

Complete Genomic Sequence of “*Thermofilum adornatus*” Strain 1910b^T, a Hyperthermophilic Anaerobic Organotrophic Crenarchaeon

I. N. Dominova,^a I. V. Kublanov,^b O. A. Podosokorskaya,^b K. S. Derbikova,^b M. V. Patrushev,^a S. V. Toshchakov^a

Immanuel Kant Baltic Federal University, Kaliningrad, Russia^a; Winogradsky Institute of Microbiology, Russian Academy of Sciences, Moscow, Russia^b

The complete genomic sequence of a novel hyperthermophilic crenarchaeon, strain 1910b^T, was determined. The genome comprises a 1,750,259-bp circular chromosome containing single copies of 3 rRNA genes, 43 tRNA genes, and 1,896 protein-coding sequences. *In silico* genome-genome hybridization suggests the proposal of a novel species, “*Thermofilum adornatus*” strain 1910b^T.

Received 14 August 2013 Accepted 19 August 2013 Published 12 September 2013

Citation Dominova IN, Kublanov IV, Podosokorskaya OA, Derbikova KS, Patrushev MV, Toshchakov SV. 2013. Complete genomic sequence of “*Thermofilum adornatus*” strain 1910b^T, a hyperthermophilic anaerobic organotrophic crenarchaeon. *Genome Announc.* 1(5):e00726-13. doi:10.1128/genomeA.00726-13.

Copyright © 2013 Dominova et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to S. V. Toshchakov, stepan.toshchakov@gmail.com.

Thermofilaceae (1) is one of two families of the crenarchaeal order *Thermoproteales*, represented by a sole valid species, *Thermofilum pendens*, as well as several nonvalidated strains and uncultured clones. This family represents a deep phylogenetic lineage within the order *Thermoproteales*, possibly being an individual order in a crenarchaeal class *Thermoprotei*. Isolated from a solfatara in Iceland (2), *T. pendens* is a hyperthermophilic, moderately acidophilic, sulfur-dependent anaerobic heterotroph, which has an obligatory need for the *Thermoproteus tenax* polar lipid fraction for growth. Results of its genome analysis (3) revealed a number of proteins involved in peptide and saccharide utilization, while biosynthetic pathways for purines, most amino acids, and cofactors were absent, indicating the adaptation to life in organic-rich environments and dependence on other organisms providing lacking nutrients. Notably, the nonvalidated species “*Thermoproteus librum*,” which has a 16S rRNA gene sequence that is 100% identical, does not require the addition of cell components of other organisms (4).

Strain 1910b^T was isolated from a mud sample from a black mud pit (86°C, pH 5.5) located near Pauzhetka (Kamchatka Peninsula, Russia). The strain grows optimally at 92°C and pH 6.0 to 6.5 in the presence of *Fervidicoccus fontis* strain 1910a culture broth with glucose and peptone as the substrates.

A BLAST search (5) revealed 97.3% 16S rRNA gene sequence similarity between strain 1910b^T and *T. pendens* strain Hvv3^T.

For the genome sequencing of strain 1910b^T, we used a combination of fragment and mate-paired library approaches. The fragment library was sequenced with the Illumina MiSeq system, and the mate-paired library, with an average insert length of 2,200 bp, was sequenced with the Life Technologies PGM system. A total of 0.5 million Illumina paired-end reads were trimmed and corrected with the Quake sequencing error correction tool (6), and 3.1 million 200-bp PGM mate-paired reads were split and trimmed by use of the CLC Genomics Workbench and then were subjected to error correction with the SAET tool (7).

Reads were assembled with CLC Assembler using recommended parameters (8). Obtained contigs were scaffolded with

SSPACE (9) and remaining gaps were closed with the GapFiller tool (10), resulting in one 1.75-Mb contig. For circularization, we split the contig into 2 parts and joined the resulting sequences in reverse order with a 500-bp gap. This “artificial” gap was successfully closed with GapFiller, proving that the obtained contig corresponds to the circular chromosome of strain 1910b^T.

The genome of strain 1910b^T is comprised of a 1,750,259-bp circular chromosome with a G+C content of 46.5%. Annotation of the genome was performed with the NCBI Prokaryotic Genomes Annotation Pipeline with subsequent manual curation.

The chromosome of strain 1910b^T contains single rRNA copies, of which the 16S and 23S rRNAs were found together, while the 5S rRNA was found to be located in another region, as was also shown for the *T. pendens* strain Hrk5 genome (3). Forty-three tRNA genes and 1,896 protein-coding sequences were found. *In silico* genome-genome hybridization performed using the GGDC 2.0 algorithm (11) revealed a 0% probability that strain 1910b^T and *T. pendens* strain Hrk5 (= strain Hvv3^T) represent the same species. Based on this evidence, we assume that strain 1910b^T represents a novel species “*Thermofilum adornatus*.” Its complete description will be published elsewhere.

Nucleotide sequence accession number. The genome sequence of “*Thermofilum adornatus*” strain 1910b^T has been deposited in NCBI GenBank under the accession number CP006646.

ACKNOWLEDGMENTS

This work was supported by the Russian Federal Targeted Program for Research and Development, grant ID 14.512.11.0070, and by RFBR grant 13-04-00049. The work of S.V.T. and M.V.P. was supported by the RF President Fellowship for Young Scientists.

REFERENCES

1. Burggraf S, Huber H, Stetter KO. 1997. Reclassification of the crenarchaeal orders and families in accordance with 16S rRNA sequence data. *Int. J. Syst. Bacteriol.* 47:657–660.
2. Zillig W, Gierl A, Schreiber G, Wunderl S, Janekovic D, Stetter KO, Klenk HP. 1983. The archaeobacterium *Thermofilum pendens* represents a novel genus of the thermophilic, anaerobic sulfur respiring *Thermoproteales*. *Syst. Appl. Microbiol.* 4:79–87.

3. Anderson I, Rodriguez J, Susanti D, Porat I, Reich C, Ulrich LE, Elkins JG, Mavromatis K, Lykidis A, Kim E, Thompson LS, Nolan M, Land M, Copeland A, Lapidus A, Lucas S, Detter C, Zhulin IB, Olsen GJ, Whitman W, Mukhopadhyay B, Bristow J, Kyrpides N. 2008. Genome sequence of *Thermophilum pendens* reveals an exceptional loss of biosynthetic pathways without genome reduction. *J. Bacteriol.* **190**:2957–2965.
4. Stetter KO. 1986. Diversity of extremely thermophilic archaeobacteria, p 39–74. *In* Brock TD (ed), *Thermophiles: general, molecular and applied microbiology*. John Wiley & Sons, New York, NY.
5. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
6. Kelley DR, Schatz MC, Salzberg SL. 2010. Quake: quality-aware detection and correction of sequencing errors. *Genome Biol.* **11**:R116.
7. Brinza D, Hyland F. 2011. Workshop: error correction methods in next generation sequencing, p 268. *In* Proceedings of the 2011 IEEE 1st International Conference on Computational Advances in Bio and Medical Sciences. IEEE Computer Society, Washington, DC.
8. CLC Bio. 2012. White Paper on de novo assembly in CLC Assembly Cell 4.0. CLC bio A/S, Aarhus, Denmark.
9. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**: 578–579.
10. Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol.* **13**:R56.
11. Auch AF, von Jan M, Klenk H-P, Göker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomics Sci.* **2**:117–134.