Draft Genome Sequences of Two Novel *Salmonella enterica* subsp. *enterica* Strains Isolated from Low-Moisture Foods with Applications in Food Safety Research

Devon R. Radford,a Carlos G. Leon-Velarde,b Shu Chen,b Amir M. Hamidi Oskouei,c Sampathkumar Balamurugana

aGuelph Research and Development Centre, Agriculture and Agri-Food Canada, Guelph, Ontario, Canada
bAgriculture and Food Laboratory, Laboratory Services Division, University of Guelph, Guelph, Ontario, Canada
cAgri-Neo Inc., Etobicoke, Ontario, Canada

ABSTRACT The genomes of two strains of *Salmonella enterica* subsp. *enterica* serovar Cubana and serovar Muenchen, isolated from dry hazelnuts and chia seeds, respectively, were sequenced using the Illumina MiSeq platform, assembled *de novo* using the overlap-layout-consensus method, and aligned to their respective most identical sequence genome scaffolds using MUMMER and BLAST searches.

*S*almonella enterica* is a leading cause of bacterial foodborne illnesses worldwide (1). Multiple outbreaks of *Salmonella* infection have also been linked to low-water-activity (a_w) foods, such as nuts and seeds (2). However, the mechanisms utilized by this bacterium to ensure their survival in dry conditions have not been fully elucidated (3). We present here the draft genome sequences of two novel strains of *S. enterica* subsp. *enterica*: strain 17-030379-0001, isolated from dry hazelnuts ($a_w = 0.483 \pm 0.003$) and serotyped as serovar Muenchen with antigenic properties 23:z29, and strain 17-030379-0002, isolated from chia seeds ($a_w = 0.601 \pm 0.003$) and serotyped as serovar Cubana with antigenic properties 6,8:d:1,2.

Genomic DNA was extracted from an overnight culture using the DNeasy blood and tissue kit (Qiagen, Mississauga, ON, Canada), following the manufacturer's protocol. Sequencing libraries were prepared from the DNA using the Nextera XT DNA library preparation kit (Illumina, San Diego, CA, USA), following the procedures described in the kit's reference guide. Sequencing was conducted using a MiSeq sequencer with a MiSeq V2 reagent kit (Illumina) and 2×250 paired-end cycles, according to the manufacturer's protocol. Raw sequence reads were filtered using the MiSeq sequencer system software (Illumina) to remove low-quality sequences and trimmed to remove adaptor sequences. Sequences passing a quality score of 30 were assembled *de novo* following an overlap-layout-consensus method using GeneStudio. In the case of the serovar Muenchen genome, this yielded 33 contigs for a total of 4,795,679 bp with a GC content of 51.29% at $1\times$ coverage. The assembled serovar Cubana genome yielded 71 contigs for a total of 5,041,200 bp with a GC content of 51.3% at $1\times$ coverage. The scaffolds of these genomes were mapped by alignment to *S. enterica* subsp. *enterica* serovars Yovokome (GenBank accession number CP019418) and Cubana (GenBank accession number CP006055) with 99.4 and 100% scaffold coverages, respectively. Assembly quality was evaluated using QUAST (4), which identified 3.37 mismatches/100 kb for the serovar Muenchen genome and 6.70 mismatches/100 kb for the serovar Cubana genome. MUMMER plots (5) were examined for contig internal inversions as markers of misassembly, and gaps were used to measure the extent of the genome coverage. Moderate gaps relative to the reference genomes were observed, but no evidence of misassembly was observed. Gene calling and annotation were carried out using the Rapid Annotations using Subsystems Technology (RAST) server (6). A total

of 5,004 features were predicted in the serovar Cubana genome, falling into 578 subsystem categories, with 4,904 predicted coding sequences and 100 RNAs. RAST predicted 4,737 features in the serovar Muenchen genome, spanning 575 subsystem categories, with 4,649 predicted coding sequences and 88 RNAs.

These assemblies will allow researchers to further explore the genomic diversity of *Salmonella* serovars adapted to low-moisture environments, impacting basic and applied food safety research.

Accession number(s). The genome sequences reported here have been deposited at DDBJ/EMBL/GenBank under the accession numbers PJAL00000000, for *Salmonella* serovar Muenchen strain 17-030379-0001, and PJAK00000000, for *Salmonella* serovar Cubana strain 17-030379-0002.

REFERENCES

1. Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A, Hoekstra RM. 2010. The global burden of nontyphoidal *Salmonella* gastroenteritis. Clin Infect Dis 50:882–889. https://doi.org/10.1086/650733.
2. Fong K, Wang S. 2016. Strain-specific survival of *Salmonella enterica* in peanut oil, peanut shell, and Chia seeds. J Food Prot 79:361–368. https://doi.org/10.4315/0362-028X.JFP-15-419.
3. Finn S, Condell O, McClure P, Amézquita A, Fanning S. 2013. Mechanisms of survival, responses and sources of *Salmonella* in low-moisture environments. Front Microbiol 4:331. https://doi.org/10.3389/fmicb.2013.00331.
4. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
5. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.
6. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.