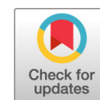
AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES



Complete Genome Sequence and Methylome Analysis of *Thermoactinomyces vulgaris* 2H

Houda Mankai,[a] Brian P. Anton,[b] Victoria Wu,[b,c] Tamas Vincze,[b] Richard J. Roberts,[b] Ferid Limam,[a] Alexey Fomenkov[b]

[a]Laboratoire des Substances Bioactives, Centre de Biotechnologies de Borj Cédria, Hammam Lif, Tunisia
[b]New England Biolabs, Inc., Ipswich, Massachusetts, USA
[c]Wellesley College, Wellesley, Massachusetts, USA

ABSTRACT Here, we report the complete genome sequence and full methylome analysis of a newly isolated, aerobic, thermophilic, Gram-positive actinomycete, a strain of *Thermoactinomyces vulgaris* designated strain 2H.

An aerobic, Gram-positive, thermophilic actinomycete, a strain of *Thermoactinomyces vulgaris* designated strain 2H, was isolated from a soil sample collected from the southwestern part of Tunisia (Tamaghza, Tozeur). The strain isolation was carried out using a dilution agar plating method on tryptic soy agar (TSA) medium (pH 7.3 ± 0.2). Ten grams of soil sample was suspended in 100 ml sterile distilled water and mixed. Further 10-fold dilutions were prepared, and a 0.1-ml sample from each dilution was spread onto the TSA plates with slight modifications (1). After incubation at 55°C for 5 days, a *Thermoactinomyces*-like strain was isolated and purified by streaking on new plates on the same medium; it was stored at 4°C.

The 16S rRNA gene sequence analysis of strain 2H (GenBank accession number KJ817372) using the EzBioCloud database (2) showed that the organism was closely related to *Thermoactinomyces vulgaris* DSM 43016^T (AF138739) (97.90% similarity) and *Thermoactinomyces intermedius* DSM 43846^T (AF138734) (97.41%). DNA-DNA hybridization confirmed that strain 2H belongs to the species *T. vulgaris*, and we have designated it *T. vulgaris* 2H. Strains with this designation are a potential source of thermostable enzymes such as proteases (3).

Genomic DNA from a culture of *T. vulgaris* 2H in tryptic soy broth (pH 7.3 ± 0.2), in shaken flasks (120 rpm) at 55°C, was purified with a GF-1 bacterial DNA extraction kit (Vivantis Technologies Sdn. Bhd., Malaysia) and sequenced using the Pacific Biosciences (PacBio) RS II sequencing platform. Briefly, SMRTbell libraries were constructed from a genomic DNA sample sheared to ~20 kb using the g-TUBE protocol (Covaris, Woburn, MA, USA), end repaired, and ligated to PacBio hairpin adapters. Incompletely formed SMRTbell templates and linear DNAs were digested with a combination of exonuclease III and exonuclease VII (New England Biolabs, Ipswich, MA, USA). DNA qualification and quantification were performed using a Qubit fluorometer (Invitrogen, Eugene, OR, USA) and a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). One 12-kb SMRTbell library was prepared according to modified PacBio sample preparation protocols, including additional separation on a BluePippin instrument (Sage Science, Beverly, MA, USA), and sequenced with C4-P6 chemistry. Two single-molecule real-time (SMRT) cells were sequenced, one with a non-size-selected (12-kb) library and one with a size-selected 14-kb library, with a 300-minute collection time. The sequencing reads (123,762), with a mean subread length of 8,746 bp and yield of 1.082 Gb, were *de novo* assembled using the HGAP_Assembly.3 version 2.3.0 with default quality and read length parameters and polished using Quiver (4). The polished assembly generated 2 closed-circular genome elements with 48% GC content for the main chromosome

Citation Mankai H, Anton BP, Wu V, Vincze T, Roberts RJ, Limam F, Fomenkov A. 2019. Complete genome sequence and methylome analysis of *Thermoactinomyces vulgaris* 2H. Microbiol Resour Announc 8:e00657-19. https://doi.org/10.1128/MRA.00657-19.

Editor Vincent Bruno, University of Maryland School of Medicine

Copyright © 2019 Mankai et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Houda Mankai, houda.mankai@googlemail.com, or Alexey Fomenkov, fomenkov@neb.com.

Received 31 May 2019
Accepted 16 July 2019
Published 8 August 2019

Downloaded from http://mra.asm.org/ on April 15, 2020 by guest

TABLE 1 Summary of genome elements, methyltransferase genes, and motifs identified in *Thermoactinomyces vulgaris* 2H[a]

Genetic element of *Thermoactinomyces vulgaris* 2H	Accession no.	Genome size (bp)	Genome coverage (×)	Methylase (RM system) name	Recognition motif[b]	Methylation, RM type
Chromosome	CP039710	2,619,280	328.15	RM.Tvu2HI	GG**C**C	5mC, II
				RM.Tvu2HII	C**C**WGG	5mC, II
				M.Tvu2HOR8305P	Inactive system	6mA, I
				M.Tvu2HORF1205P	Frameshift	5mC, II
Plasmid pTvu19	CP039711	18,747	606.42	None		

[a] RM, restriction-modification.
[b] Modified bases are shown in bold, or the base opposite to them is underlined.

(2,619,280 bp) and 41.9% GC content for the plasmid (18,747 bp). The assembled sequence was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (5, 6).

One advantage of the SMRT sequencing platform is its ability to detect the epigenetic state of sequenced DNA (7–9). Three apparent DNA methyltransferase recognition motifs, CCTGGVDCR, GGCCANB, and GGCCBBVNY, were detected by SMRT motif and modification analysis version 2.3.0, and each was reported to contain m4C modifications. However, this is a common artifact of PacBio motif calling; in fact, these motifs must originate from the genuine motifs CCWGG and GGCC as determined separately by genomewide m5C methylation analysis (B. P. Anton, unpublished data). Additional scanning with the SeqWare program (10) showed that the strain does not bear any m6A or m4C methyltransferase genes that might otherwise have been responsible for these motifs. All the m5C motifs were matched with the responsible methyltransferases, and the results are shown in Table 1 and have been deposited in REBASE (11).

Data availability. The complete genome sequence of *Thermoactinomyces vulgaris* 2H is available in GenBank under the accession numbers CP039710 and CP039711. The original sequence reads have been deposited at NCBI under the SRA accession numbers SRR8979896 and SRR8979897. The BioProject accession number is PRJNA534300.

ACKNOWLEDGMENTS

This project was supported by New England Biolabs, a company that sells research reagents, including restriction enzymes and DNA methyltransferases, to the scientific community.

A.F., B.P.A., T.V., V.W., and R.J.R. work for New England Biolabs.

REFERENCES

1. Yallop CA, Edwards C, Williams ST. 1997. Isolation and growth physiology of novel thermoactinomycetes. J Appl Microbiol 83:685–692. https://doi.org/10.1046/j.1365-2672.1997.00282.x.
2. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA and whole genome assemblies. Int J Syst Evol Microbiol 67:1613–1617. https://doi.org/10.1099/ijsem.0.001755.
3. Wang L, Qian Y, Cao Y, Huang Y, Chang Z, Huang H. 2017. Production and characterization of keratinolytic proteases by a chicken feather-degrading thermophilic strain, *Thermoactinomyces* sp. YT06. J Microbiol Biotechnol 27:2190–2198. https://doi.org/10.4014/jmb.1705.05082.
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. https://doi.org/10.1038/nmeth.2474.
5. 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.
6. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068.
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. https://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. https://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. https://doi.org/10.1016/j.sbi.2012.04.002.
10. 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. https://doi.org/10.1093/nar/gks891.
11. 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.

Downloaded from http://mra.asm.org/ on April 15, 2020 by guest