CrossMark
← click for updates

# Usefulness of High-Quality Core Genome Single-Nucleotide Variant Analysis for Subtyping the Highly Clonal and the Most Prevalent *Salmonella enterica* Serovar Heidelberg Clone in the Context of Outbreak Investigations

S. Bekal,[a,b] C. Berry,[c] A. R. Reimer,[c] G. Van Domselaar,[c] G. Beaudry,[a] E. Fournier,[a] F. Doualla-Bell,[a,d] E. Levac,[e] C. Gaulin,[f] D. Ramsay,[g] C. Huot,[h] M. Walker,[c] C. Sieffert,[c] C. Tremblay[a,b]

Laboratoire de santé publique du Québec, Québec, Canada[a]; University of Montreal, Department of Microbiology and Immunology, Québec, Canada[b]; National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada[c]; Department of Medicine, Division of Experimental Medicine, McGill University, Montréal, Québec, Canada[d]; Agence de la santé et des services sociaux de la Montérégie, Québec, Canada[e]; Ministère de la Santé et des Services Sociaux, Québec, Canada[f]; Ministère de l'Agriculture, des Pêcheries, et de l'Alimentation du Québec, Québec, Canada[g]; Direction de santé publique de la Capitale-Nationale, Québec, Canada[h]

*Salmonella enterica* serovar Heidelberg is the second most frequently occurring serovar in Quebec and the third-most prevalent in Canada. Given that conventional pulsed-field gel electrophoresis (PFGE) subtyping for common *Salmonella* serovars, such as *S.* Heidelberg, yields identical subtypes for the majority of isolates recovered, public health laboratories are desperate for new subtyping tools to resolve highly clonal *S.* Heidelberg strains involved in outbreak events. As PFGE was unable to discriminate isolates from three epidemiologically distinct outbreaks in Quebec, this study was conducted to evaluate whole-genome sequencing (WGS) and phylogenetic analysis as an alternative to conventional subtyping tools. Genomes of 46 isolates from 3 Quebec outbreaks (2012, 2013, and 2014) supported by strong epidemiological evidence were sequenced and analyzed using a high-quality core genome single-nucleotide variant (hqSNV) bioinformatics approach (SNV phylogenomics [SNVphyl] pipeline). Outbreaks were indistinguishable by conventional PFGE subtyping, exhibiting the same PFGE pattern (SHEXAI.0001/SHEBNI.0001). Phylogenetic analysis based on hqSNVs extracted from WGS separated the outbreak isolates into three distinct groups, 100% concordant with the epidemiological data. The minimum and maximum number of hqSNVs between isolates from the same outbreak was 0 and 4, respectively, while >59 hqSNVs were measured between 2 previously indistinguishable outbreaks having the same PFGE and phage type, thus corroborating their distinction as separate unrelated outbreaks. This study demonstrates that despite the previously reported high clonality of this serovar, the WGS-based hqSNV approach is a superior typing method, capable of resolving events that were previously indistinguishable using classic subtyping tools.

N ontyphoidal *Salmonella enterica* strains are important bacterial agents of salmonellosis in humans and animals (1) and represent up to 125,000 cases annually of foodborne gastroenteric disease arising from sporadic and outbreak events in Canada (2). More than 2,500 *Salmonella enterica* serovars have been described, but only a few have been associated with cases of human illness (3, 4). *Salmonella* Heidelberg ranks third and fourth among serovars causing human illness in Canada (5) and the United States (6), respectively, and is commonly detected in retail meat samples and food animals. While the majority of *Salmonella* infections are mild and self-limiting, *S.* Heidelberg can cause more severe diseases, including septicemia, myocarditis, extraintestinal infections, and death (7, 8).

Pulsed-field gel electrophoresis (PFGE) is the gold standard method used by Canadian public health laboratories for the molecular typing of *S.* Heidelberg, following standardized procedures set out by the PulseNet Canada guidelines. A well-recognized limitation of this classic typing method is that strains bearing highly common PFGE patterns occasionally render PFGE ineffective at detecting foodborne outbreaks from background sporadic cases, thus limiting the strength of laboratory evidence to support case linkages. Specifically, the *S.* Heidelberg PFGE pattern SHEXAI.0001/SHEBNI.0001 is the most common in all of North America.

In the province of Quebec (QC) and according to the Labora-

toire de santé publique du Québec (LSPQ) surveillance program, *S.* Heidelberg was the second most prevalent serovar recovered from clinical cases between 2005 and 2014. PFGE analysis alone revealed little genetic diversity among *S.* Heidelberg isolates. Between 2004 and 2014, 70% of QC isolates exhibited pulsotype 2 (SHXAI.0001/SHBNI.0001). Approximately 23% of the isolates were isolated from blood, supporting the increased invasive rate associated with this serovar, compared to 7% for *S. enterica* serovar Enteritidis and 5% for *S. enterica* serovar Typhimurium.

With decreasing costs and increasing feasibility, whole-genome sequencing (WGS) technology has emerged as an attractive

method for the real-time investigation of bacterial foodborne outbreaks and for routine surveillance purposes (9–11). The increasing readiness of WGS technology combined with the widely hypothesized and previously reported high concordance of WGS-based typing approaches with epidemiological data have resulted in the widespread adoption of WGS surveillance and outbreak response by global contemporaries responsible for public health and food safety, and this method is positioned to replace PFGE-based molecular epidemiology of foodborne pathogens (10, 12–14). This revolution in global disease surveillance has created the urgency for Canadian public health laboratories to apply WGS technology to routine public health activities.

Prior to implementation of a new technology that will inform public health and food safety decision-making, it is critical that validation studies be performed to (i) determine if the method indeed demonstrates increased discriminatory power relative to traditional subtyping tools in resolving S. Heidelberg events and (ii) understand how to responsibly interpret the data for public health decisions. Imperative to robust validation, WGS analyses must be applied to outbreak events wherein strong epidemiological evidence allowed the resolution of outbreaks from endemic strains.

In this study, we assessed the usefulness of a high-quality core genome single-nucleotide variant (hqSNV) analysis to distinguish between three epidemiologically defined outbreaks typed by PFGE and sharing the most commonly occurring PFGE pattern (SHEXAI.0001/SHEBNI.0001) of S. Heidelberg cases in North America.

## MATERIALS AND METHODS

**Bacterial strains and molecular typing.** The isolates included in this study were collected through the provincial surveillance program implemented in 2003 to monitor incidences of human salmonellosis, thus enabling detection of rapid clustering and/or outbreaks. Food isolates were received from the Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec (MAPAQ) Laboratoire d'expertises et d'analyses alimentaires following food-poisoning investigations. Strains were grown and maintained on triple sugar iron agar (TSI) at 37°C and stored in Trypticase soy broth with 10% glycerol. *Salmonella* serotyping was performed at the LSPQ following conventional agglutination methods, while phage typing was done at the Public Health Agency of Canada National Microbiology Laboratory (PHAC-NML) in Winnipeg, Canada. For this study, isolates were selected from three retrospective outbreaks involving the same PFGE pattern, occurring in 2012 ($n = 10$, 8 human clinical and 2 food isolates), 2013 ($n = 8$ human clinical isolates), and 2014 ($n = 28$, 12 human clinical and 16 food isolates), designated outbreaks 1, 2, and 3, respectively. Outbreaks 1 and 3 belonged to phage type (PT) 19, and those of outbreak 2 belonged to PT 26. Thirteen background clinical isolates were added to the study, 11 of which had the same PFGE pattern (SHEXAI.0001/SHEBNI.0001) and PTs 18, 19, 26, and 29; 1 isolate with a closely related PFGE pattern (SHXAI.0009/SHBNI.0025); and 1 isolate with a highly variant PFGE pattern (SHXAI.0197/SHBNI.0077). The subtyping and metadata for the isolates used in this study are summarized in Table 1.

**Whole-genome sequencing.** Whole-genome sequencing (WGS) of 59 selected S. Heidelberg isolates was performed at the PHAC-NML Core Genomics facility. Genomic DNA was extracted using the Epicentre metagenomic DNA isolation kit for water on bacteria cultured overnight in brain heart infusion broth, and sample libraries were prepared using the MiSeq Nextera XT library preparation kit (Illumina, Inc., San Diego, CA). Sequencing was performed on the Illumina MiSeq platform with 250-bp paired-end reads acquired using the MiSeq reagent kit V2 (500 cycles) to obtain an average genome coverage of >50 times for all isolates. Sequence

reads were assembled *de novo* into contigs using SPAdes (15) and annotated with Prokka (16). SPAdes-assembled contigs of <1 kb were filtered and removed from the analysis.

**Core genome hqSNV analysis.** To evaluate the capacity of WGS-based approaches for outbreak investigation, we used a phylogenomic approach using the PHAC-NML hqSNV pipeline (https://github.com/apetkau/core-phylogenomics). Briefly, paired-end sequencing reads were mapped against the complete reference genome S. Heidelberg SL476 using SMALT v 0.7.0.1 (http://www.sanger.ac.uk/resources/software/smalt/) with a k-mer size of 13, a step size of 6, and a minimum alignment fraction of 0.5. High-quality variants were called using FreeBayes v 0.9.8 (17) at a minimum mapping quality of 30, minimum base quality of 30, 75% minimum alternate fraction of variant bases in agreement, and minimum depth of coverage of 15. Positions not called were evaluated using the BCFtools component included in the SAMtools package (18), and positions with a discrepancy between the two variant callers were subsequently dropped from the analysis.

Illumina MiSeq raw reads were done by reference mapping to the fully closed S. Heidelberg SH-SL476 genome (GenBank accession no. NC_011083.1) using SMALT (ftp://ftp.sanger.ac.uk/pub/resources/software/smalt/smalt-manual-0.7.4.pdf).

**Phylogenomic relationship of isolates.** In this study, 394 hqSNVs were identified and used to generate the final multiple alignment. The goeBURST algorithm within PHYLOViZ (19) was used to infer the evolutionary relationship of the isolates and obtain a graphic minimum spanning tree (MST) representation of the evolutionary relationship of the isolates. PHYLOViZ is a user-friendly software that is well adapted for the analysis of sequence-based typing (SBT) data. PHYLOViZ does not directly accept pseudo alignment files as input; therefore, this file was reformatted beforehand using the GitHub repository for the pipeline and flow chart (https://github.com/apetkau/microbial-informatics-2014/tree/master/labs/core-snp). This method resulted in the generation of a table containing all sequence types (STs) and corresponding allelic profiles and a strain data file that links isolates to their respective ST.

**Nucleotide sequence accession number.** The sequence data supporting the results of this article have been deposited in the NCBI Sequence Read Archive under accession number SRP067504.

## RESULTS

**Epidemiological investigation of the three outbreaks.** Human clinical and food isolates with confirmed epidemiological links to one of three unrelated foodborne outbreaks occurring in the province of QC between 2012 and 2014 were subjected to further WGS-based analysis in this study.

Outbreak 1 occurred in May 2012 and was epidemiologically linked to a wedding event with 300 attendees. In the days following the celebration, 30.6% of attendees reported feeling ill to a primary care physician. The main symptoms reported were diarrhea (100%), fever (68%), and headache (55%). A cohort study was initiated to identify the cause of the outbreak, and stool cultures were collected from attendees who had presented to an emergency room. *Salmonella* serovar Heidelberg was identified from 10 stool samples collected from attendees who presented with symptoms of salmonellosis or reported feeling ill. The cohort study identified a single meal served at the wedding as the probable cause of the outbreak. Leftovers of the meal were cultured by MAPAQ's laboratory and subsequently tested positive for S. Heidelberg.

Outbreak 2 occurred in November 2013, when 14 cases of salmonellosis were reported to one public health regional authority in QC. The incidence rate of S. Heidelberg infection was 3.86 per 100,000 cases in 2013, which was statistically higher than the provincial or national rate for the previous 5 years (2008 to 2012,

TABLE 1 Epidemiological and microbiological information for the 59 *S.* Heidelberg human clinical and food isolates used in this study

| Strain | Source | Isolation date (mo-yr) | Outbreak no. | PFGE pattern | PT | Canadian PFGE pattern |
|--------|--------|------------------------|--------------|--------------|-----|-----------------------|
| SH12-001 | Human | 5-2012 | 1 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-002 | Human | 5-2012 | 1 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-003 | Human | 5-2012 | 1 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-004 | Human | 5-2012 | 1 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-005 | Human | 5-2012 | 1 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-006 | Human | 5-2012 | 1 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-007 | Human | 5-2012 | 1 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-008 | Human | 5-2012 | 1 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-009 | Food | 5-2012 | 1 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-010 | Food | 5-2012 | 1 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH13-001 | Human | 11-2013 | 2 | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH13-002 | Human | 11-2013 | 2 | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH13-003 | Human | 11-2013 | 2 | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH13-004 | Human | 11-2013 | 2 | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH13-005 | Human | 11-2013 | 2 | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH13-006 | Human | 11-2013 | 2 | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH13-007 | Human | 11-2013 | 2 | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH13-008 | Human | 11-2013 | 2 | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH14-001 | Human | 7-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-002 | Human | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-003 | Human | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-004 | Human | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-005 | Human | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-006 | Human | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-007 | Human | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-008 | Human | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-009 | Human | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-010 | Human | 8-2014 | 3 | 2 | 17 | SHXAI.0001/SHBNI.0001 |
| SH14-011 | Human | 8-2014 | 3 | 2 | 17 | SHXAI.0001/SHBNI.0001 |
| SH14-012 | Human | 8-2014 | 3 | 2 | ATHE-35 | SHXAI.0001/SHBNI.0001 |
| SH14-013 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-014 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-015 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-016 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-017 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-018 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-019 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-020 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-021 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-022 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-023 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-024 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-025 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-026 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-027 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH14-028 | Food | 8-2014 | 3 | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-011 | Human | 5-2012 | NA[a] | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH10-001 | Human | 8-2010 | NA | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH11-001 | Human | 7-2011 | NA | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH10-002 | Human | 2-2010 | NA | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH12-012 | Human | 6-2012 | NA | 2 | 18 | SHXAI.0001/SHBNI.0001 |
| SH12-014 | Human | 8-2012 | NA | 52 | 10 | SHXAI.0009/SHBNI.0025 |
| SH12-013 | Human | 2-2012 | NA | 87 | 32 | SHXAI.0197/SHBNI.0077 |
| SH10-014 | Human | 8-2010 | NA | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH11-002 | Human | 2-2011 | NA | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH10-015 | Human | 9-2010 | NA | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH08-001 | Human | 12-2008 | NA | 2 | 19 | SHXAI.0001/SHBNI.0001 |
| SH09-29 | Human | 8-2009 | NA | 2 | 26 | SHXAI.0001/SHBNI.0001 |
| SH10-30 | Human | 7-2010 | NA | 2 | 29 | SHXAI.0001/SHBNI.0001 |

[a] NA, not available.

**TABLE 2** Range of genetic distances between isolate pairs grouped by outbreak or by background (sporadic) isolates

| Outbreak or isolate | Genetic distance (hqSNV) for: | | |
| --- | --- | --- | --- |
| | Outbreak 1: SHXAI.0001/SHBNI.0001, PT 19 | Outbreak 2: SHXAI.0001/SHBNI.0001, PT 26 | Outbreak 3: SHXAI.0001/SHBNI.0001, PT 19 |
| Outbreak | | | |
| 1 | 0–4 | | |
| 2 | 82–86 | 0–2 | |
| 3 | 63–67 | 83–87 | 0–4 |
| | | | |
| Background isolate (PFGE pattern, phage type) | | | |
| Sporadic SHXAI.0001/SHBNI.0001, PT 19 | 8–73 | 76–85 | 57–74 |
| Sporadic SHXAI.0001/SHBNI.0001, PT 26 | 61–83 | 4–82 | 6–84 |
| Sporadic SHXAI.0001/SHBNI.0001, PT 18 | 75–78 | 87–88 | 68–71 |
| Sporadic SHXAI.0001/SHBNI.0001, PT 29 | 64–67 | 84–85 | 9–12 |
| Sporadic SHXAI.0009/SHBNI.0025, PT 10 | 76–77 | 88–89 | 77–80 |
| Sporadic SHXAI.0197/SHBNI.0077, PT 32 | 100–103 | 112–113 | 93–94 |
| | | | |
| Reference isolate SH-SL476 | 98–101 | 108–109 | 99–100 |

2.34/100 000). Stools from all cases were positive for *S. Heidelberg*, but only 8 of the people infected were interviewed through a questionnaire. All cases reported eating at a common restaurant that was identified as a potential source of the outbreak. MAPAQ's investigations did not enable identification of a contaminated food product associated with this restaurant. The 8 epidemiologically linked strains of *S. Heidelberg* obtained from the interviewees were included in the study.

Outbreak 3 occurred in August 2014, when 23 isolates were recovered within a 3-day time span, versus a weekly average of 1 or 2 isolates recovered at a regional laboratory, alerting public health officials to the outbreak. A total of 25 clinical cases were identified, of which 13 were confirmed and 12 remained probable. The definition of a confirmed case included a positive culture of *S. Heidelberg* and confirmation of exposure to the source. A probable case was defined as one having either a positive culture of *S. Heidelberg* or confirmed exposure to the source. An epidemiological follow-up tracing questionnaire identified a single restaurant as the potential source of this outbreak. The MAPAQ isolated 16 isolates from multiple suspected foods. Laboratory cultures identified *S. Heidelberg* as a causative agent, and, supported by epidemiological data, the outbreak was linked to the restaurant.

**PFGE and phage type analysis of outbreak strains.** PFGE analysis revealed that all human clinical and food isolates from the 3 outbreaks exhibited the most common PFGE pattern, pulsotype 2 (SHXAI.0001/SHBNI.0001). All isolates from outbreak 1 had the same phage type, PT 19, which is the most prevalent phage type displayed among *S. Heidelberg* isolates (Table 1).

All isolates associated with outbreak 2 exhibited PT 26, which is the third most commonly displayed phage type associated with pulsotype SHXAI.0001/SHBNI.0001. Most of the human clinical and food isolates from outbreak 3 belonged to PT 19, the most prevalent phage type associated with this PFGE pattern combination. Two human isolates displayed different phage types, PT 17 and ATHE-35 (a new phage type in the database). Thus, using PFGE typing alone, the three outbreaks were indistinguishable from one another by restriction digestion with the enzymes XbaI and BlnI, and outbreaks 1 and 3 were indistinguishable by both PFGE and phage type analysis. Given the large predominance of PTs 19 and 26 among *S. Heidelberg* isolates, phage typing is an

insufficient investigative tool that lacks the discriminatory power required to distinguish between outbreak isolates and cocirculating sporadic cases in the absence of a strong epidemiological linkage.

**Genomic diversity using hqSNV analysis.** Within each of the three outbreaks, the number of hqSNVs identified between human and food isolate pairs ranged between 0 and 4. Seven isolates (70%) from outbreak 1 (6 human clinical and 1 food) displayed 0 hqSNVs, while the second food strain, SH12-010, displayed a distance of 1 hqSNV to the human clinical isolates and 1 hqSNV to the other food isolate, SH12-009. Human isolates SH12-002 and SH12-007 displayed 1 and 3 hqSNVs, respectively, to the other isolates within outbreak 1 (Table 2). For outbreak 2, 6 human isolates had 0 hqSNVs between them and were considered genetic matches to one another, whereas isolates SH13-008 and SH13-004 were distinguished from these isolates by 1 hqSNV. No food strain was identified during the investigation of outbreak 2. For outbreak 3, 22 isolates (8 from human clinical and 14 from food samples) had 0 hqSNVs between them, forming a large group shown in Fig. 1. Three additional human isolates associated with outbreak 3 and 2 food isolates were separated from this main group of outbreak 3 isolates by 1 hqSNV. The human clinical isolate SH14-012 designated with a new phage type (ATHE-35) was separated by 3 and 4 hqSNVs from the other outbreak 3-associated isolates, confirming the epidemiological evidence that this clinical case was within the outbreak 3 case definition. Food isolates associated with outbreaks 1 and 3 were distanced from the human clinical isolates of the respective outbreak by 0 and 1 hqSNV, confirming that the suspected food product was the exposure source leading to the outbreak events.

The core genome phylogeny estimated by the SNV phylogenomics [SNVPhyl] pipeline grouped isolates belonging to each outbreak on separate distinct lineages (Fig. 1). This observation was consistent with the expected increase in discriminatory power of WGS-based methods relative to PFGE alone.

The outbreak 1 isolates (SHXAI.0001/SHBNI.0001/PT 19) displayed differences of between 82 and 86 SNVs from the outbreak 2 isolates (SHXAI.0001/SHBNI.0001/PT 26) and between 63 and 67 SNVs from the outbreak 3 isolates (SHXAI.0001/SHBNI.0001/PT 19) (Table 2). Isolates belonging to outbreak 2 (SHXAI.0001/
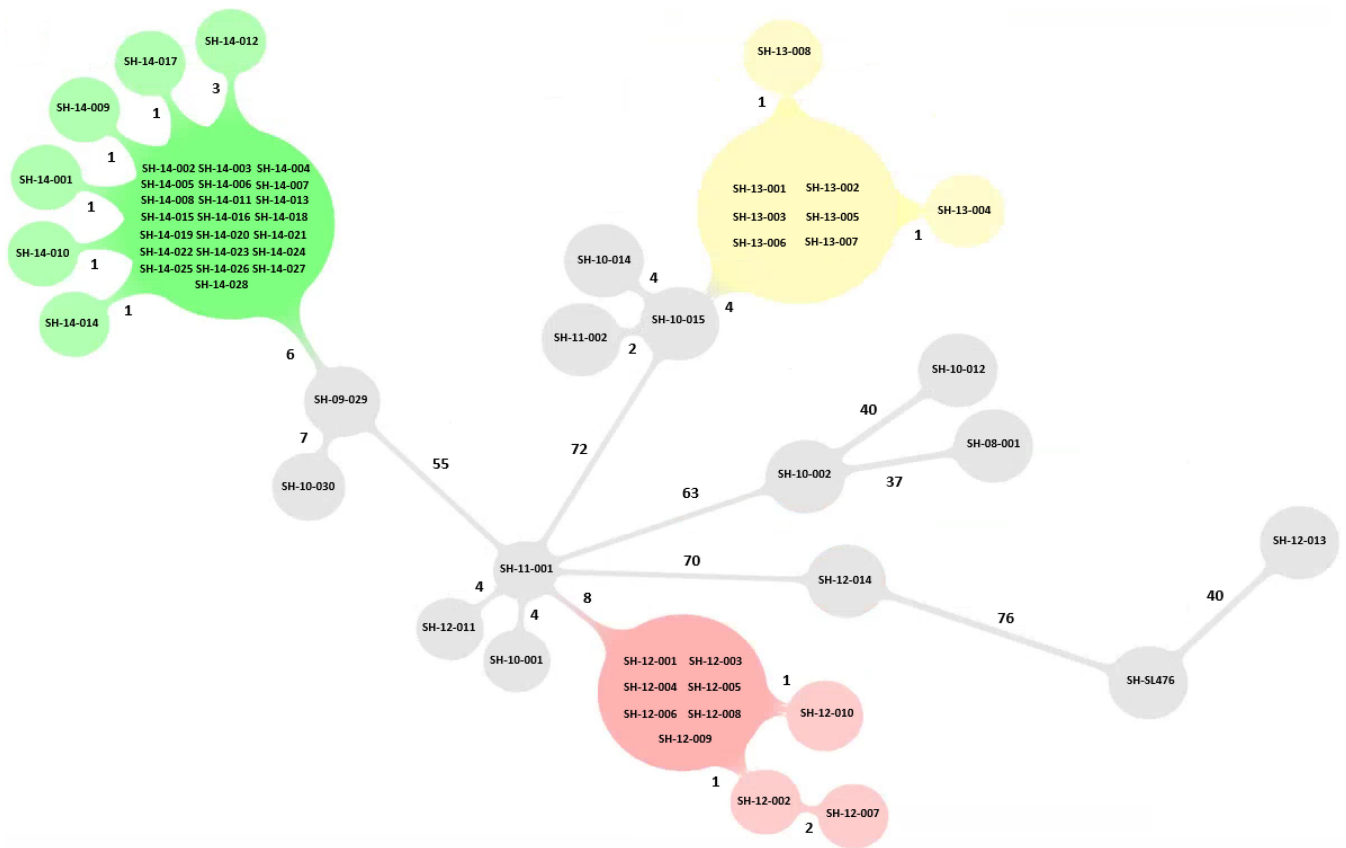
**FIG 1** Graphic representation of minimum spanning tree (MST) resulting from WGS typing of the 59 genomes involved in the 3 outbreaks using hqSNV analysis. Nonrelated outbreak isolates are shown in gray. The number on the branch denotes the number of hqSNV differences.

SHBNI.0001/PT 26) displayed differences of between 83 and 87 SNVs from isolates belonging to outbreak 3 (SHXAI.0001/ SHBNI.0001/PT 19).

To further inform the genetic context of outbreaks, 13 background isolates expected to be epidemiologically unrelated to the outbreaks were included in this study. These sporadic isolates were selected for their indistinguishable, close, or distinct PFGE patterns. Five isolates with pattern SHXAI.0001/SHBNI.0001 and PT 19 were collected in 2008 ($n = 1$), 2010 ($n = 2$), 2011 ($n = 1$), and 2012 ($n = 1$) as background isolates for outbreaks 1 and 3. These isolates displayed between 8 and 73 hqSNVs to the outbreak 1 isolates and between 57 and 74 hqSNVs to the outbreak 3 isolates. Isolate SH11-001, with a relatively low number of hqSNVs ($n = 8$) to outbreak 1, was recovered from a clinical case occurring in 2011 in a region geographically distant from the location of outbreak 1, thus decreasing the likelihood of an epidemiological linkage between these isolates.

Four background isolates for outbreak 2 (SHXAI.0001 /SHBNI.0001 and PT 26) were selected from clinical cases occurring between 2009 and 2011, thus predating the outbreak 2 event (2009 [$n = 1$], 2010 [$n = 2$], and 2011 [$n = 1$]). Phylogenetic analysis revealed that 3 of the 4 isolates (SH10-014, SH10-015, and SH11-002) isolated in 2010 and 2011 had genetic distances between 4 and 9 hqSNVs from the outbreak 2 group, suggesting a possible low level of genetic diversity within this PFGE/PT combination. In contrast, within the phylogenetic tree, the 2009 isolate (SH09-29) grouped closer to the outbreak 3 isolates (6 to 9

hqSNVs) than to the outbreak 2 group isolates (81 to 82 hqSNVs). The human clinical isolate SH10-30, belonging to the second most prevalent PFGE/PT combination seen in QC (SHXAI.0001/ SHBNI.0001/PT 29) was distant (on a separate branch) from outbreak 1 and outbreak 2 (>64 hqSNVs). However, this isolate fell on a branch distanced by 9 hqSNVs from the outbreak 3 branch carrying isolates with PT 19.

One sporadic background isolate included in this comparison, SH12-012, with a rarely recovered PFGE and phage type combination (SHXAI.0001/SHBNI.0001) and PT 18 (SH12-012), was grouped away from both the outbreak and other background isolates (>70 hqSNVs), confirming that it was a sporadic background isolate, genetically unrelated to the other isolates analyzed in this study.

hqSNV analysis revealed that the human clinical isolate SH12-014, included as a background with a PFGE pattern combination (SHXAI.0009/SHBNI.0025) highly similar to the outbreak (1, 2, and 3) isolates by 1 missing band on BlnI digestion (20.5 to 28.8 kb) and 1 missing band on XbaI digestion (33.3 kb), PT 10, was clearly distinct from the isolates with PFGE pattern SHXAI.0001/ SHBNI.0001 (>76 hqSNVs). The strain SH12-13, displaying a highly distinct PFGE/PT combination (SHXAI.0197/SHBNI.0077/PT 32) by 9 bands difference to the PFGE pattern SHXAI.0001/SHBNI.0001, was clearly not genetically related to the outbreaks (>93 hqSNVs).

The reference isolate, SH-SL476, displayed >100 hqSNVs to all QC isolates analyzed in this study.

## DISCUSSION

In this study, we undertook the WGS and analysis of three retrospective and epidemiologically characterized *S.* Heidelberg outbreak events that were indistinguishable by PFGE typing alone to assess whether WGS combined with hqSNV analysis has the discriminatory capacity required to distinguish among these events. *S.* Heidelberg is a highly clonal serovar in Canada and North America, dominated by the two-enzyme PFGE pattern combination SHXAI.0001/SHBNI.0001 (Canadian PulseNet designation) corresponding to JF6X01.0022/JF6A26.0001 (PulseNet USA designation). This study confirms the high genetic clonality of this serovar and demonstrates that predominance of the PFGE pattern SHXAI.0001/SHBNI.0001 is not attributable to the ubiquitous nature of this pattern but rather to the limited discriminatory power of PFGE typing. During outbreak investigations, factors defining the inclusion and exclusion of isolates from the event are based on a combination of laboratory and epidemiological evidence. Because the three outbreaks occurred as distinguished and delineated sources and events (i.e., restaurants and wedding), the epidemiological linkage of isolates to a specific event was clearly established by epidemiological trace-back investigation. However, because each of the selected outbreak isolates was associated with the most prevalent PFGE pattern encountered in North American clinical cases, the laboratory evidence is considered weak and does not strongly support isolate inclusion or exclusion in an event (20). As a result of these weakened criteria, the combined evidence does not enable source attribution during a foodborne outbreak. The pervasiveness of food distribution in Canada and beyond consequently requires a public health and food safety system that critically relies on laboratory evidence to detect and investigate foodborne disease outbreaks. In other words, outbreaks attributed to food products that are not consumed locally (at a single restaurant or social event, for example) will go undetected, thus preventing strong laboratory evidence to link human cases to each other and to potential sources. Recent studies have demonstrated the strength of WGS subtyping of outbreaks using hqSNV analysis for several foodborne outbreaks or outbreaks attributed to other serovars of *Salmonella* (4, 10, 12, 14, 21), including the previously problematic clonal *S.* Enteritidis (13). This study yielded very strong laboratory evidence of genetic relatedness between epidemiologically related isolates where previous conventional laboratory evidence obtained by PFGE and phage type was weak at best. Both WGS and hqSNV analysis clearly delineated each outbreak as a well-separated, distinct event on the maximum-likelihood phylogenetic tree, demonstrating 100% concordance with the epidemiological data. Isolates within an outbreak event were separated by a small number of hqSNVs, varying from 0 to 4 hqSNVs. This result indicates that these isolates are very close genetic matches, and these findings are highly promising for the application of WGS-based methods to future *S.* Heidelberg surveillance activities and outbreak investigations. In contrast to our findings, Leekitcharoenphon et al. (10) reported slightly increased single-nucleotide polymorphism (SNP) distances separating *S.* Heidelberg outbreak-related isolates (4 to19 SNPs). The slightly lower SNP distances revealed among *S.* Heidelberg isolates may be indicative of inherent levels of genetic diversity across the *Salmonella* serovars, with different serovars consisting of populations demonstrating variable genetic diversity, with no two serovars exactly alike. An additional possibility is that the previously reported findings are more reflective of the SNP calling methods applied to the analysis of the isolates and the high stringency with which hqSNVs are extracted and qualified using the SNVPhyl pipeline, which uses two variant callers to predict SNVs, SAMtools/BCFtools and FreeBayes. Another factor affecting SNP calls is the choice of a reference genome, with the differences in selection of a reference genome impacting the reference mapping and subsequent variant calling stages of the analysis. In this study, SL476, a complete *S.* Heidelberg genome, was used as the reference, whereas in the study by Leekitcharoenphon et al., isolates were mapped to a less closely related isolate, *S.* Enteritidis P125109 (10).

In this study, phylogenomic analysis revealed great hqSNV distances between outbreaks occurring at different periods. However, our analysis also revealed that certain background sporadic isolates had closer-than-expected hqSNV distances to outbreak-associated isolates. In these instances, the sporadic isolates were excluded from the event solely on the basis of their isolation outside the defined time frame of the outbreak event. This finding lends support to previous reports that genomic data should be considered in conjunction with epidemiological evidence to determine the inclusion and exclusion of isolates from outreach events (10, 22, 23). This study did not reveal a correlation between the number of different bands observed by PFGE subtyping and the number of hqSNV differences between isolates with different PFGE patterns. Indeed, isolates with indistinguishable PFGE patterns can have high genetic distances, and isolates that would have been routinely considered related in fact have no genetic matches. A consistent correlation between the phage type and the number of hqSNVs was also observed in this study, despite the use of only the core genome, thus confirming the discriminatory power of core hqSNV analysis.

This study demonstrates the increased discriminatory power of a WGS-based method and hqSNV analysis relative to a conventional typing method to distinguish between outbreak-associated isolates. hqSNV analysis enabled source attribution of a foodborne event. Moreover, although PFGE subtyping pointed to a highly clonal genetic population, WGS permitted the unveiling of additional genetic diversity, thus facilitating the resolution of isolates with a matching PFGE pattern into phylogenetically defined and epidemiologically concordant groups. Since our analysis revealed that isolates recovered across geography and time can also be genetic matches, we strongly recommend that WGS-based typing results be used to inform *S.* Heidelberg outbreak event case definitions along with the consideration of epidemiological evidence. This WGS-based analysis will next be extended to cover increased PFGE diversity and to map the genetic population of *S.* Heidelberg isolates responsible for human disease across North America.

## REFERENCES

1. **Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A, Hoekstra RM, International Collaboration on Enteric Disease "Burden of Illness" Studies.** 2010. The global burden of nontyphoidal *Salmonella* gastroenteritis. Clin Infect Dis **50:**882–889. http://dx.doi.org/10.1086/650733.
2. **Thomas MK, Murray R, Flockhart L, Pintar K, Pollari F, Fazil A, Nesbitt A, Marshall B.** 2013. Estimates of the burden of foodborne illness in Canada for 30 specified pathogens and unspecified agents, circa 2006. Foodborne Pathog Dis **10:**639–648. http://dx.doi.org/10.1089/fpd.2012.1389.
3. **Hohmann EL.** 2001. Nontyphoidal salmonellosis. Clin Infect Dis **32:**263–269. http://dx.doi.org/10.1086/318457.

4. **Leekitcharoenphon P, Lukjancenko O, Friis C, Aarestrup FM, Ussery DW.** 2012. Genomic variation in *Salmonella enterica* core genes for epidemiological typing. BMC Genomics **13**:88. http://dx.doi.org/10.1186/1471-2164-13-88.

5. **The National Microbiology Laboratory (NML), Centre for Foodborne, Environmental and Zoonotic Infectious Diseases (CFEZID), Public Health Agency of Canada, Provincial Public Health Microbiology Laboratories.** 2014. National Enteric Surveillance Program (NESP): annual summary 2012. Public Health Agency of Canada, Ottawa, Ontario, Canada.

6. **Han J, David DE, Deck J, Lynne AM, Kaldhone P, Nayak R, Stefanova R, Foley SL.** 2011. Comparison of *Salmonella enterica* serovar Heidelberg isolates from human patients with those from animal and food sources. J Clin Microbiol **49**:1130–1133. http://dx.doi.org/10.1128/JCM.01931-10.

7. **Burt CR, Proudfoot JC, Roberts M, Horowitz RH.** 1990. Fatal myocarditis secondary to *Salmonella septicemia* in a young adult. J Emerg Med **8**:295–297. http://dx.doi.org/10.1016/0736-4679(90)90009-K.

8. **Vugia DJ, Samuel M, Farley MM, Marcus R, Shiferaw B, Shallow S, Smith K, Angulo FJ, Emerging Infections Program FoodNet Working Group.** 2004. Invasive *Salmonella* infections in the United States, FoodNet, 1996-1999: incidence, serotype distribution, and outcome. Clin Infect Dis **38**(Suppl 3):S149–S156.

9. **Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H.** 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PLoS One **6**:e22751. http://dx.doi.org/10.1371/journal.pone.0022751.

10. **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**:e87991. http://dx.doi.org/10.1371/journal.pone.0087991.

11. **Dunne WM, Jr, Westblade LF, Ford B.** 2012. Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. Eur J Clin Microbiol Infect Dis **31**:1719–1726. http://dx.doi.org/10.1007/s10096-012-1641-7.

12. **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.

13. **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 for *Salmonella enterica* serotype Enteritidis. J Clin Microbiol **53**:212–218. http://dx.doi.org/10.1128/JCM.02332-14.

14. **Hoffmann M, Luo Y, Lafon PC, Timme R, Allard MW, McDermott PF, Brown EW, Zhao S.** 2013. Genome sequences of *Salmonella enterica* serovar Heidelberg isolates isolated in the United States from a multistate outbreak of human *Salmonella* infections. Genome Announc **1**:pii=e00004-12. http://dx.doi.org/10.1128/genomeA.00004-12.

15. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19**:455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

16. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30**:2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.

17. **Garrison E, Marth G.** 2012. Haplotype-based variant detection from short-read sequencing. arXiv **arXiv:**1207.3907v2 [q-bio.GN]. http://arXiv.org/abs/1207.3907v2.

18. **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.

19. **Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carrico JA.** 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics **13**:87. http://dx.doi.org/10.1186/1471-2105-13-87.

20. **Health Canada, Public Health Agency of Canada, Canadian Food Inspection Agency.** January 2011. Weight of evidence: factors to consider for appropriate and timely action in a foodborne illness outbreak investigation. Health Canada, Ottawa, Ontario, Canada.

21. **Schurch 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.

22. **Ashton PM, Peters T, Ameh L, McAleer R, Petrie S, Nair S, Muscat I, de Pinna E, Dallman T.** 2015. Whole genome sequencing for the retrospective investigation of an outbreak of *Salmonella typhimurium* DT 8. PLoS Curr **7**:pii=ecurrents.outbreaks.2c05a47d292f376afc5a6fcdd8a7a3b6. http://dx.doi.org/10.1371/currents.outbreaks.2c05a47d292f376afc5a6fcdd8a7a3b6.

23. **Octavia S, Wang Q, Tanaka MM, Kaur S, Sintchenko V, Lan R.** 2015. Delineating community outbreaks of *Salmonella enterica* serovar Typhimurium by use of whole-genome sequencing: insights into genomic variability within an outbreak. J Clin Microbiol **53**:1063–1071. http://dx.doi.org/10.1128/JCM.03235-14.