Complete genome sequence of Hydrogenobacter thermophilus type strain (TK-67)

Ahmet Zeytun1,3, Johannes Sikorski2, Matt Nolan1, Alla Lapidus1, Susan Lucas1, James Han1, Hope Tice1, Jan-Fang Cheng1, Roxanne Tapia1,3, Lynne Goodwin1,3, Sam Pitluck1, Konstantinos Liolios1, Natalia Ivanova1, Konstantinos Mavromatis1, Natalia Mikhailova1, Galina Ovchinnikova1, Amrita Pati1, Amy Chen1, Krishna Palaniappan1, Olivier D. Ngatchou-Djao5, Miriam Land1,6, Loren Hauser1,6, Cynthia D. Jeffries1,6, Cliff Han1,3, John C. Detter1,3, Susanne Übler7, Manfred Rohde5, Brian J. Tindall2, Markus Göker2, Reinhard Wirth7, Tanja Woyke1,1, James Bristow1, Jonathan A. Eisen1,8, Victor Markowitz4, Philip Hugenholtz1,9, Hans-Peter Klenk2, and Nikos C. Kyrpides1*

1 DOE Joint Genome Institute, Walnut Creek, California, USA
2 DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany
3 Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
4 Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
5 HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany
6 Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
7 Archaea Centre - University of Regensburg, Regensburg, Germany
8 University of California Davis Genome Center, Davis, California, USA
9 Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

*Corresponding author: Nikos C. Kyrpides

Keywords: strictly thermophilic, obligately chemolithoautotrophic, Gram-negative, aerobic, hydrogen-oxidizing, nonmotile, non sporeforming, rod shaped, Aquificaceae, Aquificae, GEBA

Hydrogenobacter thermophilus Kawasumi et al. 1984 is the type species of the genus Hydrogenobacter. H. thermophilus was the first obligate autotrophic organism reported among aerobic hydrogen-oxidizing bacteria. Strain TK-67 is of interest because of the unusually efficient hydrogen-oxidizing ability of this strain, which results in a faster generation time compared to other autotrophs. It is also able to grow anaerobically using nitrate as an electron acceptor when molecular hydrogen is used as the energy source, and able to aerobically fix CO2 via the reductive tricarboxylic acid cycle. This is the fifth completed genome sequence in the family Aquificaceae, and the second genome sequence determined from a strain derived from the original isolate. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 1,742,932 bp long genome with its 1,899 protein-coding and 49 RNA genes is a part of the Genomic Encyclopedia of Bacteria and Archaea project.

Introduction

Strain TK-67 (= DSM 6534 = JCM 7687 = NBRC 102181) is the type strain of Hydrogenobacter thermophilus, which in turn is the type species of the genus Hydrogenobacter [1]. Currently, there are four species in the genus Hydrogenobacter, one of which has subsequently been reclassified as Hydrogenobaculum acidophilum. Strain TK-67 was previously isolated by Kawasumi in 1980 [2]. The genus name Calderobacterium Kryukov et al. 1984 is, based on page priority, a later heterotypic synonym of Hydrogenobacter Kawasumi et al. 1984 [3], because of similar genetic, phenotypic and biochemical properties between the type strains of H. thermophilus and Calderobacterium hydrogenophilum. Despite the relatively high degree of 16S rRNA gene sequence similarity between the two species, DNA-DNA hybridization [4] indicates that they may be considered to be different species within the genus...
**Hydrogenobacter** [3]. The genus name **Hydrogenobacter** is derived from the Latin words *hydrogenum*, meaning 'that which produces water' and *bacter*, referring to a rod that forms water when exposed to oxygen. The species epithet *thermophilus* derives from the Greek words *therme*, heat, and *philus*, loving, meaning a heat-loving organism. Strain TK-6T was isolated from hot springs located on the Izu peninsula in Japan [1]. Some strains of *H. thermophilus* were also isolated from a geothermal spring in Tuscany, Italy [5,6]. Other strains similar to *H. thermophilus* have been isolated from different environments, including a saline hot spring in Japan for *‘H. halophilus’* [7], and a volcanic area in Iceland for **Hydrogenobacter** strain H-1 [8], strains T3, T13 and T171 [5]. Until 1985, *H. thermophilus* was the only obligate autotroph among all aerobic hydrogen-oxidizing bacteria reported so far [9,10]. The activities of enzymes such as NADH:ferredoxin reductase (EC 1.18.1.3) and NAD-reducing hydrogenase (EC 1.12.1.2) were studied extensively in strain TK-6T [11]. Another genome sequence of a strain derived from the original isolate, presumably held in the lab of one of the co-authors, has been published recently without much metadata [12]. Here we present a summary classification and a set of features for *H. thermophilus* strain TK-6T, together with the description of the complete genomic sequencing and annotation.

**Classification and features**

The 16S rRNA gene sequence of the strain TK-6T (Z30214) shows the highest degree of sequence identity, 97%, to the type strain of **Hydrogenobacter** [6]. Further analysis shows 96% 16S rRNA gene sequence identity with an uncultured *Aquificales* bacterium clone pKA (AF453505) from a near-neutral thermal spring in Kamchatka, Russia. The single genomic 16S rRNA sequence of *H. thermophilus* was compared with the most recent release of the Greengenes database [13] using NCBI BLAST under default values and the relative frequencies of taxa and keywords, weighted by BLAST scores, were determined. The five most frequent genera were **Hydrogenobacter** (52.4%), **Thermocrinis** (18.8%), **Aquifex** (10.3%), **Sulfurihydrogenibium** (6.2%) and **Hydrogenivirga** (5.7%). Regarding hits to sequences from other members of the genus, the average identity within HSPs (high-scoring segment pairs) was 96.1%, whereas the average coverage by HSPs was 93.5%. The species yielding the highest score was *H. hydrogenophilus*. The five most frequent keywords within the labels of environmental samples which yielded hits were ‘hot’ (6.5%), ‘yellowstone’ (5.8%), ‘spring’ (5.6%), ‘national/park’ (5.4%) and ‘microbial’ (3.9%). These keywords corroborate with what is known from the ecology and physiology of strain TK-6T [1,2]. The most frequent keywords within the labels of environmental samples which yielded hits of a higher score than the highest scoring species were ‘aquificales’ (34.1%) and ‘hot/spring’ (32.9%).

Figure 1 shows the phylogenetic neighborhood of *H. thermophilus* TK-6T in a 16S rRNA based tree. The sequence of the single 16S rRNA gene in the genome differs by one nucleotide from the previously published 16S rRNA sequence (Z30214), which contains 31 ambiguous base calls.

Cells of strain TK-6T are Gram-negative, nonmotile straight rods of 0.3 to 0.5 µm by 2.0 to 3.0 µm occurring singly or in pairs [1] (Figure 2 and Table 1). Molecular oxygen is used as an electron acceptor for respiratory metabolism [1]. Interestingly, cytochrome C552 of TK-6T is extremely thermostable and can restore its conformation even after being autoclaved for 10 minutes at 121ºC [30]. One of the denitrification enzymes of the strain TK-6T, cytochrome *cd*$_1$ nitrite reductase has been isolated and analyzed [38]. Optimum temperature for the activity of this enzyme was found to range between 70°C-75°C [38]. Moreover, this enzyme was found to be of the heme *cd*-type [33]. Ammonium and nitrate were utilized as nitrogen sources [1,33], but not urea and N$_2$. Growth was inhibited by nitrite [1]. Nitrate reduction and peroxidase were positive, while urease...
was negative [1]. Strain TK-6\textsuperscript{T} could not utilize any of the following as sole sources of energy or carbon: glucose, fructose, galactose, maltose, sucrose, xylose, raffinose, \textit{L}-rhamnose, \textit{D}-mannose, \textit{D}-trehalose, mannitol, starch, formate, acetate, propionate, pyruvate, succinate, malate, citrate, fumarate, maleate, glycolate, gluconate, \textit{D}-lactate, \textit{\alpha}-ketoglutarate, \textit{p}-hydroxybenzoate, \textit{Dl}-polyhydroxybutyrate, betaine, methanol, ethanol, methanol, dimethylamine, trimethylamine, glycine, \textit{L}-glutamate, \textit{L}-aspartate, \textit{L}-serine, \textit{L}-leucine, \textit{L}-valine, \textit{L}-tryptophan, \textit{L}-histidine, \textit{L}-alanine, \textit{L}-lysine, \textit{L}-proline, \textit{L}-arginine, nutrient broth, yeast extract-malt extract medium, and brain heart infusion [1]. Strain TK-6\textsuperscript{T} showed no growth under an atmosphere containing 90% CO\textsubscript{2}, 5% CO\textsubscript{2}, and 5% O\textsubscript{2} [1]. No heterotrophic growth was observed in the presence of glucose, fructose, pyruvate, citrate, \textit{\alpha}-ketoglutarate, succinate, fumarate, maleate, acetate, and ethanol with and without yeast extract or carbon dioxide at different concentrations (0.02, 0.05, and 0.1% wt/vol) [1]. \textit{H. thermophilus} TK-6\textsuperscript{T} was recently reported to grow on formate and formamid [39]. Malate dehydrogenase, isocitrate dehydrogenase and glucose-6-phosphate isomerase were also detected in the strain TK-6\textsuperscript{T} [1]. Enzymes of the reductive tricarboxylic acid cycle and some related enzymes in cell-free extracts of strain TK-6\textsuperscript{T} were detected and their specific activities were found to increase with the temperature, the enzymes being more active at 70°C, as compared to lower temperatures (50°C and 30°C) [10]. In \textit{H. thermophilus}, ATP-dependent citrate cleavage is catalyzed by two enzymes, citryl-CoA synthetase and citryl-CoA lyase, which catalyze ATP-dependent formation of citryl-CoA from citrate and CoA and the subsequent cleavage of citryl-CoA into acetyl-CoA and oxaloacetate, respectively [40,41]. The biochemistry of key enzymes of the reductive tricarboxylic acid cycle, such as fumarate reductase, ATP citrate lyase, pyruvate:ferredoxin oxireductase and oxoglutarate:ferredoxin oxireductase, have been studied in some detail in strain TK-6\textsuperscript{T} [10,37,42]. Strain TK-6\textsuperscript{T} lacks some important enzyme activities in the central carbon metabolic pathways [43]. For example, activities of phosphofructokinase, pyruvate kinase, \textit{\alpha}-phospho gluconate aldolase, which are key enzymes of the Embden-Meyerhof and the Entner-Doudoroff pathways, and activity of \textit{\alpha}-ketoglutarate dehydrogenase of the tricarboxylic acid cycle could not be detected in cell-free extracts of strain TK-6\textsuperscript{T} [43]. This is in accord with the findings from the genome sequencing where none of these genes were found in the genome. These metabolic deficits were considered to be partially responsible for the obligate autotrophy of the strain TK-6\textsuperscript{T} [44]. Activities of phosphoenolpyruvate synthetase and pyruvate carboxylase were also detected [10]. The reverse reactions (dehydrogenase reactions) of \textit{\alpha}-ketoglutarate synthase and pyruvate synthase could be detected by using methyl viologen as an electron acceptor [10]. Cloning experiments of the hydrogenase genes from the strain TK-6\textsuperscript{T} revealed that this strain has at least four clusters of hydrogenase genes [35]. Strain TK-6\textsuperscript{T} assimilates ammonium using glutamine synthetase (GS type I) [45]. Anisomycin, cycloheximide and emetine (100 µg/ml each) do not inhibit protein biosynthesis and therefore growth of strain TK-6\textsuperscript{T} [46]. But the inhibitors of protein biosynthesis streptomycin, kanamycin, chloramphenicol, erythromycin, oleandomycin and virginiamycin were found to suppress growth of strain TK-6\textsuperscript{T} at concentrations below 20 µg/ml [46]. No growth was observed when cell wall synthesis inhibitors were used, (D-cycloserine, fosfomycin, cephalosporin C, penicillin G, oxacillin and ampicillin) at the concentration even below 20 µg/ml [46]. Strain TK-6\textsuperscript{T} could grow in the presence of menensin, lasalosid, valionycin, nonactin and polymyxin B [46].

**Chemotaxonomy**

The major cellular fatty acids found in strain TK-6\textsuperscript{T} were \textit{C\textsubscript{18:0}} and \textit{C\textsubscript{20:1}} [1,47]. These two fatty acids comprised about 80% of the total cellular fatty acids [1,47]. The minor components detected were \textit{C\textsubscript{16:0}}, \textit{C\textsubscript{16:1}} and \textit{C\textsubscript{18:1}}, C14:0 acids (indicative of the presence of a lipopolysaccharide) and a \textit{C\textsubscript{21:0}} cyclopropane acid, representing less than 10% of the total cellular fatty acids [1,47]. The detailed fatty acid composition of the strain TK-6\textsuperscript{T} is available in [27] and [47]. The main respiratory lipoquinone is an unusual sulfur-containing quinone, a 2-methylthio-3-VI, VII-tetrahydroheptaprenyl-1,4-naphthoquinone (i.e., methionaquinone 7, MTK-7) [48,49]. Strain TK-6\textsuperscript{T} contains glycerol-ether based lipids, as well as acyl glycerides [47]. It should be noted that the ether lipids are not of the type found in members of the \textit{Archaea}, since the side chains are alkyl straight chain and not isoprenoid. The presence of glycerol monoethers (GME) (1.2 µ mol/g dwt) is a characteristic feature of the strain TK-6\textsuperscript{T}, the main one being GME-18:0 (82.7% wt) [27,47]. GME-20:1 (11.1% wt), GME-20:0 (3.5 wt), and GME-18:1 (2.7% wt) were also detected in strain TK-6\textsuperscript{T} [27,47]. No glycerol diether (GDE) was detected [27,47].
Investigations of the polar lipids has shown that the polar lipids comprise phosphatidylglycerol, phosphatidylinositol, phosphatidylaminopentantetrol and a small amount of an unidentified phospholipid. The sum of these chemotaxonomic features appear to be characteristic of members of the genus Hydrogenobacter, with features such as the presence of methionaquinone, a polar lipid pattern containing phosphatidylglycerol, phosphatidylinositol and phosphatidylaminopentantetrol and the presence of C18:0 and C20:1 fatty acids being taxonomic and evolutionary markers for at least members of the genera Hydrogenobacter, Hydrogenobaculum, Aquifex and Thermonocrinis. This has been discussed in a previous SIGS paper [50].
Table 1. Classification and general features of *H. thermophilus* TK-6ᵀ according to the MIGS recommendations [22]

| MIGS ID  | Property                  | Term                                      | Evidence code |
|---------|---------------------------|-------------------------------------------|---------------|
|         | Domain                    | Bacteria                                  | TAS [23]      |
|         | Phylum                    | Aquificae                                 | TAS [24,25]   |
|         | Class                     | Aquificae                                 | TAS [24,26]   |
| Current classification | Order Aquificales          | Aquificae                                 | TAS [24,27,28]|
|         | Family                    | Aquificaceae                              | TAS [24,29]   |
|         | Genus                     | Hydrogenobacter                           | TAS [1]       |
|         | Species                   | Hydrogenobacter thermophilus              | TAS [1]       |
|         | Type strain               | TK-6                                      | TAS [1]       |
|         | Gram stain                | negative                                  | TAS [1]       |
|         | Cell shape                | straight rods                             | TAS [1]       |
|         | Motility                  | non-motile                                | TAS [1]       |
|         | Sporulation               | no                                        | TAS [1]       |
|         | Temperature range         | 50°C–78°C                                 | TAS [30]      |
|         | Optimum temperature       | 70°C-75°C                                 | TAS [1]       |
|         | Salinity                  | not reported                              | NAS           |
| MIGS-22 | Oxygen requirement        | aerobic                                   | TAS [1]       |
|         | Carbon source             | CO₂                                       | TAS [1]       |
|         | Energy source             | H₂, thiosulfate, obligate chemolithoautotrophic | TAS [1] |
| MIGS-6  | Habitat                   | soil near hot spring                      | TAS [1]       |
| MIGS-15 | Biotic relationship       | free living                              | NAS           |
| MIGS-14 | Pathogenicity             | not reported                              | NAS           |
|         | Biosafety level           | 1                                        | TAS [31]      |
|         | Isolation                 | hot spring                                | TAS [1]       |
| MIGS-4  | Geographic location       | Izu peninsula, Japan                      | TAS [1,30]    |
| MIGS-5  | Sample collection time    | 1980 or before                           | TAS [1,2]     |
| MIGS-4.1| Latitude                  | approx. 34.9                             | NAS           |
| MIGS-4.2| Longitude                 | 138.9                                    |              |
| MIGS-4.3| Depth                     | not reported                              |              |
| MIGS-4.4| Altitude                  | not reported                              |              |

Evidence codes - IDA: Inferred from Direct Assay (first time in publication); 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 [32]. If the evidence code is IDA, then the property was directly observed by one of the authors or an expert mentioned in the acknowledgements.

**Genome sequencing and annotation**

**Genome project history**

This organism was selected for sequencing on the basis of its phylogenetic position [51], and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [52]. The genome project is deposited in the Genome On Line Database [19] 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.

**Growth conditions and DNA isolation**

*H. thermophilus* TK-6ᵀ, DSM 6534, was grown in DSMZ medium 533 (Thermophilic hydrogen bacteria medium) [53] with 5% oxygen at 72°C. DNA was isolated from 0.5-1 g of cell paste using Qia- gen Genomic 500 DNA Kit (Qiagen, Hilden, Germany) following the standard protocol as recommended by the manufacturer. DNA is available through the DNA Bank Network [54].

http://standardsingenomics.org
Hydrogenobacter thermophilus type strain (TK-6T)

Table 2. Genome sequencing project information

| MIGS ID | Property                  | Term                                                                 |
|---------|---------------------------|----------------------------------------------------------------------|
| MIGS-31 | Finishing quality         | Finished                                                             |
| MIGS-28 | Libraries used            | One 454 pyrosequence standard library, one 454 PE (20kb insert size) and one Illumina standard library |
| MIGS-29 | Sequencing platforms      | 454 GS FLX Titanium, Illumina GAii                                   |
| MIGS-31.2| Sequencing coverage      | 82.1× pyrosequence, 264.4× Illumina                                 |
| MIGS-30 | Assemblers                | Newbler version 2.3-PreRelease-10-21-2009-gcc-4.1.2, phrap           |
| MIGS-32 | Gene calling method       | Prodigal 1.4, GenePRIMP                                               |
|         | INSDC ID                  | CP0002221                                                            |
|         | Genbank Date of Release   | October 15, 2010                                                     |
|         | GOLD ID                   | Gc01411                                                              |
|         | NCBI project ID           | 41547                                                                |
|         | Database: IMG-GEBA        | 2502957034                                                           |
| MIGS-13 | Source material identifier| DSM 6534                                                              |
|         | Project relevance         | Tree of Life, GEBA                                                    |

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 [55]. Pyrosequencing reads were assembled using the Newbler assembler version 2.3-PreRelease-10-21-2009-gcc-4.1.2-threads (Roche). The initial Newbler assembly consisted of 19 contigs in one scaffold which was converted into a phrap assembly by making fake reads from the consensus, collecting the read pairs in the 454 paired end library. Illumina GAii sequencing data (449.5 Mb) was assembled with Velvet [56] 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 143.2 MB 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -I 350 -g -m -ml 20. The Phred/Phrap/Consed software package [57] 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 [55], Dupfinisher, or sequencing cloned bridging PCR fragments with subcloning or transposon bombing (Epicentre Biotechnologies, Madison, WI) [58]. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F.Chang, unpublished). A total of 24 additional Sanger 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 [59]. 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 346.5 × coverage of the genome. Final assembly contains 454,097 pyrosequence and 12,484,847 Illumina reads.

Genome annotation

Genes were identified using Prodigal [60] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [61]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRfam, 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 [62].

Genome properties

The genome consists of a 1,742,932 bp long chromosome with a 44.0% GC content (Table 3 and Figure 3). Of the 1,948 genes predicted, 1,899 were protein-coding genes, and 49 RNAs; thirty pseudogenes were also identified. The majority of the protein-coding genes (97.5%) were assigned with 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)              | 1,742,932  | 100.00%    |
| DNA coding region (bp)        | 1,666,175  | 95.60%     |
| DNA G+C content (bp)          | 766,905    | 44.00%     |
| Number of replicons           | 1          |            |
| Extrachromosomal elements     | 0          |            |
| Total genes                   | 1,948      | 100.00%    |
| RNA genes                     | 49         | 2.52%      |
| rRNA operons                  | 1          |            |
| Protein-coding genes          | 1,899      | 97.48%     |
| Pseudo genes                  | 30         | 1.54%      |
| Genes with function prediction| 1,361      | 69.87%     |
| Genes in paralog clusters     | 183        | 9.39%      |
| Genes assigned to COGs        | 1,441      | 73.97%     |
| Genes assigned Pfam domains   | 1,501      | 77.05%     |
| Genes with signal peptides    | 287        | 14.73%     |
| Genes with transmembrane helices| 381      | 19.56%     |
| CRISPR repeats                | 1          |            |

Insights into the genome

While the sequencing of the genome described in this paper was underway, Arai et al. from University of Tokyo published the first version of the *H. thermophilus* TK-6T genome [19, AP011112]. We take the opportunity to compare the two completed genome sequences, because the history of the two strains designated TK-6T might differ since the original isolation of the strain by Kawasum et al. [1], more than a 25 years ago. The first of the two genomes was published by a team of researchers located at the same place where the strain was originally analyzed, with Yasuo Igarashi participating in both, the original description of the strain and the genome analysis. According to personal information by Dr. Arai Hiroyuki (lead author in [19]), the genome was sequenced from clone and fosmid libraries generated by a strain subcultured in the lab since the time of the initial isolation. A fresh culture of the strain from JCM was used for final gap filling and error checking. The DSM 6534 version of the genome was generated from cryopreserved material, which DSMZ received in 1991 from Tohru Kodama of University of Tokyo, and the strain was preserved by storage in liquid nitrogen since it was accessed.

A comparison of the two TK-6T genomes using the genome-to-genome-distance calculation [63-65] in conjunction with NCBI-BLASTN yielded a distance of 0.0001 with formula 1, 0.0100 with formula 2 and 0.0101 with formula 3. That is, 99.99% of the total genome length was covered by HSPs, 99.0% of the positions within the HSPs held identical bases, and 98.99% of the total genome length corresponded to such identical base pairs within HSPs. The synteny of the two TK-6T genome sequences based on a DNA blot was confirmed (data not shown), whereas Table 5 provides a comparison of the basic genome statistics.

The Japanese strain has 1,868 (out of 1,893) protein coding genes identical to the DSMZ strain which is 98.7% of the genome. This means there are 25 genes in the Japanese strain that are not in the DSMZ strain, all except L34P are hypothetical genes. L34P is however present in the version of the genome as presented in this paper, but was missed from the ORF calling/annotation. We also identified 24 genes in the genome sequenced from the DSMZ strain that were missing in the Arai et al. strain. Also most of these were again hypothetical genes. The abundance profiles for both genomes were almost identical, with glycosyltransferase (COG0438) being the most frequent gene in both versions (eleven copies), followed by seven copies of an outer membrane protein (COG1538), each. The DSM 6534 genome contains seven copies of transposase IS605 OrfB (COG0675), whereas Tokyo contains five copies of it.
The DSM 6534 version of the genome also contains more copies of cation transport ATPase (COG0474, 4 vs. 2), nitrogenase molybdenum-iron protein, alpha and beta chains (COG2710, 4 vs. 2), acetyl/propionyl-CoA carboxylase, alpha subunit (COG4779, 4 vs. 3), Fe-S oxidoreductases (COG0474, 3 vs. 2), catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases (COG0664, 3 vs. 2), cation transport ATPase (COG2217, 3 vs. 2), DNA modification methylase (COG0862, 2 vs. 1), hemolysins and related proteins containing CBS domains (COG1253, 2 vs. 1), phosphoketolase (COG3957), an uncharacterized MobA-related protein (COG2068) and an uncharacterized conserved protein (COG4121) were identified in one copy, each, in the DSM 6534 genome, but absent in the U Tokyo version. The U Tokyo version contains more copies of selenocysteine-containing anaerobic dehydrogenases, (COG0243, 5 vs. 1), as well as, 1-acyl-sn-glycerol-3-phosphate acyltransferase (COG02043) and K+ -transporting ATPase, A chain (COG2060, 2 vs. 1, each).
Table 4. Number of genes associated with the general COG functional categories

| Code | value | %age | Description                                           |
|------|-------|------|-------------------------------------------------------|
| J    | 134   | 8.6  | Translation, ribosomal structure and biogenesis       |
| A    | 0     | 0.0  | RNA processing and modification                       |
| K    | 52    | 3.3  | Transcription                                         |
| L    | 85    | 5.4  | Replication, recombination and repair                  |
| B    | 2     | 0.1  | Chromatin structure and dynamics                       |
| D    | 19    | 1.3  | Cell cycle control, cell division, chromosome partitioning |
| Y    | 0     | 0.0  | Nuclear structure                                     |
| V    | 21    | 1.3  | Defense mechanisms                                    |
| T    | 53    | 3.4  | Signal transduction mechanisms                        |
| M    | 128   | 8.2  | Cell wall/membrane/envelope biogenesis                |
| N    | 23    | 1.5  | Cell motility                                         |
| Z    | 1     | 0.0  | Cytoskeleton                                          |
| W    | 0     | 0.0  | Extracellular structures                              |
| U    | 56    | 3.6  | Intracellular trafficking and secretion, and vesicular transport |
| O    | 74    | 4.7  | Posttranslational modification, protein turnover, chaperones |
| C    | 182   | 11.6 | Energy production and conversion                      |
| G    | 58    | 3.7  | Carbohydrate transport and metabolism                 |
| E    | 118   | 7.5  | Amino acid transport and metabolism                   |
| F    | 52    | 3.3  | Nucleotide transport and metabolism                   |
| H    | 107   | 6.8  | Coenzyme transport and metabolism                     |
| I    | 43    | 2.7  | Lipid transport and metabolism                        |
| P    | 78    | 5.0  | Inorganic ion transport and metabolism                |
| Q    | 15    | 1.0  | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 167   | 10.7 | General function prediction only                      |
| S    | 100   | 6.4  | Function unknown                                     |
| -    | 507   | 26.3 | Not in COGs                                           |

Table 5. Comparison of Genome Statistics

| Attribute                      | DSM 6534   | U of Tokyo | difference |
|--------------------------------|------------|------------|------------|
| Genome size (bp)               | 1,742,932  | 1,744,135  | +1,203     |
| DNA coding region (bp)         | 1,666,175  | 1,669,712  | +3,537     |
| DNA G+C content (bp)           | 766,905    | 766,984    | +79        |
| Number of replicons            | 1          | 1          | 1          |
| Extrachromosomal elements      | 0          | 0          | 0          |
| Total genes                    | 1,948      | 1,941      | -7         |
| RNA genes                      | 49         | 48         | -1         |
| rRNA operons                   | 1          | 1          | 1          |
| Protein-coding genes           | 1,899      | 1,893      | -6         |
| Pseudo genes                   | 30         | 0          | -30        |
| Genes with function prediction | 1,361      | 1,349      | -12        |
| Genes in paralog clusters      | 183        | 175        | -8         |
| Genes assigned to COGs         | 1,441      | 1,430      | -11        |
| Genes assigned Pfam domains    | 1,501      | 1,489      | -12        |
| Genes with signal peptides     | 287        | 528        | +241       |
| Genes with transmembrane helices| 381        | 385        | +4         |
| CRISPR repeats                 | 1          | 2          | +1         |
Acknowledgements

We would like to gratefully acknowledge the help of Susanne Schneider (DSMZ) for DNA extraction and quality analysis. This work was performed under the auspices of the US Department of Energy 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 599/1-1.

References

1. Kawasumi T, Igarashi Y, Kodama T, Minoda Y. Hydrogenobacter thermophilus gen. nov. sp. nov., an extremely thermophilic, aerobic, hydrogen-oxidizing bacterium. Int J Syst Bacteriol 1984; 34:5-10. doi:10.1099/00207713-34-1-5

2. Kawasumi T, Igarashi Y, Kodama T, Minoda Y. Isolation of strictly thermophilic and obligately autotrophic hydrogen bacteria. Agric Biol Chem 1980; 44:1985-1986.

3. Stöhr R, Waberski A, Völker H, Tindall BJ, Thomm M. Hydrogenothermus marinus gen. nov., sp. nov., a novel thermophilic hydrogen-oxidizing bacterium, recognition of Calderobacterium hydrogenophilum as a member of the genus Hydrogenobacter and proposal of the reclassification of Hydrogenobacter acidiphilus as Hydrogenobaculum acidophilum gen. nov., comb. nov., in the phylum Hydrogenobacter/Aquifex. Int J Syst Evol Microbiol 2001; 51:1853-1862. PubMed

4. Shima S, Suzuki K. Hydrogenobacter acidophilus sp. nov., a thermoacidophilic, aerobic, hydrogen-oxidizing bacterium requiring elemental sulfur for growth. Int J Syst Bacteriol 1993; 43:703-708. doi:10.1099/00207713-43-4-703

5. Bonjour F, Aragno M. Growth of thermophilic, obligatorily chemolithoautotrophic hydrogen-oxidizing bacteria related to Hydrogenobacter with thiosulfate and elemental sulfur as electron donors. Int J Syst Bacteriol 1994; 44:620-626. PubMed doi:10.1099/00207713-44-4-620

6. Pitulle C, Yang Y, Marchiani M, Moore ERB, Sieffert JL, Aragno M, Jurtshuk P, Jr., Fox GE. Phylogenetic position of the genus Hydrogenobacter. Int J Syst Bacteriol 1994; 44:620-626. PubMed doi:10.1099/00207713-44-4-620

7. Nishihara H, Igarashi Y, Kodama T. A new isolate of Hydrogenobacter, an obligately chemolithoautotrophic, thermophilic, halophilic and aerobic hydrogen-oxidizing bacterium from seaside saline hot spring. Arch Microbiol 1990; 153:294-298. doi:10.1007/BF00249085

8. Kristjansson VK, Ingason A, Alfredsson GA. Isolation of thermophilic obligately autotrophic hydrogen-oxidizing bacteria, similar to Hydrogenobacter thermophilus, from Icelandic hot springs. Arch Microbiol 1985; 140:321-325. doi:10.1007/BF00446971

9. Bowien B, Schlegel HG. Physiology and biochemistry of aerobic hydrogen-oxidizing bacteria. Annu Rev Microbiol 1981; 35:405-452. PubMed doi:10.1146/annurev.mi.35.100181.002201

10. Shiba H, Kawasumi T, Igarashi Y, Kodama T, Minoda Y. The CO2 assimilation via the reductive tricarboxylic acid cycle in an obligately autotrophic, aerobic hydrogen-oxidizing bacterium, Hydrogenobacter thermophilus. Arch Microbiol 1985; 141:198-203. doi:10.1007/BF00408058

11. Yoon KS, Ueda Y, Ishii M, Igarashi Y, Kodama T. NADH:ferredoxin reductase and NAD-reducing hydrogenase activities in Hydrogenobacter thermophilus strain TK-6. FEBS Microbiol Lett 1996; 139:139-142.

12. 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 doi:10.1128/JB.00158-10

13. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie E, Keller K, Huber T, Dalevi D, Hu P, Andersen G, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006; 72:5069-5072. PubMed doi:10.1128/AEM.03006-05

14. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 2000; 17:540-552. PubMed

15. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. Bioinformatics 2002; 18:452-464. PubMed doi:10.1093/bioinformatics/18.3.452

Hydrogenobacter thermophilus type strain (TK-6T)
16. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. Syst Biol 2008; 57:758-771. PubMed doi:10.1080/10635150802429642

17. Yarza P, Richter M, Peplies J, Euzéby J, Amann R, Schleifer KH, Ludwig W, Glöckner FO, Rossello-Móra R. The All-Species Living Tree project: A 16S rRNA-based phylogenetic tree of all sequenced type strains. Syst Appl Microbiol 2008; 31:241-250. PubMed doi:10.1016/j.syapm.2008.07.001

18. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? Lect Notes Comput Sci 2009; 5541:184-200. doi:10.1007/978-3-642-02008-7_13

19. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res 2008; 36:D475-D479. PubMed doi:10.1093/nar/gkm884

20. Reysenbach AL, Hamamura N, Podar M, Griffiths E, Ferreira S, Hochstein R, Heidelberg J, Johnson J, Mead D, Pohorille A, et al. Genome announcement. Complete and draft genome sequences of six members of the Aquificales. In Aquificales. J Bacteriol 2009; 191:1992-1993. PubMed doi:10.1128/JB.01645-08

21. 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 (HI 11/12T). Stand Genomic Sci 2010; 2:194-202. PubMed doi:10.4056/sigs.761490

22. 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 doi:10.1038/nbt.1360

23. 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 doi:10.1073/pnas.87.12.4576

24. Validation list 85: Validation of publication of new names and new combinations previously effectively published outside the IJSEM. Int J Syst Evol Microbiol 2002; 52:685-690. PubMed doi:10.1099/ijs.0.02358-0

25. Reysenbach AL. 2001. Phylum Bl. Aquificae. In: Garrity GM, Boone DR, Castenholz RW (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, p. 359-367.

26. Reysenbach AL. 2001. 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, p. 359.

27. L’Haridon S, Reysenbach AL, Tindall BJ, Schönheit P, Banta A, Johnsen U, Schumann P, Gambacorta A, Stackebrandt E, Jeannot C. Desulfurobacterium 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 doi:10.1099/ijs.0.63994-0

28. Reysenbach AL. 2001. 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, p. 360.

29. Reysenbach AL. 2001. Family I. Aquificaceae fam. nov. In: Garrity GM, Boone DR, Castenholz RW (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, p. 925–938.

30. Huber R, Eder W. 2006. Aquificales. In: M Dworkin, S Falkow, E Rosenberg, KH Schleifer E Stackebrandt (eds), The Prokaryotes, 3. ed, vol. 7. Springer, New York, p. 925-938.

31. Classification of bacteria and archaea in risk groups. http://www.baua.de TRBA 466.

32. 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 doi:10.1038/75556

33. Suzuki M, Cui ZJ, Ishii M, Igarashi Y. Nitrate respiratory metabolism in an obligately autotrophic hydrogen-oxidizing bacterium, Hydrogenobacter thermophilus TK-6. Arch Microbiol 2001; 175:75-78. PubMed doi:10.1007/s0020300000230

34. Ishii M, Igarashi Y, Kodama T. Colony formation of Hydrogenobacter thermophilus on a plate sulfured with GELRITE. Agric Biol Chem 1987; 51:3139-3141.
35. Ueda Y, Yamamoto M, Urasaki T, Arai H, Ishii M, Igarashi Y. Sequencing and reverse transcription-polymerase chain reaction (RT-PCR) analysis of four hydrogenase gene clusters from an obligately autotrophic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus* TK-6. *J Biosci Bioeng* 2007; 104:470-475. PubMed doi:10.1263/jbb.104.470

36. Beh M, Strauss G, Huber R, Stetter KO, Fuchs G. Enzymes of the reductive citric acid cycle in the autotrophic eubacterium *Aquifex pyrophilus* and in the archaebacterium *Thermoproteus neutrophilus*. *Arch Microbiol* 1993; 160:306-311. doi:10.1007/BF00292082

37. Ikeda T, Ochiai T, Morita S, Nishiyama A, Yamada E, Arai H, Ishii M, Igarashi Y. Anabolic five subunit-type pyruvate:ferredoxin oxidoreductase from *Hydrogenobacter thermophilus* TK-6. *Biochem Biophys Res Commun* 2006; 340:76-82. PubMed doi:10.1016/j.bbrc.2005.11.155

38. Suzuki M, Cui ZJ, Ishii M, Igarashi Y. Purification, characterization, and gene cloning of thermophilic cytochrome cd1 nitrite reductase from *Hydrogenobacter thermophilus* TK-6. *J Biosci Bioeng* 2006; 101:391-397. PubMed doi:10.1263/jbb.101.391

39. Eder W, Huber R. New isolates and physiological properties of the *Aquificales* and description of *Thermocrinis albus* sp. nov. *Extremophiles* 2002; 6:309-318. PubMed doi:10.1007/s00792-001-0259-y

40. Aoshima M, Ishii M, Igarashi Y. A novel enzyme, citryl-CoA synthetase, catalysing the first step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6. *Mol Microbiol* 2004; 52:751-761. PubMed doi:10.1111/j.1365-2958.2004.04009.x

41. Aoshima M, Ishii M, Igarashi Y. A novel enzyme, citryl-CoA lyase, catalysing the second step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6. *Mol Microbiol* 2004; 52:763-770. PubMed doi:10.1111/j.1365-2958.2004.04010.x

42. Yun NR, Yamamoto M, Arai H, Ishii M, Igarashi Y. A novel five-subunit-Type 2-oxoglutarate:ferredoxin oxidoreductases from *Hydrogenobacter thermophilus* TK-6. *Biochem Biophys Res Commun* 2002; 292:280-286. PubMed doi:10.1006.bbrc.2002.6651

43. Shiba H, Kawasumi T, Igarashi Y, Kodama T, Minoda Y. The deficient carbohydrate metabolic pathways and the incomplete tricarboxylic acid cycle in an obligately autotrophic hydrogen-oxidizing bacterium. *Agric Biol Chem* 1982; 46:2341-2345.

44. Shiba H, Kawasumi T, Igarashi Y, Kodama T, Minoda Y. Effect of organic compounds on the growth of an obligately autotrophic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus* TK-6. *Agric Biol Chem* 1984; 48:2809-2813.

45. Kameya M, Arai H, Ishii M, Kodama T. The drug resistance of *Hydrogenobacter thermophilus* strain TK-6. *FEMS Microbiol Lett* 1988; 49:179-182. doi:10.1111/j.1574-6968.1988.tb02711.x

46. Igarashi Y, Sanbongi Y, Ishii M, Kodama T. The drug resistance of *Hydrogenobacter thermophilus* strain TK-6. *FEMS Microbiol Lett* 1988; 49:179-182. doi:10.1111/j.1574-6968.1988.tb02711.x

47. Jahnke LL, Eder W, Huber R, Hope JM, Hinrichs KU, Hayes JM, Des Marais DJ, Cady SL, Summons RE. Composition of hydrothermal vent microbial communities as revealed by analyses of signature lipids, stable carbon isotopes and *Aquificales* cultures. *Appl Environ Microbiol* 2001; 67:5179-5189. PubMed doi:10.1128/AEM.67.11.5179-5189.2001

48. Ishii M, Kawasumi T, Igarashi Y, Kodama T, Minoda Y. 2-Methylthio-1,4-naphthoquinone, a new quinone from an extremely thermophilic hydrogen bacterium. *Agric Biol Chem* 1983; 47:167-169.

49. Ishii M, Kawasumi T, Igarashi Y, Kodama T, Minoda Y. 2-Methylthio-1,4-naphthoquinone, a unique sulfur-containing quinone from a thermophilic hydrogen-oxidizing bacterium, *Hydrogenobacter thermophilus*. *J Bacteriol* 1987; 169:2380-2384. PubMed

50. Wirth R, Sikorski J, Brambilla E, Misra M, Lapidus A, Copeland A, Nolan M., Lucas S, Chen F, Tice H, Cheng JF, et al. Complete genome sequence of *Thermocrinis albus* type strain (HI 11/12T). *Stand Genomic Sci* 2010; 2:194-202. PubMed doi:10.4056/sigs.761490

51. Klenk HP, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst Appl Microbiol* 2010; 33:175-182. PubMed doi:10.1016/j.syapm.2010.03.003

52. 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*
53. List of growth media used at DSMZ: http://www.dsmz.de/microorganisms/media_list.php.

54. 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. *Biopreservation and Biobanking* 2011; 9:51-55. doi:10.1089/bio.2010.0029

55. JGI website. http://www.jgi.doe.gov

56. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; 18:821-829. PubMed doi:10.1101/gr.074492.107

57. The Phred/Phrap/Consed software package. http://www.phrap.com

58. Sims D, Brettin T, Detter J, Han C, Lapidus A, Copeland A, Glavina Del Rio T, Nolan M, Chen F, Lucas S, et al. Complete genome sequence of *Kytococcus sedentarius* type strain (541T). *Stand Genomic Sci* 2009; 1:12-20. PubMed doi:10.4056/sigs.761

59. 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.

60. 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 doi:10.1186/1471-2105-11-119

61. Pati A, Ivanova N, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: A gene prediction improvement pipeline for microbial genomes. *Nat Methods* 2010; 7:455-457. PubMed doi:10.1038/nmeth.1457

62. Markowitz VM, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; 25:2271-2278. PubMed doi:10.1093/bioinformatics/btp393

63. Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci* 2010; 2:142-148. PubMed doi:10.4056/sigs.541628

64. Auch AF, Von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2010; 2:117-134. PubMed doi:10.4056/sigs.531120

65. Auch AF, Henz S, Holland B, Göker M. Genome blast distance phylogenies inferred from whole plastid and whole mitochondrion genome sequences. *BMC Bioinformatics* 2006; 7:350. PubMed doi:10.1186/1471-2105-7-350