

Complete genome sequence of *Thermaerobacter composti* strain Ins1, a spore-forming filamentous bacterium isolated from a deep geothermal reservoir

Danae Bregnard,¹ Diego Gonzalez,¹ Eva Di Francesco,¹ Naïma Mangia,¹ Simona Regenspurg,² Pilar Junier¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the complete genome sequence of *Thermaerobacter composti* strain Ins1, a gram-positive filamentous spore-forming bacterium, isolated from deep geothermal fluids used for electricity production. This is the first complete (circular) genome assigned to the species *Thermaerobacter composti*.

KEYWORDS complete genome, geothermal energy, bacteria

Members of the *Thermaerobacter* genus are thermophilic bacteria isolated from high-temperature environments including compost (*Thermaerobacter composti* JCM 15650 strain Ni80^T (1), the Mariana Trench (*Thermaerobacter marianensis* 7p75a^T) (2), shallow hydrothermal vents (*Thermaerobacter nagasakiensis* Ts1a^T) (3), a bore run-off channel (*Thermaerobacter subterraneus* C21^T) (4), and a hydrothermal beach (*Thermaerobacter litoralis* KW1^T) (5). Strain Ins1 was isolated from fluids of the production well of the deep geothermal reservoir (163°C; 3,650 m depth) of an electricity production plant in the Upper Rhine Graben (49.1571411 N, 8.1480501 E). After enrichment in Marine Broth 2216 (BD Difco, USA) at 60°C without agitation, only one bacterium grew upon isolation on solid Marine Broth 2216 (1.5% agar).

Genomic DNA was extracted from a 10-day-old culture using the Wizard HMW DNA Extraction Kit (Promega, USA). Sequencing was performed on a Pacific Biosciences Sequel II instrument after DNA shearing, size selection, and HiFi SMRTbell Library construction (chemistry version: 11.1.0.154383; average insert size: 7,718 bp). A genome was assembled from 34,272 HiFi reads using *flye* (v. 2.9.2) (6) with 2,000 bp minimum overlap and three polishing rounds (average coverage 89×); the chromosome was circularized by *flye* based on the assembly graph. One consistently methylated motif (GC^{m6}ANNNNNNGTT/A^{m6}ACNNNNNNTGC, >99% detection rate) was identified by *Smrtlink* (v. 11). Assembly quality was assessed using *CheckM* (estimated completeness: 97.03%; contamination rate: 0.53%). The circular assembled genome corresponded to 3,001,239 bp with a 73% G + C content.

The genome was annotated using NCBI Prokaryotic Genome Annotation Pipeline (PGAP v.6.5). It included 2,363 protein-coding genes, 2 copies of each ribosomal RNA gene, 47 tRNAs, and 4 non-coding RNAs (signal recognition particle sRNA, 6S RNA, RNA component of RNase P, and the transfer-messenger RNA); 54 additional putative protein-coding genes were predicted to be non-functional (pseudogenes).

Strain Ins1 was assigned to the *Thermaerobacter composti* species. The 16S rRNA gene was 99.94% identical with the sequence of *T. composti* JCM 15650 based on pairwise alignment (NCBI *blast*). Digital DNA-DNA hybridization (<https://ggdc.dsmz.de>, dDDH, formula 2) was 86.2%, which falls within the limits of the same bacterial species (7). A phylogenetic tree based on 28 concatenated ribosomal/conserved proteins (elongation factor G/elongation factor Tu/L14/L15/L16/L17/L18/L2/L22/L23/L24/L29/L3/L4/L5/L6/

Editor Elinne Becket, California State University San Marcos, San Marcos, California, USA

Address correspondence to Pilar Junier, pilarjunier@unine.ch.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 18 January 2024

Accepted 2 March 2024

Published 13 March 2024

Copyright © 2024 Bregnard et al. 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/).

S10/S11/S12/S13/S14/S17/S19/S3/S4/S5/S7/S8) aligned using *muscle* (v. 3.8.31) (8) was constructed by *iqtree* (v. 2.2.5) (9) with automated model selection, confirming that strain Ins1 clustered within the *Thermaerobacter* genus, closest to strain FW80 (10).

The amino acid composition of the full proteome showed a limited ERK bias around −15 (11), suggesting that the strain is moderately thermophilic [50°C–60°C predicted optimal growth temperature; 3.08 hours doubling time based on ribosomal protein genes, by *gRodon* v. 2.3.0 (12)]. Presence in geothermal water is likely due to the production of heat-resistant endospores. Accordingly, the genome encodes 58 out of 66 proteins considered essential for sporulation in Clostridiales (13), and endospores were observed in old cultures (1). Metabolic pathway analysis using *gapseq* v. 1.2 (14) suggested that the strain requires exogenous thiamine and may be auxotrophic for several amino acids (Try, His, Ser, Phe, Pro, Thr, Tyr, Val).

ACKNOWLEDGMENTS

The authors want to thank Vulcan Energie Ressourcen GmbH (DE) for providing access to the site and to Hydroisotop GmbH (DE) for help during sampling.

This work was supported by the European Union's Horizon 2020 research and innovation program within the framework of the REFLECT project under the grant agreement No 850626.

AUTHOR AFFILIATIONS

¹Laboratory of Microbiology, University of Neuchâtel, Neuchâtel, Switzerland

²Geoenergy, GFZ German Research Centre for Geosciences, Potsdam, Germany

AUTHOR ORCIDs

Danae Bregnard  <http://orcid.org/0000-0002-4399-5453>

Pilar Junier  <http://orcid.org/0000-0002-8618-3340>

FUNDING

Funder	Grant(s)	Author(s)
EC Horizon 2020 Framework Programme (H2020)	850626	Danae Bregnard Diego Gonzalez Eva Di Francesco Naïma Mangia Simona Regenspurg Pilar Junier

AUTHOR CONTRIBUTIONS

Danae Bregnard, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing | Diego Gonzalez, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review and editing | Eva Di Francesco, Investigation, Writing – review and editing | Naïma Mangia, Investigation, Writing – review and editing | Simona Regenspurg, Funding acquisition, Writing – review and editing | Pilar Junier, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

This whole-genome shotgun project has been deposited at NCBI under the accession [CP132508](https://www.ncbi.nlm.nih.gov/submit/seq/sra/); SRAs are available under accession [SRR27458461](https://www.ncbi.nlm.nih.gov/submit/seq/sra/).

REFERENCES

1. Yabe S, Kato A, Hazaka M, Yokota A. 2009. *Thermaerobacter composti* sp. nov., a novel extremely thermophilic bacterium isolated from compost. *J Gen Appl Microbiol* 55:323–328. <https://doi.org/10.2323/jgam.55.323>
2. Takai K, Inoue A, Horikoshi K. 1999. *Thermaerobacter marianensis* gen. nov., sp. nov., an aerobic extremely thermophilic marine bacterium from the 11000 m deep Mariana Trench. *Int J Syst Bacteriol* 49:619–628. <https://doi.org/10.1099/00207713-49-2-619>
3. Nunoura T, Akihara S, Takai K, Sako Y. 2002. *Thermaerobacter nagasakiensis* sp. nov., a novel aerobic and extremely thermophilic marine bacterium. *Arch Microbiol* 177:339–344. <https://doi.org/10.1007/s00203-002-0398-2>
4. Spanevello MD, Yamamoto H, Patel BKC. 2002. *Thermaerobacter subterraneus* sp. nov., a novel aerobic bacterium from the Great Artesian Basin of Australia, and emendation of the genus *Thermaerobacter*. *Int J Syst Evol Microbiol* 52:795–800. <https://doi.org/10.1099/00207713-52-3-795>
5. Tanaka R, Kawaichi S, Nishimura H, Sako Y. 2006. *Thermaerobacter litoralis* sp. nov., a strictly aerobic and thermophilic bacterium isolated from a coastal hydrothermal field. *Int J Syst Evol Microbiol* 56:1531–1534. <https://doi.org/10.1099/ijs.0.64203-0>
6. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>
7. Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. 2022. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 50:D801–D807. <https://doi.org/10.1093/nar/gkab902>
8. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
9. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
10. Nam JH, KimTU, LeeDH. 2019. Complete genome sequence of *Thermaerobacter* sp. FW80, an aerobic extremely thermophilic bacterium isolated from a food waste compost. *Microbiol Soc Korea* 55:462–464. <https://doi.org/10.7845/kjm.2019.9148>
11. Wang GZ, Lercher MJ. 2010. Amino acid composition in endothermic vertebrates is biased in the same direction as in thermophilic prokaryotes. *BMC Evol Biol* 10:263. <https://doi.org/10.1186/1471-2148-10-263>
12. Weissman JL, Hou S, Fuhrman JA. 2021. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc Natl Acad Sci U S A* 118:e2016810118. <https://doi.org/10.1073/pnas.2016810118>
13. Galperin MY, Mekhedov SL, Puigbo P, Smirnov S, Wolf YI, Rigden DJ. 2012. Genomic determinants of sporulation in *Bacilli* and *Clostridia*: towards the minimal set of sporulation-specific genes. *Environ Microbiol* 14:2870–2890. <https://doi.org/10.1111/j.1462-2920.2012.02841.x>
14. Zimmermann J, Kaleta C, Waschina S. 2021. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol* 22:81. <https://doi.org/10.1186/s13059-021-02295-1>