# Complete Genome Sequence of a Hyperthermophilic Archaeon, *Thermosphaera* sp. Strain 3507, Isolated from a Chilean Hot Spring

Kseniya S. Zayulina,ᵃ Alexander G. Elcheninov,ᵃ Stepan V. Toshchakov,ᵇ Ilya V. Kublanovᵃ

ᵃWinogradsky Institute of Microbiology, Research Center of Biotechnology RAS, Moscow, Russia
ᵇNational Research Center Kurchatov Institute, Moscow, Russia

**ABSTRACT** A complete genome sequence of a hyperthermophilic archaeon, *Thermosphaera* sp. strain 3507, which was isolated from a Chilean hot spring, is presented. The genome is 1,305,106 bp with a G+C content of 47.6%. Twenty-seven carbohydrate-active enzyme genes were identified, which is in accordance with the ability of the strain to grow on various polysaccharides.

The genus *Thermosphaera* is affiliated with the *Desulfurococcaceae* family (1) of the *Crenarchaeota* phylum and is currently represented by a single species, *Thermosphaera aggregans* M11TLᵀ (2). Strain 3507 was isolated from a sample of mud and water collected from a hot spring (temperature, 83°C [pH 6.3]; 34°57.518′S, 70°26.331′W) located in the Termas del Flaco area within the Tinguiririca volcano thermal zone in Chile (3). Strain 3507 was isolated by a serial dilution technique from a binary enrichment culture obtained by incubation of the sample for 7 days at 85°C in anaerobic Pfennig medium (4) with twice-reduced salt concentrations and supplemented with lichenan (1 g liter⁻¹) (pH 6.5).

For genomic sequencing, the strain was cultured for 3 days at 85°C at pH 6.5 in medium (4) supplemented with lichenan (1 g liter⁻¹). Genomic DNA isolation was performed using a Genomic-tip 20/G (Qiagen), according to the manufacturer's instructions. Approximately 100 ng of isolated DNA was used for fragment library preparation with the Nextera DNA Flex library preparation kit (Illumina), according to the manufacturer's protocol. The library was sequenced with the Illumina MiSeq system, using a $2 \times 150$-bp sequencing kit; 1,159,189 read pairs were obtained from a sequencing run. Reads were subjected to quality filtering and trimming with the Trim Reads tool of CLC Genomics Workbench v20.0.4 (Qiagen), using zero maximum ambiguities and 0.01 error probability. Trimming of sequencing adapters and merging of overlapping read pairs were performed with the SeqPrep tool (https://github.com/jstjohn/SeqPrep). A total of 710,804 read pairs and 396,658 merged reads were used for *de novo* assembly with the SPAdes v3.14.1 assembler in the "--isolate" mode (5). One contig of 1,306,603-bp length was obtained. Circularization was performed by broken read pair analysis with CLC Genomics Workbench v20.0.4 and the CLC Genome Finishing Module (Qiagen). Finally, one circular 1,305,106-bp chromosome was obtained. The start of the chromosome was set to the origin of replication predicted by the OriFinder 2 tool (6). Genome annotation was performed with PGAP (7). The average amino acid identity (AAI) and average nucleotide identity (ANI) were calculated using the AAI.rb script (8) and the pyani module v0.2.8 (9), respectively. Carbohydrate-active enzymes (CAZymes) were identified with dbCAN2 (10). Amino acid biosynthetic pathways were predicted by GapMind (11).

The final assembly of the strain 3507 genome comprises a single circular chromosome with a length of 1,305,106 bp and a G+C content of 47.6%. In total, 1,458 genes

were predicted, including 1,399 protein-coding genes, 50 RNA genes (3 rRNA genes, 45 tRNA genes, and 2 noncoding RNA genes), and 9 pseudogenes. A BLAST search revealed 99.67% 16S rRNA sequence identity with *Thermosphaera aggregans* M11TL$^T$; however, pairwise AAI and ANI values were 86.5% and 83.2%, respectively, which are below the species-level thresholds. The capability of polysaccharide utilization by the strain is in accordance with the presence of the CAZyme genes in its genome. The CAZymes were represented by 10 glycosidases (glycoside hydrolase 1 [GH1], GH13, GH57, and GH122 families) and 17 glycosyltransferases; 13 of them are predicted to be secreted. The genome analysis revealed probable arginine, histidine, lysine, methionine, proline, serine, branched-chain amino acid, and aromatic amino acid auxotrophy.

**Data availability.** The whole-genome sequence was deposited in DDBJ/ENA/GenBank under the accession number CP063144.1. The BioProject, BioSample, and SRA accession numbers are PRJNA668939, SAMN16428067, and SRR12969965, respectively.

## REFERENCES

1. Huber R, Dyba D, Huber H, Burggraf S, Rachel R. 1998. Sulfur-inhibited *Thermosphaera aggregans* sp. nov., a new genus of hyperthermophilic archaea isolated after its prediction from environmentally derived 16S rRNA sequences. Int J Syst Bacteriol 48:31–38. https://doi.org/10.1099/00207713-48-1-31.
2. Zillig W, Stetter KO, Prangishvilli D, Schäfer W, Wunderl S, Janekovic D, Holz I, Palm P. 1982. *Desulfurococcaceae*, the second family of the extremely thermophilic, anaerobic, sulfur-respiring *Thermoproteales*. Zentralbl Bakteriol Mikrobiol Hyg C 3:304–317. https://doi.org/10.1016/S0721-9571(82)80044-6.
3. Clavero J, Pineda G, Mayorga C, Giavelli A, Aguirre I, Simmons S, Martini S, Soffia J, Arriaza R, Polanco Valenzuela E, Achurra L. 2011. Geological, geochemical, geophysical and first drilling data from Tinguiririca Geothermal Area, Central Chile. Geotherm Resour Counc Trans 35:731–734.
4. Kochetkova TV, Mardanov AV, Sokolova TG, Bonch-Osmolovskaya EA, Kublanov IV, Kevbrin VV, Beletsky AV, Ravin NV, Lebedinsky AV. 2020. The first crenarchaeon capable of growth by anaerobic carbon monoxide oxidation coupled with H$_2$ production. Syst Appl Microbiol 43:126064. https://doi.org/10.1016/j.syapm.2020.126064.
5. Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol 20:714–737. https://doi.org/10.1089/cmb.2013.0084.
6. Luo H, Zhang CT, Gao F. 2014. Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. Front Microbiol 5:482. https://doi.org/10.3389/fmicb.2014.00482.
7. 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.
8. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr 4:e1900v1. https://doi.org/10.7287/peerj.preprints.1900v1.
9. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods 8:12–24. https://doi.org/10.1039/C5AY02550H.
10. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418.
11. Price MN, Deutschbauer AM, Arkin AP. 2020. GapMind: automated annotation of amino acid biosynthesis. mSystems 5:e00291-20. https://doi.org/10.1128/mSystems.00291-20.