



Closed Genome Sequences of Three *Salmonella enterica* Strains Belonging to Serovars Saintpaul, Weltevreden, and Thompson, Isolated from Mexico

 Narjol Gonzalez-Escalona,^a J. R. Aguirre-Sánchez,^b J. R. Ibarra-Rodríguez,^b C. Chaidez-Quiroz,^b Jaime Martinez-Urtaza^c

^aDivision of Microbiology, Office of Regulatory Science, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, Maryland, USA

^bCentro de Investigación en Alimentación y Desarrollo, A. C. Laboratorio Nacional para la Investigación en Inocuidad Alimentaria (LANIA), Sinaloa, Mexico

^cCentre for Environment, Fisheries and Aquaculture Science (CEFAS), Weymouth, United Kingdom

ABSTRACT Here, we report the genome sequences of three *Salmonella enterica* strains belonging to serovars Weltevreden (CFSAN047349), Saintpaul (CFSAN047351), and Thompson (CFSAN047352), isolated from river water in Sinaloa, Mexico. The genomes were closed by a combination of long-read and short-read sequencing. The strain sequence types (STs) are ST365, ST50, and ST26, respectively.

Salmonella enterica is one of the most important foodborne pathogens and is implicated in illnesses worldwide (1). There are almost 200,000 *Salmonella* genomes in the NCBI encompassing isolates from many countries. *Salmonella* spp. have been isolated from many sources, like the environment, animals, and food. Water can also be a reservoir of *Salmonella* spp. and has often been implicated as a source of contamination of food and fresh produce (2). Several *Salmonella* strains were isolated from rivers in Sinaloa, Mexico, during a study of *Salmonella* prevalence in river waters (2008 to 2010) (3). Here, we sequenced the complete genomes of three *Salmonella* strains (CFSAN047349, CFSAN047351, and CFSAN047352) isolated during that period using a combination of long-read and short-read sequencing technologies. The availability of these closed genomes will be useful for future outbreak investigations.

The strains were grown overnight in Luria-Bertani (LB) medium at 35°C, and the DNA was extracted with the DNeasy blood and tissue kit (Qiagen). The long reads for each strain were generated through MinION sequencing (Nanopore, Oxford, UK). The sequencing library was prepared using the rapid barcoding sequencing kit (SQK-RBK004). The sequencing library contained DNA fragmented randomly by a transposase present in the fragmentation mix of the SQK-RBK004 kit, rendering fragments of >30 kb. This library was run in a FLO-MIN106 (R9.4.1) flow cell, according to the manufacturer's instructions, for 48 h. The run was base called live using default settings in MinKNOW v18.12 and Guppy v1.8.7. The sequencing output was 1.6 Gb (199,000 reads, but only reads above 5 kb were used for the downstream analyses [151,258 reads]), for an estimated genome average coverage of 25 to 58×. The short-read whole-genome sequence for each strain was generated by MiSeq Illumina sequencing with the MiSeq V3 kit using 2 × 250-bp paired-end chemistry (Illumina, San Diego, CA), according to the manufacturer's instructions, at 80 to 660× coverage. The libraries were constructed using 100 ng of genomic DNA using the Nextera DNA Flex kit (Illumina), according to the manufacturer's instructions.

The final genome was achieved by using a pipeline already described (4). Briefly, the genome was obtained by *de novo* assembly, using Nanopore data and default settings within the Canu program v1.7 (5). A second assembly was generated using a SPAdes (6) hybrid assembly (with default settings) using both Nanopore and MiSeq data generated for each strain. The final corrected assembly was generated

Citation Gonzalez-Escalona N, Aguirre-Sánchez JR, Ibarra-Rodríguez JR, Chaidez-Quiroz C, Martinez-Urtaza J. 2019. Closed genome sequences of three *Salmonella enterica* strains belonging to serovars Saintpaul, Weltevreden, and Thompson, isolated from Mexico. Microbiol Resour Announc 8:e00656-19. <https://doi.org/10.1128/MRA.00656-19>.

Editor Christina A. Cuomo, Broad Institute
This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Narjol Gonzalez-Escalona, narjol.gonzalez-escalona@fda.hhs.gov.

Received 7 June 2019

Accepted 13 June 2019

Published 11 July 2019

TABLE 1 Metadata for the three environmental *S. enterica* strains reported in this study

CFSAN no.	GenBank accession no. (size [bp]) for:		GC content (%)	Hybrid assembly genome coverage (×)	BioSample accession no.	SRA accession no.	Serotype	ST
	Chromosome	Plasmid						
CFSAN047349	CP040701 (4,994,320)	CP040702 (104,768)	52	40	SAMN10261265	SRR9099568, SRR9099595	Weltevreden	365
CFSAN047351	CP040700 (4,767,334)		52	41	SAMN10261267	SRR9099567, SRR9099594	Saintpaul	50
CFSAN047352	CP040699 (4,709,901)		52	327	SAMN10261268	SRR9099569, SRR9099596	Thompson	26

by comparing the SPAdes hybrid and Canu assemblies using Mauve (7). The genomes and plasmid (if present) were confirmed as being circular closed by finding the contig end overlap and trimming the overlap. If the two assemblies agreed in synteny and size, the SPAdes hybrid assembly was used as the final assembly. Each *Salmonella* genome was of a distinct size (Table 1). The genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP; https://www.ncbi.nlm.nih.gov/genome/annotation_prok/) (8).

In silico multilocus sequence typing (MLST) analyses (<https://enterobase.warwick.ac.uk/species/index/senterica>) showed that each strain belonged to a different sequence type (ST), with CFSAN047349 belonging to ST365, CFSAN047351 belonging to ST50, and CFSAN047352 belonging to ST26. *In silico* serotyping using SeqSero (9) (<http://www.denglab.info/SeqSero>), a tool to infer serovar from the genes that determine antigenic structure, showed that the strains belonged to serovars Weltevreden, Saintpaul, and Thompson, respectively. The GC content was 52%, similar to that of other salmonellae. Only CFSAN047349 carried a plasmid of 104,768 bp.

Data availability. The accession numbers for the genome sequences are listed in Table 1.

ACKNOWLEDGMENTS

This study was supported by funding from the MCMi Challenge Grants Program proposal number 2018-646 and the FDA Foods Program Intramural Funds.

REFERENCES

- González-Escalona N, Hammack TS, Russell M, Jacobson AP, De Jesus AJ, Brown EW, Lampel KA. 2009. Detection of live *Salmonella* sp. cells in produce by a TaqMan-based quantitative reverse transcriptase real-time PCR targeting invA mRNA. *Appl Environ Microbiol* 75:3714–3720. <https://doi.org/10.1128/AEM.02686-08>.
- Walters SP, Gonzalez-Escalona N, Son I, Melka DC, Sassoubre LM, Boehm AB. 2013. *Salmonella enterica* diversity in central Californian coastal waterways. *Appl Environ Microbiol* 79:4199–4209. <https://doi.org/10.1128/AEM.00930-13>.
- Jiménez M, Martínez-Urtaza J, Rodríguez-Alvarez MX, Leon-Felix J, Cháidez C. 2014. Prevalence and genetic diversity of *Salmonella* spp. in a river in a tropical environment in Mexico. *J Water Health* 12:874–884. <https://doi.org/10.2166/wh.2014.051>.
- Gonzalez-Escalona N, Haendiges J, Miller JD, Sharma SK. 2018. Closed genome sequences of two *Clostridium botulinum* strains obtained by nanopore sequencing. *Microbiol Resour Announc* 7:e01075-18. <https://doi.org/10.1128/MRA.01075-18>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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. <https://doi.org/10.1089/cmb.2012.0021>.
- Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44: 6614–6624. <https://doi.org/10.1093/nar/gkw569>.
- 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. <https://doi.org/10.1128/JCM.00323-15>.