AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Complete Genome Sequence of *Acetobacter tropicalis* Oregon-R-modENCODE Strain BDGP1, an Acetic 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 *Acetobacter tropicalis* Oregon-R-modENCODE strain BDGP1 was isolated from *Drosophila melanogaster* for functional host-microbe interaction studies. The complete genome comprises a single chromosomal circle of 3,988,649 bp with a G+C content of 56% and a conjugative plasmid of 151,013 bp.

Received 16 August 2017 **Accepted** 17 October 2017 **Published** 16 November 2017

Citation Wan KH, Yu C, Park S, Hammonds AS, Booth BW, Celniker SE. 2017. Complete genome sequence of *Acetobacter tropicalis* Oregon-R-modENCODE strain BDGP1, an acetic acid bacterium found in the *Drosophila melanogaster* gut. Genome Announc 5:e01020-17. https://doi.org/10.1128/genomeA.01020-17.

Address correspondence to Susan E. Celniker, secelniker@lbl.gov.

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

In *Drosophila*, *Acetobacter* species are one of the major commensals of the gut microbiota and contribute to larval growth (1). Furthermore, monocolonization of *Drosophila* with *Acetobacter* species significantly reduces host development (2). The first draft sequence of *A. tropicalis* from *Drosophila*, published in 2014, consisted of 129 contigs (3). We report here the complete genome sequence consisting of a single chromosome and a conjugative plasmid.

A. tropicalis Oregon-R-modENCODE strain BDGP1 was isolated from a fecal swab. Bacteria were streaked onto nutrient broth agar (BD catalog number 213000) plates, single colonies were amplified in culture, and an aliquot was used for 16S V1 and V4 PCR (4) and sequence identification (5). DNA for sequencing was isolated (6), and whole-genome DNA sequencing was performed by the National Center for Genome Resources (NCGR), Santa Fe, New Mexico, USA, using Pacific Biosciences (PacBio) long reads on the RS II instrument (7). A single-molecule real-time (SMRT) cell library was constructed with 5 to 10 μg of DNA using the PacBio 20-kb protocol. The library was sequenced using P6 polymerase and C4 chemistry with 6-h movie times. Sequencing yielded a total of 78,825 reads with a filtered mean read length of 17,469 bp, totaling 1,377,023,046 bp (>200- and >400-fold coverage for the chromosome and plasmid, respectively). A *de novo* assembly was constructed using the hierarchical genome assembly process (HGAP2) protocol from SMRT Analysis version 2.0 (8–10). 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 (ORFs) and noncoding RNAs (ncRNAs) were predicted using the Rapid Annotations using Subsystems Technology (RAST) tool (11) and the GenBank annotation pipeline (12).

The chromosomal genome annotation predicts a total of 3,645 genes, including 3,446 protein-coding genes, 77 RNA genes (including 5 rRNA operons, 58 tRNAs, 1 transfer-messenger RNA, and 3 ncRNAs), and 122 pseudogenes. Of the 3,446 protein-coding genes, 82 are contained within two candidate prophages (29 kb, 31 genes, and 39 kb, 51 genes) characterized by genes encoding tailed phage morphogenetic, portal, and terminase large subunit proteins and bounded by genes encoding phage lysin (lysozyme) and integrase. The candidate prophages are only 1.7% of the genome. Additionally, the genome contains a single plasmid, pAtBDGP1A (151,013 bp), with a

predicted G+C content of 52%. The plasmid contains 145 candidate protein-coding genes. Among them are genes essential for conjugation (*traA*, *traG*, *traW*, and *traY*), plasmid replication (*repA*, *repB*, and *repC*), and likely members of a bacterial transport system (*tral/dotC*, *dotG*, *dotH*, *icmB/dotO*, and *icmL/dotI*) (13). The plasmid also encodes virulence proteins for arsenic resistance.

Our sequence is significantly similar to that of *A. senegalensis* 108B, having a 97% average nucleotide identity (14) with 77% coverage of the genome. *A. senegalensis* originated from a spontaneous cocoa bean fermentation process (GenBank accession number LN606600). Despite the sequence similarity, phenotypically, our strain belongs to the *A. tropicalis* species based on its ability to grow on maltose and not grow on 10% ethanol or on yeast extract supplemented with 30% D-glucose (15).

Accession number(s). The chromosome and plasmid sequences of *A. tropicalis* Oregon-R-modENCODE strain BDGP1 have been deposited in GenBank under the accession numbers CP022699 and CP022700, respectively.

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. We thank Eliza Paneru and William Fisher for phenotyping. The 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 the Lawrence Berkeley National Laboratory under U.S. Department of Energy contract DE-AC02-05CH11231.

REFERENCES

1. Broderick NA, Lemaitre B. 2012. Gut-associated microbes of *Drosophila melanogaster*. Gut Microbes 3:307–321. https://doi.org/10.4161/gmic.19896.
2. Newell PD, Douglas AE. 2014. Interspecies interactions determine the impact of the gut microbiota on nutrient allocation in *Drosophila melanogaster*. Appl Environ Microbiol 80:788–796. https://doi.org/10.1128/AEM.02742-13.
3. Newell PD, Chaston JM, Wang Y, Winans NJ, Sannino DR, Wong AC, Dobson AJ, Kagle J, Douglas AE. 2014. *In vivo* function and comparative genomic analyses of the *Drosophila* gut microbiota identify candidate symbiosis factors. Front Microbiol 5:576. https://doi.org/10.3389/fmicb.2014.00576.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Ciufo S, Li W (ed). 2013. Prokaryotic genome annotation pipeline. *In* The NCBI handbook. NCBI, Bethesda, MD. https://www.ncbi.nlm.nih.gov/books/NBK174280.
13. Zink SD, Pedersen L, Cianciotto NP, Abu-Kwaik Y. 2002. The Dot/Icm type IV secretion system of *Legionella pneumophila* is essential for the induction of apoptosis in human macrophages. Infect Immun 70:1657–1663. https://doi.org/10.1128/IAI.70.3.1657-1663.2002.
14. Luo C, Rodriguez-R LM, Konstantinidis KT. 2013. A user's guide to quantitative and comparative analysis of metagenomic datasets. Methods Enzymol 531:525–547. https://doi.org/10.1016/B978-0-12-407863-5.00023-X.
15. Ndoye B, Cleenwerck I, Engelbeen K, Dubois-Dauphin R, Guiro AT, Van Trappen S, Willems A, Thonart P. 2007. *Acetobacter senegalensis* sp. nov., a thermotolerant acetic acid bacterium isolated in Senegal (sub-Saharan Africa) from mango fruit (*Mangifera indica* L.). Int J Syst Evol Microbiol 57:1576–1581. https://doi.org/10.1099/ijs.0.64678-0.