JCM
Journals.ASM.org

Characterization of Foodborne Outbreaks of *Salmonella enterica* Serovar Enteritidis with Whole-Genome Sequencing Single Nucleotide Polymorphism-Based Analysis for Surveillance and Outbreak Detection

Angela J. Taylor,[a] Victoria Lappi,[a] William J. Wolfgang,[b] Pascal Lapierre,[b] Michael J. Palumbo,[b] Carlota Medus,[a] David Boxrud[a]

Minnesota Department of Health, Saint Paul, Minnesota, USA[a]; Wadsworth Center/New York State Department of Health, Albany, New York, USA[b]

Salmonella enterica **serovar Enteritidis is a significant cause of gastrointestinal illness in the United States; however, current molecular subtyping methods lack resolution for this highly clonal serovar. Advances in next-generation sequencing technologies have made it possible to examine whole-genome sequencing (WGS) as a potential molecular subtyping tool for outbreak detection and source trace back. Here, we conducted a retrospective analysis of *S.* Enteritidis isolates from seven epidemiologically confirmed foodborne outbreaks and sporadic isolates (not epidemiologically linked) to determine the utility of WGS to identify outbreaks. A collection of 55 epidemiologically characterized clinical and environmental *S.* Enteritidis isolates were sequenced. Single nucleotide polymorphism (SNP)-based cluster analysis of the *S.* Enteritidis genomes revealed well supported clades, with less than four-SNP pairwise diversity, that were concordant with epidemiologically defined outbreaks. Sporadic isolates were an average of 42.5 SNPs distant from the outbreak clusters. Isolates collected from the same patient over several weeks differed by only two SNPs. Our findings show that WGS provided greater resolution between outbreak, sporadic, and suspect isolates than the current gold standard subtyping method, pulsed-field gel electrophoresis (PFGE). Furthermore, results could be obtained in a time frame suitable for surveillance activities, supporting the use of WGS as an outbreak detection and characterization method for *S.* Enteritidis.**

Foodborne bacterial pathogen characterization, surveillance, and outbreak detection is an important function of the public health laboratory (1). Current practices involve time- and labor-intensive phenotypic typing, including biochemical profiling, phage typing, serotyping, and antimicrobial susceptibility testing. In addition, a variety of species-specific molecular methods for advanced characterization are utilized, including pulsed-field gel electrophoresis (PFGE), multiple-locus variable number tandem repeat analysis (MLVA), and virulence gene typing (2). Often, it is necessary to combine results from multiple techniques to provide an adequate level of discrimination in order to identify outbreak clusters within routine clinical surveillance isolates.

Salmonella is an important foodborne pathogen that is estimated to be responsible for approximately 1 million cases of illness and more than 450 deaths annually in the United States (3). *Salmonella enterica* serovar Enteritidis is responsible for 36% of *Salmonella* outbreaks in the United States, and in 1990, it replaced *Salmonella enterica* serovar Typhimurium as the most frequently reported serotype of *Salmonella* worldwide (4, 5). It is estimated that approximately 64% of *S.* Enteritidis clinical cases are attributable to contaminated eggs and 18% to poultry products (5, 6).

There is limited genetic variation between the strains of *S.* Enteritidis, which reduces the utility of current subtyping methods (7–9). For example, PFGE, the gold standard subtyping method implemented in all PulseNet laboratories, and MLVA often do not provide the resolution to differentiate between outbreak and sporadic samples in this serovar (2, 10, 11). In Minnesota, 74% of *S.* Enteritidis isolates are comprised of three CDC PulseNet PFGE pattern subtypes: JEGX01.0004, JEGX01.0002, and JEGX01.0005. PFGE analysis utilizing multiple enzymes increases the discriminatory power of PFGE compared to that of single enzyme analysis; however, this increases the cost and time of analysis, and the resolution remains suboptimal. MLVA has shown enhanced resolution compared to that of PFGE, but concerns remain about cost, usability, and turnaround time, precluding its widespread adoption. In addition, PulseNet protocols for MLVA exist only for *S.* Enteritidis, *S.* Typhimurium, and *Escherichia coli* O157:H7, further limiting its utility (10).

Next-generation sequencing technologies and simplified sample preparation have made it possible to complete whole-genome sequencing (WGS) of bacterial and viral isolates in less than 48 h. This speed and potential for improved cluster resolution make it an attractive alternative to conventional subtyping methods in clinical and public health laboratories (12–14). Improved outbreak resolution and source trace back have been shown in retrospective studies involving *Klebsiella pneumoniae* (15), *Campylobacter* spp. (16), *Escherichia coli* O104:H4 (17), *Legionella* spp.

Received 15 May 2015 **Returned for modification** 19 June 2015
Accepted 7 August 2015

Accepted manuscript posted online 12 August 2015

Citation Taylor AJ, Lappi V, Wolfgang WJ, Lapierre P, Palumbo MJ, Medus C, Boxrud D. 2015. 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 53:3334–3340. doi:10.1128/JCM.01280-15.

Editor: D. J. Diekema

Address correspondence to Angela J. Taylor, Angie.taylor@state.mn.us.

doi:10.1128/JCM.01280-15

(18), *Listeria monocytogenes* (19), *Vibrio cholerae* (20), carbapenem-resistant *Enterobacter cloacae* (18), *Mycobacterium tuberculosis* (21–23), methicillin-resistant *Staphylococcus aureus* (24–26), *Acinetobacter baumannii* (27), and *Clostridium difficile* (25). Studies using WGS for real-time surveillance and outbreak detection are more limited but have begun to show the potential value of this technique to subtype and characterize isolates during outbreaks for public health purposes (28, 29).

Sequence-based analysis of *Salmonella* organisms from next-generation sequencing data has been used to examine outbreak clusters of *Salmonella enterica* serovar Montevideo (30–32) and *S.* Enteritidis (7, 11, 29). Outbreak-related organisms in these two highly homologous serovars are distinguished by fewer than 20 pairwise single nucleotide polymorphisms (SNPs) compared to the sequence of a reference strain.

In our retrospective study, next-generation sequence data from a total of 55 outbreak-related and sporadic isolates were generated and analyzed. Detailed epidemiological data were used to define outbreak, sporadic, and suspect samples in order to address the discriminatory ability of WGS compared to that of current typing methods. We demonstrate that SNP-based analysis of WGS has improved discrimination compared to that of PFGE. This SNP-based approach reliably clustered outbreak samples and could quantitate the genetic distance between outbreak- and non-outbreak-associated isolates. This study contributes to the development of cluster definitions for prospective implementation of WGS-based outbreak surveillance and detection for *S.* Enteritidis in Minnesota.

MATERIALS AND METHODS

Bacterial isolates. The 55 strains sequenced for this study were selected from clinical and environmental *S.* Enteritidis isolates that were previously characterized by PFGE in Minnesota and Ohio from 2001 to 2014. The cohort contains 28 isolates from seven distinct foodborne outbreaks and 27 sporadic isolates. All isolates are from patients that were interviewed using the Minnesota standard enteric interview form (updated version from June 2013 [http://mnfoodsafetycoe.umn.edu/interview-forms-2/]) to attain demographic information, exposure history (including food consumption history and locations), and travel history. Isolate metadata (Table 1) and epidemiological details on outbreaks (Table 2) are provided.

Isolates were characterized as "outbreak" or "sporadic" based on the epidemiological information obtained through interviews of the case patient and the isolate's PFGE profile. Isolates were defined as outbreak isolates if they were part of a confirmed outbreak that contained two or more cases with an identified common exposure. Four isolates were collected over a span of 5 weeks from a single individual in outbreak 2. Isolates were defined as sporadic if they were not part of a recognized outbreak. An effort was made to select some sporadic isolates with temporal or PFGE pattern similarity to outbreak isolates. Epidemiological information for four isolates (labeled "suspect" in Table 1) led epidemiologists to believe that these isolates were related to one of the outbreaks in our study; however, no common exposure was found.

PFGE. Pulsed-field gel electrophoresis was performed at MDH with XbaI (Roche) using standardized methods for PulseNet laboratories (33). PFGE patterns were uploaded to the PulseNet national database, and pattern designations were assigned by the Centers for Disease Control and Prevention (CDC).

Whole-genome sequencing. Whole-genome sequencing was performed at the Minnesota Department of Health. Genomic DNA was obtained from a single colony streak incubated at 36°C overnight on tryptic soy blood agar. DNA was extracted using the Qiagen QIAcube and the DNeasy blood and tissue kit (Qiagen). DNA concentrations were quantitated using the Qubit double-stranded-DNA high-sensitivity (HS) assay kit (Thermo Fisher Scientific). Sequencing library preparation for multiplexed paired-end libraries was completed by following the manufacturer's guidelines for the Nextera XT DNA sample prep kit (Illumina) and Nextera XT index kit (Illumina). Sequencing was performed using V2 chemistries on the Illumina MiSeq following standard FASTQ-only generation protocols to produce 250-bp paired-end reads.

Sequence assembly and analysis. Sequence analysis was performed by the Wadsworth Center/New York State Department of Health bioinformatics core facility. *S.* Enteritidis strain P125109 was used as a reference genome to map the sequence reads and find positions with SNPs. The raw reads were mapped over the reference genome using BWA-MEM version 0.7.5a-r405 (34) with default parameters. The reads were sorted, and duplicate reads were removed using Picard-tools version 1.27. Read-mapping statistics were extracted with Samtools flagstat version 0.1.19-44428cd (35), and final coverage statistics were retrieved using genomeCoverageBed from the Bedtools (36) package version 2.17.0. A final read pileup was generated using Samtools mpileup, and the variant call file (VCF) was produced with BCFtools version 0.1.19-44428cd (35), ignoring indels. Each individual genome position in the VCF file (variant and wild-type [wt] positions) was assessed to determine the exact nucleotide state in the sequenced genome and to create a high quality consensus sequence. To identify an SNP, a genome position was required to have at least 20× depth of coverage of high quality mapped reads with 95% of the reads in agreement, as determined by the DP4 field in the VCF file. Positions that failed these requirements or that mapped over phage-associated islands and repeat regions were marked as unknown-state Ns in the consensus sequence. Genomic coordinates corresponding to phage sequences and repetitive elements in the reference genome were determined using Phast (37) and Mummer (38), respectively.

The SNP alignment was created by comparing all the resulting consensus sequences and by retrieving positions where at least one of the sequences experienced a nucleotide change compared to the sequence of the reference genome. The maximum-likelihood phylogenetic tree was calculated with PhyML (39) using a K80 (K2P) model, no gamma, and the SPR tree search algorithm. An SNP heatmap was calculated in R version 3.1.2 with the Package gplots using the ratio (number of SNP differences/total number of non-N positions) between any pairwise consensus sequence comparisons.

Accession numbers. Raw sequence reads for all isolates can be found at NCBI. SRA accession numbers are listed in Table 1. All draft genomes were deposited at NCBI under BioProject record PRJNA237212, called GenomeTrakr real-time SE: Minnesota Department of Health (http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA237212).

RESULTS

A total of 55 isolates from 51 different patients and one environmental site were sequenced. Pairwise comparison of all *Salmonella* isolate genomes to the reference genome yielded a total of 2,580 genome positions with SNPs, average reference genome coverage of 99.24%, and average sequencing depth of 99×.

Outbreak, sporadic, and suspect isolate WGS clustering. All outbreak isolates were found to be closely related to other isolates from the same outbreak in the SNP-based phylogenetic tree, which is concordant with epidemiological data (Fig. 1). The number of isolates from each outbreak ranged from two to seven. All outbreak isolates varied by three SNPs or fewer from other isolates within the outbreak (range, 0 to 3 SNPs; median, 0.9) (Table 2). In comparison, outbreak isolate clusters differed by an average of 42.4 SNPs (sample standard deviation [*s*] = 34.8; range, 18 to 119) from the nearest nonoutbreak neighbor isolate, and sporadic isolates by an average of 68.0 SNPs ($s = 64$; range, 18 to 249) from the nearest sporadic isolate. Suspect samples (MDH-2014-00208, MDH-2014-00213, MDH-2014-00241, and MDH-2014-00243)

TABLE 1 Metadata and isolate details for *S.* Enteritidis isolates

Outbreak (blank for sporadics)	MN WGS ID[a] for isolate	Collection date (mo/day/yr)	State	Source	NCBI accession no.	CDC PFGE pattern[b]
1	MDH-2014-00209	9/16/2000	MN	Feces	SRS569751	JEGX01.0004
1	MDH-2014-00210	9/14/2000	MN	Feces	SRS569711	JEGX01.0004
1	MDH-2014-00211	9/18/2000	MN	Feces	SRS569685	JEGX01.0004
1	MDH-2014-00212	9/19/2000	MN	Feces	SRS569765	JEGX01.0004
2	MDH-2014-00218	5/8/2001	MN	Feces	SRS569723	JEGX01.0050
2	MDH-2014-00219	5/9/2001	MN	Feces	SRS569721	JEGX01.0050
2	MDH-2014-00220	5/9/2001	MN	Feces	SRS569701	JEGX01.0050
2	MDH-2014-00222	5/25/2001	MN	Feces	SRS569753	JEGX01.0050
2	MDH-2014-00223	5/25/2001	MN	Feces	SRS570367	JEGX01.0050
2	MDH-2014-00225	6/10/2001	MN	Feces	SRS569694	JEGX01.0050
2	MDH-2014-00228	7/6/2001	MN	Feces	SRS569758	JEGX01.0050
3	MDH-2014-00227	6/28/2001	MN	Feces	SRS569799	JEGX01.0021
3	MDH-2014-00229	7/4/2001	MN	Feces	SRS569790	JEGX01.0021
3	MDH-2014-00230	7/5/2001	MN	Feces	SRS569707	JEGX01.0021
3	MDH-2014-00251	7/16/2001	MN	Feces	SRS569677	JEGX01.0021
4	MDH-2014-00234	11/23/2003	MN	Feces	SRS569755	JEGX01.0004
4	MDH-2014-00252	11/20/2003	MN	Feces	SRS569683	JEGX01.0004
4	MDH-2014-00253	12/1/2003	MN	Feces	SRS569798	JEGX01.0004
4	MDH-2014-00254	11/27/2003	MN	Feces	SRS569759	JEGX01.0004
5	MDH-2014-00238	8/15/2011	MN	Feces	SRS569681	JEGX01.0004
5	MDH-2014-00239	8/22/2011	MN	Feces	SRS569674	JEGX01.0004
5	MDH-2014-00240	8/25/2011	MN	Feces	SRS569695	JEGX01.0004
5	MDH-2014-00242	9/13/2011	MN	Feces	SRS569722	JEGX01.0004
5	MDH-2014-00244	10/14/2011	MN	Egg Farm Environmental Swab	SRS569702	JEGX01.0004
6	MDH-2014-00249	1/31/2014	MN	Feces	SRS569706	JEGX01.0034
6	MDH-2014-00250	1/30/2014	MN	Feces	SRS569696	JEGX01.0034
7	MDH-2014-00255	1/1/2014	OH	Feces	SRS569752	JEGX01.0034
7	MDH-2014-00256	2/1/2014	OH	Feces	SRS569675	JEGX01.0034
	MDH-2014-00202	6/28/2000	MN	Feces	SRS569705	JEGX01.0004
	MDH-2014-00203	7/5/2000	MN	Feces	SRS569710	JEGX01.0002
	MDH-2014-00204	7/13/2000	MN	Feces	SRS569744	JEGX01.0009
	MDH-2014-00205	8/18/2000	MN	Feces	SRS569692	JEGX01.0004
	MDH-2014-00206	8/18/2000	MN	Feces	SRS569717	JEGX01.0002
	MDH-2014-00207	8/27/2000	MN	Feces	SRS569716	JEGX01.0002
Suspect in outbreak 1	MDH-2014-00208	9/19/2000	MN	Feces	SRS569796	JEGX01.0004
Suspect in outbreak 1	MDH-2014-00213	9/27/2000	MN	Feces	SRS569714	JEGX01.0004
	MDH-2014-00214	3/4/2001	MN	Feces	SRS569699	JEGX01.0002
	MDH-2014-00215	4/15/2001	MN	Blood	SRS569693	JEGX01.1077
	MDH-2014-00216	4/27/2001	MN	Blood	SRS569682	JEGX01.0005
	MDH-2014-00217	4/28/2001	MN	Feces	SRS569794	JEGX01.0002
	MDH-2014-00221	5/6/2001	MN	Feces	SRS569724	JEGX01.0009
	MDH-2014-00224	6/8/2001	MN	Feces	SRS569673	JEGX01.0005
	MDH-2014-00226	6/16/2001	MN	Feces	SRS569739	JEGX01.0021
	MDH-2014-00231	6/25/2001	MN	Feces	SRS569680	JEGX01.0004
	MDH-2014-00232	10/11/2001	MN	Feces	SRS569679	JEGX01.0034
	MDH-2014-00233	12/3/2001	MN	Blood	SRS569766	JEGX01.0004
	MDH-2014-00235	9/26/2005	MN	Feces	SRS569712	JEGX01.0004
	MDH-2014-00236	5/3/2011	MN	Feces	SRS569708	JEGX01.0019
	MDH-2014-00237	6/17/2011	MN	Feces	SRS569703	JEGX01.0019
Suspect in outbreak 5	MDH-2014-00241	9/10/2011	MN	Feces	SRS569718	JEGX01.0004
Suspect in outbreak 5	MDH-2014-00243	10/11/2011	MN	Feces	SRS569676	JEGX01.0004
	MDH-2014-00245	6/20/2012	MN	Feces	SRS569719	JEGX01.0034
	MDH-2014-00246	7/24/2012	MN	Feces	SRS569800	JEGX01.0004
	MDH-2014-00247	7/30/2012	MN	Arm swab	SRS569704	JEGX01.0094
	MDH-2014-00248	6/4/2013	MN	Feces	SRS569795	JEGX01.0019

[a] ID, identifier.
[b] XbaI pattern only.

were 0, 131, 18, and 148 SNPs, respectively, from the outbreak isolates they were suspected of being related to.

PFGE pattern distribution. In Fig. 1, group I and group II show the distribution of JEGX01.0004 and JEGX.0002, two of the three most common PFGE patterns in Minnesota. There were averages of 141 SNP ($s = 71$; range, 19 to 224) differences within the JEGX01.0004 isolates and 58 SNP ($s = 8.7$; range, 46 to 74) differences within the JEGX01.0002 isolates (only one representative from each outbreak is included in this statistic). There was an average of 458.4 SNP ($s = 22.3$) differences between group I and group II isolates, with a minimum distance of 403 SNPS.

DISCUSSION

S. Enteritidis became a significant source of illness in the United States in the early 1990s and has always been a challenge for traditional molecular subtyping techniques (40). In comparison to PFGE, where 53% of *S.* Enteritidis isolates in Minnesota fall into a single PFGE XbaI pattern (JEGX01.0004), whole-genome sequencing is able to delineate related and nonrelated samples with exceptional resolution (11). The objective of this study was to determine the characteristics of *S.* Enteritidis outbreak clusters in Minnesota and how they compare to those of sporadic isolates using WGS SNP-based phylogenetic analysis. We established that *S.* Enteritidis isolates within the same outbreak were within three SNPs of each other, while the nearest nonoutbreak isolate to an outbreak group differed by an average of 42.4 SNPs. This result provides necessary context for understanding and interpreting the genetic diversity found by WGS of *S.* Enteritidis isolates from Minnesota for surveillance and outbreak detection.

The utility of WGS in a real-time setting was apparent during the time frame of this study. While investigating outbreak 6 in early 2014, a PulseNet national database query showed that the Ohio Department of Health was investigating an outbreak with the same primary and secondary PFGE pattern. Two isolates from this Ohio outbreak (outbreak 7) were sequenced and included in our pairwise SNP comparison. While the outbreak 6 isolates differed from each other by 0 SNPs and the outbreak 7 isolates differed from each other by only 1 SNP, the outbreak 6 and outbreak 7 isolates differed by 32.5 SNPs. The analysis of the other outbreaks in this study provided context for comparing outbreaks 6 and 7, indicating that these two sets of isolates were likely not epidemiologically related. Epidemiological evidence to link these clusters was never found, supporting the hypothesis that these outbreaks did not share a source.

We investigated the genetic diversity and stability of *S.* Enteritidis in the human host during a prolonged infection. Isolates MDH-2014-0022, MDH-2014-0023, MDH-2014-0025, and MDH-2014-0028 were collected from the same individual over a 5-week time frame. All isolates from this patient were within two SNPs of each other. Additionally, we also observed no correlation between isolate collection date and SNP differences among all of the outbreaks investigated. Although the sample size of isolates was small, these observations support a hypothesis that *S.* Enteritidis shows little genetic diversity in the host over time and that varying collection dates have little effect on SNP differences between outbreak isolates. This finding is supported by other WGS analyses of *Salmonella* strains and suggests that outbreak clades are genetically stable (7, 41–43).

A limitation to this study is the sample selection method. The outbreaks were selected based on the quality of the epidemiolog-

TABLE 2 Epidemiological details and SNP characteristics of *S.* Enteritidis outbreak isolates

Outbreak	Location (state)	Time frame	Setting	No. of cases	No. with laboratory confirmation	Vehicle	No. of isolates sequenced in this study	PFGE pattern	No. of SNPs among outbreak isolates		Minimum SNP difference from nearest neighbor[c]
									Range[a]	Median[b]	
1	MN	Sept 2000	Restaurant, table service	14	10	Multiple items	4	JEGX01.0004	0–2	1	30
2	MN	May 2001	Banquet, hotel	54	18	Eggs benedict	7	JEGA26.0002	0–2	1	31
3	MN	June–July 2001	Restaurant, table service	16	12	Eggs	4	JEGA26.0005	0–1	0.5	45
4	MN	Oct–Nov 2003	Restaurant	30	27	Eggs	4	JEGX01.0004	0–2	1	119
5	MN	Aug–Oct 2011	Commercially distributed organic eggs	8	7	Eggs	4	JEGX01.0004	0–3	2	18
6	MN	Jan–Feb 2014	Restaurant, table service	36	15	Unknown	2	JEGX01.0034	0	0	22
7	OH	Jan–Feb 2014	Cafeteria	8	7	Unknown	2	JEGX01.0034	1	1	32

[a] Range of SNP differences between isolates within the same outbreak.
[b] Median number of SNPs between isolates within outbreak. The average (standard deviation) for all outbreaks was 0.9 (0.6).
[c] Minimum number of SNPs between outbreak isolates and nearest neighbor isolate. The average (standard deviation) for all outbreaks was 42.4 (34.8).

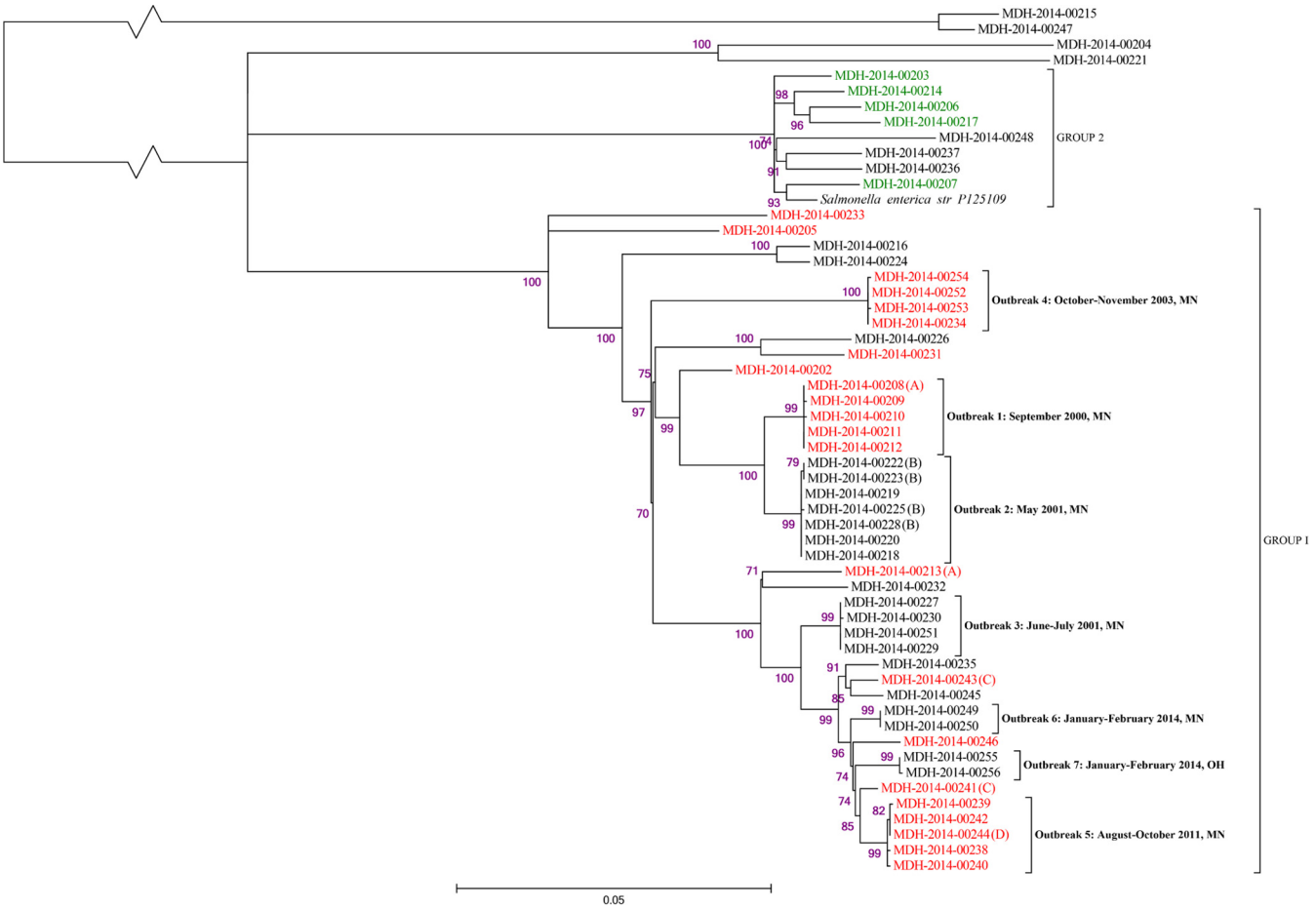

FIG 1 Maximum-likelihood tree of *S.* Enteritidis isolates produced by SNP analysis, showing outbreak clusters and PFGE pattern distribution. Outbreak isolates are indicated by brackets and descriptions giving the outbreak time frame (month[s] and year) and state. Sporadic and suspect isolates constitute all isolates not indicated as belonging to an outbreak (except for MDH-2014-00208, which was a suspect isolate that clusters with outbreak isolates). Isolates with CDC PulseNet PFGE pattern JEGX01.0004 are colored red, and isolates with pattern JEGX01.0002 are colored green. Purple values at the bases of the nodes are approximate likelihood-ratio test values displayed as percentages. The scale bar indicates the average number of substitutions per site. The letters in parentheses at the end of some isolate labels indicate the following: (A) suspect isolate in same time frame and PFGE pattern as outbreak 1; (B) all isolates are from the same patient; (C) suspect isolate in the same time frame and PFGE pattern as outbreak 5; (D) environmental isolate from outbreak 5.

ical data and included only isolates that were conclusively epidemiologically linked to the outbreak. Other selection biases include the following: (i) only 55 isolates were chosen for sequencing due to limited availability of resources (ii); our study examined isolates with the most common PFGE patterns in Minnesota and, therefore, does not represent the natural frequency and full diversity of PFGE patterns currently or historically encountered during surveillance; and (iii) the selection of sporadic isolates was done to intentionally exclude isolates that could have been a part of the studied outbreaks in order to have unambiguous WGS results to begin our interpretation of *S.* Enteritidis outbreaks. Analysis of all sporadic isolates collected during an outbreak time frame may reveal smaller SNP differences between sporadic and outbreak isolates.

Before SNP-based WGS analysis can be used routinely for *S.* Enteritidis outbreak surveillance, the method must be effective in real-time cluster detection. The developing technology of SNP-based whole-genome outbreak analysis must be examined and validated for each bacterial species and subtype and, potentially, by geographic region also, as lineages of serotypes vary spatially (11). Additionally, the development of standardized cluster definitions, sample preparation methods, and bioinformatics analysis and algorithms must be agreed upon for accurate surveillance use. The SNP-based analysis used in this study identified all outbreak isolates as distinct by their variation of less than four SNPs within the outbreak. Using this bioinformatics pipeline, future analyses could hypothetically start with a cluster definition of less than 10 SNPs; however, this would be dependent on further evaluation and on congruent results from other bioinformatics techniques. Different analysis pipelines would most likely affect the specific guidelines for SNP differences and cutoff values.

While it is the most common method used thus far, the SNP-based analysis used in this study may not prove to be the best method of utilizing high-throughput sequencing data for surveillance and outbreak detection. SNP-based WGS methods are time consuming, rely on the appropriate selection of a high-quality reference genome, require intensive computational infrastructure and adequate bioinformatics training, and do not result in stable sequence type nomenclature (16, 41). However, as WGS costs decrease and software is specifically designed for SNP-based sub-

typing and outbreak tracing, these limitations may be overcome. Additionally, sequencing analysis can reveal new targets for PCR-based subtyping techniques, such as SNP-based genotyping using PCR methods or targeted amplicon sequencing (43). Other methods that are being explored include whole-genome or core genome multilocus sequence typing (wgMLST and cgMLST, respectively) and qualitative and quantitative k-mer tree analysis (16, 28, 41, 44). These methods allow for less demanding computation and/or provide categorical/nomenclature-amenable outputs that could simplify analysis and allow for meaningful epidemiological interpretation and easy communication of results. Unlike SNP- and k-mer-based methods, wgMLST will also provide other characterizations, such as serotype, virulence type, and predicted antimicrobial resistance profile.

Barriers to implementing a real-time WGS outbreak detection approach in a public health laboratory include the cost of equipment and reagents, the need for a bioinformatics specialist/pipeline/software for analysis of sequencing data, and education of laboratory and epidemiology professionals for interpretation of the significance of WGS results (12, 45, 46). Despite the obstacles, WGS remains an improvement to current *S.* Enteritidis characterization methods and may become an even more attractive option as costs decrease. As WGS becomes faster and less expensive for public health laboratory use, in addition to subtyping, it can replace other phenotypic characterization methods, such as serotyping and antibiotic resistance testing (47).

In situations where there is a PFGE match between isolates but epidemiological links are not evident, it is difficult to determine the intensity of outbreak investigation necessary. This retrospective analysis has provided support for the reliability of WGS results when isolates show little diversity, which in turn provides guidance and support for targeted epidemiological investigations. However, it must be noted that WGS cannot be the sole basis of a determination that isolates originated at a common source or transmission type. Epidemiological information, including quality exposure data, must still support and enhance WGS results.

ACKNOWLEDGMENTS

This work was supported in part through a cooperative agreement with the Centers for Disease Control and Prevention Emerging Infections Program (grant 5U50CK000204). Also, this project was supported in part by an appointment to the Research Participation Program at the Center for Food Safety and Applied Nutrition administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration.

We would like to extend thanks to the FDA-CFSAN GenomeTrakr program for providing the sequencing instrumentation, reagents, supplies, training, and a staff ORISE internship. We thank Amy Saupe from the Minnesota Department of Health Foodborne Disease Unit for her epidemiological investigations and contributions. We thank the laboratorians in the PFGE laboratory at the Minnesota Department of Health for their work and reviews. We also acknowledge Ellen Salehi from the Bureau of Investigation and Eric Brandt from the Bureau of Public Health Laboratories at the Ohio Department of Health for their contribution of isolates and epidemiological data that are included in this study.

REFERENCES

1. **Witt-Kushner J, Astles JR, Ridderhof JC, Martin RA, Wilcke B, Jr, Downes FP, Inhorn SL, Kelley HP, Kimsey PB, Mills DE, Salfinger M, Shult PA, Verma MP, Becker SJ, Drabkowski DJ.** 2002. Core functions and capabilities of state public health laboratories: a report of the Association of Public Health Laboratories. MMWR Recommend Rep **51:**1–8.
2. **Foley SL, Zhao SH, Walker RD.** 2007. Comparison of molecular typing methods for the differentiation of *Salmonella* foodborne pathogens. Foodborne Pathog Dis **4:**253–276. http://dx.doi.org/10.1089/fpd.2007.0085.
3. **Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM.** 2011. Foodborne illness acquired in the United States—major pathogens. Emerg Infect Dis **17:**7–15. http://dx.doi.org/10.3201/eid1701.P11101.
4. **Guard-Petter J.** 2001. The chicken, the egg and *Salmonella enteritidis*. Environ Microbiol **3:**421–430. http://dx.doi.org/10.1046/j.1462-2920.2001.00213.x.
5. **Centers for Disease Control and Prevention.** 2013. Surveillance for foodborne disease outbreaks—United States, 1998-2008. MMWR Surveill Summ **62:**1–34.
6. **Painter JA, Hoekstra RM, Ayers T, Tauxe RV, Braden CR, Angulo FJ, Griffin PM.** 2013. Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998-2008. Emerg Infect Dis **19:**407–415. http://dx.doi.org/10.3201/eid1903.111866.
7. **Allard MW, Luo Y, Strain E, Pettengill J, Timme R, Wang C, Li C, Keys CE, Zheng J, Stones R, Wilson MR, Musser SM, Brown EW.** 2013. On the evolutionary history, population genetics and diversity among isolates of *Salmonella* Enteritidis PFGE pattern JEGX01.0004. PLoS One. **8**(1): e55254. http://dx.doi.org/10.1371/journal.pone.0055254.
8. **Saeed AM, Walk ST, Arshad M, Whittam TS.** 2006. Clonal structure and variation in virulence of *Salmonella* Enteritidis isolated from mice, chickens, and humans. J AOAC Int **89:**504–511.
9. **Botteldoorn N, Van Coillie E, Goris J, Werbrouck H, Piessens V, Godard C, Scheldeman P, Herman L, Heyndrickx M.** 2010. Limited genetic diversity and gene expression differences between egg- and non-egg-related *Salmonella* Enteritidis strains. Zoonoses Public Health **57:** 345–357. http://dx.doi.org/10.1111/j.1863-2378.2008.01216.x.
10. **Boxrud D, Pederson-Gulrud K, Wotton J, Medus C, Lyszkowicz E, Besser J, Bartkus JM.** 2007. 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 **45:**536–543. http://dx.doi.org/10.1128/JCM.01595-06.
11. **Deng X, Shariat N, Driebe EM, Roe CC, Tolar B, Trees E, Keim P, Zhang W, Dudley EG, Fields PI, Engelthaler DM.** 2015. Comparative analysis of subtyping methods against a whole-genome-sequencing standard in *Salmonella enterica* serotype Enteritidis. J Clin Microbiol **53:**212–218. http://dx.doi.org/10.1128/JCM.02332-14.
12. **Köser CU, Ellington MJ, Cartwright EJP, Gillespie SH, Brown NM, Farrington M, Holden MTG, Dougan G, Bentley SD, Parkhill J, Peacock SJ.** 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog **8**(8):e1002824. http://dx.doi.org/10.1371/journal.ppat.1002824.
13. **Cheung MK, Kwan HS.** 2012. Fighting outbreaks with bacterial genomics: case review and workflow proposal. Public Health Genomics **15:**341–351. http://dx.doi.org/10.1159/000342770.
14. **Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW.** 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat Rev Genet **13:**601–612. http://dx.doi.org/10.1038/nrg3226.
15. **Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Henderson DK, Palmore TN, Segre JA, NISC Comparative Sequencing Program Group** 2012. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. Sci Transl Med **4**(148):148ra116. http://dx.doi.org/10.1126/scitranslmed.3004129.
16. **Cody AJ, McCarthy ND, van Rensburg MJ, Isinkaye T, Bentley SD, Parkhill J, Dingle KE, Bowler I, Jolley KA, Maiden MCJ.** 2013. Real-time genomic epidemiological evaluation of human *Campylobacter* isolates by use of whole-genome multilocus sequence typing. J Clin Microbiol **51:** 2526–2534. http://dx.doi.org/10.1128/JCM.00066-13.
17. **Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, FitzGerald M, Godfrey P, Haas BJ, Murphy CI, Russ C, Sykes S, Walker BJ, Wortman JR, Young S, Zeng Q, Abouelleil A, Bochicchio J, Chauvin S, DeSmet T, Gujja S, McCowan C, Montmayeur A, Steelman S, Frimodt-Møller J, Petersen AM, Struve C, Krogfelt KA, Bingen E, Weill F-X, Lander ES, Nusbaum C, Birren BW, Hung DT, Hanage WP.** 2012. Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. Proc Natl Acad Sci U S A **109:**3065–3070. http://dx.doi.org/10.1073/pnas.1121491109.
18. **Reuter S, Ellington MJ, Cartwright EJP, Köser CU, Török ME, Gouliouris T, Harris SR, Brown NM, Holden MTG, Quail M, Parkhill J,**

Smith GP, Bentley SD, Peacock SJ. 2013. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. JAMA Intern Med **173:**1397–1404. http://dx.doi.org/10.1001/jamainternmed.2013.7734.

19. **Gilmour M, Graham M, Van Domselaar G, Tyler S, Kent H, Trout-Yakel K, Larios O, Allen V, Lee B, Nadon C.** 2010. High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. BMC Genomics **11:**120. http://dx.doi.org/10.1186/1471-2164-11-120.
20. **Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK.** 2011. The origin of the Haitian cholera outbreak strain. N Engl J Med **364:**33–42. http://dx.doi.org/10.1056/NEJMoa1012928.
21. **Gardy JL, Johnston JC, Sui SJH, Cook VJ, Shah L, Brodkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJM, Brinkman FSL, Brunham RC, Tang P.** 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med **364:**730–739. http://dx.doi.org/10.1056/NEJMoa1003176.
22. **Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE.** 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect Dis **13:**137–146. http://dx.doi.org/10.1016/S1473-3099(12)70277-3.
23. **Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, Blom J, Wirth T, Jaenicke S, Schuback S, Rüsch-Gerdes S, Supply P, Kalinowski J, Niemann S.** 2013. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. PLoS Med **10**(2)**:**e1001387. http://dx.doi.org/10.1371/journal.pmed.1001387.
24. **Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ.** 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med **366:**2267–2275. http://dx.doi.org/10.1056/NEJMoa1109910.
25. **Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CLC, Wilson DJ, Didelot X, O'Connor L, Lay R, Buck D, Kearns AM, Shaw A, Paul J, Wilcox MH, Donnelly PJ, Peto TEA, Walker AS, Crook DW.** 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. BMJ Open **2**(3)**:**e001124. http://dx.doi.org/10.1136/bmjopen-2012-001124.
26. **Harrison EM, Paterson GK, Holden MTG, Larsen J, Stegger M, Larsen AR, Petersen A, Skov RL, Christensen JM, Zeuthen AB, Heltberg O, Harris SR, Zadoks RN, Parkhill J, Peacock SJ, Holmes MA.** 2013. Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel *mecA* homologue *mecC*. EMBO Mol Med **5:**509–515. http://dx.doi.org/10.1002/emmm.201202413.
27. **Lewis T, Loman NJ, Bingle L, Jumaa P, Weinstock GM, Mortiboy D, Pallen MJ.** 2010. High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. J Hosp Infect **75:**37–41. http://dx.doi.org/10.1016/j.jhin.2010.01.012.
28. **Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM.** 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. J Clin Microbiol **52:**1501–1510. http://dx.doi.org/10.1128/JCM.03617-13.
29. **den Bakker HC, Allard MW, Bopp D, Brown EW, Fontana J, Iqbal Z, Kinney A, Limberger R, Musser KA, Shudt M, Strain E, Wiedmann M, Wolfgang WJ.** 2014. Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar Enteritidis. Emerg Infect Dis **20:**1306–1314. http://dx.doi.org/10.3201/eid2008.131399.
30. **Allard MW, Luo Y, Strain E, Li C, Keys CE, Son I, Stones R, Musser SM, Brown EW.** 2012. High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. BMC Genomics **13:**32. http://dx.doi.org/10.1186/1471-2164-13-32.
31. **den Bakker HC, Switt AIM, Cummings CA, Hoelzer K, Degoricija L, Rodriguez-Rivera LD, Wright EM, Fang R, Davis M, Root T, Schoonmaker-Bopp D, Musser KA, Villamil E, Waechter H, Kornstein L, Furtado MR, Wiedmann M.** 2011. A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp *enterica* serovar Montevideo pulsed-field gel electrophoresis type. Appl Environ Microbiol **77:**8648–8655. http://dx.doi.org/10.1128/AEM.06538-11.
32. **Lienau EK, Strain E, Wang C, Zheng J, Ottesen AR, Keys CE, Hammack TS, Musser SM, Brown EW, Allard MW, Cao GJ, Meng JH, Stones R.** 2011. Identification of a salmonellosis outbreak by means of molecular sequencing. N Engl J Med **364:**981–982. http://dx.doi.org/10.1056/NEJMc1100443.
33. **Ribot EM, Fair MA, Gautom R, Cameron DN, Hunter SB, Swaminathan B, Barrett TJ.** 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli O157:H7*, *Salmonella*, and Shigella for PulseNet. Foodborne Pathog Dis **3:**59–67. http://dx.doi.org/10.1089/fpd.2006.3.59.
34. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.
35. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing Subgroup.** 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:**2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.
36. **Quinlan AR, Hall IM.** 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26:**841–842. http://dx.doi.org/10.1093/bioinformatics/btq033.
37. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: A fast phage search tool. Nucleic Acids Res **39:**W347–W352. http://dx.doi.org/10.1093/nar/gkr485.
38. **Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL.** 2004. Versatile and open software for comparing large genomes. Genome Biol **5:**.
39. **Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol **59:**307–321. http://dx.doi.org/10.1093/sysbio/syq010.
40. **Braden CR.** 2006. *Salmonella enterica* serotype Enteritidis and eggs: a national epidemic in the United States. Clin Infect Dis **43:**512–517. http://dx.doi.org/10.1086/505973.
41. **Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM.** 2014. Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. PLoS One **9**(2)**:**e87991. http://dx.doi.org/10.1371/journal.pone.0087991.
42. **Okoro CK, Kingsley RA, Quail MA, Kankwatira AM, Feasey NA, Parkhill J, Dougan G, Gordon MA.** 2012. High-resolution single nucleotide polymorphism analysis distinguishes recrudescence and reinfection in recurrent invasive nontyphoidal *Salmonella* Typhimurium disease. Clin Infect Dis **54:**955–963. http://dx.doi.org/10.1093/cid/cir1032.
43. **Ogunremi D, Kelly H, Dupras AA, Belanger S, Devenish J.** 2014. Development of a new molecular subtyping tool for *Salmonella enterica* serovar Enteritidis based on single nucleotide polymorphism genotyping using PCR. J Clin Microbiol **52:**4275–4285. http://dx.doi.org/10.1128/JCM.01410-14.
44. **Kohl TA, Diel R, Harmsen D, Rothgänger J, Walter KM, Merker M, Weniger T, Niemann S.** 2014. Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach. J Clin Microbiol **52:**2479–2486. http://dx.doi.org/10.1128/JCM.00567-14.
45. **McPherson JD.** 2009. Next-generation gap. Nat Methods **6:**S2–S5. http://dx.doi.org/10.1038/nmeth.f.268.
46. **Schürch AC, Siezen RJ.** 2010. Genomic tracing of epidemics and disease outbreaks. Microb Biotechnol **3:**628–633. http://dx.doi.org/10.1111/j.1751-7915.2010.00224.x.
47. **Zhang S, Yin Y, Jones MB, Zhang Z, Deatherage Kaiser BL, Dinsmore BA, Fitzgerald C, Fields PI, Deng X.** 2015. *Salmonella* serotype determination utilizing high-throughput genome sequencing data. J Clin Microbiol **53:**1685–1692. http://dx.doi.org/10.1128/JCM.00323-15.