AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Whole-Genome Sequence and Annotation of *Salmonella enterica* subsp. *enterica* Serovar Enteritidis Phage Type 8 Strain EN1660

Benjamin J. Perry,[a] Stephen F. Fitzgerald,[a]* Carsten Kröger,[b] Andrew D. S. Cameron[a]

Department of Biology, University of Regina, Regina, Saskatchewan, Canada[a]; Department of Microbiology, School of Genetics and Microbiology, Moyne Institute of Preventive Medicine, Trinity College Dublin, Dublin, Ireland[b]

ABSTRACT The genome of *Salmonella enterica* subspecies *enterica* serovar Enteritidis phage type 8 strain EN1660, isolated from an outbreak in Thunder Bay, Canada, was sequenced to 46-fold coverage using an Illumina MiSeq with 300-bp paired-end sequencing chemistry to produce 28 contigs with an N_{50} value of 490,721 bp.

Received 12 November 2016 **Accepted** 22 November 2016 **Published** 26 January 2017

Citation Perry BJ, Fitzgerald SF, Kröger C, Cameron ADS. 2017. Whole-genome sequence and annotation of *Salmonella enterica* subsp. *enterica* serovar Enteritidis phage type 8 strain EN1660. Genome Announc 5:e01517-16. https://doi.org/10.1128/genomeA.01517-16.

Address correspondence to Andrew D. S. Cameron, andrew.cameron@uregina.ca.

* Present address: Stephen F. Fitzgerald, Royal (Dick) School of Veterinary Studies, The Roslin Institute, University of Edinburgh, Scotland.

The bacterial pathogen *Salmonella enterica* presents a significant health burden globally and remains a persistent challenge in North America, particularly for the agriculture and food production industries that work diligently to prevent *Salmonella* contamination. Nontyphoidal *Salmonella* is estimated to cost the U.S. economy $3.7 billion per year due to hospitalizations and premature deaths (1). *Salmonella enterica* subsp. *enterica* serovar Enteritidis is the most common clinical isolate in the United States, yet Salmonella Enteritidis genomes are under-represented in DNA sequence databases. Here, we report the whole-genome sequencing of Salmonella Enteritidis strain EN1660, a clinical isolate from a 1992 community outbreak in Thunder Bay, Ontario, Canada, associated with consumption of contaminated poultry products. EN1660 was classified as phage type 8 (PT8) by the Ontario Laboratory Centre for Disease Control and was determined to be motile, capable of forming biofilm, and highly invasive in Caco-2 cell culture (2). EN1660 is classified as *Salmonella* sequence type 11 (ST11) by MLST version 1.8 (3).

Genomic DNA isolation was performed using a BioBasic Canada DNA isolation kit. DNA sequencing libraries were prepared using the NEBNext Ultra DNA library prep kit, with a final library size selection of 700 bp. Template sequencing was performed using an Illumina MiSeq and V3 300-bp paired-end sequencing chemistry. Demultiplexed sequencing data were filtered for phiX sequences using Bowtie2 (4) and the *Enterobacteria* phage phiX174 *sensu lato* reference sequence (RefSeq: NC_001422.1) for comparison. Filtered sequencing reads were then quality trimmed using Trimmomatic (5), removing GC% skews from the beginnings and ends of the reads and trimming at an average *Q* score below 20 in a sliding window of 5 bp. This resulted in 985,195 high-quality paired-end reads for use in genome assembly.

The genome was assembled using "careful" mode with a *k*-mer size of 127 in SPAdes version 1.35 (6). The resulting assembly had a consensus length of 4,701,324 bp spanning 28 contigs, with an N_{50} value of 490,721 bp and an L_{75} value of five contigs. The average genome coverage was 46× with a GC % of 52.1%. Annotation of the draft genome sequence was conducted using the NCBI Prokaryotic Genome Annotation Pipeline (7). Genome annotation predicted 4,478 coding sequences, eight rRNA operons, 78 tRNAs, 17 ncRNAs, and two CRISPR arrays.

The EN1660 draft genome was screened for antibiotic resistance genes using the Antibiotic Resistance Gene Database (8), which predicted the presence of four resistance gene homologs: *mdtM*, *pbp*2, *acrB*, and *mdtL*. The draft genome sequence was also analyzed for prophage content using the PHAge Search Tool server (9). PHAST analysis revealed six prophage regions, of which two are intact (similar to Gifsy-2 and RE-2010); one more is potentially fully intact (similar to *Pseudomonas* phage B3); and three are incomplete. We used BLAST to confirm the presence of the broadly distributed *Salmonella* pathogenicity islands SPI-1, SPI-2, SPI-3, SPI-4, SPI-5, SPI-6, SPI-9, SPI-11, SPI-12, SPI-13, SPI-14, and SPI-16 in strain EN1660.

Accession number(s). The draft whole-genome sequence for *Salmonella enterica* subsp. *enterica* serovar Enteritidis EN1660 was deposited into the NCBI GenBank database under the accession number LUUA00000000.

ACKNOWLEDGMENTS

We are grateful to Ken Sanderson and Orsi Cseh at the Salmonella Genome Stock Centre at the University of Calgary for providing strain EN1660 and background information. Funding for DNA sequencing and analysis was provided to Andrew D. S. Cameron by Natural Sciences and Engineering Research Council of Canada Discovery grant 435784-2013 and Saskatchewan Health Research Foundation Establishment grant 2867.

REFERENCES

1. Economic Research Service (ERS), U.S. Department of Agriculture (USDA). 7 October 2014. Cost estimates of foodborne illnesses. http://ers.usda.gov/data-products/cost-estimates-of-foodborne-illnesses.aspx.
2. Shah DH, Zhou X, Addwebi T, Davis MA, Call DR. 2011. *In vitro* and *in vivo* pathogenicity of *Salmonella enteritidis* clinical strains isolated from North America. Arch Microbiol 193:811–821. https://doi.org/10.1007/s00203-011-0719-4.
3. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total genome sequenced bacteria. J Clin Microbiol 50:1355–1361. https://doi.org/10.1128/JCM.06094-11.
4. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
6. 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.
7. 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.
8. Liu B, Pop M. 2009. ARDB—Antibiotic Resistance Genes Database. Nucleic Acids Res 37:D443–D447. https://doi.org/10.1093/nar/gkn656.
9. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39:W347–W352. https://doi.org/10.1093/nar/gkr485.