





Complete Genome Resequencing of *Thermus thermophilus* Strain TMY by Hybrid Assembly of Long- and Short-Read Sequencing Technologies

 Kentaro Miyazaki,^{a,b}  Natsuko Tokito^{a,b}

^aBioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

^bDepartment of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan

ABSTRACT Complete genome resequencing was conducted for *Thermus thermophilus* strain TMY by hybrid assembly of Oxford Nanopore Technologies long-read and MGI short-read data. Errors in the previously reported genome sequence determined by PacBio technology alone were corrected, allowing for high-quality comparative genomic analysis of closely related *T. thermophilus* genomes.

Thermus thermophilus is an aerobic, thermophilic bacterium that grows optimally at around 70 to 75°C. Since the first isolation of this species from Mine Hot Spring in Japan in 1968 (1, 2), many *T. thermophilus* strains have been isolated from various thermal areas worldwide (3–11). Among them, strains HB8 (type strain) and HB27 have been rigorously studied biochemically (12, 13), structurally (14, 15), and genetically (16, 17).

So far, 13 complete genome sequences have been determined for *T. thermophilus* strains (<https://www.ncbi.nlm.nih.gov/genome/browse#!/prokaryotes/461/>). One of the strains, TMY, was isolated from the Otake geothermal power plant in Japan (6), and its genome sequence has been reported previously (18). However, because the sequencing was performed using PacBio technology alone, it contained numerous errors (frameshifted proteins), resulting in exclusion from the RefSeq database. For fine-scale comparative analysis of closely related *T. thermophilus* genomes, here, we reinvestigated the genome of TMY by combining Oxford Nanopore Technologies (ONT) long-read and MGI short-read sequencing technologies.

Freeze-dried TMY cells (JCM 10668) obtained from JCM (RIKEN, Japan) were inoculated into 5 ml of *Thermus* medium (ATCC 697), containing 0.4 mM MgCl₂ and 0.35 mM CaCl₂. After 24 h of cultivation at 70°C, genomic DNA was purified from pelleted cells using a blood and cell culture DNA midi kit (Qiagen). For long-read sequencing, unsheared genomic DNA (1 μg) was pretreated using a short-read eliminator kit (Circulomics) to remove fragments of <10 Kbp, and a library was constructed using a ligation sequencing kit (ONT). Sequencing was performed using a GridION X5 system on a FLO-MIN106 R9.41 flow cell (ONT). Base calling was conducted using Guppy v.4.0.11 to generate 375,245 reads (average, 4,578 bases; total, 1.72 Gb). For all software, default parameters were used. The raw sequencing data were filtered (Q < 10; length, <1,000 bases) using NanoFilt v.2.7.1 (19), yielding 213,707 reads (longest read, 163,227 bases; N₅₀, 9,231 bases; total, 1.15 Gb). For short-read sequencing, a library was constructed using an MGIEasy FS PCR-free DNA library prep set (MGI) with a ~400 to 500-bp insert. Paired-end sequencing was then performed on a DNBSEQ-400 instrument (MGI), yielding 8,454,552 paired-end reads (2 × 150 bases). The raw sequencing data were filtered (Q < 30; length, <10 bases) using fastp v.0.20.1 (20), yielding 5,182,387 paired-end reads (average, 150 bases; total, 2.53 Gb). The trimmed long- and short-read data were assembled using Unicycler v.0.4.8 (21), and the assembly was polished using Pilon v.1.23 (22), generating a single circular chromosome and a single circular plasmid. Automatic annotation was conducted using DFAST v.1.4.0 (23), and the genomic features are summarized in Table 1.

Editor Frank J. Stewart, Montana State University

Copyright © 2021 Miyazaki and Tokito. 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 Kentaro Miyazaki, miyazaki-kentaro@aist.go.jp.

Received 1 October 2021

Accepted 2 November 2021

Published 18 November 2021

TABLE 1 Genome statistics and features of *Thermus thermophilus* strain TMY

Genetic element	Length (bp)	GC content (%)	No. of coding DNA sequences	No. of rRNAs	No. of tRNAs	GenBank accession no.
Chromosome	2,151,326	69.0	2,309	6	52	AP025158
Plasmid	19,144	67.4	26	0	1	AP025159

Comparative sequence analysis between the previously reported sequence (chromosome, GenBank accession number [AP017920.1](#); plasmid pTMY, [AP017921.1](#)) (18) and the present result revealed 98.2% pairwise identity. The majority of differences were single-nucleotide deletions in the PacBio sequence, while the presence of two large insertions (~30 Kbp total) in our sequence was experimentally confirmed by Sanger sequencing.

Data availability. The complete genome sequence of *T. thermophilus* TMY is available from DDBJ/EMBL/GenBank under the accession numbers summarized in Table 1. The raw sequencing data were deposited in the SRA database under the accession numbers [DRR313875](#) (Nanopore) and [DRR313876](#) (DNBSEQ).

ACKNOWLEDGMENT

This work was partly supported by the following grants awarded to K.M. from the Japan Society for the Promotion of Science (JSPS): Grant-in-Aid for Scientific Research (A) (19H00936) and Grant-in-Aid for Challenging Research (Pioneering) (19H05538).

REFERENCES

- Oshima T, Imahori K. 1971. Isolation of an extreme thermophile and thermostability of its transfer ribonucleic acid and ribosomes. *J Gen Appl Microbiol* 17:513–517. <https://doi.org/10.2323/jgam.17.513>.
- Oshima T, Imahori K. 1974. Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. *Int J Sys Evol Microbiol* 24:102–112. <https://doi.org/10.1099/00207713-24-1-102>.
- Manaia CM, Hoste B, Gutierrez MC, Gillis M, Ventosa A, Kersters K, Da Costa MS. 1995. Halotolerant *Thermus* strains from marine and terrestrial hot springs belong to *Thermus thermophilus* (ex Oshima and Imahori, 1974) nom. rev. emend. *Syst Appl Microbiol* 17:526–532. [https://doi.org/10.1016/S0723-2020\(11\)80072-X](https://doi.org/10.1016/S0723-2020(11)80072-X).
- Beffa T, Blanc M, Lyon PF, Vogt G, Marchiani M, Fischer JL, Aragno M. 1996. Isolation of *Thermus* strains from hot composts (60 to 80 degrees C). *Appl Environ Microbiol* 62:1723–1727. <https://doi.org/10.1128/aem.62.5.1723-1727.1996>.
- Lyon PF, Beffa T, Blanc M, Auling G, Aragno M. 2000. Isolation and characterization of highly thermophilic xylanolytic *Thermus thermophilus* strains from hot composts. *Can J Microbiol* 46:1029–1035. <https://doi.org/10.1139/w00-075>.
- Fujino Y, Kawatsu R, Inagaki F, Umeda A, Yokoyama T, Okaue Y, Iwai S, Ogata S, Ohshima T, Doi K. 2008. *Thermus thermophilus* TMY isolated from silica scale taken from a geothermal power plant. *J Appl Microbiol* 104:70–78. <https://doi.org/10.1111/j.1365-2672.2007.03528.x>.
- Miyazaki K. 2019. Complete genome sequencing of *Thermus thermophilus* strain HC11, isolated from Mine Geyser in Japan. *Microbiol Resour Announc* 8:e00873-19. <https://doi.org/10.1128/MRA.00873-19>.
- Miyazaki K, Tomariguchi N. 2019. Complete genome sequences of *Thermus thermophilus* strains AA2-20 and AA2-29, isolated from Arima Onsen in Japan. *Microbiol Resour Announc* 8:e00820-19. <https://doi.org/10.1128/MRA.00820-19>.
- Miyazaki K, Moriya T, Nemoto N, Oshima T, Yura K, Bessho Y. 2021. Complete genome sequence of *Thermus thermophilus* strain HB5018, isolated from Mine Hot Spring in Japan. *Microbiol Resour Announc* 10:e00039-21. <https://doi.org/10.1128/MRA.00039-21>.
- Miyazaki K, Moriya T, Tokito N, Oshima T, Yura K, Bessho Y. 2021. Complete genome sequences of *Thermus thermophilus* strains HB5002 and HB5008, isolated from Mine Hot Spring in Japan. *Microbiol Resour Announc* 10:e00272-21. <https://doi.org/10.1128/MRA.00272-21>.
- Miyazaki K, Tomariguchi N, Ueno Y. 2021. Complete genome sequences of four halophilic *Thermus thermophilus* strains isolated from Arima Hot Spring in Japan. *Microbiol Resour Announc* 10:e00874-21. <https://doi.org/10.1128/MRA.00874-21>.
- Miyazaki K. 2005. A hyperthermophilic laccase from *Thermus thermophilus* HB27. *Extremophiles* 9:415–425. <https://doi.org/10.1007/s00792-005-0458-z>.
- Yamada T, Akutsu N, Miyazaki K, Kakinuma K, Yoshida M, Oshima T. 1990. Purification, catalytic properties, and thermal stability of threo-Ds-3-isopropylmalate dehydrogenase coded by *leuB* gene from an extreme thermophile, *Thermus thermophilus* strain HB8. *J Biochem* 108:449–456. <https://doi.org/10.1093/oxfordjournals.jbchem.a123220>.
- Imada K, Sato M, Tanaka N, Katsube Y, Matsuura Y, Oshima T. 1991. Three-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of *Thermus thermophilus* at 2.2 Å resolution. *J Mol Biol* 222:725–738. [https://doi.org/10.1016/0022-2836\(91\)90508-4](https://doi.org/10.1016/0022-2836(91)90508-4).
- Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, Kyogoku Y, Miki K, Masui R, Kuramitsu S. 2000. Structural genomics projects in Japan. *Nat Struct Biol* 7:943–945. <https://doi.org/10.1038/80712>.
- Tamakoshi M, Yaoi T, Oshima T, Yamagishi A. 1999. An efficient gene replacement and deletion system for an extreme thermophile, *Thermus thermophilus*. *FEMS Microbiol Lett* 173:431–437. <https://doi.org/10.1111/j.1574-6968.1999.tb13535.x>.
- Miyazaki K, Tomariguchi N. 2019. Occurrence of randomly recombined functional 16S rRNA genes in *Thermus thermophilus* suggests genetic interoperability and promiscuity of bacterial 16S rRNAs. *Sci Rep* 9:12333. <https://doi.org/10.1038/s41598-019-47807-z>.
- Fujino Y, Nagayoshi Y, Ohshima T, Ogata S, Doi K. 2017. Complete genome sequence of *Thermus thermophilus* TMY, isolated from a geothermal power plant. *Genome Announc* 5:e01596-16. <https://doi.org/10.1128/genomeA.01596-16>.
- De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
- 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>.
- Tanizawa Y, Fujisawa T, Kaminuma E, Nakamura Y, Arita M. 2016. DFAST and DAGA: Web-based integrated genome annotation tools and resources. *Biosci Microbiota Food Health* 35:173–184. <https://doi.org/10.12938/bmfh.16-003>.