



Molecular Epidemiology of Salmonellosis in Florida, USA, 2017–2018

Nitya Singh¹, Xiaolong Li², Elizabeth Beshearse¹, Jason L. Blanton³, Jamie DeMent^{4,5} and Arie H. Havelaar^{1*}

¹ Animal Sciences Department, Emerging Pathogens Institute, Food Systems Institute, University of Florida, Gainesville, FL, United States, ² Department of Environmental and Global Health, Emerging Pathogens Institute, University of Florida, Gainesville, FL, United States, ³ Bureau of Public Health Laboratories, Florida Department of Health, Jacksonville, FL, United States, ⁴ Independent Researcher, Orlando, FL, United States, ⁵ Food and Waterborne Disease Program, Florida Department of Health, Tallahassee, FL, United States

OPEN ACCESS

Edited by:

Qinning Wang,
New South Wales Health
Pathology, Australia

Reviewed by:

Yasser M. Sanad,
University of Arkansas at Pine Bluff,
United States
Zhenwen Zhou,
Guangzhou Medical University, China

*Correspondence:

Arie H. Havelaar
ariehavelaar@ufl.edu

Specialty section:

This article was submitted to
Infectious Diseases - Surveillance,
Prevention and Treatment,
a section of the journal
Frontiers in Medicine

Received: 21 January 2021

Accepted: 22 March 2021

Published: 22 April 2021

Citation:

Singh N, Li X, Beshearse E,
Blanton JL, DeMent J and
Havelaar AH (2021) Molecular
Epidemiology of Salmonellosis in
Florida, USA, 2017–2018.
Front. Med. 8:656827.
doi: 10.3389/fmed.2021.656827

The state of Florida reports a high burden of non-typhoidal *Salmonella enterica* with approximately two times higher than the national incidence. We retrospectively analyzed the population structure and molecular epidemiology of 1,709 clinical isolates from 2017 and 2018. We found 115 different serotypes. Rarefaction suggested that the serotype richness did not differ between children under 2 years of age and older children and adults and, there are ~22 well-characterized dominant serotypes. There were distinct differences in dominant serotypes between Florida and the USA as a whole, even though *S. Enteritidis* and *S. Newport* were the dominant serotypes in Florida and nationally. *S. Javiana*, *S. Sandiego*, and *S. IV 50:z4, z23:-* occurred more frequently in Florida than nationally. Legacy Multi Locus Sequence Typing (MLST) was of limited use for differentiating clinical *Salmonella* isolates beyond the serotype level. We utilized core genome MLST (cgMLST) hierarchical clusters (HC) to identify potential outbreaks and compared them to outbreaks detected by Pulse Field Gel Electrophoresis (PFGE) surveillance for five dominant serotypes (*Enteritidis*, *Newport*, *Javiana*, *Typhimurium*, and *Bareilly*). Single nucleotide polymorphism (SNP) phylogenetic-analysis of cgMLST HC at allelic distance 5 or less (HC5) corroborated PFGE detected outbreaks and generated well-segregated SNP distance-based clades for all studied serotypes. We propose “combination approach” comprising “HC5 clustering,” as efficient tool to trigger *Salmonella* outbreak investigations, and “SNP-based analysis,” for higher resolution phylogeny to confirm an outbreak. We also applied this approach to identify case clusters, more distant in time and place than traditional outbreaks but may have been infected from a common source, comparing 176 Florida clinical isolates and 1,341 non-clinical isolates across USA, of most prevalent serotype *Enteritidis* collected during 2017–2018. Several clusters of closely related isolates (0–4 SNP apart) within HC5 clusters were detected and some included isolates from poultry from different states in the US, spanning time periods over 1 year. Two SNP-clusters within the same HC5 cluster included isolates with the same multidrug-resistant profile from both humans and poultry, supporting the epidemiological link. These clusters likely reflect the vertical transmission of *Salmonella* clones from higher levels in the breeding pyramid to production flocks.

Keywords: hierarchical clustering, outbreak detection, phylogeny, SNP, cgMLST, mlst, whole genome sequencing, *Salmonella enterica*

INTRODUCTION

Nontyphoidal *Salmonella enterica* (hereafter referred to as *Salmonella*) is a significant public health problem and one of the most frequent cause of bacterial gastroenteritis worldwide (1, 2). *Salmonella* is a leading cause of foodborne illness in the United States, causing about 1.2 million annual cases annually (3). The state of Florida has consistently observed incidence rates of salmonellosis that are approximately twice as high as the national average (4). Annually, 6,000–7,000 cases of salmonellosis are confirmed in the state, causing a high burden on surveillance systems including detection of outbreaks and implementation of control measures.

Several well-established and standardized methods have been commonly used for the subtyping of *Salmonella*, including sero- and phage typing, pulsed-field gel electrophoresis (PFGE), multi-locus variable numbers tandem repeat analysis (MLVA), and multi-locus sequence typing (MLST). PFGE has been the gold standard method for the subtyping of *Salmonella* (5), and gained widespread adoption under PulseNet USA, a surveillance network of public health laboratories across the United States and other countries (6). However, the limited resolution of PFGE for closely related strains, specifically those involved in outbreaks has impeded its utility as an effective subtyping tool. For example, for *S. enterica* serotype Enteritidis, reported as one of the most prevalent serotypes among laboratory confirmed cases (7), a considerable number of isolates (45%) received by PulseNet USA belonged to indistinguishable PFGE patterns using the XbaI restriction enzyme (8), whereas, 74% of *S. Enteritidis* isolates were found to comprise of only three unique PFGE patterns in another study (9). Although, additional restriction enzymes could be used to increase the discriminatory power of PFGE (10), the laboratory protocol would also become more time consuming and expensive. Similarly, despite offering better discrimination, MLVA has had limited use because of its inability to differentiate common patterns (11), and standardized PulseNet protocols have been established for serotypes Enteritidis and Typhimurium only. Legacy (7-gene based) MLST has been a valuable tool for inferring evolutionary descent and clonal diversification among genetic lineages within bacterial population structures (12, 13). However, the use of only seven housekeeping genes fails to capture genetic heterogeneity within strains of the same serotype and suffers from lower discriminatory power (14).

In recent years, decreasing costs of sequencing technologies and advances in bioinformatics tools have greatly facilitated the development of whole-genome sequencing (WGS) methods that have been transformational for the surveillance of foodborne pathogens in the public health infrastructure (15). With the ultimate subtyping resolution delivered by interrogating entire genomes, an expansion of studies have demonstrated the utility of WGS to enhance our capacity to detect outbreaks (16–18), conduct surveillance (19) and understand the genomic population structure of *Salmonella* (8, 14).

Core genome MLST (cgMLST) offers a reliably fast method for identifying closely related strains within the same serotypes and is used e.g., by the Centers for Disease Control and Prevention (CDC) for outbreak detection (20). Precompiled assemblies and

genotyped strains for all *Salmonella* WGS reads submitted to NCBI are available at Enterobase (<http://enterobase.warwick.ac.uk/>), an online resource for analyzing and visualizing genomic variation within enteric bacteria (14, 21). This offers an attractive solution to identify closely related genetic isolates, which can be further evaluated as potential outbreaks via more extensive SNP-based protocols. WGS data can also be highly valuable in source identification of clinical cases through comparison with isolates from food, animals and other non-clinical sources which are made available to NCBI mainly through GenomeTrakr (22).

The Florida Department of Health's Bureau of Public Health Laboratories (BPHL) started to implement WGS in 2016, with full coverage since 2019 and employed PFGE-based subtyping for molecular surveillance through 2018 in collaboration with CDC within PulseNet USA. In 2019, PulseNet transitioned to WGS-based methods and BPHL also ceased the use of PFGE for *Salmonella* isolate typing. WGS data are routinely shared with CDC, who analyze the data for multistate outbreaks, while detecting outbreaks restricted to the state of Florida is the responsibility of the Florida Department of Health (FDOH). An efficient tool for analyzing the high volume of sequence data is needed for this purpose. Phylogenetic methods, while highly specific, are using computationally intensive and time-consuming methods and thus present a challenge for efficient outbreak detection in the state. In the current work we are proposing a two-step “combination approach” (cgMLST-SNP phylogeny) which can be applied for different purposes like outbreak detection and identification of case- clusters and putative sources. As a first step, cgMLST can be used to narrow down the number of genetically related isolates at allelic level and in a second step, SNP-based phylogeny can identify genetic relatedness of isolates at a more detailed level.

Against this background, the research objectives of this study were to:

- 1) Understand the population structure of human *Salmonella* isolates in Florida, using molecular serotyping and multi-locus sequence typing (7-gene (legacy) MLST, core genome (cg) MLST);
- 2) Evaluate the potential of whole genome sequencing data analysis to improve detection of foodborne disease outbreaks for all serotypes of *Salmonella* while accounting for the large volume of isolates in Florida;
- 3) Understand the genetic relationship between clinical isolates from putative outbreaks and non-clinical *Salmonella* isolates from Florida and elsewhere in the United States collected during 2017–2018.

MATERIALS AND METHODS

Data

Whole genome sequencing data used for the current work were downloaded from public repositories (NCBI and Enterobase), while metadata used for data collection and management were collected from FDOH and the US Department of Agriculture Food Safety and Inspection Service (USDA FSIS).

WGS Data Generation

BPHL in Jacksonville, FL receives *Salmonella* isolates from all diagnostic laboratories in Florida under updated Rule 64D-3.029 and generates WGS data (since 2016) for surveillance purposes, following a standardized CDC sequencing protocol (23) as follows: isolates were grown on trypticase soy agar plates with 5% sheep blood (Remel, Lenexa, KS, USA). DNA extraction was conducted using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). DNA libraries were constructed using Nextera XT DNA Library Prep Kit (Illumina, Inc., San Diego, CA, USA). The genomic libraries were sequenced on the Illumina MiSeq system using a combination of v2 (500 cycle) and v3 (600 cycle) chemistries. BPHL stores all sequence and related metadata in a Bionumerics database. All sequenced data are shared with CDC for surveillance purposes. These data are submitted to the NCBI BioSample database (<https://www.ncbi.nlm.nih.gov/biosample/>), under Bioproject number PRJNA230403 with limited metadata information.

Metadata Collection and Management

An integrated and deidentified metadata set from BPHL Bionumerics and FDOH Merlin database was made available by FDOH under a data sharing agreement, details are provided in Li et al. (4). We selected isolates from metadata received from FDOH, with a date of sample collection by the clinical laboratory from January 2017 to December 2018 and downloaded sequence data from Enterobase (<https://enterobase.warwick.ac.uk>) using PNUSA numbers specified in the metadata (https://dataverse.harvard.edu/dataverse/Salmonella_FL). After data cleaning [as detailed in Li et al. (4), involved removal of isolates with missing data or duplicated in the database] a curated set of 1,709 isolates was available for further analysis. We additionally included WGS data from Enterobase of 230 isolates collected from non-clinical sources (including soil/water, food, poultry, animal sources) in Florida and 1,317 *S. Enteritidis* isolates collected from poultry sources from other states during same time period, using identifiers from NCBI. We segregated clinical and non-clinical isolates into four different isolate sets, comprising:

- 1) “*all clinical*,” a set of all 1,709 clinical isolates for exploring outbreak detection by cgMLST clustering and SNP based phylogeny;
- 2) “*sporadic clinical*,” a set of 1,632 isolates including 1,595 sporadic cases and 37 first occurrences of isolated from outbreaks flagged by PFGE surveillance for exploring serotype, serotype diversity, and sequence type population structure;
- 3) “*sporadic clinical and non-clinical*,” a set of 1,862 isolates, obtained by adding non-clinical isolates (230) to the sporadic clinical isolates, for exploring any connection between clinical and non-clinical isolates;
- 4) “*Enteritidis poultry*,” a set of 1,341 isolates of *S. Enteritidis* collected during 2017–2018, from poultry sources across the USA (including 24 Florida isolates), for exploring possible genetic links between clinical isolates of *S. Enteritidis* in Florida and poultry isolates across USA.

Additional metadata (day, month, year of collection) for all non-clinical isolates were provided by the US Department

of Agriculture Food Safety and Inspection Service through a Freedom of Information Act request.

All datasets were cleaned, organized, and visualized in the statistical language R version 3.6.0.

Serotype Diversity

All isolates were serotyped by employing the *in-silico* typing resource SISTR (<https://lfz.corefacility.ca/sistr-app/>), as implemented in Enterobase. We submitted ambiguous serotype assignments to the developers, who resolved them using SISTR ver. 2 (24).

Rarefaction analysis was performed on “*sporadic clinical*” isolates set, based on Hill numbers (qD) to represent serotype diversity (25) for $q = 0, 1, \text{ and } 2$ using R package “*rareNMtests*.” Hill numbers are also referred as *order of diversity* and differ in sensitivity of diversity values to rare vs. abundant serotypes. The exponent q modulates the type of weighted mean and with $q = 0 \sim$ weighted harmonic mean (serotype richness), $q = 1 \sim$ geometric mean (exponential Shannon index representing “typical” serotypes) and $q = 2 \sim$ arithmetic mean (inverse Simpson index representing effective number of “dominant” serotypes). Rarefaction was done for all cases and separately for cases <2 and ≥ 2 years of age.

A tanglegram was created to compare the rankings of top 20 serotypes in Florida and in the whole USA. The USA data was obtained from the National Enteric Disease Surveillance: Salmonella Annual Report, 2016 (26), which is the most recent available report.

Multilocus Sequence Typing MLST-Based Phylogeny

Enterobase implements both the traditional MLST approach based on sequence profiles of seven housekeeping genes (legacy MLST) and core genome MLST (cgMLST). Based on legacy MLST, eBurst groups (eBG) are defined as clusters of isolates sharing at least 6 common allelic sites. A minimum spanning tree (MSTree) based on legacy MLST was generated using the Enterobase MSTreeV2 algorithm and was visualized in GrapeTree (21) for the “*sporadic clinical*” isolate set with color-coding by eBG and labeled by SISTR predicted serotype. We also cross-tabulated the eBG groups associated with effective number of serotypes in the state as identified by rarefaction and corresponding legacy MLST Sequence Types.

Core genome MLST profiles were created using the Enterobase platform (14) for all four samples sets and MSTrees were created for first three isolates sets and visualized as above, while isolate set four was further filtered and used in SNP based analysis as described in sections Association of Clinical and Non-clinical Isolates and Florida Clusters. For tree construction, the distance between genomes was calculated using the number of shared cgMLST alleles and genomes were linked using single-linkage clustering. The MSTree for the “*sporadic clinical*” isolates was color-coded by SISTR predicted serotypes and the MSTree for “*sporadic clinical + non-clinical*” isolates set was color coded by source niche (clinical and non-clinical sources, *viz.*, poultry, soil/water, raw chicken, avian, food etc.), as available at NCBI. If undefined at NCBI, the source niche field for clinical isolates was

updated as human based on the FDOH metadata. Hierarchical clusters (HC) of closely associated isolates having cgMLST allelic difference of $\leq 2, 5, 10,$ and 20) were generated and stable cluster group numbers referred as HC2, HC5, HC10, and HC20, respectively, were assigned to isolates in the “*all clinical*” set.

cgMLST Hierarchical Clustering for Outbreak Detection

We compared PFGE-based outbreak detection as implemented in Florida in 2017–2018 with the potential results that could have been obtained with cgMLST-based hierarchical clustering analysis. This comparison was restricted to the five most commonly detected serotypes in the “*all clinical*” set. HC5 clusters with >2 isolates were selected for further analysis and joined with all outbreak-associated isolates for selected serotypes, as detected by PFGE surveillance. The selected serotypes included: Enteritidis (100 isolates), Newport (50 isolates), Javiana (13 isolates), Bareilly (38 isolates), and Typhimurium including the monophasic variant I 4, [5],12:i- (32 isolates).

SNP-Based Phylogeny of cgMLST Hierarchical Clusters

SNP based phylogenies were constructed for all selected serotypes in the previous step. Core-genome single-nucleotide polymorphisms (core-SNPs) were identified using the Snippy pipeline v4.3.8 (<https://github.com/tseemann/snippy>). All isolates of selected serotypes belonged to same eBG type, justifying the choice of one reference genome except *S. Newport* for which we chose a reference genome close to the dominant eBG type in the sample set. Reference genomes for each selected serotype were *S. Enteritidis* (P125109, accession number: AM933172), *S. Newport* (AML0800465, accession number: AANRFQ010000001), *S. Bareilly* (CFSAN000189, accession number: CP006053), *S. Javiana* (CFSAN001992, accession number: NC_020307), *S. Typhimurium* (LT2, accession number: NC_003197), and their respective annotated GenBank files were downloaded from NCBI. Assembled genomes for all serotypes were mapped against the corresponding reference genome using *BWA-mem* (27) and *minimap2* (28), and SNPs were called using *SAMtools* (29) and *FreeBayes* (<https://github.com/ekg/freebayes>). SNPs located in repetitive and recombinogenic regions were removed prior to phylogenetic analysis using *Gubbins* v2.3.4. (30) Only parsimony informative polymorphic sites extracted from the *Gubbins* output were used for further analysis. SNP distance matrices were calculated using the *snp-dists* tool (<https://github.com/tseemann/snp-dists>) and imported in R along with metadata information about HC5 cluster labels, isolate dates, PFGE profile, and zip codes. Maximum Likelihood Trees with 1,000 bootstrap replications were generated using *iqTree* V1.6.10 package with best fitting nucleotide evolutionary model identified by *ModelFinder* (31). Bootstrap supported trees were imported into R and were annotated/color-coded using metadata to reflect isolate date and HC5 cluster and were visualized using package *ggtree* (32). Further, for all major HC5 clusters SNP distance-based heatmaps and annotated trees were generated using *ComplexHeatmap* (33), *ape* (34), and *ggtree*. A detailed manual comparison was done

between outbreak detection based on PFGE profiling and HC5 clusters, while taking into account phylogenetic relationships as defined by SNP analysis.

Spatial Distribution of HC5 Clusters

Isolates assigned to HC5 clusters of size >10 in the “*all clinical*” set were plotted on a map of Florida at the zip-code level (only available for clinical isolates) to explore their spatial distribution. Standard distance was used to quantify the spatial distribution of zip-code areas involved in each HC5 cluster. Similar to the concept of standard deviation, standard distance is a metric in spatial statistics measuring the dispersion of features around their geographical mean center which is constructed from the mean x-coordinate and mean y-coordinate (35). The standard distance is given as:

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n} + \frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n}} \quad (1)$$

where x_i and y_i are the coordinates of feature i and \bar{X} and \bar{Y} represent the mean center of features. Euclidean distances between each pair of centroids of zip-code areas within clusters were calculated, and the distance matrices were used to plot heatmaps for visualization.

Association of Clinical and Non-clinical Isolates

MSTrees for the “*sporadic clinical and non-clinical*” isolates were constructed as described in section WGS Data Generation to aid the comparative visualization of possible shared sources of isolates. We compared HC5 cluster profiles of 1,862 Florida isolates to identify any matching non-clinical isolates in HC5 clusters of clinical isolates.

As a proof of principle, we applied the proposed combination approach for investigating genetic links between clinical and non-clinical isolates of *S. Enteritidis* (the most prevalent serotype in Florida and across the USA) (26). One seventy six clinical and 24 non-clinical *S. Enteritidis* isolates collected from Florida were first filtered for shared HC5 cluster profiles and were then further investigated by SNP based phylogenetic analysis, as above.

Guided by the results of this step, we additionally created HC5 cluster profiles of 1,317 poultry isolates of *S. Enteritidis* collected during 2017–2018 from other states and included them in above comparison to explore the infection sources across the US. Additionally, presence of antimicrobial resistance (AMR) genes was investigated using NCBI’s AMRFinder (36), using online server (37), for clinical and poultry isolates that shared HC5 cluster profiles.

RESULTS

Serotype Diversity

The serotype diversity was calculated using the “*sporadic clinical*” set of 1,632 clinical isolates (Figure 1). There were 115 serotypes in the set (0D), the rarefaction curve suggests that significantly more serotypes are circulating in Florida as the Hill curve for $q = 0$ continued to increase with increasing number of isolates in the

set. Many of the observed serotypes were rare, as demonstrated by the much lower number of ~ 36 typical serotypes ($q = 1$), while there were ~ 22 dominant serotypes ($q = 2$). Among 1,155 cases of age ≥ 2 years, there were 95 serotypes ($q = 0$), ~ 32 typical serotypes ($q = 1$) and ~ 18 dominant serotypes ($q = 2$), while for 477 cases of age < 2 years, there were 76 serotypes ($q = 0$), with ~ 32 typical serotypes ($q = 1$) and ~ 20 dominant serotypes ($q = 2$).

The diversity of *Salmonella* isolates in Florida appeared to be higher than in the US as a whole (Figure 2). In our dataset, the top-5 serotypes comprised 27% of all isolates, whereas, in the US this was 47% (26). There is also a striking difference in the dominance of serotypes between Florida and the US. The top two serotypes were the same (Enteritidis and Newport) while Javiana ranked third in Florida vs. fourth in the US. Sandiego, the fourth ranking serotype in Florida did not even appear in the top-20 serotypes in the US. Likewise, serotype IV 50:z4, z23:- (formerly Flint) was a frequent occurrence in Florida but not nationally. On the other hand, serotypes I 4, [5],12:i:- and Infantis ranked relatively low in Florida.

MultiLocus Sequence Typing

Legacy MLST Analysis

We detected 186 Sequence Types (ST) and 110 eBG among 1,632 sporadic isolates (Figure 3). The top 24 eBurst Groups (eBGs) associated with the 22 dominant serotypes and contributing STs

are shown in Table 1. Many eBGs included only one serotype, and in most cases, one ST dominated in the eBG. For example, all except 2 *S. Enteritidis* isolates were ST11, *S. Sandiego* was dominated by ST126 while *S. Javiana* consisted mainly of ST24. *S. Newport* had most divergent ST types, including ST118, ST46, ST45, and ST5, and was spanning over two eBG, namely two and three. Note that *S. Typhimurium* and its monophasic variant (I 4, [5],12:i:-) were included in the same eBG and both shared STs 19 and 34 interchangeably. eBG 23 included seven serotypes from *Salmonella enterica* subsp. *houtenae* (group IV). Serotypes like Rubislaw, Poona, and Saintpaul were grouped in two eBG.

Core Genome MLST Analysis

Even at the cgMLST level, most serotypes were clustered together (Figure 4). Hierarchical clustering at HC2, HC5, HC10, and HC20 level identified 1,476, 1,360, 1,100, and 692 clusters, respectively. Of these, 29, 45, 77, and 104 included >2 isolates. MSTrees for all HC levels are provided in Supplementary Figure 1. The number of hierarchical clusters, even at the HC2 level, exceeded the number of detected outbreaks during the study period. This raised the question, which we investigated further, whether HC clustering can be a useful tool to select closely related isolates and reduce the size of isolate sets for more time consuming and computationally expensive SNP-based phylogenetic analysis. HC10 resulted in large clusters, and we considered this level as insufficiently discriminatory for outbreak

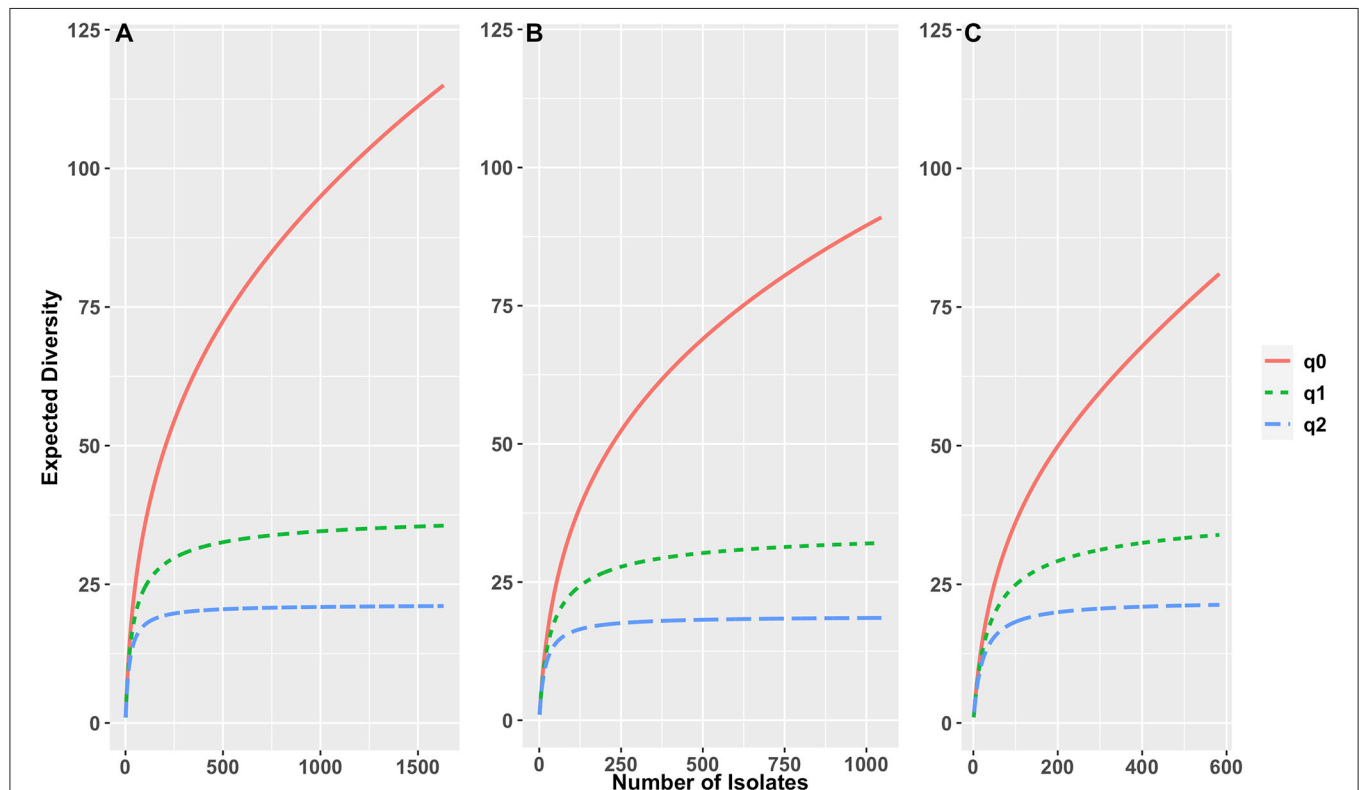
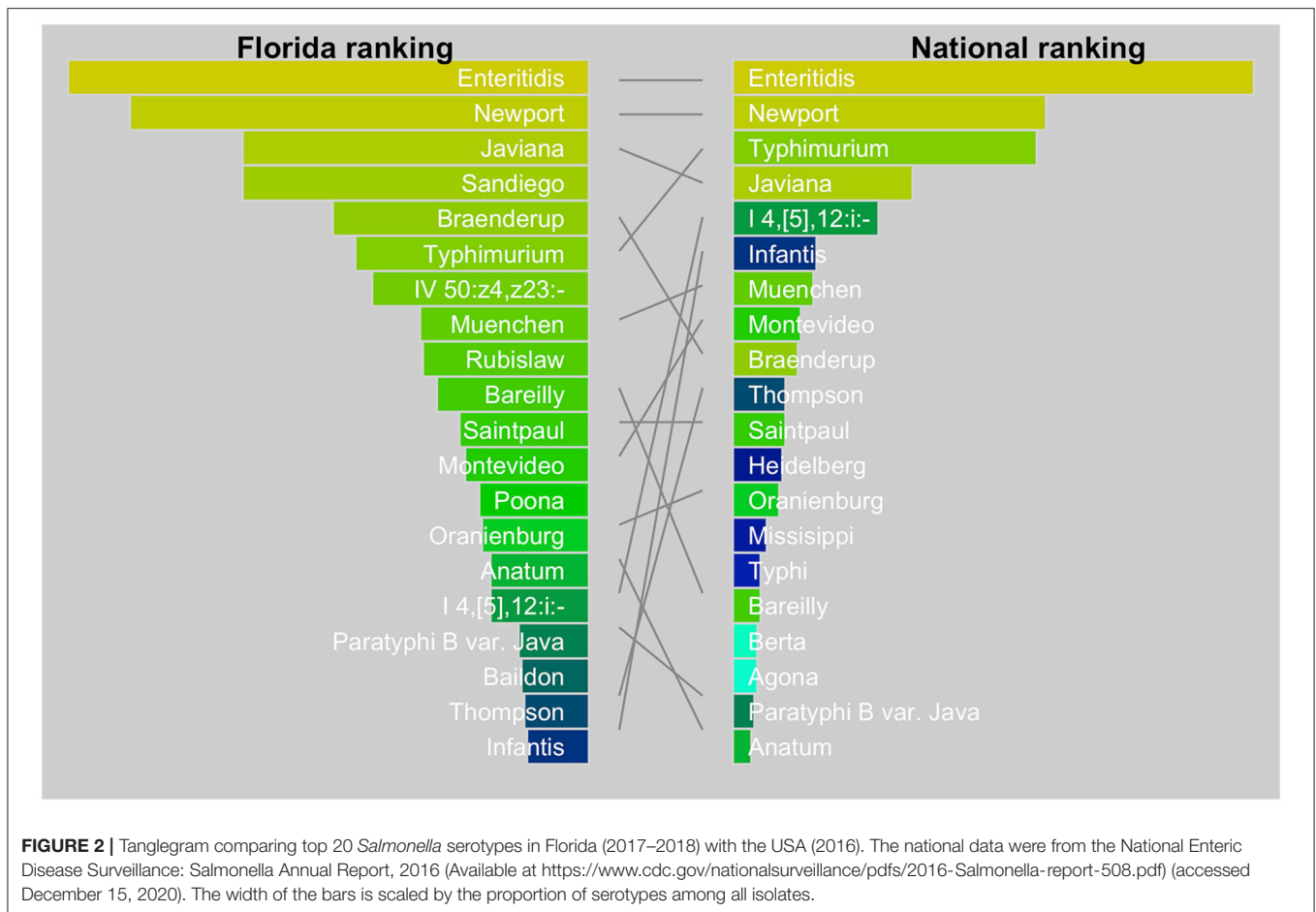


FIGURE 1 | Rarefaction curves based on sample size-based diversity accumulation, for (A) 1,632 *S. enterica* isolates from all ages, (B) 1,155 isolates of age ≥ 2 years, (C) 477 isolates of age < 2 years, from Florida, USA 2017–2018. X-axis shows number of isolates, y axis shows expected diversity using three Hill numbers indicating the total number of serotypes in the set ($q = 0$), the number of “typical” serotypes ($q = 1$) and the number of “dominant” serotypes ($q = 2$).

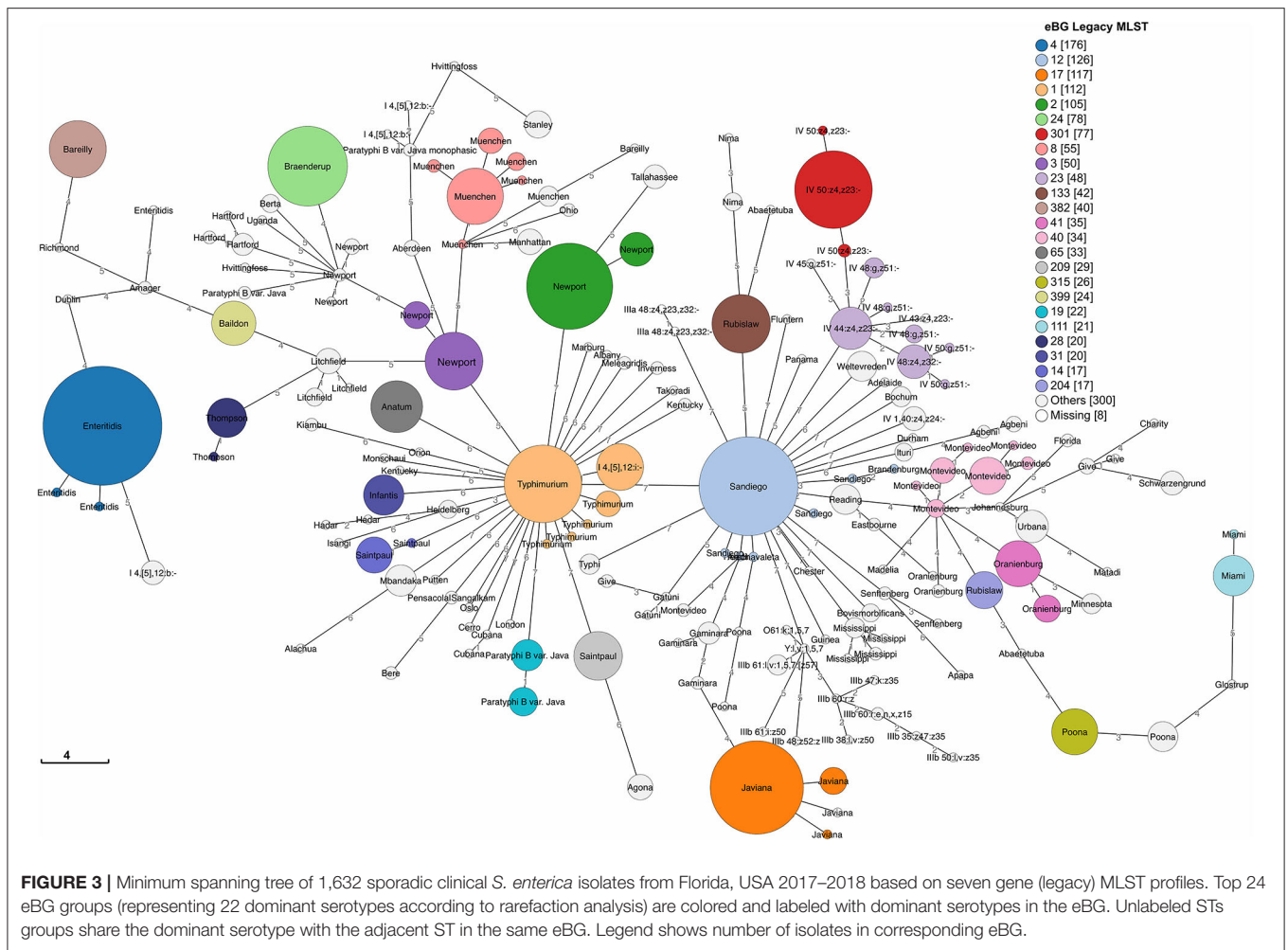


detection, and based our analysis on HC5 clustering. HC2 clustering was evaluated after having performed phylogenetic analysis on all isolates in HC5 clusters (see section SNP Based Phylogenies and cgMLST Clustering for Outbreak Detection).

SNP Based Phylogenies and cgMLST Clustering for Outbreak Detection

We compared PFGE-based outbreak detection as implemented in Florida in 2017–2018 with the potential results that could have been obtained with cgMLST-based hierarchical cluster analysis. As a proof of principle, this comparison was restricted to the five most commonly detected serotypes while adding the monophasic variant of *S. Typhimurium* (I i4, [5], 12:i-). **Figure 5** shows phylogenetic trees for each serotype; detailed SNP-distance trees for HC5 clusters are provided in the **Supplementary Figures 2–6**. The phylogenetic tree of 100 isolates of *S. Enteritidis* showed good agreement with HC5 clustering, including 14 well-segregated clades, with 5 large clusters (87, 150, 182, 614, 10584) comprising 13–14 isolates and 9 smaller clusters with 3–4 isolates. Isolates in all five large HC5 clusters had different SNP ranges, 87: 0–24 SNPs, 150: 0–21 SNPs, 182: 0–28 SNPs, 614: 0–16 SNPs, and 10584: 0–12 SNPs. In all smaller clusters the isolates were 0–10 SNPs apart, ranging from 0–2 for HC5 58778 and up to 4–10 for 50773. In contrast, three

smaller clades had multiple HC5 clusters: one, most distant from others, had 2 isolates in different HC5 clusters 19036, 40641 with 0 SNPs difference; second at the very bottom, had 4 isolates, 3 in HC5 26508 within 2–5 SNPs and 1 at 12–14 SNPs from the rest in HC5 536574; third clade had HC5 cluster 58778 with 4 isolates within 0–2 SNPs and HC5 cluster 81880 at 70 SNPs from rest. Interestingly HC5 clusters 36574 and 81880 were labeled as same PFGE and were 33 SNPs apart. For *S. Newport*, three distant clades were observed corroborating their different eBG types. Among 50 isolates, 36 isolates corresponding to dominant eBG type clustered in one clade corresponding to HC5 cluster 63415. The second clade included three HC5 clusters, 113949 and 13124 clustered together within 1–6 SNPs and the third cluster 75478 was ~250 SNPs apart from these. The third clade included five isolates, from two outbreaks (10–15 SNPs apart) and one related isolate. Three clades were observed for 13 isolates of *S. Javiana*; one clade included six isolates from six different HC5 clusters (2–31 SNPs apart) and two other clades included four and three isolates from the same HC5 cluster. Within these latter clusters, isolates were 8–27 SNPs different. Thirty-eight isolates of *S. Bareilly* clustered in two clades. One larger clade included 35 isolates from one HC cluster and included 12 isolates from recognized outbreaks. This cluster was quite heterogeneous with SNP distances ranging between 2 and 27. Interestingly, two



isolates from one of these outbreaks and a related isolate were assigned to another HC5 cluster that was also clearly separated by SNP analysis. 17 isolates of *S. Typhimurium* clustered separately from 15 isolates of the monophasic variant *S. I 4, [5],12:i:-*, while one *S. Typhimurium* isolate clustered with the monophasic variant by both HC5 clustering and SNP analysis and *vice versa*. 8 *S. Typhimurium* isolates grouped in one clade, all belonging to the same HC5 cluster while the other isolates were singlets or doublets. Fifteen isolates of *S. I 4, [5],12:i:-* grouped in two clades supported by HC5 clustering, with one exception.

After comparing HC2 and HC5 cluster profiles (available in **Supplementary Data Sheet 2**) in all SNP -phylogenies, we observed that HC2 was too discriminatory, while HC5 clusters showed well-segregated clades, in SNP- based phylogenetic trees. Rationale for this can be given as 2 or less allelic difference at HC2 level was not sufficient to pick up enough SNP signals for creating reliable SNP phylogenies, while clustering at HC5 level provided an optimum solution for sample size and discriminatory power by SNP phylogeny.

We assigned isolates from clusters detected by either PFGE surveillance or HC5 clustering to different categories, depending on the degree of correspondence between different methods

(**Table 2**). PFGE surveillance detected 12 clusters involved in recognized outbreaks (6 *S. Enteritidis*, 2 *S. Newport*, 1 *S. Javiana*, 2 *S. Typhimurium*, and 1 *S. Bareilly*).

1. Two outbreaks (1 *S. Enteritidis*, 1 *S. Newport*) included isolates within one HC5 cluster, and a maximum of 3 SNP distance.
2. Two outbreaks (1 *S. Enteritidis*, 1 *S. Newport*) were characterized by isolates with the same PFGE profile, and 0–2 SNP difference but nevertheless, belonged to different HC5 clusters. Isolates from the *S. Enteritidis* outbreak (1704MLJEG-2) were in the same HC10 cluster. Noticeably these two isolates were very distinct from the other 98 in *S. Enteritidis* isolates in the phylogenetic tree.
3. Four outbreaks (2 *S. Enteritidis*, 1 *S. Javiana*, 1 *S. Typhimurium*) included isolates with one PFGE profile, but two or more HC5 clusters that were supported by SNP analysis. Note that isolates from the two *S. Enteritidis* outbreaks (occurring ~2 months apart) had the same PFGE profile but belonged to three HC5 clusters.
4. One *S. Typhimurium* outbreak and one *S. Bareilly* outbreak included isolates that were different by PFGE profile, HC5 clusters, and SNP analysis. The *S. Bareilly* included isolates

TABLE 1 | eBurst Groups (eBG) and sequence types (ST) corresponding to 22 dominant serotypes among 1,632 sporadic clinical *Salmonella* isolates collected during 2017–2018 in Florida, based on 7 gene MLST profiles.

eBG	#Isolates in eBG	Dominant serotype*	Other serotypes*	Sequence type (#isolates in ST)
4	176	Enteritidis		11 (174), 1925 (1), 1973 (1)
12	12	Sandiego	Arechavaleta Brandenburg Reading	126 (120), 2818 (1), 4026 (1), 5823 (1); 238 (1); 65 (1); 412 (1)
17	117	Javiana	-	24 (107), 1674 (9), 4750 (1)
1	112	Typhimurium	I 4, (5),12:i-	19 (65), 34 (1), 213 (1), 313 (1), 3975 (8), 99 (1); 34 (26), 19 (9)
2	105	Newport	-	5 (14), 118 (91)
24	78	Braenderup	-	22 (78)
301	77	IV 50:z4, z23:-		2053 (74), 2405 (2), 4008 (1)
8	55	Muenchen		82 (1), 112 (39), 1567 (4), 2870 (8), 3795 (2), 4195 (1)
3	50	Newport		45 (9), 46 (41)
23	48	IV 48:g, z51:-	IV 48:z4,z32:- IV 44:z4,z23:- IV 50:g,z51:-	433 (14), 454 (1), 782 (1), 1869 (5); 2058 (14); 433 (8); 454 (3), 1209 (1), 2197 (1)
133	42	Rubislaw		94 (42)
382	40	Bareilly		2553 (40)
41	35	Oranienburg		23 (26), 1675 (9)
40	34	Montevideo		4 (8), 81 (17), 195 (4), 316 (2), 699 (1), 1531 (1), 4451 (1)
65	33	Anatum		64 (33)
209	29	Saintpaul		95 (29)
315	26	Poona		964 (26)
399	24	Baildon		2166 (24)
19	22	Paratyphi B		88 (10), 127 (12)
111	21	Miami	Sendai	80 (15), 4198 (1); 80 (5)
28	20	Thompson		26 (19), 4384 (1)
31	20	Infantis		32 (20)
14	17	Saintpaul		27 (1), 50 (16)
204	16	Rubislaw		1575 (16)
46	13	Poona		447 (11), 308 (1), 1069 (1)

*Serotyped using SISTR (24).

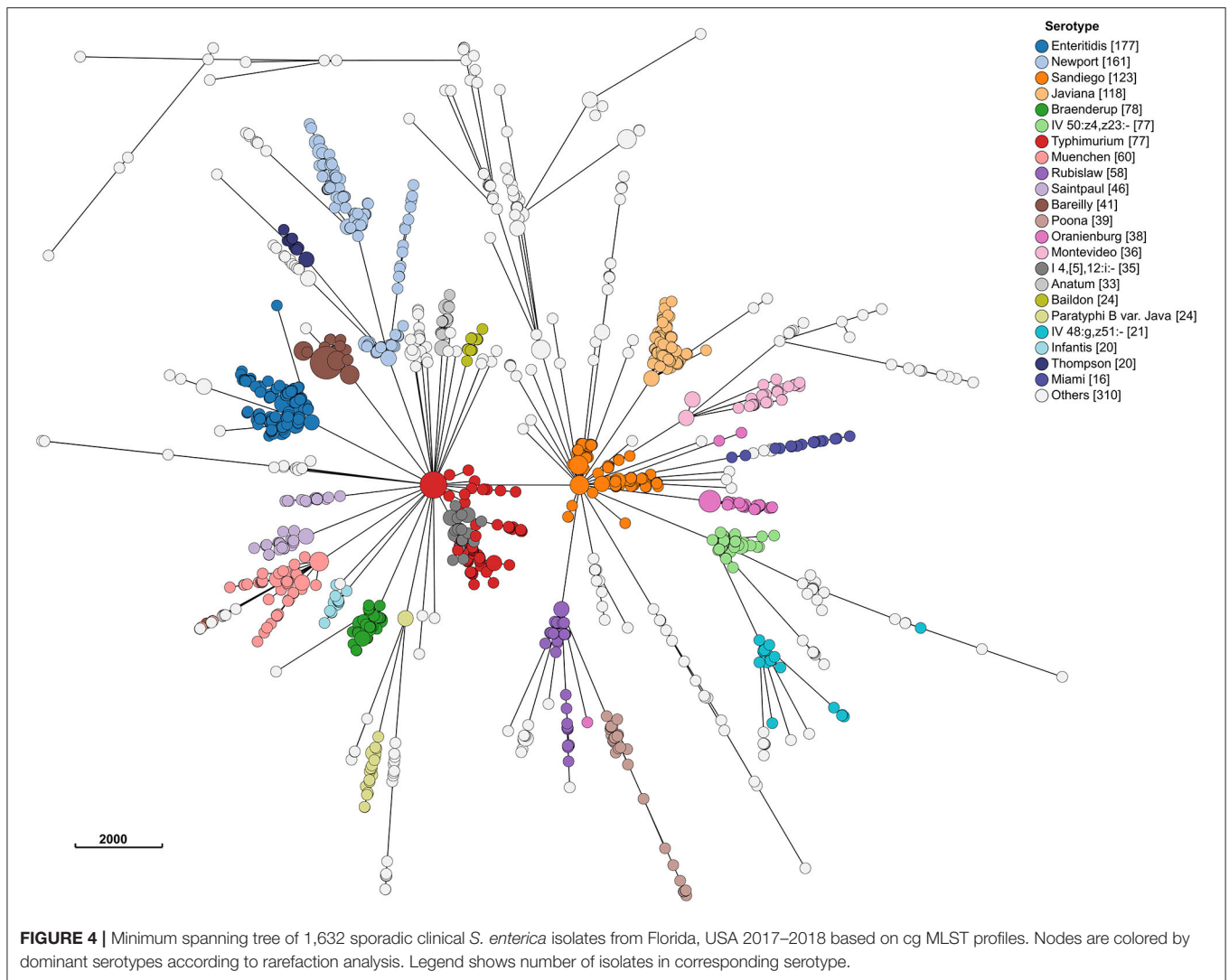
from patients up to 485 days apart. Within one HC5 cluster in this outbreak, there was one dominant PFGE profile, and isolates were up to 15 SNP different. The second HC5 cluster had the same PFGE profile but a different HC5; these isolates were more than 40 SNP different from the core cluster.

- All isolates from five HC5 clusters were collected within a 60-day window (range 1–28 days) which is commonly used for detecting outbreaks of *Salmonella* (20). Three of these clusters showed more than one PFGE profile.
- There were two *S. Enteritidis* outbreaks with multiple PFGE profiles but one HC5 cluster, supported by phylogenetic analysis. The HC5 clusters included 10 and 13 additional isolates, all closely related by SNP distance. Isolates in these clusters were obtained over a time span of 451 and 558 days, respectively.

- We detected 14 HC5 clusters [8 *S. Enteritidis*, 3 *S. Newport*, 1 *S. Javiana*, and 2 *S. i4, (5),12:i-*] that were not included in previously recognized outbreaks, all supported by phylogenetic analysis. These clusters were prolonged in time (range between first and last isolate in the cluster 63 – 558 days) and 12 of these clusters showed more than 1 PFGE profile.

Spatial Distribution of HC5 Clusters

Geospatial analysis of eight large (more than 10 isolates) prolonged case series of clinical isolates, detected by HC5 clustering, and confirmed by SNP analysis, suggested that while most clusters were confined to a specific region in Florida, there were considerable distances between the cases (Table 3, Figure 6, and Supplementary Figure 7). Clusters involved between 10



and 26 zip-code areas, and the standard distance of zip-code areas in each cluster varied between 66 and 244 km. Only three of these clusters were detected as outbreaks by PFGE surveillance.

Association of Clinical and Non-clinical Isolates

Florida Clusters

Source and serotype based color-coded MSTrees for the isolate set “sporadic clinical + non-clinical” is shown in **Supplementary Figure 8**. We detected 12 HC5 clusters that included isolates from clinical and non-clinical sources representing six serotypes (**Table 4**).

Four clusters belonged to *S. Enteritidis* and all were prolonged case series as described above. All non-clinical isolates in these *S. Enteritidis* clusters were isolated from poultry. In total, these clusters (>2 clinical isolates) comprised 55 clinical and 21 non-clinical isolates and were investigated by SNP-based phylogeny as shown in **Figure 7**. A SNP distance heatmap for all 4 clusters

is included as **Supplementary Figure 9**. Two clusters (150 and 182) were quite heterogeneous with SNP distances up to 30, while in one cluster (10584) SNP distances up to 13 were observed. Within each of these clusters, smaller subclusters with highly related human and poultry isolates were detected. One cluster (614) separated into two subclusters of which one included a poultry isolate.

Multistate *S. Enteritidis* Clusters

Because a large proportion of poultry meat consumed in Florida is produced in other states, we extended our HC5 profile comparison of 176 clinical *S. Enteritidis* isolates from Florida with 1,341 non-clinical isolates from across the USA (“*Enteritidis poultry*” set) and detected 8 HC5 clusters (>2 clinical isolates) including 67 clinical and 513 poultry isolates (21: Florida, 492: other states). Four of these clusters, 10584, 150, 182, and 614, were already identified as including poultry isolated from Florida, and increased with an additional 1, 171, 237, and 31 isolates from other states, respectively. Four additional HC5 clusters: 165 (3

TABLE 2 | Comparison of cgMLST and PFGE based outbreak detection for *Salmonella* in Florida, 2017–2018.

Outbreak	HC5 cluster	Serotype	PFGE	#Ob	#Nob	First date	Last date	Duration (days)	Comments
1. Outbreak detected by PFGE, confirmed by HC5 and SNP									
1703FLJEG-1	58778	Enteritidis	JEGX01.1057	4	0	20170119	20170120	2	1-2 SNPs
1809MLJJP-1WGS	47842	Newport	JJPX01.0010	2	1	20180606	20180718	43	1 additional case detected by HC5 (2–5 SNPs)
2. Outbreak detected by PFGE, confirmed by SNP, more than one HC5									
1704MLJEG-2	40641	Enteritidis	JEGX01.0049	1	0	20170420	20170512	23	0 SNPs
	19036		JEGX01.0049	1	0				
1701MLJJP-1	938	Newport	JJPX01.0010	1	0	20170206	20170624	139	1 SNPs
	106562		JJPX01.0010	1	0				
3. Outbreak detected by PFGE, not confirmed by HC5 and SNP									
1703MLJEG-1	26508	Enteritidis	JEGX01.0031	1	0	20170328		12	2 HC5, 12 SNPs
	36574		JEGX01.0031	1	0	20170408			
1708MLJEG-4	81880	Enteritidis	JEGX01.0031	1	0	20170605		8	2 HC5, 46 SNP. One related isolate (3 SNPs) detected approx. 1 year later
	26508		JEGX01.0031	1	0	20170612			
	26508		JEGX01.0031	0	1	20180530			
1803FLJGG-1	105287	Javiana	JGGX01.0094	1	0	20171102	20180305	124	6 HC5, 12–31 SNPs with 1 exception (2 SNP)
	11615		JGGX01.0094	1	0				
	12146		JGGX01.0094	1	0				
	115499		JGGX01.0094	1	0				
	109542		JGGX01.0094	1	0				
	182313		JGGX01.0094	1	0				
1807MLJJP-1	167788	Typhimurium	JPXX01.0003	1	0	20180406	20180606	62	3 HC5, 5–16 SNP
	168142		JPXX01.0003	1	0				
	169033		JPXX01.0003	1	0				
4. Outbreak detected, not confirmed by PFGE, HC5 and SNP									
1704MLTDK-1	69749	Typhimurium	JPXX01.0033	2	0	20170602	20170612	42	2 PFGE, 2 HC5, 2 SNP clusters (333 SNPs apart)
	76749		JPXX01.0033	1		20170524			
	57152		JPXX01.0351	1		20170502			
1705MLJAP-1	59624	Bareilly	JAPX01.0064	12	21	20170207	20180606	485	2 PFGE, 2 HC5, > 47 SNPs between clusters
	67069		JAPX01.0216		2				
			NA		1				
5. Outbreak not detected by PFGE, one cluster by HC5 and supported by SNP, all cases within 60 days									
-&	85	Enteritidis	JEGX01.0004	0	1	20170417	20170514	28	2 PFGE, 1–4 SNPs
			JEGX01.0021	0	2				
-	2705	Enteritidis	JEGX01.0004	0	2	20170516	20170516	1	2 PFGE, 1–2 SNPs
			JEGX01.0034	0	1				
-	113949	Newport	JJPX01.0036	0	2	20171124	20171124	1	2 PFGE, 1 SNP
			JJPX01.0038		1				
-	127322	Javiana	JGGX01.0012	0	4	20180314	20180314	1	1 PFGE, 1 HC5, 8–27 SNPs
-	73033	Typhimurium	JPXX01.5110	0	7	20170509	20170517	9	1 PFGE, 0–4 SNPs
			NA		1				
6. Outbreak detected, multiple PFGE profiles, one HC5 cluster supported by SNP; prolonged case series									
1706MLJEG-1	614	Enteritidis	JEGX01.0023	1	10	20170312	20180605	451	1 recognized outbreak isolate (5–8 SNPs from 8 isolates), 4 PFGE
			JEGX01.0021	0	2				
			JEGX01.0004	0	1				
			JEGX01.0034	0					
1708MLJEG-1	87	Enteritidis	JEGX01.0002	3	6	20170101	20180712	558	3 recognized outbreak isolates (2 at 0–7 SNPs from 6 isolates), 5 PFGE
			JEGX01.0020		1				
			JEGX01.0024		1				
			JEGX01.0184		1				
			JJPX01.0001		1				

(Continued)

TABLE 2 | Continued

Outbreak	HC5 cluster	Serotype	PFGE	#Ob	#Nob	First date	Last date	Duration (days)	Comments
7. Outbreak not detected, multiple PFGE profiles, one HC5 cluster supported by SNP; prolonged case series									
-	62607	Enteritidis	JEGX01.0021	0	3	20170426	20180530	400	1 PFGE, 2–7 SNPs
-	182	Enteritidis	JEGX01.0021	0	12	20170323	20180530	434	3 PFGE, multiple small SNPs clusters, maximum distance 28 SNPs
			JEGX01.0023	0	1				
			JEGX01.0004	0	1				
-	19852	Enteritidis	JEGX01.0021	0	2	20170107	20170527	141	2 PFGE, 0 SNPs
			JEGX01.0004	0	1				
-	165	Enteritidis	JEGX01.0004	0	2	20170322	20180405	380	2 PFGE, 2–6 SNPs
			JEGX01.0034	0	1				
-	150	Enteritidis	JEGX01.0004	0	11	20170327	20180612	443	2 PFGE, multiple small SNPs clusters, maximum distance 21 SNPs
			JEGX01.0034	0	3				
-	9391	Enteritidis	JEGX01.0004	0	3	20170404	20170605	63	1 PFGE (1–4 SNPs)
-	50773	Enteritidis	JEGX01.0004	0	3	20180315	20180613	91	1 PFGE (4–10 SNPs)
-	10584	Enteritidis	JEGX01.0005	0	12	20170417	20180612	422	2 PFGE (0–12 SNPs)
			JEGX01.0030		1				
-	75478	Newport	JJPX01.0384	0	4	20170627	20180623	362	1 PFGE (2–4 SNPs)
	63415	Newport	JJPX01.3716	0	28	20170126	20180623	514	≥7 PFGE profiles, many close clusters (22 isolates 0–7 SNPs), maximum distance 19 SNPs
			NA		3				
			JAA01		1				
			JEGX01.0005		1				
			JGXX01.0006		1				
			JJPX01.5786		1				
			JLXX01.0483		1				
-	13124	Newport	JJPX01.0032	0	1	20170309	20180125	323	3 PFGE, 4 SNPs
			JJPX01.0036		1				
			JJPX01.0400		1				
-	14210	Javiana	JGGX01.0012	0	1	20170411	20171105	209	3 PFGE profiles, 9–16 SNP
			JGGX01.0488		1	20170417			
			JGGX01.0013		1	20171105			
-	576	14, [5], 12:i:-	JPXX01.2583		2	20171001	20171005	179	2 PFGE, 0–10 SNP
			JPXX01.3877		2	20170410	20170524		
-	6	14, [5], 12:i:-	JPXX01.3034	0	1	20170304	20170610	468	6 PFGE, 3 SNP clusters
			JPXX01.1314		3	20170426	20180614		
			JPXX01.1314		2	20180509	20180424		
			JPXX01.3775		2	20180617			
			JPXX01.5522*		2	20180419			
					1	20180412			

#Ob, number of outbreak isolates; #Nob, number of non-outbreak isolates.

*No outbreak detected.

*1 isolate serotyped as Typhimurium.

clinical, 19 poultry), 62607 (3 clinical, 6 poultry), 85 (3 clinical, 15 poultry), and 9391 (3 clinical, 12 poultry) were detected in which all poultry isolates were from other states.

SNP-based phylogeny and maximum likelihood phylogenetic trees for 7 clusters are included in **Supplementary Figure 10**. A phylogenetic tree for HC5 cluster 614, comprising 31 poultry isolates from 6 other states and 1 isolate from Florida, along with 14 clinical isolates from Florida, is depicted in **Figure 8**. SNP-distances in this cluster were in a broad range (0–35). The majority of clinical isolates (8) were clustered separately (5–15 SNP apart) from poultry isolates, while three isolates clustered closely (0–4 SNP distance) in three smaller subclusters

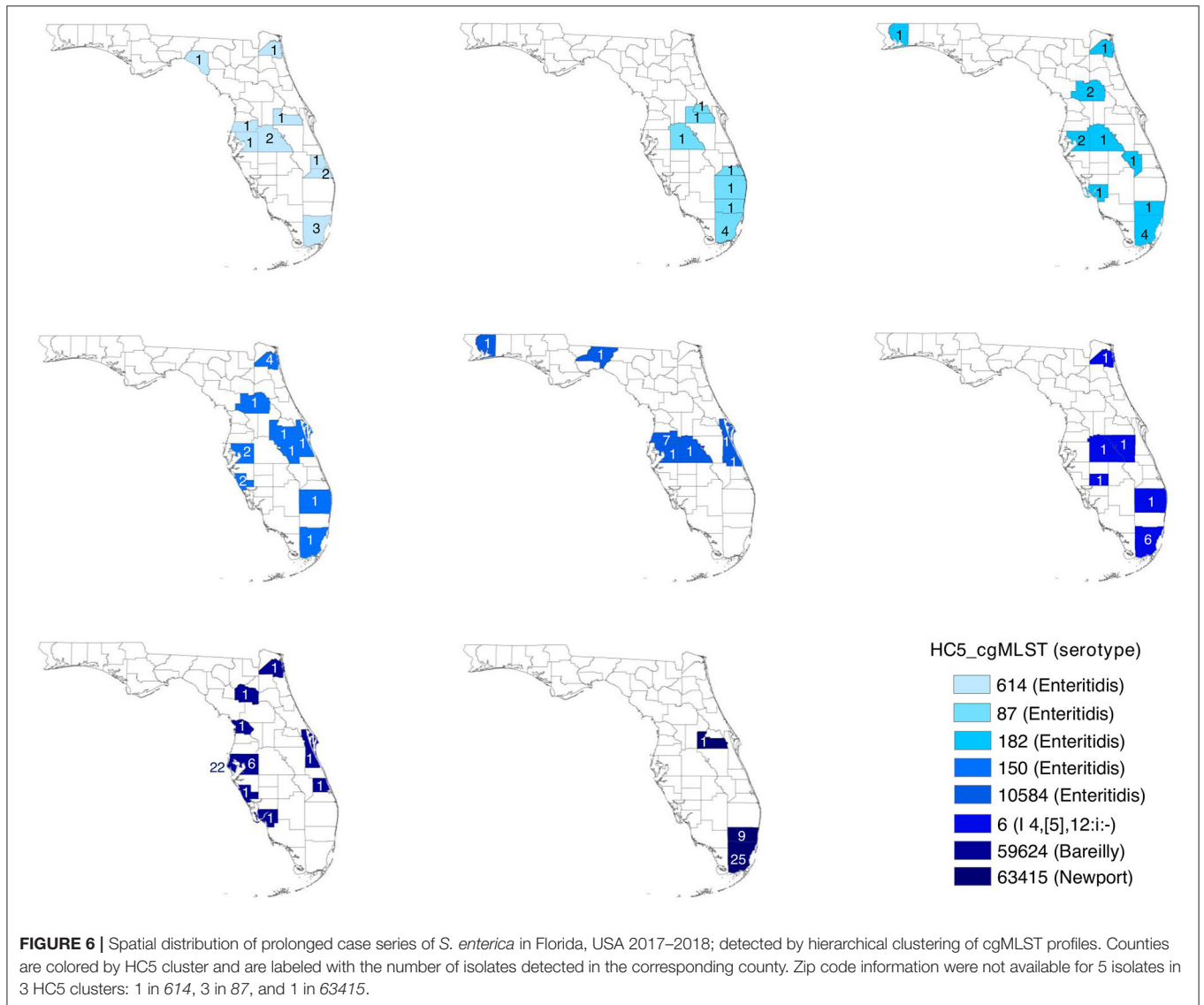
of 3–11 isolates and three isolates clustered together with two poultry isolates. All other seven clusters had variety of SNP distance ranges, commonly up to 30, while large clusters like 150 and 182 included distant isolates of >30 SNP as well, as depicted in the SNP-distance heatmaps for all eight clusters in **Supplementary Figure 11**.

As *S. Enteritidis* is a highly clonal organism, to further zoom in to highly related clusters, we used a small range of 0–4 SNP distance as cutoff and obtained 2–6 close SNP clusters within all HC5 clusters, except 62607 and 85. **Table 5** enumerates clinical and non-clinical isolates from SNP sub-clusters in each HC5 cluster and involved HC2 clusters in each SNP cluster

TABLE 3 | Geospatial distribution of 8 prolonged cases series (> 10 isolates) based on HC5 clusters in Florida, 2017–2018.

Serotype	HC5 cluster	#Isolates	#Zip code areas	Standard distance (km)
Bareilly	59624*	35	25	92
Enteritidis	87*	13	11	155
	150	14	14	173
	182	14	14	244
	614* ^{&}	14	12	177
	10584	13	10	202
I 4, [5],12:i:-	6	11	11	173
Newport	63415	36	26	66

*Outbreak detected by PFGE surveillance, additional cases detected by HC5 clustering.
 & Prolonged case series.



and **Figure 9** depicts their time-series, stratified by state and collection source. In HC5 cluster 150, all 6 SNP clusters were spread over time in 2017 and 2018 and suggested genetic

relatedness of Florida clinical isolated with poultry isolates collected from 8 states, mainly Georgia. For HC5 cluster 165, one SNP cluster was close in time, but poultry isolates were

TABLE 4 | Source matching between clinical and non-clinical isolates from Florida based on HC5 clusters.

Serotype	HC5 Cluster	Source	#Isolates
Braenderup	68954	Human	3
		Poultry	1
Enteritidis	150 ^{&c}	Human	14
		Poultry	4
	182 ^{&c}	Human	14
		Poultry	9
	614 ^{&c}	Human	14
		Poultry	1
	3213	Human	2
		Poultry	1
10584 ^{&c}	Human	13	
	Poultry	7	
I 4, [5], 12:i:-	576	Human	4
		Food	1
		Livestock	1
	118459	Human	1
		Poultry	1
Infantis	4584	Human	1
		Poultry	2
	106860	Human	1
		Poultry	1
Saintpaul	9009	Human	2
		Environment	1
Sandiego	136994	Aquatic Animal	2
		Non-clinical sponge	1

^{&c} Prolonged case series.

collected from two distant states and a second SNP cluster was spread out in time. HC5 cluster 182 had six large, interesting SNP clusters, comprising poultry, and clinical isolates collected in both 2017 and 2018. HC5 cluster 10584 had poultry isolates only from Florida in all 3 SNP clusters mainly collected in a span of 2–6 months in 2018 while HC5 cluster 9391 had three time-sparse SNP clusters with poultry isolates from Georgia and Pennsylvania. HC5 Cluster 614 is heterogenous in terms of states of poultry isolates but SNP clusters shared time proximity with clinical isolates. Of 24 detected SNP clusters, 10 included only one HC2 cluster profile, while the remaining 14 SNP clusters included between 2 and 8 HC2 clusters, confirming our choice of HC5 as the optimal level of clustering.

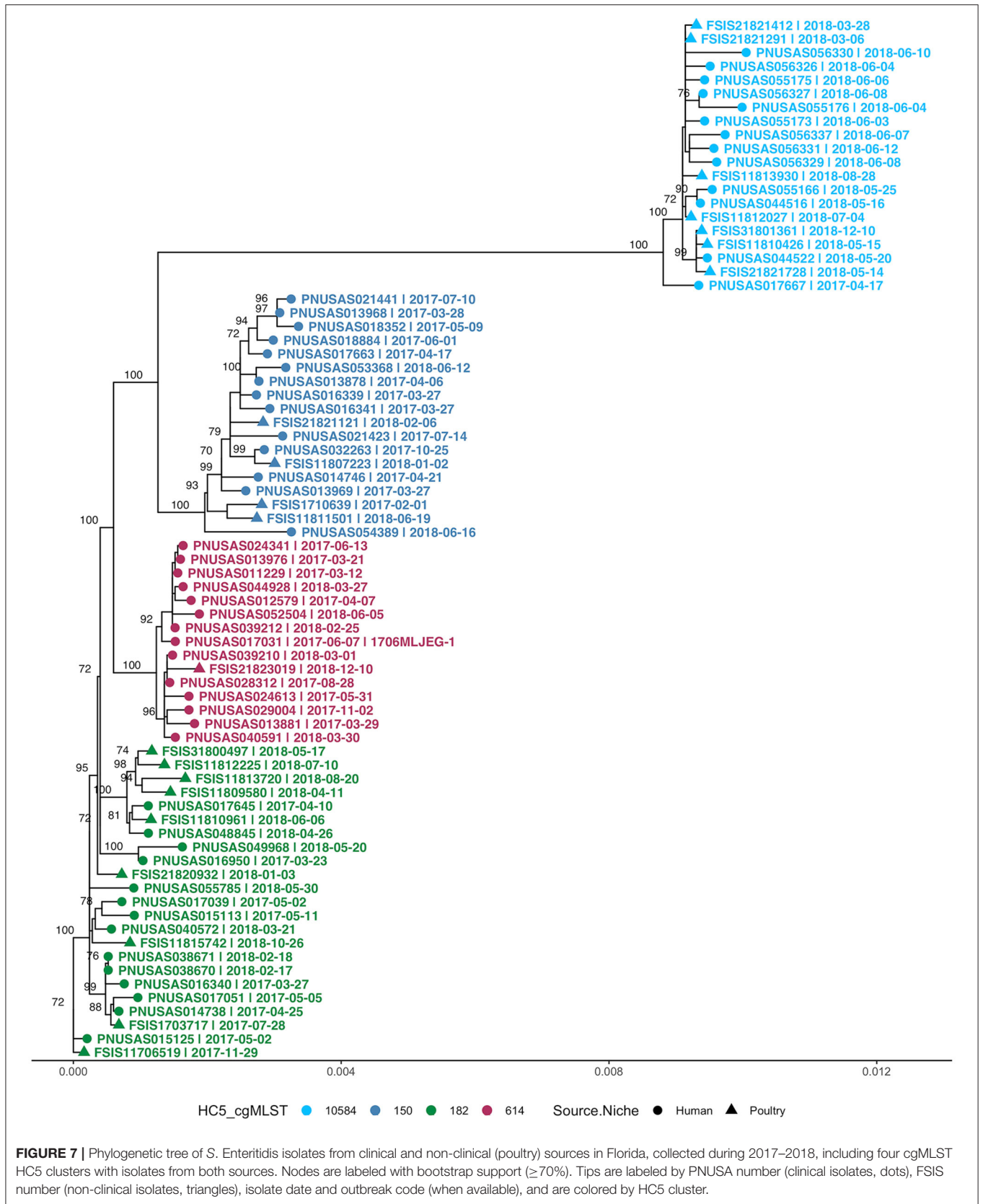
To further evaluate genetic relatedness between clinical and poultry isolates, we searched for the presence of genes conferring antimicrobial drug resistance in all 580 isolates in the eight common HC5 clusters, via AMRFinder. It is notable that AMRFinder considers complete genome information along with plasmids while HC5 clustering only considers the core genome. Twenty-four isolates carried the tetracycline resistant gene *tet(A)* and, out of these eight clustered in HC5 62607 (poultry: 6 and human: 2) and two clustered in HC 5150 and none of these 10 isolates were assigned to close SNP clusters. The other 14

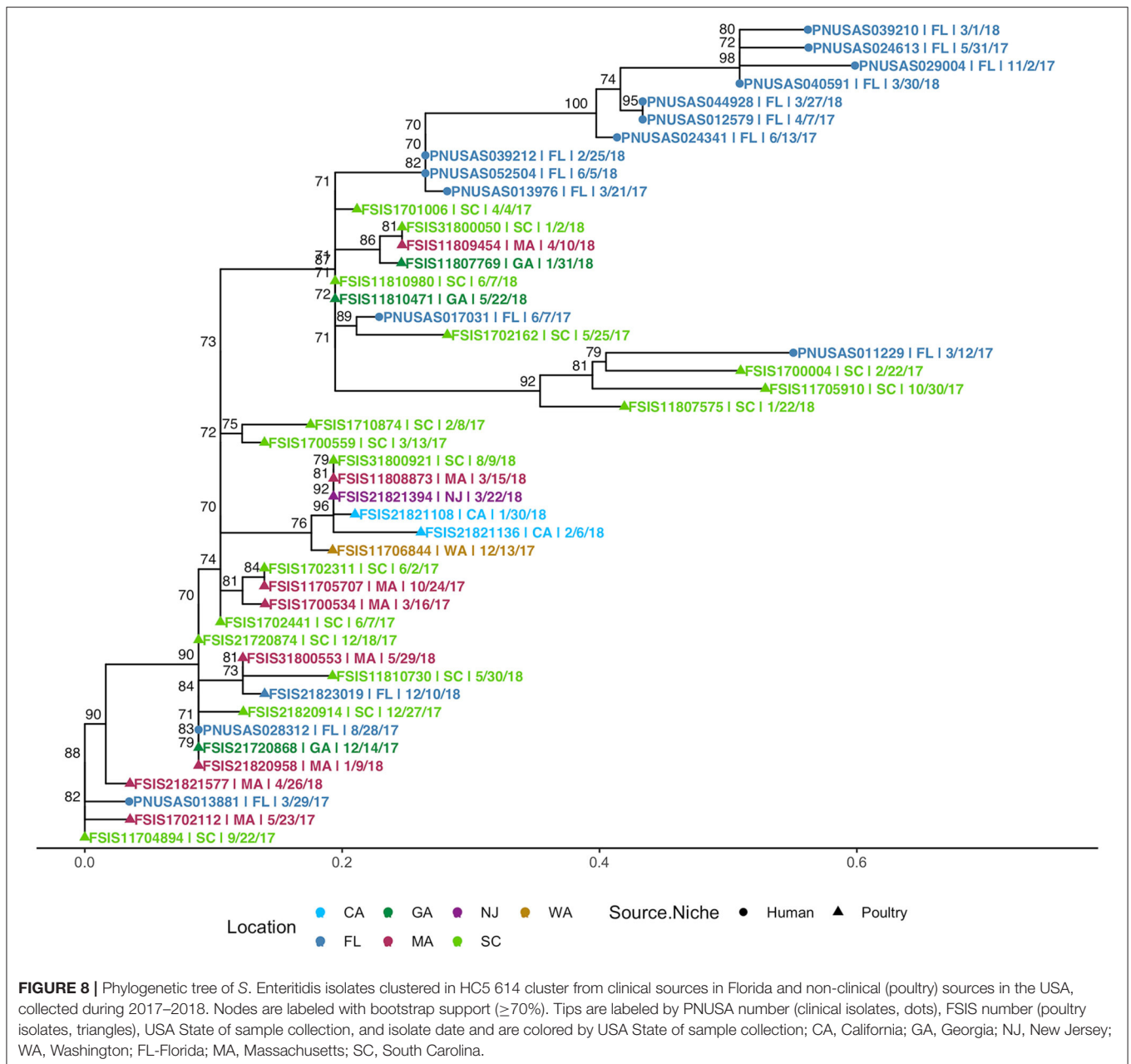
isolates carried additional genes for beta-lactamase resistance (*bla_{TEM-1}*), sulphonamide resistance (*sul2*), aminoglycoside 3'-phosphotransferase (*aph(3'')-Ib*) and aminoglycoside O-phosphotransferase (*aph(6)-Id*). All 14 isolates belonged to HC5 9391 and six multi-drug resistant isolates were assigned to two SNP clusters (one cluster with two human and two poultry isolates, and one cluster with one human and one poultry isolate, depicted in **Figure 9**). A detailed table is included in **Supplementary Data Sheet 3**.

DISCUSSION

We present a retrospective analysis of whole genome sequencing data of *Salmonella enterica* isolates from Florida in the years 2017–2018. Our earlier work has shown that the sequenced isolates in 2017–18 were a relatively unbiased sample of confirmed clinical cases (4). The population structure of *Salmonella* in Florida was found to be highly diverse with 115 serotypes observed in a set of 1,632 sporadic clinical isolates obtained over this 2-year period. A larger dataset is expected to include a much higher number of serotypes as the rarefaction curve for serotype richness was far from reaching an asymptote. However, many serotypes, including unsampled ones, were rare as demonstrated by the much lower number of 36 typical serotypes, and 22 dominant serotypes. Both of these are well-sampled as demonstrated by plateaus reached in the rarefaction curves. As little as 250 isolates appeared to be sufficient to identify the dominant serotypes. Serotype diversity was similar in young children (<2 years) as compared to older children and adults (≥2 years). These results differ from those reported by Judd et al. (38) who observed substantially higher serotype richness in children <2 years of age for the USA as a whole. No previous analysis was available to compare Florida rarefaction results with national data.

The dominant serotypes circulating in Florida were strikingly different from those in the US as a whole. *S. Javiana* has previously been identified as a “Southern” serotype (39), e.g., in the CDC *Salmonella* atlas, which covers data from 1968 to 2011. The high incidence of *S. Sandiego* was unexpected. Previous data suggest this serotype occurred in low numbers throughout the USA between 1968 and 2011, without a clear geographic pattern. *S. Sandiego* has been involved in several outbreaks in the past decade. In 2013, eight outbreaks associated with pet turtles sickened 473 persons across the country. Multiple serotypes were involved, including two outbreaks with *S. Sandiego* (124 and 7 cases, respectively with *S. Newport* also involved in the larger outbreak). None of the *S. Sandiego* cases were from Florida (40). In 2018, an outbreak involving 101 cases was reportedly caused by Spring Pasta Salad purchased at Hy-Vee grocery stores. All cases resided in, or traveled to, states where these stores are located, i.e., Iowa, Kansas, Minnesota, Nebraska, and South Dakota. No cases we recognized in Florida (41). We detected a small cluster of three isolates of *S. Sandiego* from aquatic sources in Florida, but no human isolates were included in this cluster (**Table 4**). Further, studies are recommended to elucidate the reservoirs and transmission pathways of this serotype.





As previously observed (14), there was a high level of correspondence between eBurst Group assignments and serotype designations. Due to the dominance of a limited number of sequence types within an eBG, legacy MLST was of limited use for differentiating clinical *Salmonella* isolates beyond the serotype level, and thus, for identification of putative sources of infection. We therefore explored the use of core genome MLST as an alternative for epidemiological inference, which has also been reported by others as an effective method for surveillance purposes (42).

Our analysis suggests that hierarchical clustering of cgMLST profiles can be used as an effective screening to initiate more detailed phylogenetic analysis and is an efficient way to recognize potential outbreaks or prolonged case clusters in a

large set of WGS data, as has been recognized as efficient approach in earlier studies (43). Nevertheless, comparison of results with PFGE surveillance was complex. WGS-based surveillance would have detected more putative outbreaks than PFGE surveillance, even though some outbreaks detected by PFGE surveillance would have been missed. Importantly, HC5 clustering detected additional clusters of clinical isolates that differed by isolation dates as much as 558 days. This suggests that there are persistent sources of *Salmonella* infection, that would not be detected by traditional outbreak detection, where a time window of 60 days is used to detect closely related isolates.

To investigate this hypothesis further, 14 clusters of clinical cases of *S. Enteritidis* with prolonged duration that were not

TABLE 5 | SNP clusters[†] (0–4 SNP difference) of clinical isolates from Florida and poultry isolates of *Salmonella* Enteritidis from the USA, within HC 5 clusters.

HC5 cluster	HC2 cluster (# of isolates)	SNP cluster [†]	Clinical isolates	Number of poultry isolates by state	
150	8440 (4)	1	PNUSAS014746 (FL 04/21/2017)	FSIS1700341 (MS 03/08/2017) FSIS31800585 (WA 06/04/2018) FSIS31800622 (WA 06/07/2018)	
		2	PNUSAS053368 (FL 06/12/2018) PNUSAS013878 (FL 04/06/2017)	FSIS1700715 (GA 03/22/2017) FSIS21820955 (NY 01/09/2018)	
	8440 (3), 169289 (1)	3	PNUSAS032263 (FL 10/25/2017)	FSIS1700276 (GA 03/07/2017) FSIS21720256 (GA 09/13/2017) FSIS11811144 (GA 06/07/2018) FSIS31800742 (GA 06/28/2018)	
				4	PNUSAS017663 (FL 04/17/2017) PNUSAS016341 (FL 03/27/2017)
	8440 (4), 72875 (1), 77030 (1), 88317 (1), 115025 (1), 126031 (1)	5	PNUSAS054389 (FL 06/16/2018)	FSIS31800470 (GA 05/09/2018)	
				6	PNUSAS013969 (FL 03/27/2017)
	165	165 (3)	1	PNUSAS040595 (FL 04/05/2018)	FSIS21821669 (MA 05/09/2018) FSIS31800337 (WA 04/03/2018)
			2	PNUSAS040570 (FL 03/15/2018) PNUSAS011947 (FL 04/22/2017)	FSIS11812199 (SC 07/10/2018) FSIS1701123 (SC 04/11/2017) FSIS21720358 (MA 09/26/2017)
	182	31321 (9), 135578 (1)	1	PNUSAS017645 (FL 04/10/2017)	FSIS11811483 (WA 06/09/2018) FSIS21821927 (WA 07/13/2018) FSIS11810961 (FL 06/06/2018) FSIS21820940 (IL 01/04/2018) FSIS21821386 (OK 03/21/2018) FSIS31800336 (AL 04/03/2018) FSIS11811422 (AL 06/09/2018) FSIS1700876 (AL 003/28/2017) FSIS21720176 (AL 08/29/2017)
					2
60809 (2), 83883 (3)		3	PNUSAS048845 (FL 04/26/2018)	FSIS11808374 (MA 02/21/2018) FSIS11811751 (GA 06/26/2018) FSIS21822276 (AL 09/19/2018) FSIS21822629 (AL 10/24/2018)	
				4	PNUSAS040572 (FL 03/21/2018)
117653 (5)		5	PNUSAS038671 (FL 02/18/2018) PNUSAS038670 (FL 2/17/2018) PNUSAS016340 (FL 03/27/2017) PNUSAS014738 (FL 04/25/2017) PNUSAS017051 (FL 05/05/2017)	FSIS11812210 (CA 07/10/2018) FSIS11810099 (NJ 05/10/2018) FSIS11810454 (MI 05/18/2018) FSIS31800499 (CO 05/17/2018) FSIS31800539 (AL 05/24/2018) FSIS21820981 (AL 01/10/2018) FSIS21821128 (AL 02/06/2018) FSIS11808362 (MS 02/20/2018) FSIS1703451 (MS 07/18/2017) FSIS21821090 (TX 01/31/2018) FSIS11807765 (TX 01/30/2018) FSIS11808961 (TX 03/20/2018) FSIS1701679 (TX 05/03/2017) FSIS1703717 (FL 07/28/2017) FSIS1700022 (AL 02/22/2017)	
				4	PNUSAS040572 (FL 03/21/2018)
3334 (4)		5	PNUSAS038671 (FL 02/18/2018) PNUSAS038670 (FL 2/17/2018) PNUSAS016340 (FL 03/27/2017) PNUSAS014738 (FL 04/25/2017) PNUSAS017051 (FL 05/05/2017)	FSIS11812210 (CA 07/10/2018) FSIS11810099 (NJ 05/10/2018) FSIS11810454 (MI 05/18/2018) FSIS31800499 (CO 05/17/2018) FSIS31800539 (AL 05/24/2018) FSIS21820981 (AL 01/10/2018) FSIS21821128 (AL 02/06/2018) FSIS11808362 (MS 02/20/2018) FSIS1703451 (MS 07/18/2017) FSIS21821090 (TX 01/31/2018) FSIS11807765 (TX 01/30/2018) FSIS11808961 (TX 03/20/2018) FSIS1701679 (TX 05/03/2017) FSIS1703717 (FL 07/28/2017) FSIS1700022 (AL 02/22/2017)	
				4	PNUSAS040572 (FL 03/21/2018)
41659 (18), 125926 (1), 126076 (1)		5	PNUSAS038671 (FL 02/18/2018) PNUSAS038670 (FL 2/17/2018) PNUSAS016340 (FL 03/27/2017) PNUSAS014738 (FL 04/25/2017) PNUSAS017051 (FL 05/05/2017)	FSIS11812210 (CA 07/10/2018) FSIS11810099 (NJ 05/10/2018) FSIS11810454 (MI 05/18/2018) FSIS31800499 (CO 05/17/2018) FSIS31800539 (AL 05/24/2018) FSIS21820981 (AL 01/10/2018) FSIS21821128 (AL 02/06/2018) FSIS11808362 (MS 02/20/2018) FSIS1703451 (MS 07/18/2017) FSIS21821090 (TX 01/31/2018) FSIS11807765 (TX 01/30/2018) FSIS11808961 (TX 03/20/2018) FSIS1701679 (TX 05/03/2017) FSIS1703717 (FL 07/28/2017) FSIS1700022 (AL 02/22/2017)	
				4	PNUSAS040572 (FL 03/21/2018)

(Continued)

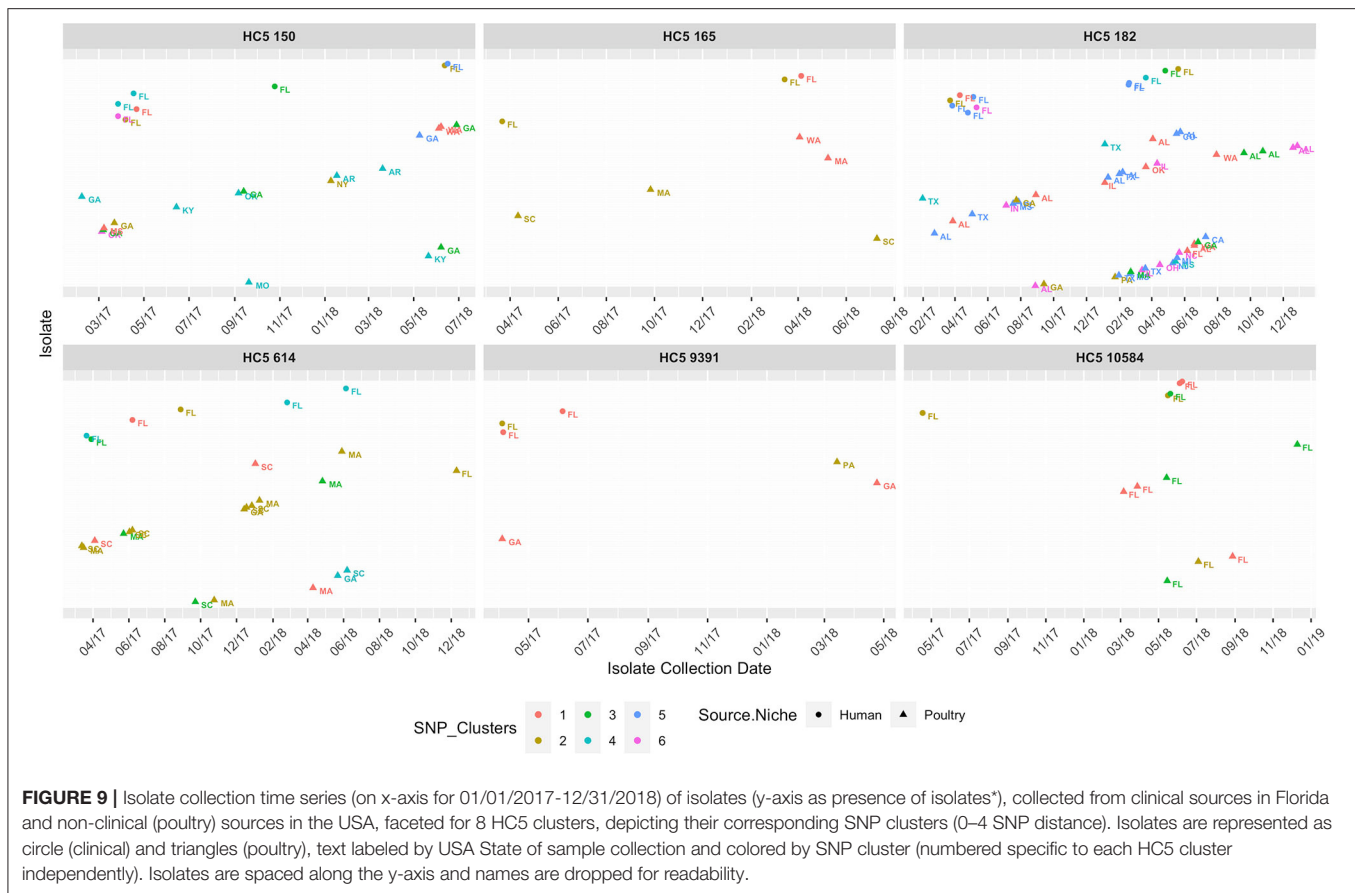
TABLE 5 | Continued

HC5 cluster	HC2 cluster (# of isolates)	SNP cluster [†]	Clinical isolates	Number of poultry isolates by state
	64255 (7), 161204 (1), 164058 (1)	6	PNUSAS015113 (FL 05/11/2017)	FSIS11704369 (AL 08/28/2017) FSIS21923137 (AL 12/19/2018) FSIS21923159 (AL 12/27/2018) FSIS11809591 (OH 04/16/2018) FSIS11810473 (NC 05/22/2018) FSIS1703098 (IN 07/05/2017) FSIS21821483 (IL 04/11/18/2018) FSIS11808858 (IL 03/14/2018)
614	614 (6)	1	PNUSAS017031 (FL 06/07/2017)	FSIS1701006 (SC 04/04/2017) FSIS11810980 (SC 06/07/2018) FSIS31800050 (SC 01/02/2018) FSIS11809454 (MA 04/10/2018) FSIS11810471 (GA 05/22/2018)
	614 (5), 68955 (1), 72379 (1), 98146 (1), 103657 (1), 106428 (1), 152255 (1), 163446 (1)	2	PNUSAS028312 (FL 08/28/2017)	FSIS11705707 (MA 10/24/2017) FSIS1700534 (MA 03/16/2017) FSIS1700559 (SC 03/13/2017) FSIS1702311 (SC 06/02/2017) FSIS1702441 (SC 06/07/2017) FSIS21720874 (SC 12/18/2017) FSIS21820914 (SC 12/27/2017) FSIS21720868 (GA 12/14/2017) FSIS21820958 (MA 01/09/2018) FSIS31800553 (MA 05/29/2018) FSIS21823019 (FL 12/10/2018)
	614 (4)	3	PNUSAS013881 (FL 03/29/2017)	FSIS11704894 (SC 09/22/2017) FSIS1702112 (MA 05/23/2017) FSIS21821577 (MA 04/26/2018)
	614 (5)	4	PNUSAS052504 (FL 06/05/2018) PNUSAS039212 (FL 02/25/2018) PNUSAS013976 (FL 03/21/2017)	FSIS11810471 (GA 05/22/2018) FSIS11810980 (SC 06/07/2018)
9391	9391 (1), 86566 (1), 119878 (1), 74731 (1)	1	PNUSAS022594 (FL 06/05/2017) PNUSAS014741 (FL 04/05/2017)	FSIS21821561 (GA 04/24/2018) FSIS1701015 (GA 04/04/2017)
	62823 (1), 78219 (1)	2	PNUSAS016449 (FL 04/04/2017)	FSIS31800280 (PA 03/14/2018)
10584	10584 (5), 12492 (1), 151970 (1)	1	PNUSAS056327 (FL 06/08/2018) PNUSAS055176 (FL 06/04/2018) PNUSAS017667 (FL 04/17/2017)	FSIS21821412 (FL 03/28/2018) FSIS21821291 (FL 03/06/2018) FSIS11812027 (FL 07/04/2018) FSIS11813930 (FL 08/28/2018)
	10584 (2), 12492 (1)	2	PNUSAS044516 (FL 05/16/2018) PNUSAS017667 (FL 04/17/2017)	FSIS11812027 (FL 07/04/2018)
	10584 (4)	3	PNUSAS044522 (FL 05/20/18)	FSIS31801361 (FL 12/10/2018) FSIS11810426 (FL 05/15/2018) FSIS21821728 (FL 05/14/2018)
85			No cluster	
62607			No cluster	

[†]Numbered independently for each HC5 cluster. AR, Arkansas; AL, Alabama; CA, California; CO, Colorado; GA, Georgia; IL, Illinois; IN, Indiana; KY, Kentucky; MA, Massachusetts; MI, Michigan; MS, Mississippi; NC, North Carolina; NJ, New Jersey; OK, Oklahoma; OH, Ohio; PA, Pennsylvania; TX, Texas; WA, Washington; FL, Florida; MA, Massachusetts; SC, South Carolina; WA, Washington.

related to detected outbreaks were investigated using SNP phylogenetic analysis. Eight of these clusters matched with non-clinical (all poultry) isolates from regulatory inspection by the United States Department of Agriculture Food Safety and Inspection Service (FSIS) through the GenomeTrakr Network. The phylogenetic results were supported by the presence of antimicrobial resistance genes in some clusters. Geospatial analysis suggested that these isolates were not only distant in time but also across locations throughout the whole US, supporting

the hypothesis that some clones of *S. Enteritidis* persist in the broiler meat chain and sporadically cause human infections. Vertical transmission from parent stock or higher in the breeding pyramid is the most likely explanation of this widespread and sporadic occurrence of genetically similar isolates from poultry sources in different states and at different time points. Modern gene-based methods have recently been used to confirm vertical transmission in a vertically integrated poultry operation in Australia (44), and turkeys in the United States (45). Our study



only included clinical isolates from Florida, and it would be interesting to explore if clinical isolates from other states can also be linked to these persistent clones. FSIS samples chicken carcasses at slaughter and finished products (chicken parts and ground chicken). Many establishments purchase carcasses for further processing into finished product, which obscures the origin of a common strain. Among 513 isolates in *S. Enteritidis* clusters, 30 were from animal-young chicken, 50 from chicken carcass, 144 from comminuted chicken, and 289 from raw intact chicken. No isolates from turkey were detected among these clusters.

Use of genomic diversity for assessing genetic and epidemiological linkages of pathogens has been widely used. SNP-based and gene-by-gene (MLST) methods (12) have gained more popularity, the choice being dependent on the intended application (46). Previous studies comparing the use of cgMLST and SNP-based phylogeny for surveillance purposes have suggested that cgMLST presents a reproducible and scalable approach for outbreak detection and is congruent with SNP analysis (43, 46–49), which uses one or several mutations across different alleles and results in intrinsically higher discriminatory power (50). Each approach has their strength and shortcomings. The gene-by-gene approach is helpful to collapse large genomic data sets into a compact and simpler form but might miss individual nucleotide variants

within alleles. SNP-based approaches have extensively been used for outbreak detection (47, 51, 52) and provide great distinguishing power, differentiating isolates even at the level of single base pair substitutions (in theory). However, they often are computationally demanding and time consuming, and are sensitive to recombination, contamination and sequence coverage (52). Both these approaches further need epidemiological data support before confirming any outbreak or prolonged case series (53). In the current work, we have proposed a “combination approach” to utilize the merits of both cgMLST (compactness) and SNP- based approaches (high discriminatory power). We suggest the use of HierCC (54) method for screening genetically closer isolates, reducing the sample size considerably for SNP-based phylogeny. We found that Hierarchical Clustering at five or less allelic difference can be used as an efficient tradeoff between inclusiveness and specificity. As genetic diversity varies by *Salmonella* serotype, more experience with this method is needed and outbreak investigations should include building SNP-based phylogenies to support conclusions about genetic relatedness of isolates and possibly adapt the level of hierarchical clustering to address the specific research question or serotype. Such work may also support establishment of serotype-specific working thresholds for outbreak detection. Hierarchical clustering also appears to be an effective method to identify case clusters that are more

distant in time and place than traditional outbreaks but may have been infected from a common source. The potential of this new approach to source identification would be strengthened by more systematic surveillance of *Salmonella* in non-human reservoirs. Current data for Florida is very limited. Isolates in NCBI were mainly deposited by FSIS, thereby biasing the available data to animal source foods. Isolates from other foods and non-clinical sources from Florida are very rare in the public domain and need more investigation.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Florida Institutional Review Board. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

NS, XL, and AH conceived the study, performed the analyses, and drafted the manuscript. JB supervised the laboratory analysis.

EB and JD performed the surveillance metadata collection and contributed to the writing. All authors critically revised the manuscript and gave final approval of the version to be published.

FUNDING

This analysis was funded in part by the Florida Department of Health through the Florida Integrated Food Safety Center of Excellence.

ACKNOWLEDGMENTS

We thank Marco Salemi and Massimiliano Tagliamonte (University of Florida, Gainesville, FL) for their support with phylogenetic analysis and Eduardo Taboado and James Robertson (Public Health Canada, Guelph, Canada) for resolving ambiguous serotype assignments with SISTR2. We thank Paul Fiorella, Marek Pawlowicz, and Matthew Schimenti (Bureau of Public Health Laboratories, Jacksonville, FL) for generating PFGE data and Amber Ginn, Jessy Motes, and Matthew Schimenti (Bureau of Public Health Laboratories, Jacksonville, FL) for generating WGS data. We thank Peter Evans (USDA/FSIS) for reviewing a draft of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2021.656827/full#supplementary-material>

REFERENCES

- Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, et al. World Health organization global estimates and regional comparisons of the burden of foodborne disease in 2010. *PLoS Med.* (2015) 12:e1001923. doi: 10.1371/journal.pmed.1001923
- Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleeschauwer B, et al. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. *PLoS Med.* (2015) 12:e1001921. doi: 10.1371/journal.pmed.1001921
- Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, et al. Foodborne illness acquired in the United States—Major pathogens. *Emerg Infect Dis.* (2011) 17:7–15. doi: 10.3201/eid1701.P11101
- Li X, Singh N, Beshearse E, Blanton JL, DeMent J, Havelaar AH. Spatial epidemiology of Salmonellosis in Florida, 2009–2018. *Front Public Health.* (2021) 8:1001. doi: 10.3389/fpubh.2020.603005
- Ribot EM, Hise KB. Future challenges for tracking foodborne diseases: PulseNet, a 20-year-old US surveillance system for foodborne diseases, is expanding both globally and technologically. *EMBO Rep.* (2016) 17:1499–1505. doi: 10.15252/embr.201643128
- Gerner-Smidt P, Hise K, Kincaid J, Hunter S, Rolando S, Hyytia-Trees E, et al. PulseNet USA: a five-year update. *Foodborne Pathog Dis.* (2006) 3:9–19. doi: 10.1089/fpd.2006.3.9
- Ran L, Wu S, Gao Y, Zhang X, Feng Z, Wang Z, et al. Laboratory-based surveillance of nontyphoidal *Salmonella* infections in China. *Foodborne Pathog Dis.* (2011) 8:921–7. doi: 10.1089/fpd.2010.0827
- Deng X, Desai PT, den Bakker HC, Mikoleit M, Tolar B, Trees E, et al. Genomic epidemiology of *Salmonella enterica* serotype Enteritidis based on population structure of prevalent lineages. *Emerg Infect Dis.* (2014) 20:1481–9. doi: 10.3201/eid2009.131095
- Taylor AJ, Lappi V, Wolfgang WJ, Lapierre P, Palumbo MJ, Medus C, et al. Characterization of foodborne outbreaks of salmonella enterica serovar enteritidis with whole-genome sequencing single nucleotide polymorphism-based analysis for surveillance and outbreak detection. *J Clin Microbiol.* (2015) 53:3334–40. doi: 10.1128/JCM.01280-15
- Zheng J, Keys CE, Zhao S, Meng J, Brown EW. Enhanced subtyping scheme for *Salmonella enteritidis*. *Emerg Infect Dis.* (2007) 13:1932–5. doi: 10.3201/eid1312.070185
- Boxrud D, Pederson-Gulrud K, Wotton J, Medus C, Lyszkowicz E, Besser J, et al. Comparison of multiple-locus variable-number tandem repeat analysis, pulsed-field gel electrophoresis, and phage typing for subtype analysis of *Salmonella enterica* serotype Enteritidis. *J Clin Microbiol.* (2007) 45:536–43. doi: 10.1128/JCM.01595-06
- Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol.* (2013) 11:728–36. doi: 10.1038/nrmicro3093
- Achtman M, Wain J, Weill F-X, Nair S, Zhou Z, Sangal V, et al. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLOS Pathog.* (2012) 8:e1002776. doi: 10.1371/journal.ppat.1002776
- Alikhan N-F, Zhou Z, Sergeant MJ, Achtman M. A genomic overview of the population structure of *Salmonella*. *PLoS Genet.* (2018) 14(4):e1007261. doi: 10.1371/journal.pgen.1007261
- Deng X, den Bakker HC, Hendriksen RS. Genomic epidemiology: whole-genome-sequencing-powered surveillance and outbreak investigation of foodborne bacterial pathogens. *Annu Rev Food Sci Technol.* (2016) 7:353–74. doi: 10.1146/annurev-food-041715-033259

16. Ashton PM, Peters T, Ameh L, McAleer R, Petrie S, Nair S, et al. Whole genome sequencing for the retrospective investigation of an outbreak of Salmonella typhimurium DT 8. *PLoS Curr.* (2015) 7. doi: 10.1371/currents.outbreaks.2c05a47d292f376afc5a6fcd8a7a3b6
17. Inns T, Ashton PM, Herrera-Leon S, Lighthill J, Foulkes S, Jombart T, et al. Prospective use of whole genome sequencing (WGS) detected a multi-country outbreak of Salmonella Enteritidis. *Epidemiol Infect.* (2017) 145:289–98. doi: 10.1017/S0950268816001941
18. Pijnacker R, Dallman TJ, Tijmsa ASL, Hawkins G, Larkin L, Kotila SM, et al. An international outbreak of Salmonella enterica serotype Enteritidis linked to eggs from Poland: a microbiological and epidemiological study. *Lancet Infect Dis.* (2019) 19:778–86. doi: 10.1016/S1473-3099(19)30047-7
19. Yachison CA, Yoshida C, Robertson J, Nash JHE, Kruczkiewicz P, Taboada EN, et al. The validation and implications of using whole genome sequencing as a replacement for traditional serotyping for a National Salmonella Reference Laboratory. *Front Microbiol.* (2017) 8:1044. doi: 10.3389/fmicb.2017.01044
20. Besser JM, Carleton HA, Trees E, Stroika SG, Hise K, Wise M, et al. Interpretation of whole-genome sequencing for enteric disease surveillance and outbreak investigation. *Foodborne Pathog Dis.* (2019) 16:504–12. doi: 10.1089/fpd.2019.2650
21. Zhou Z, Alikhan N-F, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, et al. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res.* (2018) 28:1395–404. doi: 10.1101/gr.232397.117
22. Allard MW, Strain E, Melka D, Bunning K, Musser SM, Brown EW, et al. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. *J Clin Microbiol.* (2016) 54:1975–83. doi: 10.1128/JCM.00081-16
23. *Whole Genome Sequencing (WGS) | PulseNet Methods | PulseNet | CDC.* Available online at: <https://www.cdc.gov/pulsenet/pathogens/wgs.html> (accessed March 8, 2021).
24. Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VP, Nash JH, et al. The Salmonella In Silico Typing Resource (SISTR): an open web-accessible tool for rapidly typing and subtyping draft Salmonella genome assemblies. *PLoS ONE.* (2016) 11:e0147101. doi: 10.1371/journal.pone.0147101
25. Chiu C-H, Chao A. Distance-based functional diversity measures and their decomposition: a framework based on hill numbers. *PLoS ONE.* (2014) 9:e100014. doi: 10.1371/journal.pone.0100014
26. Centers for Disease Control and Prevention (CDC). *National Enteric Disease Surveillance: Salmonella Annual Report.* (2016). Available online at: <https://www.cdc.gov/nationalsurveillance/pdfs/2016-Salmonella-report-508.pdf> (accessed December 15, 2020).
27. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* (2013). Available online at: <https://arxiv.org/abs/1303.3997>
28. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* (2018) 34:3094–100. doi: 10.1093/bioinformatics/bty191
29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* (2009) 25:2078–9. doi: 10.1093/bioinformatics/btp352
30. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* (2015) 43:e15. doi: 10.1093/nar/gku1196
31. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* (2017) 14:587–9. doi: 10.1038/nmeth.4285
32. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol.* (2017) 8:28–36. doi: 10.1111/2041-210X.12628
33. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* (2016) 32:2847–9. doi: 10.1093/bioinformatics/btw313
34. Paradis E, Schliep K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* (2019) 35:526–8. doi: 10.1093/bioinformatics/bty633
35. Mitchell A, *Esri Press | The ESRI Guide to GIS Analysis, Volume 2 | Spatial Measurements and Statistics.* Available online at: <https://esripress.esri.com/display/index.cfm?fuseaction=display&websiteID=86&moduleID=0> (accessed February 25, 2021).
36. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother.* (2019) 63:e00483–19. doi: 10.1128/AAC.00483-19
37. Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, McDermott PF. Using genomics to track global antimicrobial resistance. *Front Public Health.* (2019) 7:242. doi: 10.3389/fpubh.2019.00242
38. Judd MC, Hoekstra RM, Mahon BE, Fields PI, Wong KK. Epidemiologic patterns of human Salmonella serotype diversity in the USA, 1996–2016. *Epidemiol Infect.* (2019) 147:e187. doi: 10.1017/S0950268819000724
39. Centers for Disease Control and Prevention (CDC). *An Atlas of Salmonella in the United States, 1968–2011: Laboratory-based Enteric Disease Surveillance.* Atlanta, GA: US Department of Health and Human Services, CDC (2013).
40. Centers for Disease Control and prevention. *Salmonella Case Count Map.* (2018). Available online at: <https://www.cdc.gov/salmonella/small-turtles-03-12/map.html> (accessed December 15, 2020).
41. Centers for Disease Control and Prevention. *Salmonella (2018) Outbreaks.* (2018). Available online at: <https://www.cdc.gov/salmonella/sandiego-07-18/index.html> (accessed December 15, 2020).
42. Moura A, Tourdjman M, Leclercq A, Hamelin E, Laurent E, Fredriksen N, et al. Real-time whole-genome sequencing for surveillance of Listeria monocytogenes, France. *Emerg Infect Dis.* (2017) 23:1462–70. doi: 10.3201/eid2309.170336
43. Pearce ME, Alikhan NF, Dallman TJ, Zhou Z, Grant K, Maiden MCJ. Comparative analysis of core genome MLST and SNP typing within a European Salmonella serovar Enteritidis outbreak. *Int J Food Microbiol.* (2018) 274:1–11. doi: 10.1016/j.ijfoodmicro.2018.02.023
44. Crabb HK, Allen JL, Devlin JM, Firestone SM, Wilks CR, Gilkerson JR. Salmonella spp. transmission in a vertically integrated poultry operation: clustering and diversity analysis using phenotyping (serotyping, phage typing) and genotyping (MLVA). *PLoS ONE.* (2018) 13:e0201031. doi: 10.1371/journal.pone.0201031
45. Miller EA, Elnekave E, Flores-Figueroa C, Johnson A, Kearney A, Munoz-Aguayo J, et al. Emergence of a novel Salmonella enterica serotype reading clonal group is linked to its expansion in commercial turkey production, resulting in unanticipated human illness in North America. *mSphere.* (2020) 5:e00056-20. doi: 10.1128/mSphere.00056-20
46. Nadon C, Van Walle I, Gerner-Smidt P, Campos J, Chinen I, Concepcion-Acevedo J, et al. PulseNet international: vision for the implementation of whole genome sequencing (WGS) for global foodborne disease surveillance. *Euro Surveill.* (2017) 22:30544. doi: 10.2807/1560-7917.ES.2017.22.23.30544
47. Katz LS, Griswold T, Williams-Newkirk AJ, Wagner D, Petkau A, Sieffert C, et al. A comparative analysis of the Lyve-SET phylogenomics pipeline for genomic epidemiology of foodborne pathogens. *Front Microbiol.* (2017) 8:375. doi: 10.3389/fmicb.2017.00375
48. Jagadeesan B, Baert L, Wiedmann M, Orsi RH. Comparative analysis of tools and approaches for source tracking listeria monocytogenes in a food facility using whole-genome sequence data. *Front Microbiol.* (2019) 10:947. doi: 10.3389/fmicb.2019.00947
49. Schürch AC, Arredondo-Alonso S, Willems RJL, Goering R V. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect.* (2018) 24:350–4. doi: 10.1016/j.cmi.2017.12.016
50. Uelze L, Grützkke J, Borowiak M, Hammerl JA, Juraschek K, Deneke C, et al. Typing methods based on whole genome sequencing data. *One Heal Outlook.* (2020) 2:1–19. doi: 10.1186/s42522-020-0010-1
51. Davis S, Pettengill JB, Luo Y, Payne J, Shpuntoff A, Rand H, et al. CFSAN SNP pipeline: an automated method for constructing snp matrices from next-generation sequence data. *PeerJ Comput Sci.* (2015) 1:e20. doi: 10.7717/peerj-cs.20

52. Petkau A, Mabon P, Sieffert C, Knox NC, Cabral J, Iskander M, et al. SNVPhyl: a single nucleotide variant phylogenomics pipeline for microbial genomic epidemiology. *Microb Genom.* (2017) 3:e000116. doi: 10.1099/mgen.0.000116
53. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. Evaluation of whole genome sequencing for outbreak detection of *salmonella enterica*. *PLoS ONE.* (2014) 9:e87991. doi: 10.1371/journal.pone.0087991
54. Zhou Z, Charlesworth J, Achtman M. HierCC: A Multi-Level Clustering Scheme for Population Assignments based on Core Genome MLST. *Bioinformatics.* Available online at: <https://github.com/zheminzhou/HierCC> (accessed March 9, 2021). doi: 10.1101/2020.11.25.397539

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Singh, Li, Beshearse, Blanton, DeMent and Havelaar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.