AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

PacBio Genome Sequences of Eight *Escherichia albertii* Strains Isolated from Humans in the United States

Rebecca L. Lindsey,[a] Lori A. Rowe,[b] Dhwani Batra,[b] Peyton Smith,[a] Nancy A. Strockbine[a]

[a]Enteric Diseases Laboratory Branch, Division of Foodborne, Waterborne, and Environmental Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA
[b]Biotechnology Core Facility Branch, Division of Scientific Resources, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

ABSTRACT *Escherichia albertii* is an emerging pathogen that is closely related to *Escherichia coli* and can carry some of the same virulence genes as *E. coli*. Here, we report the release of Illumina-corrected PacBio sequences for eight *E. albertii* genomes. Two of these strains carry Shiga toxin 2f.

Escherichia albertii is an emerging pathogen that has been isolated worldwide from humans, birds, animals, and the environment; the resultant isolates that have been described in the literature since 2003 are reviewed here (1). *E. albertii* strains were implicated in an outbreak of restaurant-associated gastrointestinal disease in Japan (2), and stx_{2f}-positive *E. albertii* strains have been isolated from a bird and humans (3). Due to the limited number of complete *E. albertii* genomes publicly available (seven at NCBI as of 11 June 2018), we selected eight diverse strains (all isolated from humans, in different U.S. states, from 1954 to 2014) for PacBio sequencing. Here, we report the release of eight complete circularized *E. albertii* chromosomes and their associated plasmid sequences. Two of these strains (2012EL-1823B and 2014C-4015) carry an stx_{2f} gene.

Using standard microbial methods, strains were isolated at clinical laboratories as suspected enteric pathogens (4). Strains were received at the *Escherichia/Shigella* Reference Laboratory, CDC, and checked for purity. For initial identification and confirmation of *E. albertii*, a multiplex PCR was used to test for the presence of the *clpX*, *lysP*, and *mdh* genes (5). Additional confirmation included *rpoB* gene-based sequence analysis or a multiplex *Escherichia* species PCR (6, 7). Strain 2045-54 is an *E. albertii* strain, but it was originally described as a *Shigella boydii* strain of serotype 13 by W. H. Ewing (CDC, Atlanta, GA); in 1955, it was sent to the National Collection of Type Cultures (NCTC, Public Health England) and named NCTC 9362 (5).

Strain growth, DNA extraction, sequencing, and assembly were completed as previously described, except where noted (8). A single colony was selected from a streak of a frozen stock of a pure culture for a second streak on blood agar (Becton, Dickinson and Company, USA), and the equivalent of 5 to 6 colonies were selected from the second streak for genomic DNA extraction according to the manufacturer's protocol (Promega Wizard Genomic DNA purification kit, Promega Corporation, Madison, WI). This DNA extract was used for all sequencing. For Illumina MiSeq sequencing, libraries were prepared with a Nextera XT library prep kit (Illumina, USA) and sequenced following the manufacturer's protocols (Illumina, USA). For PacBio sequencing, DNA was sheared to 20 kb utilizing needle shearing and used to generate large SMRTbell libraries using the standard 20-kb library protocols of the Pacific Biosciences SMRTbell template prep kit 1.0 (PacBio, Menlo Park, CA). The libraries were further size selected utilizing BluePippin (Sage Scientific, Beverly, MA) with a cutoff size of 10 kb. The finished library was bound to proprietary P6v2 polymerase and sequenced on a PacBio RS II sequencer using C4v2 chemistry for 360-minute movies. Sequence reads were filtered

Citation Lindsey RL, Rowe LA, Batra D, Smith P, Strockbine NA. 2019. PacBio genome sequences of eight *Escherichia albertii* strains isolated from humans in the United States. Microbiol Resour Announc 8:e01663-18. https://doi.org/10.1128/MRA.01663-18.

Editor J. Cameron Thrash, University of Southern California

Address correspondence to Rebecca L. Lindsey, rebecca.lindsey@cdc.hhs.gov.

Received 10 December 2018
Accepted 22 January 2019
Published 28 February 2019

TABLE 1 Sequence IDs, NCBI accession numbers, GC content, and sizes of *E. albertii* sequences

E. albertii sequence ID	Assignment	NCBI accession no.	Coverage (×)	% GC content	Total size (bp)
05-3106	Chromosome	CP030778	156.90	49.77	4,719,735
p05-3106-1	Plasmid	CP030779	71.65	44.19	56,603
p05-3106-2	Plasmid	CP030780	147.98	54.23	80,632
06-3542	Chromosome	CP034162	181.25	49.80	4,709,095
p06-3542	Plasmid	CP034163	62.44	47.31	95,683
07-3866	Chromosome	CP030781	127.69	49.77	4,940,006
p07-3866	Plasmid	CP030782	62.43	48.13	104,269
2010C-3449	Chromosome	CP034212	85.04	50.02	4,923,641
2012EL-1823B	Chromosome	CP030783	64.97	49.66	4,809,821
p2012EL-1823B-1	Plasmid	CP030784	26.42	51.34	81,130
p2012EL-1823B-2	Plasmid	CP030785	20.08	47.17	100,347
p2012EL-1823B-3	Plasmid	CP030786	35.44	44.68	105,846
2013C-4143	Chromosome	CP030787	124.52	49.80	4,659,709
2014C-4015	Chromosome	CP034166	84.05	49.83	4,623,903
p2014C-4015-1	Plasmid	CP034165	20.38	44.95	63,807
p2014C-4015-2	Plasmid	CP034164	28.70	47.37	96,264
p2014C-4015-3	Plasmid	CP034167	51.69	45.17	136,645
NCTC 9362	Chromosome	CP034213	162.95	50.15	4,551,125
pNCTC 9362	Plasmid	CP034214	30.11	52.60	40,180

and assembled *de novo* utilizing the PacBio Hierarchical Genome Assembly Process version 3 and polished using Quiver (9) (minread = 1,000, genome size = 5,000,000). For 06-3542, the assembly was generated using Canu 1.6 (minReadLength = 1,000 -g5m), as it generated a circularized chromosome and plasmid (10). All PacBio sequences were Illumina corrected with unicycler_polish that uses Pilon (11, 12) with default settings.

Table 1 lists the sequence identification (ID) numbers, NCBI accession numbers, GC content, and chromosome and plasmid sizes for each *E. albertii* strain. A single chromosomal sequence (circular with overlapping ends) was obtained for all strains with a minimum of 64× coverage.

Data availability. The whole-genome sequences reported here have been deposited in DDBJ/ENA/GenBank under the accession numbers listed in Table 1. The versions described in this paper are the first versions. PacBio (SRX5170195 to SRX5170202) and Illumina (SRR8355575 to SRR8355582) sequencing reads for strains in this study have been deposited in the NCBI Sequence Read Archive (SRA) under the accession numbers listed above.

ACKNOWLEDGMENTS

We are very grateful to Devon Stoneburg, Lisley Garcia-Toledo, Pooja Patel, and Haley Martin for their expert contributions.

This work was funded by federal appropriations to the Centers for Disease Control and Prevention through the Advanced Molecular Detection Initiative line item.

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. Use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention or by the U.S. Department of Health and Human Services.

REFERENCES

1. Grillova L, Sedlacek I, Pachnikova G, Stankova E, Svec P, Holochova P, Micenkova L, Bosak J, Slaninova I, Smajs D. 2018. Characterization of four *Escherichia albertii* isolates collected from animals living in Antarctica and Patagonia. J Vet Med Sci 80:138–146. https://doi.org/10.1292/jvms.17-0492.
2. Ooka T, Tokuoka E, Furukawa M, Nagamura T, Ogura Y, Arisawa K, Harada S, Hayashi T. 2013. Human gastroenteritis outbreak associated with *Escherichia albertii*, Japan. Emerg Infect Dis 19:144–146. https://doi.org/10.3201/eid1901.120646.
3. Ooka T, Seto K, Kawano K, Kobayashi H, Etoh Y, Ichihara S, Kaneko A, Isobe J, Yamaguchi K, Horikawa K, Gomes TA, Linden A, Bardiau M, Mainil JG, Beutin L, Ogura Y, Hayashi T. 2012. Clinical significance of *Escherichia*

albertii. Emerg Infect Dis 18:488–492. https://doi.org/10.3201/eid1803.111401.
4. Jorgensen JH, Pfaller MA, Carroll KC, Funke G, Landry ML, Richter SS, Warnock DW. 2015. Manual of clinical microbiology, 11th ed. American Society for Microbiology, Washington, DC. https://doi.org/10.1128/9781555817381.
5. Hyma KE, Lacher DW, Nelson AM, Bumbaugh AC, Janda JM, Strockbine NA, Young VB, Whittam TS. 2005. Evolutionary genetics of a new pathogenic *Escherichia* species: *Escherichia albertii* and related *Shigella boydii* strains. J Bacteriol 187:619–628. https://doi.org/10.1128/JB.187.2.619-628.2005.
6. Lindsey RL, Fedorka-Cray PJ, Abley M, Turpin JB, Meinersmann RJ. 2015. Evaluating the occurrence of *Escherichia albertii* in chicken carcass rinses by PCR, Vitek analysis, and sequencing of the *rpoB* gene. Appl Environ Microbiol 81:1727–1734. https://doi.org/10.1128/AEM.03681-14.
7. Lindsey RL, Garcia-Toledo L, Fasulo D, Gladney LM, Strockbine N. 2017. Multiplex polymerase chain reaction for identification of *Escherichia coli*, *Escherichia albertii* and *Escherichia fergusonii*. J Microbiol Methods 140:1–4. https://doi.org/10.1016/j.mimet.2017.06.005.
8. Lindsey RL, Batra D, Smith P, Patel PN, Tagg KA, Garcia-Toledo L, Loparev VN, Juieng P, Sheth M, Joung YJ, Rowe LA. 2018. PacBio genome sequences of *Escherichia coli* serotype O157:H7, diffusely adherent *E. coli*, and *Salmonella enterica* strains, all carrying plasmids with an *mcr-1* resistance gene. Microbiol Resour Announc 7:e01025-18. https://doi.org/10.1128/MRA.01025-18.
9. Chin CS, 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. https://doi.org/10.1038/nmeth.2474.
10. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
11. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
12. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.