

# Complete Genome Sequences of Two Shiga Toxin-Producing *Escherichia coli* Strains from Serotypes O119:H4 and O165:H25

Rebecca L. Lindsey, Kristen Knipe, Lori Rowe, Lisley Garcia-Toledo, Vladimir Loparev, Phalasy Juieng, Eija Trees, Nancy Strockbine, Devon Stripling, Peter Gerner-Smidt

Centers for Disease Control and Prevention, Atlanta, Georgia, USA

**Shiga toxin-producing *Escherichia coli* (STEC) is an important foodborne pathogen. Here, we report complete whole-genome sequences for two STEC strains of serotypes O119:H4 and O165:H25 isolated from clinical cases in the United States.**

Received 28 October 2015 Accepted 4 November 2015 Published 17 December 2015

**Citation** Lindsey RL, Knipe K, Rowe L, Garcia-Toledo L, Loparev V, Juieng P, Trees E, Strockbine N, Stripling D, Gerner-Smidt P. 2015. Complete genome sequences of two Shiga toxin-producing *Escherichia coli* strains from serotypes O119:H4 and O165:H25. *Genome Announc* 3(6):e01496-15. doi:10.1128/genomeA.01496-15.

**Copyright** © 2015 Lindsey et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](#).

Address correspondence to Rebecca L. Lindsey, [rebecca.lindsey@cdc.hhs.gov](mailto:rebecca.lindsey@cdc.hhs.gov).

Shiga toxin-producing *Escherichia coli* (STEC) is a common cause of foodborne disease (1) that can cause severe illness in the form of bloody diarrhea and hemolytic-uremic syndrome (HUS). There are very few reference sequences available for non-O157 STEC serogroups outside the “big six” that are regarded as adulterants by the USDA. O119 and O165 are both relatively common diarrheagenic serogroups, and serogroup O165 has also been associated with an outbreak in Oregon in 2006 (2, 3). Here, we report the availability of closed whole-genome sequences for O119:H4 and O165:H25 STEC. These closed genome sequences represent STEC serotypes that did not previously have publicly available high-quality reference sequences generated by PacBio sequencing and confirmed with optical mapping.

*E. coli* genomic DNA was extracted according to the manufacturer's protocol (ArchivePure; 5 Prime, Gaithersburg, MD). DNAs were sheared to 10 kbp or 20 kbp utilizing g-Tubes (Covaris, Inc., Woburn, MA). The sheared products were further size selected utilizing 0.45× AMPure beads. Sheared DNAs were used to generate large SMRTbell libraries using the standard library protocols of the Pacific Biosciences DNA template preparation kit (Menlo Park, CA). The finished libraries were bound to proprietary P5 polymerase and sequenced on a PacBio RSII sequencer using C3 chemistry for 150-min movies. Sequence reads were filtered and assembled *de novo* utilizing the PacBio Hierarchical Genome Assembly Process version 3 (4). The resulting chromosome assemblies were confirmed using OpGen (Gaithersburg, MD) whole-genome maps (WGM). WGM were generated according to the OpGen protocol. Isolate number 2009C-3133 (O119:H4, *stx*<sub>1</sub> positive, 10-kb library), isolated from a patient in New York in 2009 (accession no. CP013025), was assembled into one closed chromosome with a length of 5,155,368 bp and a G+C content of 50.76%. There are three unnamed plasmids associated with 2009C-3133: accession no. CP013024 is 120,611 bp (closed), accession no. CP013026 is 63,800 bp (closed), and accession no. CP013027 is 174,564 bp. Isolate number 2012C-4227 (O165:H25, *stx*<sub>1</sub> and *stx*<sub>2</sub> positive, 20-kb library), isolated from a patient with diarrhea in Oregon in 2009 (accession no. CP013029), assembled

into one closed chromosome with a length of 5,202,850 bp and a G+C content of 50.67%. There are two unnamed plasmids associated with 2012C-4227: accession no. CP013028 is 74,656 bp (closed), and accession no. CP013030 is 97,704 bp (closed). The sequences were annotated with the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (5).

A detailed report on an additional analysis of the draft genome sequences will be included in a future publication.

**Nucleotide sequence accession numbers.** The annotated whole-genome *E. coli* sequences have been deposited in DDBJ/ENA/GenBank under the accession numbers CP013024 to CP013030. The versions described in this paper are the first versions.

## ACKNOWLEDGMENTS

This work was supported through U.S. government funds.

The findings and conclusions of this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

## REFERENCES

- Centers for Disease Control and Prevention. 2011. Vital signs: incidence and trends of infection with pathogens transmitted commonly through food--foodborne diseases active surveillance network, 10 U.S. sites, 1996-2010. *MMWR Morb Mortal Wkly Rep* 60:749-755.
- Luna-Gierke RE, Griffin PM, Gould LH, Herman K, Bopp CA, Strockbine N, Mody RK. 2014. Outbreaks of non-O157 Shiga toxin-producing *Escherichia coli* infection: USA. *Epidemiol Infect* 142:2270-2280. <http://dx.doi.org/10.1017/S0950268813003233>.
- Robins-Browne RM, Hartland EL. 2002. *Escherichia coli* as a cause of diarrhea. *J Gastroenterol Hepatol* 17:467-475. <http://dx.doi.org/10.1046/j.1440-1746.2002.02769.x>.
- Chin C, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563-569. <http://dx.doi.org/10.1038/nmeth.2474>.
- Klimke W, Agarwala R, Badretdin A, Chetvernin S, Ciufo S, Fedorov B, Kiryutin B, O'Neill K, Resch W, Resenchuk S, Schafer S, Tolstoy I, Tatusova T. 2009. The National Center for Biotechnology Information's protein Clusters Database. *Nucleic Acids Res* 37:D216-D223. <http://dx.doi.org/10.1093/nar/gkn734>.