

Whole-Genome Sequence of *Escherichia coli* Serotype O157:H7 Strain EDL932 (ATCC 43894)

Gaylen A. Uhlich,^a George C. Paoli,^a Chin-Yi Chen,^a Bryan J. Cottrell,^a Xinmin Zhang,^b Xianghe Yan^{a*}

Molecular Characterization of Foodborne Pathogens Research Unit, U.S. Department of Agriculture, Agriculture Research Service, Eastern Regional Research Center, Wyndmoor, Pennsylvania, USA^a; BioInfoRx, Inc., Madison, Wisconsin, USA^b

* Present address: Xianghe Yan, Environmental, Microbial, & Food Safety Laboratory, Beltsville Agriculture Research Center, USDA-ARS, Beltsville, Maryland, USA.

The genome sequence of *Escherichia coli* serotype O157:H7 EDL933, a ground beef isolate from a 1983 hemorrhagic colitis outbreak, is a standard reference for comparative genomic studies of Shiga toxin-producing *E. coli* strains. Here, we report the genome sequence of a patient stool isolate from that outbreak, strain EDL932.

Received 18 May 2016 Accepted 25 May 2016 Published 14 July 2016

Citation Uhlich GA, Paoli GC, Chen C-Y, Cottrell BJ, Zhang X, Yan X. 2016. Whole-genome sequence of *Escherichia coli* serotype O157:H7 strain EDL932 (ATCC 43894). *Genome Announc* 4(4):e00647-16. doi:10.1128/genomeA.00647-16.

Copyright © 2016 Uhlich 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 Gaylen A. Uhlich, gaylen.uhlich@ars.usda.gov.

Shiga toxin-producing *Escherichia coli* (STEC) strains cause intestinal disease characterized by hemorrhagic colitis, which may progress to severe sequelae, such as hemolytic-uremic syndrome. In the United States, O157:H7 is the most important STEC serotype, both in total numbers of individual cases and large outbreaks.

STEC O157:H7 strain EDL933 (CDC; aka ATCC 43895) is a ground beef isolate from a 1983 hemorrhagic colitis outbreak in Michigan (1). The closed EDL933 sequence (GenBank accession no. AE005174.2 [2]) and a newly annotated version (GenBank accession no. CP008957 [3]) have served as reference genomes for sequence comparisons with many STEC strains.

Individual serotype O157:H7 strains show significant genotypic variability depending on the environment from which they were isolated. This is evident with regard to biofilm-forming properties and RpoS-dependent stress resistance (4, 5). Here, we report the whole-genome sequencing (WGS) of strain EDL932 (ATCC 43894), a patient isolate from that same 1983 Michigan outbreak.

Large genomic DNA (gDNA) fragments for single-molecule real-time (SMRT) sequencing were extracted from a frozen cell pellet (5 ml of overnight culture) of EDL932, designated 43894OW, using the Qiagen Genomic-tip 100/G kit. SMRT sequencing was done at the University of Delaware Sequencing and Genotyping Center (Newark, DE) using the PacBio RSII SMRT DNA sequencing system. *De novo* assembly of the SMRT reads used Hierarchical Genome Assembly Process (version 3). Total DNA was also sent to ProteinCT Biotechnologies, LLC (Madison, WI) for complete WGS workflow. Libraries were prepared using the Nextera DNA library preparation kit (Illumina) and sequenced using Illumina MiSeq. Approximately 4.5 million 2 × 250-bp paired-end reads were generated. Raw data quality was evaluated using FastQC, and Trimmomatic (USADCELLAB.org) was used to remove adapters and low-quality sequences (<Q20). Clean data were assembled with Velvet (<https://www.ebi.ac.uk/~zerbino/velvet/> [6]). Several parameters were tested, and the

assemblies were evaluated by comparison to the EDL933 NCBI genome reference sequence (accession no. NZ_CP008957.1). The final assembly used in the report has a hash size of 69. In addition to genome assembly, clean reads were remapped to the EDL933 NCBI genome using Burrows-Wheeler Aligner (BWA) (7). Duplicated reads were removed with Picard (<http://broadinstitute.github.io/picard/>); variants were called using SAMtools (8) and annotated with SnpEff (9).

The unclosed 43894OW draft genome aligned with the 5,547,323-bp EDL933 reference genome, except for a nearly 13-kbp region with exceptional coverage depth, indicating gene duplication in strain EDL932 (Integrative Genomics Viewer [IGV]; Broad Institute, Cambridge, MA). Complete resolution of duplicated areas may expand the genome size. Variant detection identified small numbers of single-nucleotide polymorphisms (SNPs) and indels, many with low-quality scores. Included was a T-to-G transversion at position 721 of the *rpoS* coding sequence (CDS) compared to the reference EDL933, which changes a stop codon to a glutamic acid residue and extends EDL932 RpoS to 330 residues. This transversion matched the consensus for other serotype O157 strains, including strain EDL933 (ATCC 43895) *rpoS* gene sequenced in our earlier study (10), indicating variability among different outbreak isolates or a mistake during the original EDL933 sequence.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. [LPWC00000000](https://www.ncbi.nlm.nih.gov/nuclink/LPWC00000000). The version described in this paper is version LPWC02000000.

ACKNOWLEDGMENT

The mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

FUNDING INFORMATION

This work was supported by the U.S. Department of Agriculture, Agricultural Research Service, in-house project number 8072-42000-067-00D.

REFERENCES

1. Wells JG, Davis BR, Wachsmuth IK, Riley LW, Remis RS, Sokolow R, Morris GK. 1983. Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare *Escherichia coli* serotype. *J Clin Microbiol* 18:512–520.
2. Perna NT, Plunkett G III, Burland V, Mau B, Glasner JD, Rose DJ, Mayhew GF, Evans PS, Gregor J, Kirkpatrick HA, Pósfai G, Hackett J, Klink S, Boutin A, Shao Y, Miller L, Grotbeck EJ, Davis NW, Lim A, Dimalanta ET, Potamousis KD, Apodaca J, Anantharaman TS, Lin J, Yen G, Schwartz DC, Welch RA, Blattner FR. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–533. <http://dx.doi.org/10.1038/35054089>.
3. Latif H, Li HJ, Charusanti P, Palsson BØ, Aziz RK. 2014. A gapless, unambiguous genome sequence of the enterohemorrhagic *Escherichia coli* O157:H7 strain EDL933. *Genome Announc* 2(4):e00821-14. <http://dx.doi.org/10.1128/genomeA.00821-14>.
4. Van Hoek AH, Aarts HJ, Bouw E, van Overbeek WM, Franz E. 2013. The role of *rpoS* in *Escherichia coli* O157 manure-amended soil survival and distribution of allelic variations among bovine, food and clinical isolates. *FEMS Microbiol Lett* 338:18–23. <http://dx.doi.org/10.1111/1574-6968.12024>.
5. Biscola FT, Abe CM, Guth BE. 2011. Determination of adhesin gene sequences in, and biofilm formation by, O157 and non-O157 Shiga toxin-producing *Escherichia coli* strains isolated from different sources. *Appl Environ Microbiol* 77:2201–2208. <http://dx.doi.org/10.1128/AEM.01920-10>.
6. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
7. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <http://dx.doi.org/10.1093/bioinformatics/btp698>.
8. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <http://dx.doi.org/10.1093/bioinformatics/btr509>.
9. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6:80–92. <http://dx.doi.org/10.4161/fly.19695>.
10. Uhlich GA, Chen CY, Cottrell BJ, Hofmann CS, Dudley EG, Strobaugh TP, Jr, Nguyen LH. 2013. Phage insertion in *mlrA* and variations in *rpoS* limit curli expression and biofilm formation in *Escherichia coli* serotype O157:H7. *Microbiology* 159:1586–1596. <http://dx.doi.org/10.1099/mic.0.066118-0>.