

Complete Genome Sequence of a Strain of *Azospirillum thiophilum* Isolated from a Sulfide Spring

Alexey Fomenkov,^a Tamas Vincze,^a Margarita Grabovich,^b Brian P. Anton,^a Galina Dubinina,^c Maria Orlova,^b Elena Belousova,^b Richard J. Roberts^a

New England Biolabs, Ipswich, Massachusetts, USA^a; Department of Biology, Voronezh State University, Voronezh, Russia^b; Winogradsky Institute of Microbiology, Research Center of Biotechnology, Russian Academy of Sciences, Moscow, Russia^c

We report the complete, closed genome sequence and complete methylome of *Azospirillum thiophilum* strain BV-S^T.

Received 6 November 2015 Accepted 10 November 2015 Published 7 January 2016

Citation Fomenkov A, Vincze T, Grabovich M, Anton BP, Dubinina G, Orlova M, Belousova E, Roberts RJ. 2016. Complete genome sequence of a strain of *Azospirillum thiophilum* isolated from a sulfide spring. *Genome Announc* 4(1):e01521-15. doi:10.1128/genomeA.01521-15.

Copyright © 2016 Fomenkov 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 Alexey Fomenkov, Fomenkov@neb.com.

In this work, we report the complete, closed genome sequence of *A. thiophilum* strain BV-S^T (DSM 21654^T, VKM B-2513^T), which was isolated from a bacterial mat of a sulfide spring. This strain was previously described based on its morphological and biochemical characteristics (1, 2).

The genome was sequenced using the Pacific Biosciences (PacBio RSII) single-molecule real-time (SMRT) sequencing platform. Two SMRTbell libraries were prepared according to modified 10- and 20-kb PacBio sample preparation protocols, including additional separation on a BluePippin DNA size selection system (Sage Science, Beverly, MA) and sequenced using PacBio's C4 chemistry on 17 SMRT cells using either 180- or 240-min collection times. Sequencing reads were processed, mapped, and assembled by the SMRT analysis pipeline using the HGAP3 protocol and polished using Quiver (3) to give a fully closed genome with 354× coverage. The total genome size was 7,609,460 bp consisting of eight closed chromosomes of 3,037,424, 1,354,909, 878,428, 712,512, 694,460, 518,276, 330,423, and 83,026 bp, respectively, which together encoded a total of 6,620 genes. This multichromosome assembly is consistent with the structure of previously published *Azospirillum* genome sequences (4–6). The assembled sequences were annotated with the NCBI Prokaryotic Genomes Annotation Pipeline (PGAP) and were deposited at DDBJ/EMBL/GenBank.

While this manuscript was in preparation, the same genome was sequenced by another group and the whole-genome shotgun sequence was deposited in GenBank (NZ_LAEL00000000.1) (Y. Kwak and J.-H. Shin, unpublished data). Comparison of the two

assemblies using Mauve 2.4.0 indicates the following correspondence between the two sets of contigs: CP012401 and LAEL01000001, with CP012401 2,684 bp larger; LAEL01000011 is completely included within LAEL01000001 and CP012401; CP012402 and LAEL01000002 + LAEL01000008, with CP012402 3,627 bp smaller than those 2 put together; CP012403 and LAEL01000004, with CP012403 446 bp larger; CP012404 and LAEL01000003, with CP012404 4,432 bp larger; CP012405 and LAEL01000007, with CP012405 12,253 bp smaller; CP012406 and LAEL01000005, with CP012406 5,897 bp smaller; CP012407 and LAEL01000006, with CP012407 10,910 bp smaller; CP012408 and LAEL01000009 + LAEL01000010, with CP012408 7,576 bp larger than those 2 put together.

Epigenetic modification at each nucleotide position was determined using kinetic variation (KV) in the nucleotide incorporation rates, and methylated motifs were deduced from the KV data (7–9). Five methylated motifs corresponding to one m4C and four m6A modifications were detected by SMRT sequencing of untreated genomic DNA, and no additional motifs were detected in Tet2 treated DNA, suggesting an absence of m5C modification. The motifs were matched with their corresponding methyltransferase genes, and the results are shown in Table 1. These results have also been deposited in REBASE (10).

Nucleotide sequence accession numbers. The complete genome sequences of *A. thiophilum* BV-S^T (DSM 21654^T, VKM B-2513^T) are available in GenBank under the accession numbers CP012401, CP012402, CP012403, CP012404, CP012405, CP012406, CP012407, and CP012408.

TABLE 1 Summary of methyltransferases identified in *A. thiophilum* strain BV-S^T

Motif ^a	Assigned or predicted	Methylation type	Restriction modification type
GANTC	M.Ath2165I	m6A	II
RGATCY	M.Ath2165II	m6A	II
GGCGCCGG		m4C	II
GACNNNNNNCTGG	S.Ath2165ORFDR	m6A	I
GCGATCC		m6A	II

^a Modified bases are highlighted in bold.

ACKNOWLEDGMENTS

This project was supported by the Small Business Innovation Research Program (NIGMS) of the National Institutes of Health under award R44GM105125 to R.J.R.

R.J.R., A.F., B.P.A., and T.V. work for New England Biolabs, a company that sells research reagents, including restriction enzymes and DNA methyltransferases, to the scientific community.

FUNDING INFORMATION

NIH SBIR provided funding to Richard J. Roberts under grant number R44GM105125.

REFERENCES

1. Lavrinenko K, Chernousova E, Gridneva E, Dubinina G, Akimov V, Kuever J, Lysenko A, Grabovich M. 2010. *Azospirillum thiophilum* sp. nov., a diazotrophic bacterium isolated from a sulfide spring. *Int J Syst Evol Microbiol* 60:2832–2837. <http://dx.doi.org/10.1099/ijs.0.018853-0>.
2. Frolov EN, Belousova EV, Lavrinenko KS, Dubinina GA, Grabovich MY. 2013. Capacity of *Azospirillum thiophilum* for lithotrophic growth coupled to oxidation of reduced sulfur compounds. *Microbiology* 82: 271–279. <http://dx.doi.org/10.1134/S0026261713030053>.
3. Chin C, 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>.
4. Rivera D, Revale S, Molina R, Gualpa J, Puente M, Maroniche G, Paris G, Baker D, Clavijo B, McLay K, Spaepen S, Perticari A, Vazquez M, Wisniewski-Dyé F, Watkins C, Martínez-Abarca F, Vanderleyden J, Cassán F. 2014. Complete genome sequence of the model rhizosphere strain *Azospirillum brasilense* Az39, successfully applied in agriculture. *Genome Announc* 2(4):e00683-14. <http://dx.doi.org/10.1128/genomeA.00683-14>.
5. Kaneko T, Minamisawa K, Isawa T, Nakatsukasa H, Mitsui H, Kawaharada Y, Nakamura Y, Watanabe A, Kawashima K, Ono A, Shimizu Y, Takahashi C, Minami C, Fujishiro T, Kohara M, Katoh M, Nakazaki N, Nakayama S, Yamada M, Tabata S, Sato S. 2010. Complete genomic structure of the cultivated rice endophyte *Azospirillum* sp. B510. *DNA Res* 17:37–50.
6. Wisniewski-Dyé F, Borziak K, Khalsa-Moyers G, Alexandre G, Sukharnikov LO, Wuichet K, Hurst GB, McDonald WH, Robertson JS, Barbe V, Calteau A, Rouy Z, Mangenot S, Prigent-Combaret C, Normand P, Boyer M, Siguier P, Dessaux Y, Elmerich C, Condemine G, Krishnen G, Kennedy I, Paterson AH, González V, Mavingui P, Zhulin IB. 2011. *Azospirillum* genomes reveal transition of bacteria from aquatic to terrestrial environments. *PLoS Genet* 7:e1002430. <http://dx.doi.org/10.1371/journal.pgen.1002430>.
7. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7:461–465. <http://dx.doi.org/10.1038/nmeth.1459>.
8. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, Fomenkov A, Roberts RJ, Korlach J. 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res* 40:e29. <http://dx.doi.org/10.1093/nar/gkr1146>.
9. Korlach J, Turner SW. 2012. Going beyond five bases in DNA sequencing. *Curr Opin Struct Biol* 22:251–261. <http://dx.doi.org/10.1016/j.sbi.2012.04.002>.
10. Roberts RJ, Vincze T, Posfai J, Macelis D. 2010. Rebase—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* 38:D234–D236. <http://dx.doi.org/10.1093/nar/gkp874>.