

Complete genome sequence of the thermophilic sulfur-reducer *Desulfurobacterium thermolithotrophum* type strain (BSA^T) from a deep-sea hydrothermal vent

Markus Göker¹, Hajnalka Daligault², Romano Mwirichia³, Alla Lapidus⁴, Susan Lucas⁴, Shweta Deshpande⁴, Ioanna Pagani⁴, Roxanne Tapia^{2,4}, Jan-Fang Cheng⁴, Lynne Goodwin^{2,4}, Sam Pitluck⁴, Konstantinos Liolios⁴, Natalia Ivanova⁴, Konstantinos Mavromatis⁴, Natalia Mikhailova⁴, Amrita Pati⁴, Amy Chen⁵, Krishna Palaniappan⁵, Cliff Han², Miriam Land^{4,6}, Loren Hauser^{4,6}, Chongle Pan^{4,6}, Evelyne-Marie Brambilla¹, Manfred Rohde⁷, Stefan Spring¹, Johannes Sikorski¹, Reinhard Wirth⁸, John C. Detter^{2,4}, Tanja Woyke⁴, James Bristow⁴, Jonathan A. Eisen^{4,9}, Victor Markowitz⁵, Philip Hugenholtz^{4,10}, Nikos C. Kyrpides⁴, and Hans-Peter Klenk^{1*}

¹ Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

² Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

³ Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

⁴ DOE Joint Genome Institute, Walnut Creek, California, USA

⁵ Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

⁶ Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

⁷ HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany

⁸ University of Regensburg, Microbiology – Archaeenzentrum, Regensburg, Germany

⁹ University of California Davis Genome Center, Davis, California, USA

¹⁰ Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

*Corresponding author: Hans-Peter Klenk

Keywords: anaerobic, thermophilic, neutrophilic, obligately chemolithoautotrophic, Gram-negative, marine, sulfur-reducing, *Desulfurobacteriaceae*, GEBA

Desulfurobacterium thermolithotrophum L'Haridon et al. 1998 is the type species of the genus *Desulfurobacterium* which belongs to the family *Desulfurobacteriaceae*. The species is of interest because it represents the first thermophilic bacterium that can act as a primary producer in the temperature range of 45-75 °C (optimum 70°C) and is incapable of growing under microaerophilic conditions. Strain BSA^T preferentially synthesizes high-melting-point fatty acids (C₁₈ and C₂₀) which is hypothesized to be a strategy to ensure the functionality of the membrane at high growth temperatures. This is the second completed genome sequence of a member of the family *Desulfurobacteriaceae* and the first sequence from the genus *Desulfurobacterium*. The 1,541,968 bp long genome harbors 1,543 protein-coding and 51 RNA genes and is a part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

Introduction

Strain BSA^T (= DSM 11699) is the type strain of the species *Desulfurobacterium thermolithotrophum*, which is the type species of its genus *Desulfurobacterium* [1], that currently consists of three validly named species [19]. The genus name is derived from the Latin words 'de' meaning 'from', 'sulfur', and 'bacterium' meaning 'a stick, staff', yielding the Neo-Latin word '*Desulfurobacterium*' meaning 'sulfur-reducing rod-shaped bacterium' [1]. The species epithet is derived from the latinized Greek

word 'thermê' meaning 'heat', the latinized Greek word 'lithos' meaning 'stone' and the latinized Greek word 'trophos' meaning 'feeder, rearer, one who feeds', yielding the Neo-Latin word 'thermolithotrophum' meaning 'referring to its thermophilic way of life and lithotrophic metabolism' [1,2]. Strain BSA^T was collected from the Snake Pit vent field on the mid Atlantic Ridge with the help of the submersible *Nautile* at a depth of 3,500 m [1]. Although it shares most features with other

members of the *Aquificales*, it is distinct in its inability to grow under microaerophilic conditions [1]. Strain BSA^T was the first non-hyperthermophilic primary producer isolated from deep-sea vents [1]. Here we present a summary classification and a set of features for *D. thermolithotrophum* strain BSA^T, together with the description of the complete genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of *D. thermolithotrophum* BSA^T was compared using NCBI BLAST [3,4] under default settings (e.g. considering only the high-scoring segment pairs (HSPs) from the best 250 hits) with the most recent release of the Greengenes database [5] and the relative frequencies of taxa and keywords (reduced to their stem [6] were determined, weighted by BLAST scores. The most frequently occurring genera were *Desulfurobacterium* (30.3%), *Thermoanaerobacter* (18.8%), *Thermovibrio* (14.2%), *Balnearium* (11.0%) and *Persephonella* (4.1%) (80 hits in total). Regarding the two hits to sequences from members of the species, the average identity within HSPs was 98.9%, whereas the average coverage by HSPs was 92.8%. Regarding the single hit to sequences from other members of the genus, the average identity within HSPs was 98.6%, whereas the average coverage by HSPs was 64.4%. Among all other species, the one yielding the highest score was "*Desulfurobacterium crinifex*" (AJ507320), which corresponded to an identity of 98.6% and HSP coverage of 64.4%. (Note that the Greengenes database uses the INSDC (= EMBL/NCBI/DDBJ) annotation, which is not an authoritative source for nomenclature or classification.) The highest-scoring environmental sequence was AF068800 ('hydrothermal vent clone VC2.1Bac24'), which showed an identity of 99.7% and an HSP coverage of 92.7%. The most frequently occurring keywords within the labels of all environmental samples which yielded hits were 'hydrotherm' (5.4%), 'vent' (4.9%), 'microbi' (3.6%), 'water' (2.9%) and 'deep' (2.0%) (167 hits in total). The most frequently occurring keyword within the labels of those environmental samples which yielded hits of a higher score than the highest scoring species was 'hydrotherm, vent' (50.0%) (1 hit in total).

Figure 1 shows the phylogenetic neighborhood of *D. thermolithotrophum* BSA^T in a 16S rRNA based tree. The sequences of the two identical 16S rRNA gene copies in the genome differ by two nucleotides from

the previously published 16S rRNA sequence (AJ001049).

The cells of strain BSA^T are small rods, about 1-2 µm long and 0.4-0.5 µm wide and occur singly or in pairs (Figure 2) [1]. Cells stain Gram-negative and are motile via three polar flagella; spores are not produced [1]. Strain BSA^T grows between 40 and 75°C with an optimum around 70°C, while no growth is detected at 37 or 80°C after 48 h incubation [1]. Growth occurs between pH 4.4 and 8, with an optimum around pH 6.25. No growth is detected at pH 3.7 or 8.5 after 48h incubation at 70°C [1]. Growth is observed in sea salts at concentrations ranging from 15 to 70g/l, with an optimum of approximately 35g/l (corresponding to 23 g NaCl/l [1]). No growth was observed in sea salts at concentrations of 10 and 80 g/l after 48 h incubation at 70°C [1]. Under optimal growth conditions (temperature, pH and NaCl), the doubling time of strain BSA^T is around 135 min [1]. Strain BSA^T is a strictly anaerobic chemolithotrophic organism that uses sulfur as an electron acceptor in the presence of H⁺ for growth [1]. It utilizes thiosulfate, sulfite and polysulfides as alternative electron acceptors with H₂ as an electron donor. Cysteine, nitrate or nitrite are not utilized and growth on sulfur, thiosulfate, polysulfides or sulfite was accompanied by exponential H₂S production [1]. No growth was observed on acetate, formate, methanol, monomethylamine and yeast extract with N₂-CO₂ or H₂ atmosphere in the presence or absence of sulfur [1]. Nitrate, tryptone and yeast extract were used as nitrogen sources [1]. Growth of strain BSA^T was inhibited by chloramphenicol, penicillin G and rifampicin at 100 µg/ml but not by streptomycin when added before incubation at the optimum temperature [1].

Chemotaxonomy

The total lipid content of strain BSA^T is about 6% of the total dry weight and is characterized by the presence of aminophospholipids and a phospholipid at about 66%, R_f 0.7 and 30%, R_f 0.5, respectively, as well as minor compounds [1]. Gas chromatographic analysis of fatty acid components of both compounds revealed the presence of saturated and monounsaturated acyl chains [1]. The phosphoinositol contains C_{16:0} (15%), C_{18:1} (41%) identified as methyl-oleate, and C_{18:0} (44%) identified as stearate. The phosphoamino-positive compounds contained C_{16:0} (14%), C_{18:1} (43%), C_{18:0} (31%) and C_{20:0} (12%), as well as minor compounds [1].

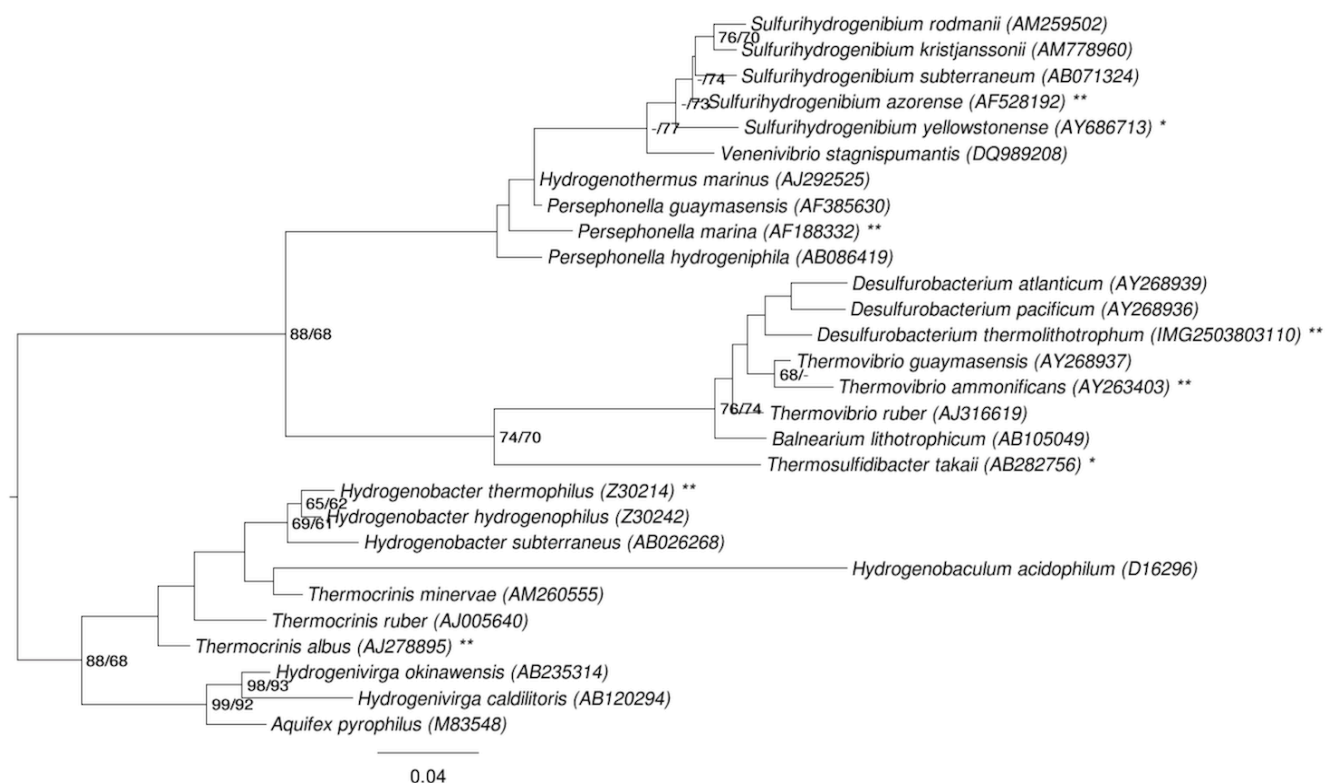


Figure 1 Phylogenetic tree highlighting the position of *D. thermolithotrophum* relative to the type strains of the other species within the order *Aquificales*. The tree was inferred from 1,422 aligned characters [7,8] of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion [9]. Rooting was done initially using the midpoint method [10] and then checked for its agreement with the current classification (Table 1). The branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches are support values from 1,000 ML bootstrap replicates [11] (left) and from 1,000 maximum parsimony bootstrap replicates [12] (right) if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [13] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks (referenced in [14-17] and CPO02444).

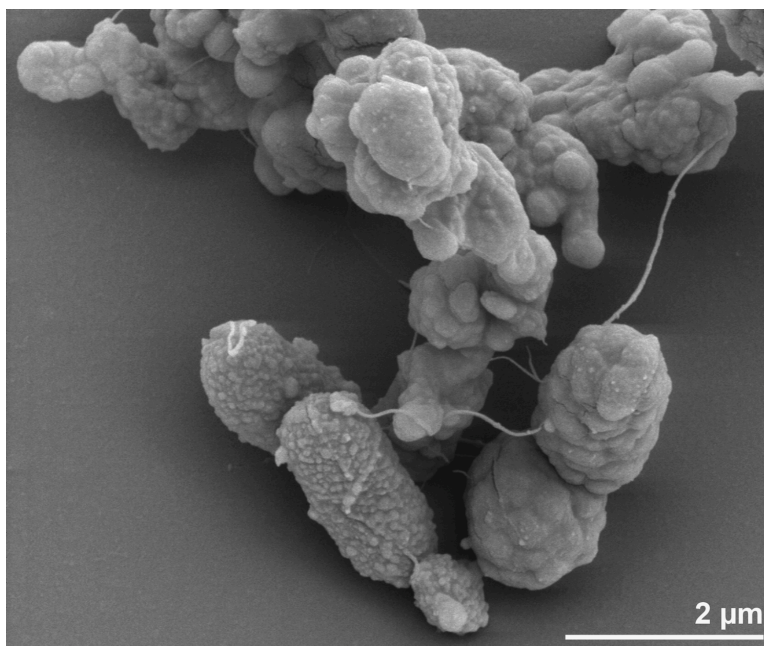


Figure 2. Scanning electron micrograph of *D. thermolithotrophum* BSA^T

Table 1. Classification and general *D. thermolithotrophum* BSA^T in accordance with the MIGS recommendations [18] and the NamesforLife database [19].

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [20]
		Phylum ' <i>Aquificae</i> '	TAS [22]
	Current classification	Class <i>Aquificae</i>	TAS [23,24]
		Order <i>Aquificales</i>	TAS [23,25,26]
		Family <i>Desulfurobacteriaceae</i>	TAS [25]
		Genus <i>Desulfurobacterium</i>	TAS [1,25,27]
		Species	<i>Desulfurobacterium thermolithotrophum</i>
MIGS-7	Strain	BSA ^T	TAS [1]
MIGS-12	Reference for biomaterial	DSM 11699	TAS [1]
	Gram stain	negative	TAS [1]
	Cell shape	rod-shaped	TAS [1]
	Motility	motile	TAS [1]
	Sporulation	non-sporulating	TAS [1]
	Temperature range	40-75°C	TAS [1]
	Optimum temperature	70°C	TAS [1]
	Salinity	15 to 70 g per l, optimum at 35 g	TAS [1]
MIGS-22	Oxygen requirement	strictly anaerobic	TAS [1]
	Carbon source	CO ₂	NAS
	Energy metabolism	chemolithoautotrophic, sulfur reduction	TAS [1]
MIGS-6	Habitat	marine	TAS [1]
MIGS-15	Biotic relationship	free-living	TAS [1]
	Biosafety level	1	TAS [28]
MIGS-19	Trophic level	level 1 primary producer	TAS [1]
MIGS-23.1	Isolation	deep-sea hydrothermal vent chimney	TAS [1]
MIGS-4	Geographic location	Snake Pit vent field, Mid-Atlantic Ridge	TAS [1]
MIGS-5	Sample collection time	November/December 1995	TAS [1,29]
MIGS-4.1	Latitude	23.36	TAS [1,29]
MIGS-4.2	Longitude	-44.93	TAS [1,29]
MIGS-4.3	Depth	3,500 m	TAS [1,29]
MIGS-4.4	Altitude	-3,500 m	TAS [1]

Evidence codes - TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from of the Gene Ontology project [30].

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [31], and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [32]. The genome project is deposited in the Genomes On Line Database [13] and

the complete genome sequence is deposited in GenBank. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.

Table 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Four genomic libraries: one 454 pyrosequence standard library, two 454 PE library (10.5 kb insert size), one Illumina library
MIGS-29	Sequencing platforms	Illumina GAii, 454 GS FLX Titanium
MIGS-31.2	Sequencing coverage	282.0 × Illumina; 40.0 × pyrosequence
MIGS-30	Assemblers	Newbler version 2.3, p Velvet version 0.7.63, phrap version SPS - 4.24
MIGS-32	Gene calling method	Prodigal 1.4, GenePRIMP
	INSDC ID	CP002543
	Genbank Date of Release	March 2, 2011
	GOLD ID	Gc01671
	NCBI project ID	51497
	Database: IMG-GEBA	2503754020
MIGS-13	Source material identifier	DSM 11699
	Project relevance	Tree of Life, GEBA

Growth conditions and DNA isolation

D. thermolithotrophum strain BSA^T, DSM 11699, was grown anaerobically in DSMZ medium 829 (*Desulfurobacterium* medium) [34] at 70°C. DNA was isolated from 0.5-1 g of cell paste using Qiagen Genomic 500 DNA Kit (Qiagen 10262) following the standard protocol as recommended by the manufacturer without modifications. DNA is available through the DNA Bank Network [35].

Genome sequencing and assembly

The genome was sequenced using a combination of Illumina and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website [36]. Pyrosequencing reads were assembled using the Newbler assembler (Roche). The initial Newbler assembly consisting of 96 contigs in one scaffold was converted into a phrap [37] assembly by making fake reads from the consensus, to collect the read pairs in the 454 paired end library. Illumina GAii sequencing data (45.0 Mb) was assembled with Velvet [38] and the consensus sequences were shredded into 1.5 kb overlapped fake reads and assembled together with the 454 data. The 454 draft assembly was based on 192.1 Mb 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -l 350 -g -m -ml 20. The Phred/Phrap/Consed software package [37] was used for sequence assembly and quality assessment in the subsequent finishing process. After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with gapResolution [36], Dupfinisher [39], or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F. Chang,

unpublished). A total of 101 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. Illumina reads were also used to correct potential base errors and increase consensus quality using a software Polisher developed at JGI [40]. The error rate of the completed genome sequence is less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided 322.0 × coverage of the genome. The final assembly contained 126,482 pyrosequence and 12,545,740 Illumina reads.

Genome annotation

Genes were identified using Prodigal [41] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [42]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGR-Fam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [33].

Genome properties

The genome consists of one circular chromosome with a total length of 1,541,968 bp and a G+C content of 35.0% (Table 3 and Figure 3). Of the 1,594 genes predicted, 1,543 were protein-coding genes, and 51 RNAs; 34 pseudogenes were also identified. The majority of the protein-coding genes (75.5%) were assigned a putative function while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 4.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	1,541,968	100.00%
DNA coding region (bp)	1,448,295	93.93%
DNA G+C content (bp)	538,896	34.95%
Number of replicons	1	
Extrachromosomal elements	0	
Total genes	1,594	100.00%
RNA genes	51	3.20%
rRNA operons	2	
tRNA genes	43	2.70%
Protein-coding genes	1,543	96.80%
Pseudo genes	34	2.13%
Genes with function prediction	1,204	75.53%
Genes in paralog clusters	600	37.64%
Genes assigned to COGs	1,330	83.44%
Genes assigned Pfam domains	1,327	83.25%
Genes with signal peptides	394	24.72%
Genes with transmembrane helices	322	20.20%
CRISPR repeats	1	

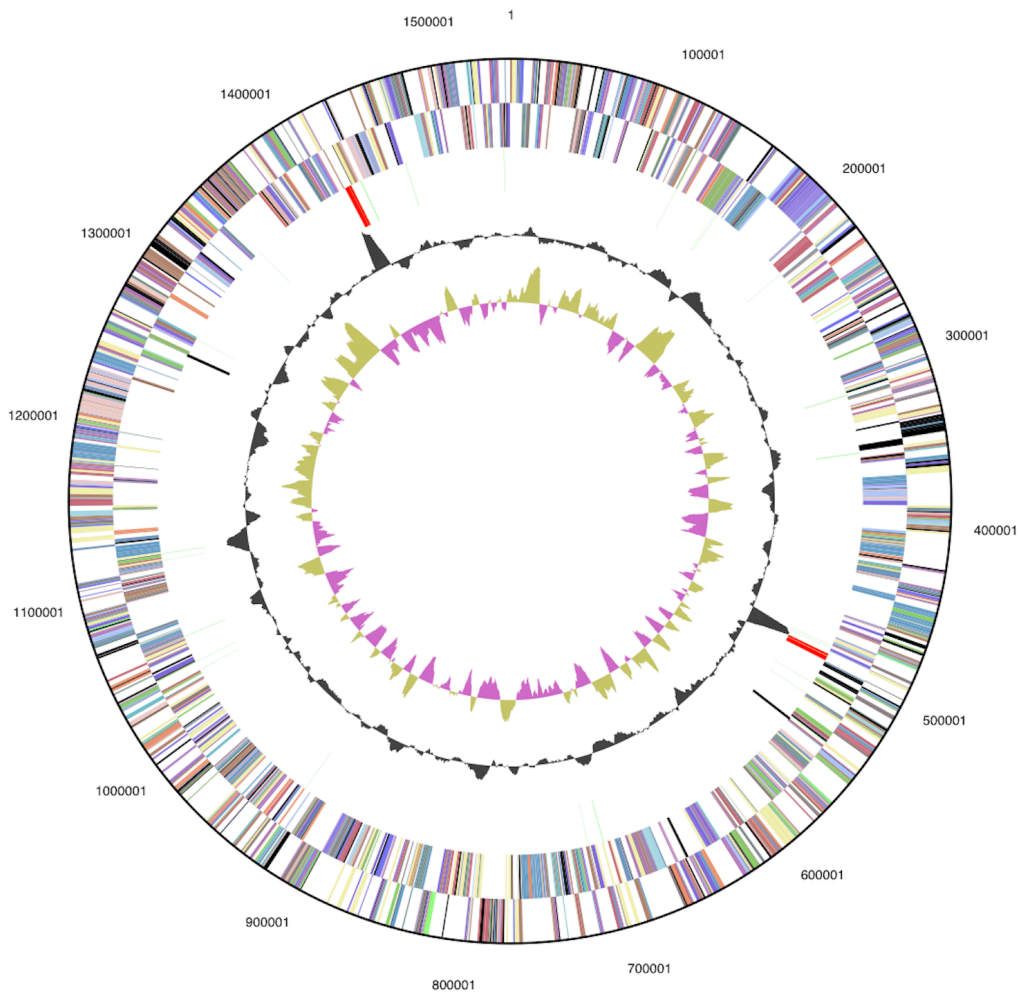


Figure 3. Graphical circular map of the genome. From bottom to top: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

Table 4. Number of genes associated with the general COG functional categories

Code	value	%age	Description
J	142	9.7	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	47	3.2	Transcription
L	126	8.6	Replication, recombination and repair
B	1	0.1	Chromatin structure and dynamics
D	20	1.4	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	13	0.9	Defense mechanisms
T	52	3.6	Signal transduction mechanisms
M	110	7.5	Cell wall/membrane/envelope biogenesis
N	67	4.6	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	74	5.1	Intracellular trafficking, secretion, and vesicular transport
O	55	3.8	Posttranslational modification, protein turnover, chaperones
C	113	7.7	Energy production and conversion
G	42	2.9	Carbohydrate transport and metabolism
E	108	7.4	Amino acid transport and metabolism
F	57	3.9	Nucleotide transport and metabolism
H	88	6.0	Coenzyme transport and metabolism
I	38	2.6	Lipid transport and metabolism
P	57	3.9	Inorganic ion transport and metabolism
Q	13	0.9	Secondary metabolites biosynthesis, transport and catabolism
R	140	9.6	General function prediction only
S	97	6.6	Function unknown
-	264	16.6	Not in COGs

Acknowledgements

We would like to gratefully acknowledge the help of Thomas Hader (University of Regensburg) for growing *D. thermolithotrophum* cultures. This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under

contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, UT-Battelle and Oak Ridge National Laboratory under contract DE-AC05-00OR22725, as well as German Research Foundation (DFG) INST 599/1-2.

References

1. L'Haridon S, Cilia V, Messner P, Raguénès G, Gambacorta A, Sleytr UB, Prieur D, Jeanthon C. *Desulfurobacterium thermolithotrophum* gen. nov., sp. nov., a novel autotrophic, sulphur-reducing bacterium isolated from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* 1998; **48**:701-711. [PubMed](http://dx.doi.org/10.1099/00207713-48-3-701) <http://dx.doi.org/10.1099/00207713-48-3-701>
2. List of Changes in Taxonomic Opinion N° 2. *Int J Syst Evol Microbiol* 2005; **55**:1403-1404. [PubMed](#)
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**:403-410. [PubMed](#)
4. Korf I, Yandell M, Bedell J. BLAST, O'Reilly, Sebastopol, 2003.

5. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microbiol* 2006; **72**:5069-5072. [PubMed](#)
<http://dx.doi.org/10.1128/AEM.03006-05>
6. Porter MF. An algorithm for suffix stripping *Program: electronic library and information systems* 1980; **14**:130-137.
7. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464. [PubMed](#)
<http://dx.doi.org/10.1093/bioinformatics/18.3.452>
8. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552. [PubMed](#)
9. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* 2008; **57**:758-771. [PubMed](#)
<http://dx.doi.org/10.1080/10635150802429642>
10. Hess PN, De Moraes Russo CA. An empirical test of the midpoint rooting method. *Biol J Linn Soc Lond* 2007; **92**:669-674. [PubMed](#)
<http://dx.doi.org/10.1111/j.1095-8312.2007.00864.x>
11. Pattengale ND, Alipour M, Bininda-Emonds OR, Moret BM, Stamatakis A. How many bootstrap replicates necessary? *J Comput Biol* 2010; **17**:337-354.
12. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0 b10. Sinauer Associates, Sunderland, 2002.
13. Liolios K, Chen IM, Mavromatis K, Tavernarakis N, Hugenholtz P, Markowitz VM, Kyrpides NC. The Genomes On Line Database (GOLD) in 2009: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2010; **38**:D346-D354. [PubMed](#)
<http://dx.doi.org/10.1093/nar/gkp848>
14. Reysenbach AL, Hamamura N, Podar M, Griffiths E, Ferreira S, Hochstein R, Heidelberg J, Johnson J, Mead D, Pohorille A, et al. Complete and draft genome sequences of six members of the *Aquificales*. *J Bacteriol* 2009; **191**:1992-1993. [PubMed](#)
<http://dx.doi.org/10.1128/JB.01645-08>
15. Zeytun A, Sikorski J, Nolan M, Lapidus A, Lucas S, Han J, Tice H, Cheng JF, Tapia R, Goodwin L, et al. Complete genome sequence of *Hydrogenobacter thermophilus* type strain (TK-6^T). *Stand Genomic Sci* 2011; **4**:131-143. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.1463589>
16. Arai H, Kanbe H, Ishii M, Igarashi Y. Complete genome sequence of the thermophilic, obligately chemolithoautotrophic hydrogen-oxidizing bacterium *Hydrogenobacter thermophilus* TK-6. *J Bacteriol* 2010; **192**:2651-2652. [PubMed](#)
<http://dx.doi.org/10.1128/JB.00158-10>
17. Wirth R, Sikorski J, Brambilla E, Misra M, Lapidus A, Copeland A, Nolan M, Lucas S, Chen F, Tice H, et al. Complete genome sequence of *Thermocrinis albus* type strain (CDC 1076^T). *Stand Genomic Sci* 2010; **2**:194-202. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.761490>
18. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](#)
<http://dx.doi.org/10.1038/nbt1360>
19. Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; **37**:9.
20. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](#)
<http://dx.doi.org/10.1073/pnas.87.12.4576>
21. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.
22. Reysenbach AL. Phylum BI. *Aquificae*. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 359-367.
23. List Editor. Validation List no. 85. Validation of publication of new names and new combinations previously effectively published outside the IJ-SEM. *Int J Syst Evol Microbiol* 2002; **52**:685-690. [PubMed](#)
<http://dx.doi.org/10.1099/ijs.0.02358-0>
24. Reysenbach AL. Class I. *Aquificae* class. nov. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 359.
25. L'Haridon S, Reysenbach AL, Tindall BJ, Schönheit P, Banta A, Johnsen U, Schumann P, Gambacorta A, Stackebrandt E, Jeanthon C. *De-*

- sulfurobacterium atlanticum* sp. nov., *Desulfurobacterium pacificum* sp. nov. and *Thermovibrio guaymasensis* sp. nov., three thermophilic members of the *Desulfurobacteriaceae* fam. nov., a deep branching lineage within the Bacteria. *Int J Syst Evol Microbiol* 2006; **56**:2843-2852. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.63994-0>
26. Reysenbach AL. Order I. *Aquificales* ord. nov. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 359.
 27. Alain K, Rolland S, Crassous P, Lesongeur F, Zbinden M, le Gall C, Godfroy A, Page A, Juniper SK, Cambon-Bonavita MA, et al. *Desulfurobacterium crinifex* sp. nov., a novel thermophilic, pinkish-streamer forming, chemolithoauto-trophic bacterium isolated from a Juan de Fuca Ridge hydrothermal vent and amendment of the genus *Desulfurobacterium*. *Extremophiles* 2003; **7**:361-370. [PubMed](#) <http://dx.doi.org/10.1007/s00792-003-0329-4>
 28. BAuA. 2010, Classification of *Bacteria* and *Archaea* in risk groups. <http://www.baua.de> TRBA 466, p. 77.
 29. Harmsen HJM, Prieur D, Jeanthon C. Distribution of microorganisms in deep-sea hydrothermal vent chimneys investigated by whole-cell hybridization and enrichments of thermophilic subpopulations. *Appl Environ Microbiol* 1997; **63**:2876-2883. [PubMed](#)
 30. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29. [PubMed](#) <http://dx.doi.org/10.1038/75556>
 31. Klenk HP, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst Appl Microbiol* 2010; **33**:175-182. [PubMed](#) <http://dx.doi.org/10.1016/j.syapm.2010.03.003>
 32. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, et al. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 2009; **462**:1056-1060. [PubMed](#) <http://dx.doi.org/10.1038/nature08656>
 33. Markowitz VM, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [PubMed](#) <http://dx.doi.org/10.1093/bioinformatics/btp393>
 34. List of growth media used at DSMZ: <http://www.dsmz.de/catalogues/catalogue-microorganisms/culture-technology/list-of-media-for-microorganisms.html>.
 35. Gemeinholzer B, Dröge G, Zetzsche H, Haszprunar G, Klenk HP, Güntsch A, Berendsohn WG, Wägele JW. The DNA Bank Network: the start from a German initiative. *Biopreserv Biobank* 2011; **9**:51-55. <http://dx.doi.org/10.1089/bio.2010.0029>
 36. JGI website. <http://www.jgi.doe.gov/>
 37. The Phred/Phrap/Consed software package. <http://www.phrap.com>
 38. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829. [PubMed](#) <http://dx.doi.org/10.1101/gr.074492.107>
 39. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. In: Proceeding of the 2006 international conference on bioinformatics & computational biology. Arabnia HR, Valafar H (eds), CSREA Press. June 26-29, 2006: 141-146.
 40. Lapidus A, LaButti K, Foster B, Lowry S, Trong S, Goltsman E. POLISHER: An effective tool for using ultra short reads in microbial genome assembly and finishing. AGBT, Marco Island, FL, 2008
 41. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-11-119>
 42. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; **7**:455-457. [PubMed](#) <http://dx.doi.org/10.1038/nmeth.1457>