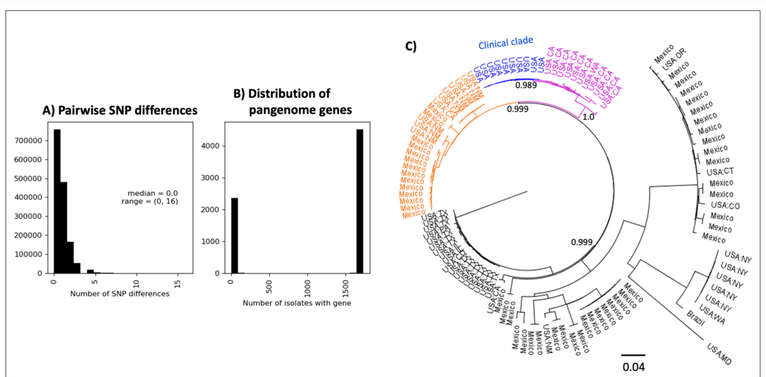
Plasmid types in red were identified in both the clinical and farm isolates.



Amongst the highly similar clinical and farm plasmids was a cryptic plasmid (~4.2 kbp, 6 genes) previously observed across the *Enterobacteriaceae*. The clinical variants differed by 6 SNPs and had been collected from two patients in Oregon and Missouri separated by 13 days. The five farm plasmids (4 *Salmonella* Corvallis, 1 *Salmonella* Liverpool) were collected on the same day at the Holtville collection site. The four Holtville *Salmonella* Corvallis plasmids differed from each other by 2 SNPs and the *Salmonella* Liverpool plasmid by 62 SNPs. The phylogeny shows that the two clinical plasmids and four Holtville *Salmonella* Corvallis plasmids belonged to sister subclades, differing by 29 to 41 SNPs.

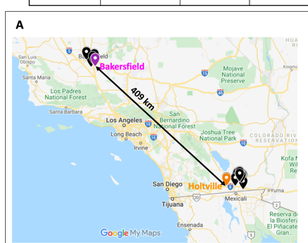


Amongst the highly similar clinical and farm plasmids were IncI1-I(gamma) plasmids (~85 kbp, ~90 genes) which have been observed across the *Enterobacteriaceae*. The pangenome of the clinical and farm plasmids and the best BLAST hits from the NCBI contained 286 genes with 28 core genes. The core gene phylogeny of these plasmids showed that eight of the clinical plasmids belonged to a subclade with the two farm isolates. Hierarchical clustering of the pangenome gene presence and absence matrix had similar groupings as the core gene phylogeny.



SNP analysis confirmed the previous findings of the FDA and CDC outbreak investigation. The clinical clade formed a single, nearly-clonal clade indicative of a single source. The closest-related environmental isolates from the NCBI Pathogen Detection database were from California, indicating ancestry going back a decade.

The core genome of the clinical clade was large (4,399 genes) compared to the median number of genes per genome (4,512 genes). Nonetheless, there was a substantial accessory genome (2,577 genes) that was sparsely distributed—genes occurred in 6% or less of the isolates, indicative of horizontal transfer.

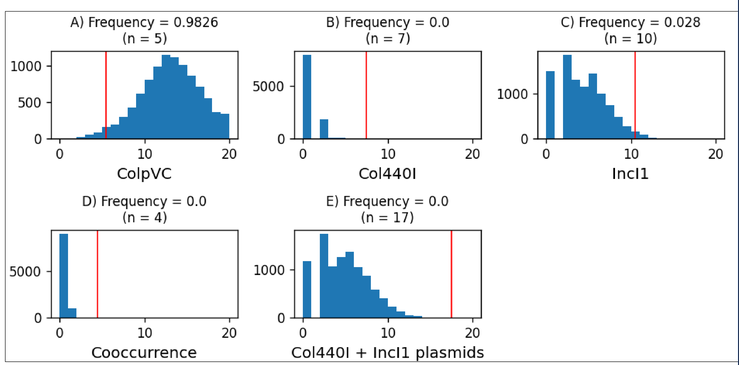


B

Isolate	Source	Serovar	Col440I	ColpVC	Uncharacterized	IncI1-I(gamma)
SRR13785991	Clinical	Newport	1	0	0	0
SRR13913413	Clinical	Newport	0	1	0	0
SRR12422240	Clinical	Newport	0	0	1	0
SRR12424118	Clinical	Newport	0	0	1	1
SRR12431014	Clinical	Newport	0	0	0	1
SRR13944980	Clinical	Newport	0	0	0	1
SRR12277395	Clinical	Newport	0	0	0	1
SRR12295683	Clinical	Newport	0	0	0	1
SRR12317091	Clinical	Newport	0	0	0	1
SRR12337985	Clinical	Newport	0	0	0	1
SRR12341064	Clinical	Newport	0	0	0	1
SRR12507421	Clinical	Newport	0	0	0	1
SRR13859564	Clinical	Newport	0	0	0	1
SRR12908953	Farm (Holtville)	Corvallis	1	1	1	0
SRR12908061	Farm (Holtville)	Corvallis	1	1	1	0
SRR12908060	Farm (Holtville)	Corvallis	1	1	1	0
SRR12966139	Farm (Holtville)	Corvallis	1	1	1	0
SRR12908098	Farm (Holtville)	Liverpool	0	0	1	1
SRR12800912	Farm (Bakersfield)	Idikan	1	0	0	0
SRR12801140	Farm (Bakersfield)	Idikan	1	0	0	0
SRR12801652	Farm (Bakersfield)	Typhimurium	0	0	0	1

We identified high similarity (≥95% identity, ≥90% alignment coverage) plasmids that might have undergone recent horizontal transfer between the clinical and farm isolates. There were 14 plasmids from 13 clinical isolates with high similarity to 17 plasmids from 8 farm isolates—comprising 3 known and 1 uncharacterized plasmid type.

The 8 farm isolates with the plasmids had been collected from 2 sampling sites separated by 409 km: 1) water samples from the New River in Seeley, California (about 30 km West of Holtville); 2) soil samples near an irrigation filling station next to the Bakersfield onion farm.



We tested if the number of highly similar plasmids in the clinical and farm isolates was due to originating from the same regional microbiota through a resampling experiment. The null hypothesis was that just as many high similarity ColpVC, Col440I, and IncI1-I(gamma) plasmids could be found in randomly sampled environmental *Salmonella enterica* isolates (those in the NCBI Pathogen Detection database, collected across the world) as were found in the farm isolates. The uncharacterized plasmid type was excluded because we could not confidently identify all instances in the clinical and farm isolates.

The frequency of observing the outbreak counts or higher for individual plasmid types was high for the ColpVC plasmid type (98.3%), but 0% for the IncI1-I(Gamma) and Col440I plasmid types. Isolates simultaneously carrying two or more highly similar plasmids was never observed. There was a low frequency (0.6%) of observing two of each plasmid type and the frequency of observing a similar sum (n = 17) of Col440I and IncI1-I(Gamma) plasmids as in the outbreak was 0%. These results led us to reject our null hypothesis.