AMERICAN SOCIETY FOR MICROBIOLOGY genomeAnnouncements™

Complete Genome Sequence of Hyperthermophilic Archaeon *Thermococcus* sp. EXT12c, Isolated from the East Pacific Rise 9°N

Damien Courtine,[a,b,c] Karine Alain,[a,b,c] Myriam Georges,[a,b,c] Nadège Bienvenu,[a,b,c] Hilary G. Morrison,[d] A. Murat Eren,[d,e] Loïs Maignien[a,b,c,d]

Université de Bretagne Occidentale (UBO, UBL), Institut Universitaire Européen de la Mer (IUEM), UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes (LM2E), Plouzané, France[a]; CNRS, IUEM–UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes (LM2E), Plouzané, France[b]; Ifremer, UMR 6197, Laboratoire de Microbiologie des Environnements Extrêmes (LM2E), Plouzané, France[c]; Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts, USA[d]; Department of Medicine, University of Chicago, Chicago, Illinois, USA[e]

ABSTRACT We report the genome sequence of *Thermococcus* sp. EXT12c isolated from a deep-sea hydrothermal vent at the East Pacific Rise 9°N. Microbes in the genus *Thermococcus* are able to grow anaerobically at high temperature, around neutral pH, and some of them under high hydrostatic pressure.

We isolated *Thermococcus* sp. EXT12c from a hydrothermal chimney rock sample collected from a 2496-m depth near the East Pacific Rise 9°N (9°50′40.2″N, 104°17′37.798″W) during the oceanographic cruise EXTREME (October 2001). *T.* sp. EXT12c is able to grow under anaerobic conditions, at 85°C and pH 6.8 in TRM medium (1). The strain is available in the UBOCC culture collection (Brest, France) under the reference no. UBOCC-M-2417.

We used a phenol-chloroform technique for DNA extraction and the TruSeq DNA PCR-free kit (Illumina, USA) to prepare paired-end sequencing libraries with an average insert size of 550 nt. Whole-genome sequencing at the Marine Biological Laboratory (Woods Hole, MA, USA) using an Illumina MiSeq machine (MiSeq reagent kit v3) produced 1,335,523 2 × 300 bp reads after quality filtering (2). Our *de novo* assembly with CLC Genomics Workbench v8.5.1 (https://www.qiagenbioinformatics.com/products/clc-genomics-workbench) resulted in a single chromosome with 2,155,760 nt and a GC content of 54.58%. This single chromosome recruited 98.8% of the short reads, with an average coverage of 350×.

The MaGe genome annotation platform (3–14) identified 2,365 coding sequences; a single 16S-23S operon; and 2 5S rRNA, 46 tRNA, and 16 miscellaneous RNA genes. InterProScan identified 1 integrase, 6 transposases, and 3 clustered regularly interspaced short palindromic repeat (CRISPR) loci associated with *cas* genes (*cas*, *cst*, and *cmr*), suggesting that the strain probably carries two types of CRISPR systems, class I type I and type III (15). These features suggest that the strain has a certain genomic plasticity.

Among the *Thermococcus* species with published genomes, *T.* sp. EXT12c is most closely related to *T. nautili* strain 30-1[T] (16). These genomes have a DNA-DNA hybridization value of 43.6% and an average nucleotide identity of 91.18%, as predicted with GGDC v2.1 (17–19) and OrthoANI v1.20 (20), respectively.

T. sp. EXT12c possesses some complete metabolic pathways like the glycolysis and amino acid biosynthesis pathways for alanine, asparagine, glycine, glutamate, and tryptophan. To date, only ten *Thermococcales*, including *T. kodakaraensis*, *T. litoralis*,

Pyrococcus furiosus, and *P. abyssi* (21–25), harbor a complete tryptophan biosynthesis pathway.

In this metabolic pathway, a single locus contains genes leading to the synthesis of chorismate, an intermediate for multiple metabolic pathways (TEXT12C_2159 to TEXT12C_2167), and tryptophan (TEXT12C_2174 to TEXT12C_2168, genes *trpCDEGFBA*). Within the *T. kodakaraensis* genome, genes that code for tryptophan biosynthesis are present in a single locus too (26). A transcriptomic study showed that the chorismate synthesis is downregulated when *T. kodakaraensis* is cultivated under high hydrostatic pressure, compared to atmospheric pressure (27). However, in the same study, the gene *trpC*, labeled TK0252 in *T. kodakaraensis*, is upregulated, indicating that the strain could continue to produce tryptophan and compensate the decrease of chorismate production under high pressure. Therefore, the regulation of tryptophan biosynthesis in *T.* sp. EXT12c under high-pressure conditions requires further investigations.

Accession number(s). This genome sequence has been deposited in DDBJ/ENA/GenBank under the accession no. LT900021. The version described in this paper is the first version.

REFERENCES

1. Zeng X, Birrien J-L, Fouquet Y, Cherkashov G, Jebbar M, Querellou J, Oger P, Cambon-Bonavita M-A, Xiao X, Prieur D. 2009. *Pyrococcus* CH1, an obligate piezophilic hyperthermophile: extending the upper pressure-temperature limits for life. ISME J 3:873–876. https://doi.org/10.1038/ismej.2009.21.

2. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML. 2013. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol 4:1111–1119.

3. Bocs S, Cruveiller S, Vallenet D, Nuel G, Médigue C. 2003. AMIGene: annotation of MIcrobial genes. Nucleic Acids Res 31:3723–3726. https://doi.org/10.1093/nar/gkg590.

4. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. 2009. Rfam: updates to the RNA families database. Nucleic Acids Res 37:D136–D140. https://doi.org/10.1093/nar/gkn766.

5. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.

6. Lowe TM, Eddy SR. 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964.

7. Claudel-Renard C, Chevalet C, Faraut T, Kahn D. 2003. Enzyme-specific profiles for genome annotation: PRIAM. Nucleic Acids Res 31:6633–6639. https://doi.org/10.1093/nar/gkg847.

8. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795. https://doi.org/10.1016/j.jmb.2004.05.028.

9. Gardy JL, Laird MR, Chen F, Rey S, Walsh CJ, Ester M, Brinkman FSL. 2005. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics 21:617–623. https://doi.org/10.1093/bioinformatics/bti057.

10. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJA, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. 2009. InterPro: the integrative protein signature database. Nucleic Acids Res 37:D211–D215. https://doi.org/10.1093/nar/gkn785.

11. Karp PD, Paley S, Romero P. 2002. The Pathway Tools software. Bioinformatics 18:S225–S232. https://doi.org/10.1093/bioinformatics/18.suppl_1.S225.

12. Sonnhammer EL, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6:175–182.

13. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4:41. https://doi.org/10.1186/1471-2105-4-41.

14. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh L-SL. 2005. The universal protein resource (UniProt). Nucleic Acids Res 33:D154–D159. https://doi.org/10.1093/nar/gki070.

15. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJ, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin EV. 2015. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol 13:722–736. https://doi.org/10.1038/nrmicro3569.

16. Oberto J, Gaudin M, Cossu M, Gorlas A, Slesarev A, Marguet E, Forterre P. 2014. Genome sequence of a hyperthermophilic archaeon, *Thermococcus nautili* 30–1, that produces viral vesicles. Genome Announc 2(2):e00243-14. https://doi.org/10.1128/genomeA.00243-14.

17. 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 Genomic Sci 2:117–134. https://doi.org/10.4056/sigs.531120.

18. Auch AF, Klenk H-P, Göker M. 2010. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand Genomic Sci 2:142–148. https://doi.org/10.4056/sigs.541628.

19. Meier-Kolthoff JP, Klenk H-P, Göker M. 2014. Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. Int J Syst Evol Microbiol 64:352–356. https://doi.org/10.1099/ijs.0.056994-0.

20. Lee I, Kim YO, Park S-C, Chun J. 2015. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. https://doi.org/10.1099/ijsem.0.000760.

21. Maeder DL, Weiss RB, Dunn DM, Cherry JL, González JM, DiRuggiero J,

Robb FT. 1999. Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. Genetics 152:1299–1305.

22. Bridger SL, Lancaster WA, Poole FL, Schut GJ, Adams MWW. 2012. Genome sequencing of a genetically tractable *Pyrococcus furiosus* strain reveals a highly dynamic genome. J Bacteriol 194:4097–4106. https://doi.org/10.1128/JB.00439-12.

23. Fukui T, Atomi H, Kanai T, Matsumi R, Fujiwara S, Imanaka T. 2005. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. Genome Res 15:352–363. https://doi.org/10.1101/gr.3003105.

24. Gardner AF, Kumar S, Perler FB. 2012. Genome sequence of the model hyperthermophilic archaeon *Thermococcus litoralis* NS-C. J Bacteriol 194:2375–2376. https://doi.org/10.1128/JB.00123-12.

25. Cohen GN, Barbe V, Flament D, Galperin M, Heilig R, Lecompte O, Poch O, Prieur D, Quérellou J, Ripp R, Thierry J-C, Van der Oost J, Weissenbach J, Zivanovic Y, Forterre P. 2003. An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. Mol Microbiol 47:1495–1512. https://doi.org/10.1046/j.1365-2958.2003.03381.x.

26. Tang X, Ezaki S, Fujiwara S, Takagi M, Atomi H, Imanaka T. 1999. The tryptophan biosynthesis gene cluster trpCDEGFBA from *Pyrococcus kodakaraensis* KOD1 is regulated at the transcriptional level and expressed as a single mRNA. Mol Gen Genet 262:815–821. https://doi.org/10.1007/s004380051145.

27. Vannier P, Michoud G, Oger P, Marteinsson VÞ, Jebbar M. 2015. Genome expression of *Thermococcus barophilus* and *Thermococcus kodakarensis* in response to different hydrostatic pressure conditions. Res Microbiol 166:717–725. https://doi.org/10.1016/j.resmic.2015.07.006.