

Genome Sequence and Methylome of Soil Bacterium *Gemmatirosa kalamazonensis* KBS708^T, a Member of the Rarely Cultivated *Gemmatimonadetes* Phylum

Jennifer M. DeBruyn,^a Mark Radosevich,^a K. Eric Wommack,^b Shawn W. Polson,^c Loren J. Hauser,^d Mariam N. Fawaz,^a Jonas Korch,^e Yu-Chih Tsai^e

Department of Biosystems Engineering & Soil Science, University of Tennessee, Knoxville, Tennessee, USA^a; Department of Plant & Soil Science, University of Delaware, Newark, Delaware, USA^b; Center for Bioinformatics & Computational Biology, University of Delaware, Newark, Delaware, USA^c; Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA^d; Pacific Biosciences, Menlo Park, California, USA^e

Bacteria belonging to the phylum *Gemmatimonadetes* are found in a wide variety of environments and are particularly abundant in soils. Here, we present the complete genome sequence and methylation pattern of the newly described *Gemmatirosa kalamazonensis* type strain.

Received 27 February 2014 Accepted 13 March 2014 Published 3 April 2014

Citation DeBruyn JM, Radosevich M, Wommack KE, Polson SW, Hauser LJ, Fawaz MN, Korch J, Tsai Y-C. 2014. Genome sequence and methylome of soil bacterium *Gemmatirosa kalamazonensis* KBS708^T, a member of the rarely cultivated *Gemmatimonadetes* phylum. *Genome Announc.* 2(2):e00226-14. doi:10.1128/genomeA.00226-14.

Copyright © 2014 DeBruyn et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Jennifer M. DeBruyn, jdebruyn@utk.edu.

Bacteria belonging to the phylum *Gemmatimonadetes* are frequently found in soils (1). To date, only two *Gemmatimonadetes* strains have been characterized: *Gemmatimonas aurantiaca* T-27 from wastewater (2) and *Gemmatirosa kalamazonensis* KBS708 isolated from soil (3). Here, we report the complete genome sequence of *G. kalamazonensis*.

G. kalamazonensis KBS708^T (ATCC BAA-2150, NCCB 100411) was grown for 10 days on VL55 minimal medium with 0.025% peptone, as previously described (3). Genomic DNA was extracted using an UltraClean microbial DNA isolation kit (Mo Bio) and randomly sheared to ~10-kb target size using G-tubes (Covaris, Inc.). Poly(dA) tails were added to the 3' ends using terminal deoxynucleotidyl transferase (TdT). The poly(dA)-tailed library was then annealed with poly(dT) sequencing primer and sequenced using DNA/polymerase binding kit 2.0 with a MagBead loading kit and 120-min sequencing time on the PacBio RS instrument (Pacific Biosciences, Inc.).

Single-molecule real-time (SMRT) sequencing data collected using the TdT library at Pacific Biosciences was combined with standard SMRT sequencing data collected at the University of Delaware for *de novo* genome assembly using the Hierarchical Genome Assembly Process (HGAP) (4). Initial output from HGAP yielded one 5,318-kb chromosome (118× coverage) and two large satellite elements (1,059 kb and 1,041 kb with 102× and 98× coverage, respectively). A collapsed 53-kb tandem repeat region in the 1,059-kb element was identified and resolved into a 1,106-kb element. A 21-kb high-copy element with sequence overlapping the chromosome was also resolved in the process (376× coverage). This manual curation process resulted in a 5,312-kb circular chromosome and three circular satellite elements (1,106 kb, 1,041 kb, and 21 kb), with 72.6% average G+C content. Final assembly was polished using the Quiver consensus algorithm included in the SMRT analysis software package. Base

modifications were identified using the base modification analysis protocol (Pacific Biosciences).

The Prodigal genome annotation pipeline at Oak Ridge National Laboratory (5, 6) was used to predict genes and provide annotation based on homology searches. A total of 6,373 candidate protein-coding genes were predicted. The 5.3-Mb chromosome contained 48 tRNA genes. Two sets of *rrn* genes were identified, with one set in an operon and the second split, as its 16S rRNA gene was located 1.28 Mb away and on the opposite strand from the 23S-5S genes. A total of 4,515 protein coding genes were predicted, 3,434 of which were assigned a function based on homology. The two large plasmids (1.1 Mb and 1.0 Mb) contain 1,026 and 798 predicted genes, respectively. The small high-copy element (21 kb) contains 34 predicted genes; the 5 genes that were assigned functions suggest it may be a phage.

Methylation analysis (7) revealed the presence of three active N⁶-methyladenine methyltransferases with the recognition sequences 5'-RGATCY-3', 5'-ATGCAC-3', and 5'-CCAGN₇TCA-3', each with >99% of the genomic positions conforming to the sequence motif detected as methylated (boldface denotes a methylated base; underlining denotes a methylated base on the opposite DNA strand).

Accession numbers. The genome sequences and methylation data are deposited at NCBI under GenBank accession no. CP007127 to CP007130 and GEO accession no. GSE55390.

ACKNOWLEDGMENTS

This work was funded by the NSF/USDA Microbial Observatories program award 2007-35319-18432 and NSF-OCE1148118.

J. Korch and Y.-C. Tsai are full-time employees at Pacific Biosciences, a company commercializing single-molecule real-time sequencing technologies.

We thank Tyson Clark for assistance with methylome data.

REFERENCES

1. DeBruyn JM, Nixon LT, Fawaz MN, Johnson AM, Radosevich M. 2011. Global biogeography and quantitative seasonal dynamics of *Gemmatimonadetes* in soil. *Appl. Environ. Microbiol.* 77:6295–6300. <http://dx.doi.org/10.1128/AEM.05005-11>.
2. Zhang H, Sekiguchi Y, Hanada S, Hugenholtz P, Kim H, Kamagata Y, Nakamura K. 2003. *Gemmatimonas aurantiaca* gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum *Gemmatimonadetes* phyl. nov. *Int. J. Syst. Evol. Microbiol.* 53:1155–1163. <http://dx.doi.org/10.1099/ijs.0.02520-0>.
3. DeBruyn JM, Fawaz MN, Peacock AD, Dunlap JR, Nixon LT, Cooper KE, Radosevich M. 2013. *Gemmatirosa kalamazoonensis* gen. nov., sp. nov., a member of the rarely-cultivated bacterial phylum *Gemmatimonadetes*. *J. Gen. Appl. Microbiol.* 59:305–312. <http://dx.doi.org/10.2323/jgam.59.305>.
4. 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. <http://dx.doi.org/10.1038/nmeth.2474>.
5. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
6. Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC. 2012. Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* 28:2223–2230. <http://dx.doi.org/10.1093/bioinformatics/bts429>.
7. Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov A, Turner SW, Korlach J, Roberts RJ. 2012. The methylomes of six bacteria. *Nucleic Acids Res.* 40:11450–11462. <http://dx.doi.org/10.1093/nar/gks891>.