AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequences of Four *Salmonella enterica* Strains Associated with Pistachios Assembled Using a Combination of Short- and Long-Read Sequencing

Julie Haendiges,[a,b] Narjol Gonzalez-Escalona,[b] Jesse D. Miller,[a] Maria Hoffmann[b]

[a]Applied Research Center, NSF International, Ann Arbor, Michigan, USA

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

ABSTRACT Here, we report the genomes of two *Salmonella enterica* subsp. *enterica* serovar Montevideo strains (CFSAN005645 and FCC0123) and two *Salmonella enterica* subsp. *enterica* serovar Senftenberg strains (FSW0104 and CFSAN087304) isolated from pistachios. The genomes were closed using a hybrid assembly method using short- and long-read sequencing technology.

Salmonella enterica is a Gram-negative bacterium that is responsible for an average of 1.2 million illnesses (1). Foods with low water activity, such as pistachios and other varieties of nuts, are becoming a vehicle for increasing numbers of these infections (2).

These four strains were selected for sequencing because they showed relatedness to each other based on results from the NCBI pathogen detection site (https://www.ncbi.nlm.nih.gov/pathogens/) and they were available in the lab. The genomes of these four strains from pistachios were sequenced to create reference strains in the event of future outbreaks. The *Salmonella enterica* subsp. *enterica* serovar Senftenberg strains, according to the NCBI pathogen detection site, belong to NCBI single nucleotide polymorphism (SNP) cluster PDS000031739. The *Salmonella enterica* subsp. *enterica* serovar Montevideo strains belong to NCBI SNP cluster PDS000032600. These clusters contain numerous isolates from pistachio and environmental samples over a 2- to 5-year timespan. These are all draft genomes with many contigs; therefore, to generate better reference genome sequences for these strains, their genomes were closed using a hybrid assembly method utilizing both short- and long-read sequence technology.

The strains were obtained by the U.S. Food and Drug Administration as part of a federal public health investigation. The strains were cultured in Trypticase soy broth (Becton, Dickinson, Franklin Lakes, NJ) overnight at 37°C. The genomic DNA was isolated using the Maxwell RSC cultured cells DNA kit (Promega, Madison, WI) following the manufacturer's protocols, with the addition of RNase A treatment. The long-read sequencing was performed using a GridION device (Oxford Nanopore Technologies, Oxford, UK). The sequencing libraries were prepared using the Rapid Barcoding Sequencing kit (SQK-RBK004) following the manufacturer's protocols. The prepared libraries were sequenced on a FLO-MIN106 (R9.4.1) flowcell for 48 hours. The long reads were base called and demultiplexed using Guppy v3.2.2. All reads below 5,000 base pairs in length were discarded. The short-read sequence data for each strain were previously generated by FDA GenomeTrakr-participating laboratories using a MiSeq instrument with 2 × 250-bp paired-end chemistry (Illumina, San Diego, CA). The short-read sequencing data were downloaded from the NCBI Sequence Read Archive. Each sample was individually analyzed by a previously published method (3) as follows: the long-read sequencing data were initially *de novo* assembled using Canu v1.7 (4) with default settings. A second assembly was produced using the SPAdes assembler (Galaxy v3.11.1)

Citation Haendiges J, Gonzalez-Escalona N, Miller JD, Hoffmann M. 2019. Complete genome sequences of four *Salmonella enterica* strains associated with pistachios assembled using a combination of short- and long-read sequencing. Microbiol Resour Announc 8:e00975-19. https://doi.org/10.1128/MRA.00975-19.

Editor Kenneth M. Stedman, Portland State University

Address correspondence to Julie Haendiges, Julie.Haendiges@fda.hhs.gov.

Received 12 August 2019
Accepted 30 August 2019
Published 19 September 2019

TABLE 1 Metadata of the strains sequenced in this study

Isolate (BioSample accession no.)	No. of long reads (>4,999 bp)	Collection yr	NCBI accession no. (genome length [bp])	G+C content (%)	Hybrid assembly coverage (×)	SRA accession no.	
						Short read	Long read
CFSAN005645 (SAMN02265272)	167,338	2009	CP040380 (4,615,193)	52.3	103	SRR958039	SRR9052981
FCC0123 (SAMN02678532)	194,831	2009	CP040379 (4,619,529)	52.3	73	SRR5384655	SRR9052982
FSW0104 (SAMN02678831)	33,310	2013	CP037894 (4,820,913)	52.2	77	SRR1257282	SRR8707510
CFSAN087304 (SAMN10498876)	43,679	2018	CP037892 (4,819,673)	52.2	25	SRR8261878	SRR8707509

(5) with the hybrid option of using both short- and long-read sequencing outputs with default settings. Overlapping regions were identified at the end of the assemblies using Gepard and trimmed from the assembly (6). After comparison of the two assemblies for synteny using Mauve v20150226 (7), a final assembly was generated. A single circular, closed contig was generated for each isolate.

The four final assemblies were annotated using the NCBI Prokaryotic Genome Annotation Pipeline and deposited at DDBJ/EMBL/GenBank. The sequencing statistics can be found in Table 1. *In silico* multilocus sequence typing (MLST) analyses (8) showed two sequence types (STs); the two *S.* Montevideo strains belong to ST138, and the two *S.* Senftenberg strains belong to ST185. The *S.* Montevideo strains belong to clade I of the four clades previously described for *S.* Montevideo and to eBurstGroup 39 (eBG 39) (9). The other three clades belong to eBG 40, which contains the majority of the *S.* Montevideo strains. The *S.* Senftenberg strains belong to eBG 30, whereas the majority of *S.* Senftenberg strains are in eBG 55 (8).

Data availability. The genome sequences of the four *S. enterica* strains are listed in Table 1.

ACKNOWLEDGMENTS

This project was supported by funding from the MCMi Challenge Grants Program proposal number 2018-646 and the FDA Foods Program Intramural Funds and the NSF International Applied Research Center.

REFERENCES

1. 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. https://doi.org/10.3201/eid1701.p11101.
2. Podolak R, Enache E, Stone W, Black DG, Elliott PH. 2010. Sources and risk factors for contamination, survival, persistence, and heat resistance of Salmonella in low-moisture foods. J Food Prot 73:1919–1936. https://doi.org/10.4315/0362-028x-73.10.1919.
3. Gonzalez-Escalona N, Aguirre-Sanchez JR, Ibarra-Rodriguez 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.
4. 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.
5. 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. https://doi.org/10.1089/cmb.2012.0021.
6. Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23:1026–1028. https://doi.org/10.1093/bioinformatics/btm039.
7. Darling AC, 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.
8. Alikhan NF, Zhou Z, Sergeant MJ, Achtman M. 2018. A genomic overview of the population structure of Salmonella. PLoS Genet 14:e1007261. https://doi.org/10.1371/journal.pgen.1007261.
9. Nguyen SV, Harhay DM, Bono JL, Smith TPL, Fields PI, Dinsmore BA, Santovenia M, Wang R, Bosilevac JM, Harhay GP. 2018. Comparative genomics of Salmonella enterica serovar Montevideo reveals lineage-specific gene differences that may influence ecological niche association. Microb Genom 4:e000202. https://doi.org/10.1099/mgen.0.000202.