Complete Genome Sequences of Two *Salmonella enterica* subsp. *enterica* Serovar Anatum Strains Isolated from Papaya

Daniela Miller,a Kuan Yao,a Brooke Whitney,b Maria Sanchez Leon,a Sandra Tallent,a Maria Hoffmanna

a Center for Food Safety and Applied Nutrition, Division of Microbiology, Office of Regulatory Science, U.S. Food and Drug Administration, College Park, Maryland, USA
b Center for Food Safety and Applied Nutrition, Coordinated Outbreak Response and Evaluation, Office of Foods and Veterinary Medicine, U.S. Food and Drug Administration, College Park, Maryland, USA

**ABSTRACT**  *Salmonella enterica* is a major global foodborne pathogen that causes gastroenteritis and, in some cases, death. *Salmonella* serovar Anatum has been increasingly associated with foodborne salmonellosis outbreaks. In this report, we announce two complete genome sequences of *Salmonella* Anatum isolated from papaya fruit.

*Salmonella* outbreaks linked to consumption of fresh papayas have caused over 350 reported cases of salmonellosis and 2 deaths in the United States since the first incidence in 2011 ([www.cdc.gov/salmonella/](http://www.cdc.gov/salmonella/)). *Salmonella enterica* strains isolated from papaya are prevalent and highly diverse (1). A 2011 investigation by the U.S. Food and Drug Administration (FDA) isolated *Salmonella* spp. from 16% of tested papaya import samples from Mexico (n = 388), comprising isolates representing over 22 different serotypes (2). Many papaya-associated *Salmonella* isolates represent less-studied non-typhoidal serovars, highlighting a need for high-quality reference genomes for these strains. In this announcement, we report the complete genome sequences of two *Salmonella enterica* serovar Anatum papaya-associated isolates (CFSAN003959 and CFSAN003961).

CFSAN003959 was isolated from fresh papaya fruit sampled from a grower in Mexico in 2012. CFSAN003961 was isolated from fresh papaya fruit collected during import from Mexico to the United States in 2011. Collection of this sample occurred during increased sampling as a result of a 2011 *Salmonella enterica* serovar Agona outbreak. The *Salmonella* isolates were originally isolated and characterized as *Salmonella* as described in the Bacteriological Analytical Manual (BAM), chapter 5, “*Salmonella*” ([https://www.fda.gov/food/laboratory-methods-food/bacteriological-analytical-manual-bam-chapter-5-salmonella](https://www.fda.gov/food/laboratory-methods-food/bacteriological-analytical-manual-bam-chapter-5-salmonella)).

Each isolate was cultured in Trypticase soy broth (Becton, Dickinson, Franklin Lakes, NJ, USA) overnight at 37°C. Genomic DNA was isolated from overnight cultures using the Promega Wizard Genomic DNA purification kit and Maxwell RSC Instrument (Promega Corporation, Madison, WI). Genomic DNA was sheared to approximately 20-kb fragments using a g-TUBE (Covaris, Inc., Woburn, MA). Sequencing libraries were prepared following the 20-kb PacBio sample preparation protocol using the BluePippin size selection system (Sage Science, Beverly, MA, USA). Each isolate was sequenced based on previously reported procedures on the PacBio RS II platform (Pacific Biosciences, Menlo Park, CA, USA) using two single-molecule real-time (SMRT) cells per isolate sequenced. Analysis of the sequence reads was implemented using SMRT Analysis 2.3.0. ([https://smrt-analysis.readthedocs.io/en/latest/SMRT-Pipe-Reference-Guide-v2.3.0/](https://smrt-analysis.readthedocs.io/en/latest/SMRT-Pipe-Reference-Guide-v2.3.0/)). The continuous long-read data from each isolate were assembled de novo into one contig using the PacBio Hierarchical Genome Assembly Process 3 (HGAP3) program with default parameters. Gepard was used for identifying overlapping regions, and the closed
genomes were rotated to start at the dnaA gene (3). Genomes were checked manually for even sequencing coverage. Then, the improved consensus sequence was uploaded in SMRT Analysis 2.3.0 to determine the final consensus and accuracy scores using the Quiver consensus algorithm. The sequencing statistics are listed in Table 1. The assembled sequences were annotated using the NCBI Prokaryotic Genome Annotation Pipeline and subsequently deposited at DDBJ/EMBL/GenBank (4).

The chromosome of Salmonella Anatum CFSAN003959 consists of 4,718,314 bp with a GC content of 52.6%. Three "intact" prophage regions were identified using PHASTER, including Salmonella phages Fels-1 and ST64T and prophage Gifsy-1 (5). One antimicrobial resistance gene, aac(6')-Iaa, was identified using ResFinder (4). aac(6')-Iaa is a chromosomally encoded aminoglycoside acetyltransferase that acetylates tobramycin, kanamycin, and amikacin (6).

The chromosome of Salmonella Anatum CFSAN003961 consists of 4,752,859 bp with a GC content of 52.6%. Four "intact" prophage regions were identified using PHASTER, including Salmonella phage Fels-1 and prophage Gifsy-1 (5). ResFinder analysis identified one antimicrobial resistance gene, aac(6')-Iaa (7).

Data availability. The complete genome sequences of Salmonella Anatum isolates CFSAN003959 and CFSAN003961 are publicly available in GenBank, and their details can be found in Table 1.

ACKNOWLEDGMENT

This project was supported by the U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition, Office of Regulatory Science.

REFERENCES

1. Thompson CP, Doak AN, Amirani N, Schroeder EA, Wright J, Kariyawasam S, Lamendella R, Shariati NW. 2018. High-resolution identification of multiple Salmonella serovars in a single sample by using CRISPR-SeroSeq Appl Environ Microbiol 84:e01859-18. https://doi.org/10.1128/AEM.01859-18.
2. Mba-Jonas A, Culpepper W, Hill T, Cantu V, Loera J, Borders J, Saathoff-Huber L, Nsubuga J, Zambrana I, Dalton S, Williams I, Neil KP. 2018. A multistate outbreak of human Salmonella agona infections associated with consumption of fresh, whole papayas imported from Mexico—United States, 2011. Clin Infect Dis 66:1756–1761. https://doi.org/10.1093/cid/cix1094.
3. Krumsieck J, Arnold R, Rattie 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.
4. 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.
5. Arndt D, Grant J, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.
6. Salipante SJ, Hall BG. 2003. Determining the limits of the evolutionary potential of an antibiotic resistance gene. Mol Biol Evol 20:653–659. https://doi.org/10.1093/molbev/msg074.
7. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261.

| Nucleotide accession no. | Strain         | No. of SMRT cells | SRA accession no. | Coverage (x) | Mean read length (bp) | No. of reads | N50 read length (bp) |
|--------------------------|----------------|-------------------|-------------------|--------------|------------------------|--------------|-----------------------|
| CP041184                 | CFSAN003959    | 2                 | SRS622107         | 290          | 13,871                 | 159,914      | 23,472                |
| CP041183                 | CFSAN003961    | 2                 | SRS478508         | 260          | 12,108                 | 137,193      | 21,206                |

TABLE 1 Sequence accession numbers and sequencing statistics