



Complete Genome Sequence of the Freshwater Bacterium *Beggiatoa leptomitiformis* Strain D-401

Alexey Fomenkov,^a Zhiyi Sun,^a Tamas Vincze,^a Galina Dubinina,^c Maria Orlova,^b Sergey V. Tarlachkov,^{d,e} Brian P. Anton,^a Margarita Y. Grabovich,^b Richard J. Roberts^a

^aNew England Biolabs, Ipswich, Massachusetts, USA

^bDepartment of Biochemistry and Cell Physiology, Voronezh State University, Voronezh, Russia

^cFederal State Institution, Federal Research Centre for Fundamentals of Biotechnology, Russian Academy of Sciences, Moscow, Russia

^dAll-Russian Collection of Microorganisms (VKM), G. K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Russian Academy of Sciences, Pushchino, Russia

^eBranch of Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Pushchino, Russia

ABSTRACT Here, we report the complete closed genome sequence and methylome analysis of *Beggiatoa leptomitiformis* strain D-401 (DSM 14945, UNIQEMU 779), which is quite different from the previously described *Beggiatoa leptomitiformis* neotype strain D-402^T (DSM 14946, UNIQEM U 779) with regard to morphology and lithotrophic growth in the presence of thiosulfate.

The taxonomy of the genus *Beggiatoa* is still a work in progress. Despite the fact that many morphotypes of *Beggiatoa* have been described previously, only two species have been validated: *Beggiatoa alba* and *Beggiatoa leptomitiformis*. Previously, we reported the genome sequence of *B. leptomitiformis* strain D-402^T (1), after which we described this strain as a representative of a new species within the genus *Beggiatoa* (2). Since the isolates were initially described based on the morphological criteria alone and named *Beggiatoa leptomitiformis*, the species name was retained but its orthography was changed based on the rules of Appendix 9 of the International Code of Nomenclature of Prokaryotes (3) and validated as *Beggiatoa leptomitiformis* sp. nov. with the type strain D-402^T. Here, we report the genome sequence of a second strain, *B. leptomitiformis* D-401. Both strains differ in their morphology and physiology, especially their ability to grow lithotrophically in the presence of thiosulfate. While *B. leptomitiformis* D-402^T is able to accumulate elemental sulfur intracellularly, D-401 is not. Therefore, comparative genomic analysis of these strains could illuminate metabolic features associated with lithotrophic growth in *Beggiatoa*.

B. leptomitiformis D-401 was sequenced using the Pacific Biosciences (PacBio) RS II sequencing platform. Briefly, SMRTbell libraries were constructed from genomic DNA sheared to a size ranging from ~10 to 20 kb using the manufacturer's instructions. DNA quality analysis and quantification were performed using the Qubit fluorimeter (Invitrogen, Eugene, OR) and 2100 Bioanalyzer (Agilent Technology, Santa Clara, CA). One 16-kb SMRTbell library was prepared according to the 20-kb PacBio sample preparation protocol, including additional separation on a BluePippin to remove fragments less than 7 kb. One size-selected and one non-size-selected library were sequenced by using C4-P6 chemistry using 2 single-molecule real-time (SMRT) cells with 240-minute collection times. Sequencing reads were processed, mapped, and assembled with the Pacific Biosciences SMRT analysis pipeline using the HGAP3 protocol and polished using Quiver (4). A total of 1.5 Gb of sequencing data was assembled into a single closed

Received 12 March 2018 Accepted 22 March 2018 Published 26 April 2018

Citation Fomenkov A, Sun Z, Vincze T, Dubinina G, Orlova M, Tarlachkov SV, Anton BP, Grabovich MY, Roberts RJ. 2018. Complete genome sequence of the freshwater bacterium *Beggiatoa leptomitiformis* strain D-401. Genome Announc 6:e00311-18. <https://doi.org/10.1128/genomeA.00311-18>.

Copyright © 2018 Fomenkov 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 Alexey Fomenkov, fomenkov@neb.com.

TABLE 1 Summary of DNA methyltransferase genes and their modified motifs identified in *B. leptomitiformis* strains D-402^T and D-401

Motif ^a	Gene predicted in Ble402	Gene predicted in Ble401	% detected in Ble402	% detected in Ble401	Methylation type	Restriction modification type
Assigned						
GATC	M.Ble402I	M.Ble401I	99.8	100	m6A	II
GRAGCAG	M.Ble402II	M.Ble401II	99.5	100	m6A	II
SAGCTS	M.Ble402III	M.Ble401III	20.1	99.4	m4C	II
ACA Y NNNNNRTGT	S.Ble402ORF7560P	S.Ble401ORFEP	96.4	Not detected ^b	m6A	I
CA A YNNNNR T TG	S.Ble402ORF6900P	S.Ble401ORFCP	71.8	72.45	m6A	I
CA G NNNNNR T AAT	S.Ble402ORF1460P	S.Ble401ORFTP	96.9	98.8	m6A	I
Unassigned						
CATCHAG			100	100	m6A	II
CGGAG			98.9	99.4	m6A	III
CGGTCA			98.2	99.2	m6A	II
DCTGGATD			97.7	99.9	m6A	II
GGCTGA			99.6	99.9	m6A	II
GTTGNAG			100	100	m6A	II
TCGA			98.7	99.9	m6A	II
5mC detected with mTet2 oxidation						
GGHCC= GG NCC	M.Ble402ORF17485P	M.Ble401ORFQP	71.56	ND ^c	5mC	II
CCDGG= CC NGG	M.Ble402ORF8490P	M.Ble401ORFFP	39.55	ND	5mC	II
GGCCNB= GG CC	M.Ble402ORF16030P	M.Ble401ORFPP	29.6	ND	5mC	II
Predicted/not detected						
RGCGCY	M.Ble402ORF12705P	M.Ble401ORFKP			5mC	II
GRCGYC	M.Ble402ORF6255P	M.Ble401ORFAP			5mC	II
GCTCCA	Ble402ORF115P	Ble401ORFRP			m6A	II
GCATGC	M.Ble402ORF1215P	M1-2.Ble401ORFSP			ND	II
AAGCTT	M.Ble402ORF12455P	M.Ble401ORFJP			ND	II
TCTAGA	M.Ble402ORF3920P	M.Ble401ORFWP			ND	II
ATGCAT	M.Ble402ORFA1P				ND	II

^aModified bases are in bold, and modified bases on an opposite strand are in bold and underlined.

^bThe S.Ble401ORFEP modified motif was not detected directly by SMRT pipeline motif and modification software in *B. leptomitiformis* strain D-401, but it was confirmed manually using PBMotStat software (T.V.).

^cND, not determined.

circular genome of 4,266,286 bp with 290.26-fold coverage. The assembled sequence was annotated using the NCBI Prokaryotic Genomes Annotation Pipeline (PGAP).

Surprisingly, comparative analysis of *B. leptomitiformis* D-401 and D-402^T showed more than 99.5% identity with only a 98-genomic loci difference between them. The differences consisted primarily of insertions and deletions, with only one single nucleotide polymorphism (SNP).

Epigenetic modification at each nucleotide position was measured as kinetic variation (KV) in the nucleotide incorporation rates, and methylated motifs were deduced from the KV data (5–7). Based on the KV patterns, we identified one DNA methyltransferase recognition motif presumed to contain m4C and nine presumed to contain m6A. Three additional motifs presumed to contain m5C were previously identified in strain D-402^T using Tet2 treatment; while we did not repeat that analysis here, we assume the results would be similar. Matching of motifs with genes of the methyltransferases responsible for each was carried out, and the results are shown in Table 1 and deposited in REBASE (8). Both the methyltransferase genes and the observed methylated motifs were identical between the two *B. leptomitiformis* strains.

Accession number(s). The complete, closed genome sequence of the *B. leptomitiformis* strain D-401 is available in DDBJ/ENA/GenBank with the accession number [CP018889](https://www.ncbi.nlm.nih.gov/nuccore/CP018889).

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. and Russian Foundation of Fundamental Investigations (RFFI) grant 1804-00556 to M.Y.G.

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

REFERENCES

1. Fomenkov A, Vincze T, Grabovich MY, Dubinina G, Orlova M, Belousova E, Roberts RJ. 2015. Complete genome sequence of the freshwater colorless sulfur bacterium *Beggiatoa leptomitoformis* neotype strain D-402. *Genome Announc* 6:e01436. <https://doi.org/10.1128/genomeA.01436-15>.
2. Dubinina G, Savvichev A, Orlova M, Gavrish E, Verborg S, Grabovich M. 2017. *Beggiatoa leptomitoformis* sp. nov., the first freshwater member of the genus capable of chemolithoautotrophic growth. *Int J Syst Evol Microbiol* 67:197–204. <https://doi.org/10.1099/ijsem.0.001584>.
3. Parker CT, Tindall BJ, Garrity GM. 20 Nov 2015. International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol*. <https://doi.org/10.1099/ijsem.0.000778>.
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. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
5. 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. <https://doi.org/10.1038/nmeth.1459>.
6. 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. <https://doi.org/10.1093/nar/gkr1146>.
7. Korlach J, Turner SW. 2012. Going beyond five bases in DNA sequencing. *Curr Opin Struct Biol* 22:251–261. <https://doi.org/10.1016/j.sbi.2012.04.002>.
8. Roberts RJ, Vincze T, Posfai J, Macelis D. 2015. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* 43:D298–D299. <https://doi.org/10.1093/nar/gku1046>.