Complete Genome Sequence of *Bdellovibrio* sp. Strain KM01, a Predatory Bacterium Isolated from Soil

Sidney C. Davis,[a] Karla J. Martinez,[a] Laura E. Williams[a]

ᵃDepartment of Biology, Providence College, Providence, Rhode Island, USA

ABSTRACT We report the complete genome sequence of the predatory bacterium *Bdellovibrio* sp. strain KM01, isolated from soil collected near a pond. The genome is 3,961,288 bp long with 45.5% GC content. Comparative genomics among *Bdellovibrio* strains will help us understand how genotypic differences affect differences in predatory phenotypes.

Because *Bdellovibrio* species can kill Gram-negative bacteria, they may be useful as clinical therapies; however, more *Bdellovibrio* genome sequences are needed to understand how genotype affects differences in predatory phenotypes. We report the genome of *Bdellovibrio* sp. strain KM01, isolated following the enrichment protocol described in reference 1 from soil collected near a pond at a Rhode Island state park.

To obtain genomic DNA, we cocultured KM01 in HM buffer (1) with 1.5 ml overnight culture of *Raoultella* sp. strain 0037. After 48 h at 28°C and 200 rpm, we passed the coculture through a 0.45-μm filter and then extracted genomic DNA from the filtrate containing *Bdellovibrio* sp. strain KM01 using the Wizard genomic DNA purification kit (Promega). The University of Maryland Institute for Genome Sciences sheared genomic DNA using a g-TUBE at 3,400 rpm and then size selected it on a Blue Pippin instrument (Sage Scientific) with an 11,000-bp cutoff. The library was prepared using a SMRTbell template prep kit v1.0 with no modifications and sequenced on a PacBio RS II instrument with P6-C4 chemistry. The University of Rhode Island Genomics and Sequencing Center sheared genomic DNA using a Covaris S-220 focused ultrasonicator. The library was prepared using a PrepX DNA library kit (TaKaRa Bio), visualized on a high-sensitivity BioAnalyzer chip (Agilent), and quantified using the KAPA Illumina quantification kit (Roche). The library was sequenced on an Illumina MiSeq instrument to generate 2 × 250-bp paired-end reads.

For *de novo* assembly of long reads, we launched an Amazon EC2 instance of SMRT Portal v2.3.0 and used Hierarchical Genome Assembly Process v3 (HGAP3) (2) with default target coverage (25×) and 4.0 Mb estimated genome size. From one single-molecule real-time (SMRT) cell, we obtained 99,389 subreads with an N_{50} value of 12,098 bp, which assembled into a 3,978,353-bp contig with 158× mean coverage. To circularize the contig, we used BLAST (3) to locate the overlap between contig ends and EMBOSS extractseq (4) to trim the overlap, yielding a 3,960,707-bp closed contig. Using extractseq, we adjusted the first position to align it with the predicted *dnaA* start codon.

To polish the closed contig, we processed 2,028,031 MiSeq read pairs with SolexaQA v3.1.4 (5) using DynamicTrim to remove bases with a quality score of <13 and then using LengthSort to discard reads of <65 bp. This yielded 1,925,830 read pairs. Using the Burrows-Wheeler Aligner MEM algorithm (BWA-MEM) v0.7.13 (6), we mapped 98% of the reads (3,775,952 of 3,851,660) to the closed contig. We sorted and indexed the alignment with SAMtools v1.3 (7) and then confirmed 99.98% of the closed contig using Pilon v1.22 (8). Pilon identified and corrected 584 small indels and 1 single-nucleotide polymorphism (SNP), and it flagged five regions as local continuity breaks, two of which it fixed. This yielded a 3,961,296-bp draft sequence.

To evaluate continuity breaks and assess changes made by Pilon, we used Canu v2.0 (9) to correct PacBio subreads and then used minimap2 v2.1 (10) to align corrected reads to the draft sequence. We sorted and indexed the alignment with SAMtools and then visualized it with Tablet v1.19.09.03 (11). Within the regions previously flagged by Pilon as local continuity breaks, we examined the alignment for variants and corrected six homopolymer tracts, generating the final 3,961,288-bp genome with 45.5% GC content and 160× mean MiSeq read coverage.

We annotated the genome with Rapid Annotations using Subsystems Technology (RAST) (12) and NCBI's Prokaryotic Genome Annotation Pipeline (PGAP) (13). PGAP identified 3,752 protein-coding genes, 44 RNA-coding genes, and 17 pseudogenes.

Data availability. The *Bdellovibrio* sp. strain KM01 genome sequence was deposited in GenBank under accession number CP058348. The MiSeq and PacBio reads were deposited in the SRA under accession numbers SRR12207263 and SRR12207262, respectively.

REFERENCES

1. Williams LE, Cullen N, DeGiorgis JA, Martinez KJ, Mellone J, Oser M, Wang J, Zhang Y. 2019. Variation in genome content and predatory phenotypes between Bdellovibrio sp. NC01 isolated from soil and B. bacteriovorus type strain HD100. Microbiology (Reading) 165:1315–1330.

2. Chin C-S, 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.

3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

4. Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16:276–277. https://doi.org/10.1016/s0168-9525(00)02024-2.

5. Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485. https://doi.org/10.1186/1471-2105-11-485.

6. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

8. 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.

9. 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.

10. Li H. 2018. minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

11. Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D. 2010. Tablet: next generation sequence assembly visualization. Bioinformatics 26:401–402. https://doi.org/10.1093/bioinformatics/btp666.

12. 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.

13. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.