



Genome Sequences of Uncommon Shiga Toxin-Producing *Escherichia coli* Serotypes

Baha Abdalhamid,^{a,b} Emily L. Mccutchen,^b Peter C. Iwen,^{a,b} Joao Carlos Gomes-Neto,^c Andrew K. Benson,^c Steven H. Hinrichs^a

^aDepartment of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska, USA

^bNebraska Public Health Laboratory, University of Nebraska Medical Center, Omaha, Nebraska, USA

^cDepartment of Food Science and Technology, University of Nebraska, Lincoln, Nebraska, USA

ABSTRACT Shiga toxin-producing *Escherichia coli* (STEC) is a foodborne disease with worldwide outbreaks. STEC serotypes O157, O26, O45, O103, O111, O121, and O145 cause the most outbreaks. There is little published information regarding the other serotypes. We report the draft genome sequences for 11 uncommon STEC serotypes from Nebraska.

Shiga toxin-producing *Escherichia coli* (STEC) can cause hemorrhagic colitis and a life-threatening hemolytic uremic syndrome (1, 2). STEC causes outbreaks that are a worldwide public health concern (2–4). The 6 non-O157 STEC serotypes, O26, O45, O103, O111, O121, and O145, in addition to the STEC O157 serotype are responsible for an overwhelming majority of outbreaks (5, 6). Information regarding the other STEC serotypes is scarce (4). We present in this study draft genome sequences for 11 different uncommon STEC serotypes collected from Nebraska.

From November 2017 to February 2019, 120 STEC isolates were collected from different laboratories in Nebraska following routine stool culture for original isolation. Stool samples were subcultured on blood, MacConkey agar, Hektoen enteric agar, and MacConkey agar with sorbitol agar medium and were incubated overnight at 37°C. Isolates were then transported to the Nebraska Public Health Laboratory (NPHL) for further characterization. At NPHL, isolates were grown overnight on blood agar plates at 37°C. A MagNA Pure compact nucleic acid isolation kit I (Roche Diagnostics, IN, USA) was used to extract genomic DNA from overnight-grown isolates using a MagNA Pure compact instrument (Roche Diagnostics) following the manufacturer's instructions. A NanoDrop 2000 UV-visible (UV-Vis) spectrophotometer (Thermo Fisher, MA, USA) and a Qubit 3.0 fluorometer (Invitrogen, CA, USA) were used to determine DNA measures qualitatively and quantitatively, respectively.

Bacterial genomic DNA libraries were constructed using the Nextera XT library prep kit (Illumina, CA, USA) as recommended by the manufacturers. Whole-genome sequencing (WGS) was performed using the Illumina MiSeq platform (Illumina, CA, USA) to generate 300-bp paired-end reads. A rate of clusters passing the filter of >80%, a Phred quality score (QS30) of >75%, and a cluster density of 600 to 1,300 were used as parameters to assess the quality of the run. FastQC 0.10, Trimmomatic 0.33, SPAdes 3.12, BBDMap 38.06, and QUAST 4.1 (7–10) were used for assessing the quality of sequence reads, trimming, *de novo* assembly, purging of contigs less than 200 bp, and determining the quality of the *de novo* assembled genomes, respectively (Table 1). These bioinformatics tools were used on GitHub. Default parameters were used for all software unless otherwise specified. The NCBI Prokaryotic Genome Annotation Pipeline (PGAP) 4.8 (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/) was used to annotate the draft genome of each strain (11).

KmerFinder 3.1 and SeroTypeFinder 2.0 were used to determine organism identification and serotypes, respectively, for the STEC isolates (12, 13). In addition, multilocus sequencing types (MLSTs) were determined using MLST 2.0 (Table 1) (14).

Citation Abdalhamid B, Mccutchen EL, Iwen PC, Gomes-Neto JC, Benson AK, Hinrichs SH. 2020. Genome sequences of uncommon Shiga toxin-producing *Escherichia coli* serotypes. *Microbiol Resour Announc* 9:e01496-19. <https://doi.org/10.1128/MRA.01496-19>.

Editor David Rasko, University of Maryland School of Medicine

Copyright © 2020 Abdalhamid et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Baha Abdalhamid, babdahamid@unmc.edu.

Received 4 December 2019

Accepted 12 February 2020

Published 5 March 2020

TABLE 1 Summary characteristics of whole-genome sequencing of uncommon serotypes of Shiga toxin-producing *E. coli*

| Strain code ^a | Serotype ^b | ST ^c | Genome size (bp) | No. of contigs | % GC content | N_{50} (bp) | No. of reads | Genome coverage (×) | SRA accession no. | GenBank accession no. |
|--------------------------|-----------------------|-----------------|------------------|----------------|--------------|---------------|--------------|---------------------|-------------------|-----------------------|
| STEC_170836 | O185:H28 | ST-517 | 5,162,389 | 112 | 50.64 | 15,2751 | 1,128,853 | 56.44 | SRX6741578 | VZEL00000000 |
| STEC_180018 | O69:H11 | ST-14 | 5,399,941 | 260 | 50.38 | 131,480 | 1,148,343 | 57.42 | SRX6741577 | VZEP00000000 |
| STEC_180309 | O118/O151:H16 | ST-73 | 5,478,013 | 293 | 50.42 | 92,792 | 968,438 | 48.42 | SRX6741575 | VZFK00000000 |
| STEC_180427 | O118/O151:H16 | ST-335 | 5,416,226 | 310 | 50.41 | 90,305 | 1,067,885 | 53.39 | SRX6741582 | VZEC00000000 |
| STEC_180452 | O55:H7 | ST-131 | 5,359,179 | 151 | 50.42 | 311,454 | 2,824,168 | 141.21 | SRX6741581 | VZDR00000000 |
| STEC_180473 | O123/O186:H2 | ST-658 | 5,493,863 | 277 | 50.42 | 137,641 | 1,534,681 | 76.92 | SRX6741580 | VZFY00000000 |
| STEC_180505 | O118/O151:H16 | ST-21 | 5,419,695 | 314 | 50.43 | 89,301 | 1,072,233 | 53.61 | SRX6741579 | VZGA00000000 |
| STEC_180563 | O151:H16 | ST-17 | 5,567,883 | 310 | 50.37 | 82,954 | 1,129,349 | 56.47 | SRX6741584 | VZGH00000000 |
| STEC_180607 | O123:H2 | ST-21 | 5,330,230 | 271 | 50.58 | 136,012 | 1,408,908 | 70.45 | SRX6741574 | VZGR00000000 |
| STEC_180727 | O88:H25 | ST-21 | 5,016,338 | 136 | 50.64 | 168,851 | 1,929,340 | 96.47 | SRX6741572 | VZDY00000000 |
| STEC_180784 | O71:H49 | ST-21 | 5,477,716 | 348 | 50.44 | 104,253 | 1,349,778 | 67.49 | SRX6741571 | VZEA00000000 |

^a All strains carried *stx*₁. STEC_170836 carried both *stx*₁ and *stx*₂.

^b Serotypes are predicted based on the whole-genome sequencing.

^c ST, sequence type determined by MLST.

The average genome size of the sequenced isolates was 5,374,679 bp, with a range of 5,016,338 to 5,567,883 bp. The average GC content was 50.5%, with a range of 50.37% to 50.64%. The average number of contigs was 253, with a low of 112 contigs and a high of 348 contigs. The highest N_{50} value was 311,454 bp, while the lowest N_{50} value was 82,954 bp, with the average being 136,163 bp.

The data presented in this study will provide the genetic basis for a more detailed analysis of virulence factors, antimicrobial resistance determinants, and phylogenetic relatedness among these strains.

Data availability. PRJNA527109 is the BioProject number for these genomic sequences at NCBI. The numbers of contigs, N_{50} values, serotypes, MLSTs, total number of reads generated for each isolate, genome coverage, GenBank accession numbers, and Sequence Read Archive (SRA) accession numbers are provided in Table 1.

ACKNOWLEDGMENT

This work was supported by a grant from the CDC (CDC-RFA-CK19-1904).

REFERENCES

- Boerlin P, McEwen SA, Boerlin-Petzold F, Wilson JB, Johnson RP, Gyles CL. 1999. Associations between virulence factors of Shiga toxin-producing *Escherichia coli* and disease in humans. *J Clin Microbiol* 37:497–503. <https://doi.org/10.1128/JCM.37.3.497-503.1999>.
- Nuesch-Inderbilen M, Morach M, Cernela N, Althaus D, Jost M, Mausezahl M, Bloomberg G, Stephan R. 2018. Serotypes and virulence profiles of Shiga toxin-producing *Escherichia coli* strains isolated during 2017 from human infections in Switzerland. *Int J Med Microbiol* 308:933–939. <https://doi.org/10.1016/j.ijmm.2018.06.011>.
- 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>.
- Lang C, Hiller M, Konrad R, Fruth A, Flieger A. 2019. Whole-genome-based public health surveillance of less common Shiga toxin-producing *Escherichia coli* serovars and untypeable strains identifies four novel O genotypes. *J Clin Microbiol* 57:e00768-19. <https://doi.org/10.1128/JCM.00768-19>.
- Mellor GE, Fegan N, Duffy LL, Mc MK, Jordan D, Barlow RS. 2016. National survey of Shiga toxin-producing *Escherichia coli* serotypes O26, O45, O103, O111, O121, O145, and O157 in Australian beef cattle feces. *J Food Prot* 79:1868–1874. <https://doi.org/10.4315/0362-028X.JFP-15-507>.
- Bosilevac JM, Koohmariaie M. 2011. Prevalence and characterization of non-O157 Shiga toxin-producing *Escherichia coli* isolates from commercial ground beef in the United States. *Appl Environ Microbiol* 77: 2103–2112. <https://doi.org/10.1128/AEM.02833-10>.
- 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>.
- 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>.
- Bushnell B. 2016. BMAP short-read aligner, and other bioinformatic tools. <http://sourceforge.net/projects/bbmap/>.
- 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>.
- 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>.
- Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy in silico serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J Clin Microbiol* 53:2410–2426. <https://doi.org/10.1128/JCM.00008-15>.
- Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, Sicheritz-Ponten T, Aarestrup FM, Ussery DW, Lund O. 2014. Benchmarking of methods for genomic taxonomy. *J Clin Microbiol* 52:1529–1539. <https://doi.org/10.1128/JCM.02981-13>.
- Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Ponten 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>.