



Complete Genome Sequence of *Bacillus kochii* Oregon-R-modENCODE Strain BDGP4, Isolated from *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 *Bacillus kochii* Oregon-R-modENCODE strain BDGP4 was isolated from the gut of *Drosophila melanogaster* for functional host microbial interaction studies. The complete genome comprised a single chromosomal circle of 4,557,232 bp with a G+C content of 37% and a single plasmid of 137,143 bp.

The *Bacillus* genus represents one of the oldest, most ubiquitous, and molecularly well-characterized bacterial clades. *Bacillus kochii*, first described in 2012, is aerobic, motile, and catalase positive (1). This report describes the first complete sequence of *B. kochii* Oregon-R-modENCODE strain BDGP4, associated with a sequenced *Drosophila melanogaster* host facilitating mechanistic studies of host–microbiome interactions.

Bacillus kochii Oregon-R-modENCODE strain BDGP4 was isolated from a fecal swab. Bacteria were streaked onto nutrient broth agar (BD catalog no. 213000) plates; single colonies were amplified in culture, and an aliquot was used for 16S V1 and V4 PCR (2) and sequence identification (3). DNA for sequencing was isolated (4), and whole-genome DNA sequencing was performed by the National Center for Genome Resources (NCGR, Santa Fe, New Mexico, USA) using Pacific Biosciences (PacBio, Menlo Park, CA, USA) long-read sequencing on the RS II instrument (5). A single-molecule real-time (SMRT) cell library was constructed with 5 to 10 μg of input DNA using the PacBio 20-kb 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 20,658 reads with a filtered mean read length of 13,027 bp, totaling 269,115,728 bp (>50-fold coverage). The files generated by the PacBio instrument were used for a *de novo* assembly constructed using the hierarchical genome assembly process (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 RAST tool (9) and the GenBank annotation pipeline (10).

The chromosomal genome annotation predicts 4,296 protein-coding genes, 10 rRNA operons, 107 tRNAs, and 5 ncRNAs. Of the 4,296 protein-coding genes, 72 are contained within candidate prophages. Our strain contains a single near-complete cryptic prophage and one partial copy (<https://omictools.com/phage-finder-tool>). The cryptic prophage is 39,935 bp and the partial copy is 13,008 bp. They constitute 1.16% of the genome. The 39-kb cryptic prophage contains genes encoding the large and small terminase subunit proteins but lacks an essential portal protein that is contained within the partial copy (11). In addition, the genome contains a single plasmid pBkBDGP4A (137,143 bp, 34% G+C) encoding 131 predicted genes, of which three are predicted

Received 29 August 2017 Accepted 1 September 2017 Published 5 October 2017

Citation Wan KH, Yu C, Park S, Hammonds AS, Booth BW, Celniker SE. 2017. Complete genome sequence of *Bacillus kochii* Oregon-R-modENCODE strain BDGP4, isolated from *Drosophila melanogaster* gut. Genome Announc 5:e01074-17. <https://doi.org/10.1128/genomeA.01074-17>.

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.

conjugal transfer proteins and one is predicted to encode a MerR protein, known to respond to environmental stimuli, such as oxidative stress, heavy metals, or antibiotics (12). The plasmid is also predicted to encode three germination proteins—GerKA, GerKB, and GerKC—that may affect the sporulation and germination processes of the host.

Accession number(s). The complete chromosome and plasmid sequences of *Bacillus kochii* Oregon-R-modENCODE strain BDGP4 have been deposited in GenBank under accession numbers [CP022983](#) (chromosome) and [CP022984](#) (plasmid).

ACKNOWLEDGMENTS

We thank J. B. Brown, J.-H. Mao, A. Snijders, S. Langley, and Nancy 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 no. DE-AC02-05CH11231.

REFERENCES

- Seiler H, Schmidt V, Wenning M, Scherer S. 2012. *Bacillus kochii* sp. nov., isolated from foods and a pharmaceuticals manufacturing site. *Int J Syst Evol Microbiol* 62:1092–1097. <https://doi.org/10.1099/ij.s.0.027771-0>.
- 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, Chetvernin V, Ciufu S, Li W. 2013. Prokaryotic genome annotation pipeline. *In* The NCBI Handbook [Internet]. National Center for Biotechnology Information, Bethesda, MD. <https://www.ncbi.nlm.nih.gov/books/NBK174280>.
- 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>.
- Brown NL, Stoyanov JV, Kidd SP, Hobman JL. 2003. The MerR family of transcriptional regulators. *FEMS Microbiol Rev* 27:145–163. [https://doi.org/10.1016/S0168-6445\(03\)00051-2](https://doi.org/10.1016/S0168-6445(03)00051-2).