

Complete Genome Sequence of *Salmonella enterica* subsp. *enterica* serovar Minnesota Strain

Hua Wang*, Maria Hoffmann, Anna Laasri, Andrew P. Jacobson, David Melka, Phillip E. Curry, Thomas S. Hammack, and Jie Zheng

Division of Microbiology, Office of Regulatory Science, Center for Food Safety and Nutrition, U.S. Food and Drug Administration, College Park, Maryland

*Corresponding author: E-mail: hua.wang@fda.hhs.gov.

Accepted: September 25, 2017

Data deposition: This project has been deposited at DDBJ/EMBL/GenBank under the accession CP017720.

Abstract

Mango has been implicated as food vehicle in several *Salmonella*-causing foodborne outbreaks. Here, *Salmonella enterica* subsp. *enterica* serovar Minnesota was isolated from fresh mango fruit imported from Mexico in 2014. The complete genome sequence of *S. Minnesota* CFSAN017963 was sequenced using single-molecule real-time DNA sequencing. Distinct prophage regions, *Salmonella* pathogenicity islands, and fimbrial gene clusters were observed in comparative genomic analysis on *S. Minnesota* CFSAN017963 with other phylogenetically closely related *Salmonella* serovars. Core genome multilocus sequencing typing analysis of all the *S. Minnesota* isolates in the Genbank and Enterobase also revealed a high genomic diversity among the genomes analyzed.

Key words: *Salmonella*, *Salmonella* Minnesota, complete genome sequence.

Introduction

Salmonella is the second most common etiology after norovirus causing foodborne illness outbreaks in United States (Could et al. 2013). Most *Salmonella* infections are associated with poultry products, however, a study of Center for Disease Control and Prevention (CDC) showed that 46% of *Salmonella* infections can be attributed various types of fresh produce (Painter et al. 2013). Therefore, fresh market produce, including mango, presents a microbial safety challenge to consumers. Mangoes are typically consumed fresh without a “kill-step” that would eliminate any harmful pathogens that may be present. In addition, fresh mangoes processed into fresh-cut or pureed products increase the risk of bacterial growth by breaking the natural exterior barrier. Mangoes have been reported to be associated with seven *Salmonella* outbreaks in United States during 1998–2015, which were caused by most common serovars of public health concern such as Newport, Saintpaul, and Branderup that were isolated from imported mangoes from Brazil, Peru, and Mexico (CDC 2016).

Salmonella Minnesota (CFSAN017963) sequenced in this study was isolated from the fresh mango fruit imported

from Mexico in 2014. This serovar has been isolated from cantaloupe imported from Mexico in 2016. *Salmonella* Minnesota ST3088 mostly occurring in poultry supply chain in Brazil was firstly whole genome sequenced multidrug-resistance CMY-2 producing strain with a novel sequence type (ST3088) resulting from the combination of the new allele *sucA433* with alleles *aroC13*, *dnaN11*, *hemD25*, *hisD197*, *purE12*, and *thrA4* (Moura et al. 2017). Based on phylogenetic analysis of seven concatenated house-keeping genes used in MLST, ST3088 was shown to be most closely related to ST548 and ST1678. In this study, we present the first complete genome sequence of *S. Minnesota* (CFSAN017963) strain and the clonal relationship of *Salmonella* Minnesota isolates of the same ST type available in the Genbank and Enterobase.

Materials and Methods

Salmonella enterica subsp. *enterica* serovar Minnesota Isolate

Salmonella enterica subsp. *enterica* serovar Minnesota (CFSAN017963) strain was isolated from imported fresh

mango during a surveillance assignment of U.S. Food and Drug Administration (FDA) in 2014.

DNA Extraction and Genome Sequencing

The isolate was cultured in Trypticase soy broth (Becton, Dickinson, Franklin Lakes, NJ) overnight at 37 °C. Genomic DNA was isolated from the overnight culture using the DNeasy blood and tissue kit (Qiagen, Inc., Valencia, CA). The DNA was sequenced using the Pacific Biosciences (PacBio) RS II sequencing platform, as previously reported (Hoffmann et al. 2015). Genomic DNA was sheared into ~10-kb fragments using g-TUBE (Covaris, Inc., Woburn, MA). The library was prepared based on the 10-kb PacBio sample preparation protocol and sequenced using P4/C2 chemistry on three single-molecule real-time (SMRT) cells with a 180-min collection time. The continuous long-read data were de novo assembled using the PacBio hierarchical genome assembly process (HGAP version 3.0) with default parameters (Chin et al. 2013).

Genomic Analysis of *S. Minnesota* CFSAN017963

Annotation of assembly was processed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) and subsequently deposited at DDBJ/EMBL/GenBank. An all-by-all BLAST comparison of *S. Minnesota* CFSAN017963 with other phylogenetically closely related serovars including *S. Montevideo*, *S. Panama*, *S. Javiana*, *S. Infantis*, *S. Cubana*, *S. Enteritidis*, and *S. Typhimurium* (accession numbers CP007530, CP012346, CP004027, LN649235, CP006055, AM933172, and AE006468, respectively), were undertaken using the BLAST Ring Image Generator (BRIG) (Alikhan et al. 2011). *Salmonella* Pathogenicity Islands (SPIs) and fimbria were identified and annotated using Artemis (Rutherford et al. 2000). Putative genomic islands containing prophage sequences and antibiotic resistance gene were identified using PHASTER, a new version of PHAST (Arndt et al. 2016) and ResFinder (Zankari et al. 2012), respectively.

MLST and cgMLST Analyses of CFSAN017963 and Other *S. Minnesota* Strains from Genbank and Enterobase

CFSAN017963 and other available *S. Minnesota* isolates which were archived in GenBank (<http://www.ncbi.nlm.nih.gov/genome/genomes/152>; last accessed October 3, 2017) and Enterobase (<https://enterobase.warwick.ac.uk/species/index/senterica>; last accessed October 3, 2017) (supplementary table S1, Supplementary Material online) all belong to the same sequence type (ST548), except for strains CFSAN024581 and 21100 which are ST285. Core genome MLST (cgMLST) of these available Minnesota genomes was performed in silico using Ridom SeqSphere+ v4.0.0 (Ridom GmbH, Germany). Briefly, as described in the software and previous study (Toro et al. 2016), a first defined cgMLST scheme was created using the

cgMLST target definer with default setting in Ridom SeqSphere software. The genome of *S. Minnesota* CFSAN017962 was used as the reference genome. All available draft genomes in the databases were used for comparison to the reference genome to establish a list of core and accessory genome genes. cgMLST performed a gene-by-gene analysis and identified SNPs within different alleles to establish genetic distance calculations. The core SNP matrix was used for maximum-likelihood (ML) phylogeny reconstruction based on the Kimura 2-parameter model in MEGA v6 software. The statistical support of the nodes in the ML tree was assessed by bootstrap resampling with 1,000 replicates.

Results and Discussion

Whole Genome Comparison

The closed *S. Minnesota* genome was sequenced with 135× coverage. A final consensus was reached with predicted accuracy at 100% for the complete genome. The size of *S. Minnesota* CFSAN017963 complete genome was 4,716,739 bp. Annotation revealed a total of 4,747 coding sequence (CDS) features in the genome. *Salmonella Minnesota* CFSAN017963 has a total of eight prophage regions encoded on the chromosome, five of which are intact (fig. 1). Prophage 1, 4, and 6 are incomplete 30.1-, 6.1-, and 13.5-kb phage remnants, respectively. All of the prophages displayed a mosaic nature, which also conferred a large genetic variation to the genomes among *Salmonella* serovars and strains.

The genome of *S. Minnesota* CFSAN017963 was compared with eight *Salmonella* genomes to determine genetic differences between Minnesota and the other phylogenetically closely related serovars (fig. 1). The whole genome comparison revealed that the *S. Minnesota* genome carries eight known *Salmonella* Pathogenicity Islands (SPIs-1 to 6, SPI-9, and SPI-16), which are conserved within other *Salmonella* genomes except SPI-6. The SPI-6 in *S. Minnesota* carries an intact SPI-6 island as in *S. Typhi* CT18, including a Type VI Secretion System (T6SS), two fimbrial gene clusters (*saf*ABCD and *tcf*ABCD) and the adhesin/invasin, PagN. Moreover, the *S. Minnesota* CFSAN017963 genome possesses a unique repertoire of fimbrial gene clusters. Twelve of them are chaperone-usher-dependent fimbrial operons, which were termed *bcf*, *sth*, *stj*, *std*, *stc*, *stb*, *fim*, *saf*, *tcf*, *sti*, *fae*, and *stk/sta* (fig. 1). Most *Salmonella enterica* subsp. I serovars share core fimbrial gene clusters (FGCs) such as *saf*, *bcf*, *fim*, *stb*, *sth*, and *std*, together with *sti*, *stf*, and *lpf*, which are prominently absent in serovar Typhi. The *S. Minnesota* CFSAN017963 carried additional *stj*, *stc*, *tcf*, *fae*, and *stk/sta* FGCs, even one FGC more than the serovars that had the highest numbers of FGCs among all the serovars reported in previous study (Yue et al. 2012). The accumulation of a large number of different FGCs may benefit survival, persistence, and transmission and broaden the host and environment range that can be colonized.

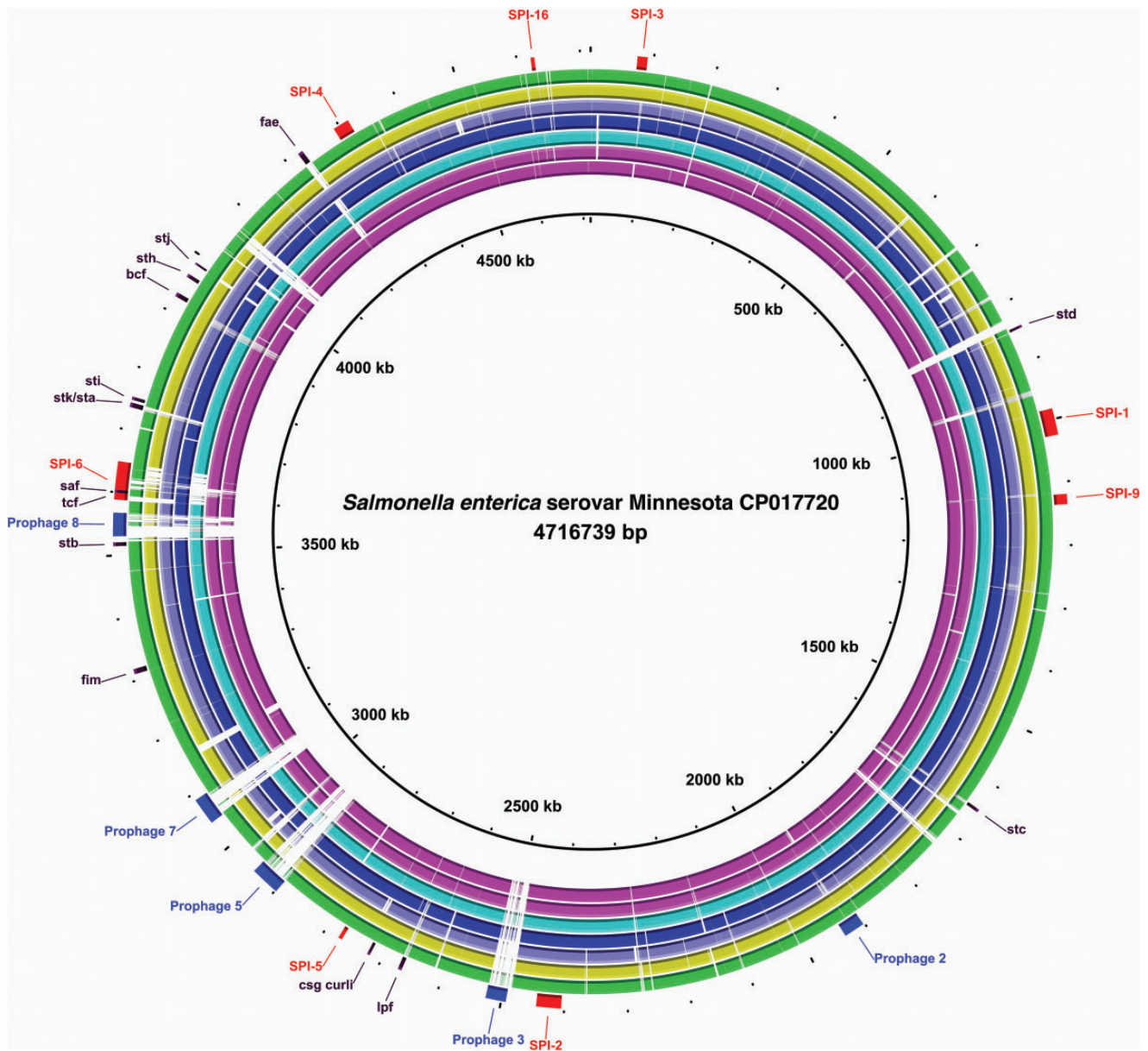


FIG. 1.—Genome comparisons of *Salmonella* Minnesota CFSAN017963 with other seven phylogenetically closely related *Salmonella* serovars. The innermost rings show *S. Minnesota* CFSAN017963 genome position (kb = kilobases). GC content (black) and GC skew (purple/green). The remaining rings (from ring 4 to ring 10) are the genomes from *S. Montevideo*, *S. Panama*, *S. Javiana*, *S. Infantis*, *S. Cubana*, *S. Enteritidis*, and *S. Typhimurium*, respectively, which showed BLASTn comparison of these *Salmonella* genomes against *S. Minnesota* CFSAN017963. BLASTn matches with an identity between 70% and 100% are colored in gradient, whereas nonmatching regions appeared as blank spaces in each ring. The outer ring contains annotations, colored according to function: regions variable in other *Salmonella* genomes such as the prophage regions (blue), *Salmonella* pathogenicity islands (SPIs) (red), and fimbrial gene clusters (purple).

Plasmids

Salmonella Minnesota CFSAN017963 carries two plasmids of 280,421 and 7,665 bp. The 280-kb plasmid is a RepFIIA virulence plasmid, containing an F-like plasmid transfer region (supplementary fig. S1, Supplementary Material online) (Carver et al. 2009). The 7-kb plasmid is a small cryptic plasmid.

MLST and cgMLST Analyses of CFSAN017963 and Other *S. Minnesota* Strains from Database

Using cgMLST with CFSAN017963 as the reference genome, 4,074 core genes were shared among all 52 Minnesota genomes. Using the scheme, a total of 1,680 genes were found to carry alleles. Single nucleotide polymorphisms (SNPs) identified were presented in supplement

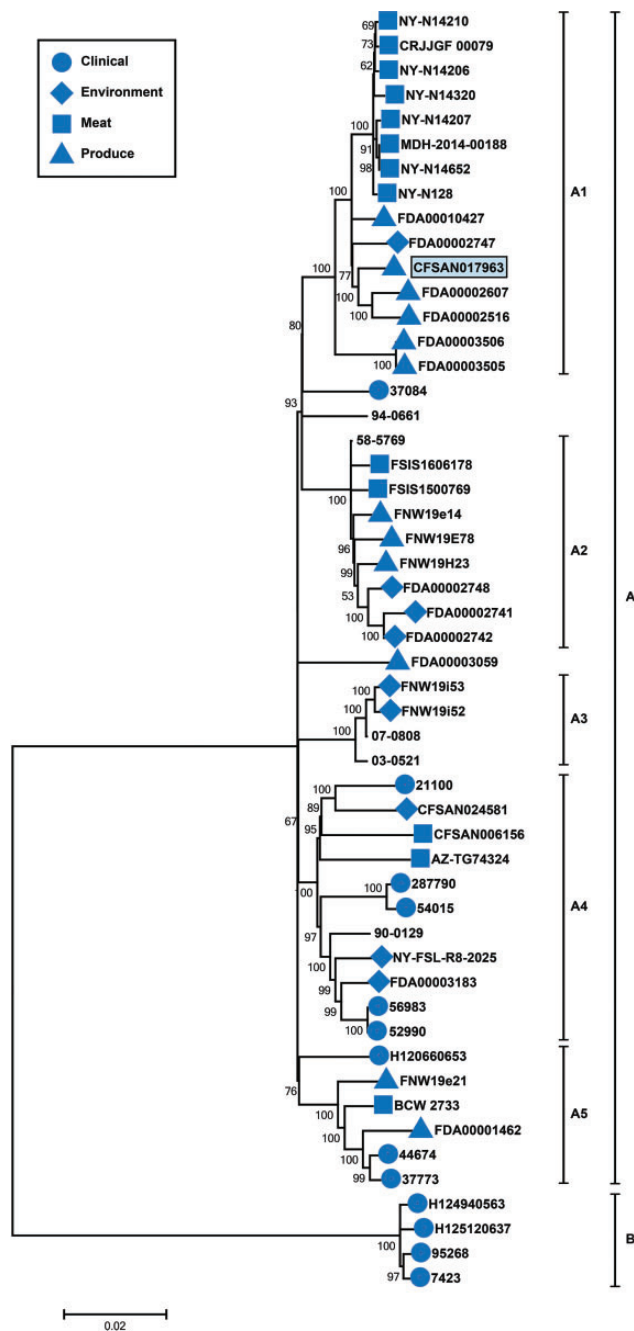


FIG. 2.—The maximum likelihood tree of 52 draft or complete *S. Minnesota* genomes including CFSAN017963 (highlighted) was constructed by MEGA v6 using SNP matrix generated from cgMLST analysis with Ridom SeqSphere+ v4.0.0.

tary table S1 in Supplementary Material online. A maximum likelihood (ML) tree using the SNPs from the 1,680 loci partitioned all 52 *S. Minnesota* strains into two main groups, with group A comprising 48 strains in five subgroups (fig. 2). All four strains included in group B are from clinical samples. CFSAN017963 was surrounded by

other four food/environment isolates from Mexico in subgroup 1. The remaining food and environmental isolates from Mexico resided in subgroup 2 except for CFSAN024581. All Asian isolates were clustered into subgroup 5 except FDA00003059 which is singlet. Meat isolates were also partitioned largely along host sources into different subgroups.

Conclusion

Salmonella Minnesota is of significant public health importance and has a very wide host range. Here, we report that *S. Minnesota* CFSAN017963 carries eight known *Salmonella* Pathogenicity Islands including a complete SPI-6 and encodes 14 different FGCs, which may explain the niche adaptation in broad host and environment range.

Acknowledgment

This work was supported by the US FDA, Center for Food Safety and Applied Nutrition, Office of Regulatory Science.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Literature Cited

Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402.

Arndt D, et al. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44(W1):W16–W21.

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25(1):119–120.

Centers for Disease Control and Prevention. 2016. Foodborne Outbreak Online Database (FOOD Tool). Available from: <https://www.cdc.gov/foodborneoutbreaks/>, last accessed October 3, 2017.

Chin CS, et al. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10(6):563–569.

Could LH, et al. 2013. Surveillance for foodborne disease outbreaks – United States, 1998–2008. *Morbidity and Mortality Weekly Report* 62:1–34.

Hoffmann M, et al. 2015. Complete genome sequence of *Salmonella enterica* subsp. *enterica* Serovar Agona 460004 2-1, associated with a multistate outbreak in the United States. *Genome Announc.* 3:e00690-15.

Moura Q, et al. 2017. Draft genome sequence of a multidrug-resistant CMY-2-producing *Salmonella enterica* subsp. *enterica* serovar Minnesota ST3088 isolated from chicken meat. *J Glob Antimicrob Resist.* 8:67–69.

Painter JA, et al. 2013. Attribution of foodborne illnesses, hospitalizations, and deaths to food commodities by using outbreak data, United States, 1998–2008. *Emerg Infect Dis.* 19(3):407–415.

Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16(10):944–945.

Toro M, et al. 2016. Whole-genome sequencing analysis of *Salmonella enterica* serovar Enteritidis isolates in Chile provides insights into possible transmission between gulls, poultry, and humans. *Appl Environ Microbiol.* 82(20):6223–6232.

Yue M, et al. 2012. Diversification of the *Salmonella* Fimbriae: A Model of Macro- and Microevolution. *PLoS ONE.* 7(6):e38596. <https://doi.org/10.1371/journal.pone.0038596>, last accessed October 3, 2017.

Zankari E, et al. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother.* 67(11):2640–2644.

Associate editor: Howard Ochman