



Complete Genome Sequence of *Enterococcus durans* Oregon-R-modENCODE Strain BDGP3, a Lactic Acid Bacterium Found in the *Drosophila melanogaster* Gut

Kenneth H. Wan, Charles Yu,* Soo Park, Ann S. Hammonds, Benjamin W. Booth, Susan E. Celniker

Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA

ABSTRACT *Enterococcus durans* Oregon-R-modENCODE strain BDGP3 was isolated from the *Drosophila melanogaster* gut for functional host-microbe interaction studies. The complete genome is composed of a single circular genome of 2,983,334 bp, with a G+C content of 38%, and a single plasmid of 5,594 bp.

Enterococcus is part of the large genus of lactic acid bacteria (LAB). In *Drosophila*, a number of species have been identified as symbionts, including *Enterococcus faecalis*, *Enterococcus faecium*, *Enterococcus gallinarum*, and *Enterococcus durans* (1). This report describes the sequence of *E. durans* Oregon-R-modENCODE strain BDGP3, associated with a sequenced *Drosophila melanogaster* host.

E. durans Oregon-R-modENCODE strain BDGP3 was isolated from *Drosophila melanogaster* guts. Bacteria were streaked onto brain heart infusion (BHI) agar plates (catalog number 241830; BD Difco). Purified colonies were amplified in liquid BHI (catalog number 211059; Difco) culture overnight, and an aliquot was used for 16S V1 and V4 PCR (2) and sequence identification by ABI 3730XL Sanger sequencing (reviewed in reference 3). Genomic DNA for sequencing was isolated by treating cells with protease, followed by cetyltrimethylammonium bromide (CTAB) extraction (4).

Whole-genome DNA sequencing was performed by the National Center for Genome Resources (NCGR), Santa Fe, NM, using Pacific Biosciences (PacBio, Menlo Park, CA) long-read sequencing on the RSII instrument (5). A single-molecule real-time (SMRT) cell library was constructed with 5 to 10 μ g of input DNA using the PacBio 20-kbp protocol. The library was loaded onto one SMRT cell and sequenced using P6 polymerase and C4 chemistry with 6-h movie times. Sequencing yielded a total of 62,492 reads, with a filtered mean read length of 8,747 bp, totaling 546,630,958 bp (>150-fold coverage for the chromosome and >50-fold coverage for the plasmid). The files generated by the PacBio instrument were used for *de novo* assembly constructed using the Hierarchical Genome Assembly Process 2 (HGAP2) protocol from SMRT Analysis version 2.0 (6, 7). This protocol relies on BLASR for alignment (8), the Celera assembler for assembly (7), and Quiver for consensus polishing (6). The final contigs were manually trimmed and reviewed to produce a single circular chromosome and a single plasmid. Annotations of protein-encoding open reading frames and noncoding RNAs (ncRNAs) were predicted using the Rapid Annotation of microbial genomes using Subsystems Technology tool (9) and the GenBank annotation pipeline (10).

The genome sequence is composed of a single circular chromosome (2,983,334 bp), with a predicted overall G+C content of 38%. The chromosomal genome is predicted to contain 2,612 protein-coding genes, 159 pseudogenes, 3 ncRNAs, 6 rRNA operons, 69 tRNA genes, and one transfer-messenger RNA (tmRNA). Of the 2,612 protein-coding

Received 11 September 2017 Accepted 18 September 2017 Published 5 October 2017

Citation Wan KH, Yu C, Park S, Hammonds AS, Booth BW, Celniker SE. 2017. Complete genome sequence of *Enterococcus durans* Oregon-R-modENCODE strain BDGP3, a lactic acid bacterium found in the *Drosophila melanogaster* gut. *Genome Announc* 5:e01041-17. <https://doi.org/10.1128/genomeA.01041-17>.

Copyright © 2017 Wan 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 Susan E. Celniker, secelniker@lbl.gov.

* Present address: Charles Yu, Genentech, Inc., South San Francisco, California, USA.

genes, 104 are contained within candidate prophages. Our strain contains a single integrated likely prophage and two partial copies (Phage_Finder; Omic Tools). The single copy contains 36,659 bp, and the 2 partial copies contain 13,567 and 16,827 bp. They constitute 2.24% of the genome. Two cornerstone highly conserved prophage proteins are the large terminase subunit and the portal protein (reviewed in reference 11). Only the 36-kb copy contains genes that encode these proteins. In addition, the genome contains a single plasmid, pEdBDGP3A (5,594 bp, 33% G+C content). pEdBDGP3A contains a candidate gene for the plasmid replication initiation protein RepB.

Accession number(s). The complete chromosome and plasmid sequences of *Enterococcus durans* Oregon-R-modENCODE strain BDGP3 are deposited in GenBank under the accession numbers [CP022930](#) (chromosome) and [CP022931](#) (plasmid).

ACKNOWLEDGMENTS

We thank J. B. Brown, J.-H. Mao, A. Snijders, S. Langley, and N. Bonini for scientific discussions. We thank Faye Schilkey, Jennifer Jacobi, and Nicholas Devitt of the NCGR for PacBio sequencing. The Oregon-R-modENCODE fly strain initially obtained from the Bloomington *Drosophila* Stock Center (25211) has been cultured at the BDGP for 6 years.

This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy contract DE-AC02-05CH11231.

REFERENCES

- Cox CR, Gilmore MS. 2007. Native microbial colonization of *Drosophila melanogaster* and its use as a model of *Enterococcus faecalis* pathogenesis. *Infect Immun* 75:1565–1576. <https://doi.org/10.1128/IAI.01496-06>.
- Yu Z, Morrison M. 2004. Comparisons of different hypervariable regions of *rrs* genes for use in fingerprinting of microbial communities by PCR-denaturing gradient gel electrophoresis. *Appl Environ Microbiol* 70:4800–4806. <https://doi.org/10.1128/AEM.70.8.4800-4806.2004>.
- Slatko BE, Kieleczawa J, Ju J, Gardner AF, Hendrickson CL, Ausubel FM. 2011. “First generation” automated DNA sequencing technology. *Curr Protoc Mol Biol* Chapter 7:Unit 7.2. <https://doi.org/10.1002/0471142727.mb0702s96>.
- Wilson K. 2001. Preparation of genomic DNA from bacteria. *Curr Protoc Mol Biol* Chapter 2:Unit 2.4. <https://doi.org/10.1002/0471142727.mb0204s56>.
- Korlach J, Bjornson KP, Chaudhuri BP, Cicero RL, Flusberg BA, Gray JJ, Holden D, Saxena R, Wegener J, Turner SW. 2010. Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol* 472: 431–455. [https://doi.org/10.1016/S0076-6879\(10\)72001-2](https://doi.org/10.1016/S0076-6879(10)72001-2).
- 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. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
- Koren S, Harhay GP, Smith TP, Bono JL, Harhay DM, McVey SD, Radune D, Bergman NH, Phillippy AM. 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol* 14: R101. <https://doi.org/10.1186/gb-2013-14-9-r101>.
- Chaisson MJ, Tesler G. 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics* 13:238. <https://doi.org/10.1186/1471-2105-13-238>.
- Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvermin V, Ciufu S, Li W. 2013. Prokaryotic Genome Annotation Pipeline. The NCBI handbook, 2nd ed. National Center for Biotechnology Information, Bethesda, MD.
- Casjens S. 2003. Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol* 49:277–300. <https://doi.org/10.1046/j.1365-2958.2003.03580.x>.