Complete genome sequence of ‘Thermobaculum terrenum’ type strain (YNP1)<sup>T</sup>

Hajnalka Kiss<sup>1</sup>, David Cleland<sup>2</sup>, Alla Lapidus<sup>3</sup>, Susan Lucas<sup>3</sup>, Tijana Glavina Del Rio<sup>3</sup>, Matt Nolan<sup>1</sup>, Hope Tice<sup>1</sup>, Cliff Han<sup>1</sup>, Lynne Goodwin<sup>1,3</sup>, Sam Pitluck<sup>3</sup>, Konstantinos Liolios<sup>3</sup>, Natalia Ivanova<sup>3</sup>, Konstantinos Mavromatis<sup>3</sup>, Galina Ovchinnikova<sup>3</sup>, Amrita Pati<sup>3</sup>, Amy Chen<sup>3</sup>, Krishna Palaniappan<sup>3</sup>, Miriam Land<sup>1,3</sup>, Loren Hauser<sup>1,3</sup>, Yun-Juan Chang<sup>1,5</sup>, Cynthia D. Jeffries<sup>3,5</sup>, Megan Lu<sup>1</sup>, Thomas Brettin<sup>1</sup>, John C. Detter<sup>1</sup>, Markus Göker<sup>6</sup>, Brian J. Tindall<sup>6</sup>, Brian Beck<sup>2</sup>, Timothy R. McDermott<sup>7</sup>, Tanja Woyke<sup>3</sup>, James Bristow<sup>3</sup>, Jonathan A. Eisen<sup>1,8</sup>, Victor Markowitz<sup>4</sup>, Philip Hugenholtz<sup>3</sup>, Nikos C. Kyrpides<sup>3</sup>, Hans-Peter Klenk<sup>6</sup>* and Jan-Fang Cheng

1 Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
2 ATCC- American Type Culture Collection, Manassas, Virginia, USA
3 DOE Joint Genome Institute, Walnut Creek, California, USA
4 Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
5 Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
6 DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany
7 Thermal Biology Institute, Montana State University, Bozeman, Montana, USA
8 University of California Davis Genome Center, Davis, California, USA

*Corresponding author: Hans-Peter Klenk

Keywords: extreme thermal soil, thermoacidophile, Gram-positive, nonmotile, non-spore-forming, obligate aerobe, Incertae sedis, Chloroflexi, GEBA

‘Thermobaculum terrenum’ Botero et al. 2004 is the sole species within the proposed genus ‘Thermobaculum’. Strain YNP1<sup>T</sup> is the only cultivated member of an acid tolerant, extremely thermophilic species belonging to a phylogenetically isolated environmental clone group within the phylum Chloroflexi. At present, the name ‘Thermobaculum terrenum’ is not yet validly published as it contravenes Rule 30 (3a) of the Bacteriological Code. The bacterium was isolated from a slightly acidic extreme thermal soil in Yellowstone National Park, Wyoming (USA). Depending on its final taxonomic allocation, this is likely to be the third completed genome sequence of a member of the class Thermomicrobia and the seventh type strain genome from the phylum Chloroflexi. The 3,101,581 bp long genome with its 2,872 protein-coding and 58 RNA genes is a part of the Genomic Encyclopedia of Bacteria and Archaea project.

Introduction

Strain YNP1<sup>T</sup> (= ATCC BAA-798 = CCMEE 7001) is the proposed type strain of the not yet validly published species ‘Thermobaculum terrenum’, which represents the type species of the not yet validly published genus ‘Thermobaculum’ [1]. The strain was cultivated from a moderately acidic (pH 3.9) extreme thermal soil in Yellowstone National Park (YNP), Wyoming (USA) for which a thorough chemotaxonomic characterization was published by Botero et al. in 2004 [1]. Although the biological characteristics of the novel strain fulfill all criteria required for the type strain of a novel genus, the proposed name ‘Thermobaculum terrenum’ (= hot small rod belonging to earth/soil) has not yet been validly published (= included in one of the updates of the Validation List that is regularly published in Int J Syst Evol Bacteriol), because rule 30 (3a) of the Bacteriological Code (1990 Revision), which re-
requires that as of 1st January 2001 the description of a new species [...] must include the designation of a type strain, and a viable culture of that strain must be deposited in at least two publicly accessible service collections in different countries from which subcultures must be available [2]. Strain YNP1T is currently deposited only in two US culture collections. Here we present a summary classification and a set of features for ‘T. terrenum’ strain YNP1T, together with the description of the complete genomic sequencing and annotation.

**Classification and features**

Based on analyses of 16S rRNA gene sequences, strain YNP1T is the sole cultured representative of the genus *Thermobaculum*. It has no close relatives among the validly described species within the *Chloroflexi*. The type strain of *Sphaerobacter thermophilus* [3] shares the highest pairwise similarity (84.9%), followed by *Thermoleophilum album* and *T. minutum* [4-6], the two sole members of the actinobacterial order *Thermoleophilia* [7] with 83.6% sequence identity, and three type strains from the clostridial genus *Thermaeobacter* (83.2-83.5%) [8], that are currently not placed within a named family. Only four uncultured bacterial clones in GenBank share a higher degree of sequence similarity with strain YNP1T than the type strain of the ‘closest’ related species, *S. thermophilus*. These are clone DRV-SSB031 from rock varnish in the Whipple Mountains, California (92.1%) [9], and clones AY6.14 (FJ891044), AY6.27 (FJ891057) and AY6.18 (FJ891048) from quartz substrates in the hyperarid core of the Atacama Desert (86.9-87.9%). No phylotypes from environmental screening or metagenomic surveys could be linked to ‘T. terrenum’, indicating a rather rare occurrence in the habitats screened thus far (as of September 2010). A representative genomic 16S rRNA sequence of ‘T. terrenum’ YNP1T was compared using BLAST with the most recent release of the Greengenes database [10] and the relative frequencies of taxa and keywords, weighted by BLAST scores, were determined. The three most frequent genera were *Thermobaculum* (81.2%), *Sphaerobacter* (10.3%) and *Conexibacter* (8.4%). The five most frequent keywords within the labels of environmental samples which yielded hits were ‘microbial’ (3.6%), ‘waste’ (3.3%), ‘soil’ (3.3%), ‘simulated’ (3.2%) and ‘level’ (3.1%). The five most frequent keywords within the labels of environmental samples which yielded hits of a higher score than the highest scoring species were ‘soil’ (4.5%), ‘structure’ (3.3%), ‘simulated’ (3.2%), ‘level/site/waste’ (2.9%) and ‘core’ (2.1%).

Figure 1 shows the phylogenetic neighborhood of ‘T. terrenum’ strain YNP1T in a 16S rRNA based tree. The sequences of the two identical 16S rRNA gene copies in the genome do not differ from the previously published 1,333 nt long partial sequence generated from ATCC BAA-798 (AF391972).

![Figure 1. Phylogenetic tree highlighting the position of ‘T. terrenum’ strain YNP relative to the type strains of the other species within the phylum *Chloroflexi*. The trees were inferred from 1,316 aligned characters [11,12] of the 16S rRNA gene sequence under the maximum likelihood criterion [13] and rooted in accordance with the current taxonomy. The branches are scaled in terms of the expected number of substitutions per site. Numbers above the branches are support values from 1,000 bootstrap replicates [14] if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [15] are shown in blue, published genomes [16] and GenBank records [CP000804,CP000875,CP000909,CP001337] in bold, e.g. the GEBA genome *S. thermophilus* [17].](image-url)
The cells of strain YNP1\textsuperscript{T} are 1-1.5 × 2-3 μm long, non-motile rods (Figure 2 and Table 1), enveloped by a thick cell wall external to a cytoplasmic membrane [1]. YNP1\textsuperscript{T} cells occur singly or in pairs, stain Gram-positive in the exponential growth-phase, are obligately aerobic, and non-spore-forming [1]. Colonies are pink-colored and growth occurs best at pH 6-8 (pH\textsubscript{opt} 7) and 67°C, with a possible temperature range of 41-75°C [1]. Culture doubling time at \( T_{\text{opt}} \) was 4 hours and increases sharply above 70°C, whereas growth at the temperature extremes was relatively poor [1]. Cells grow best in complex media containing 0.5% NaCl and yeast extract (for growth factors) [1], but also on sucrose, fructose, glucose, ribose, xylose, sorbitol, and xylitol [1]. Strain YNP1\textsuperscript{T} was positive for catalase, urease, and nitrate reduction, but tested negative for oxidases, and was also negative for fermentation of glucose or lactose [1]. No anaerobic growth was observed in the presence of sulfate, nitrate, ferric iron, or arsenate as possible electron acceptors [1]. No chemolithoautotrophic growth was observed in an experimental matrix that included the electron donors \( \text{H}_2, \text{H}_2\text{S}, \) or \( \text{S}_0 \) with oxygen as the electron acceptor. Surprisingly, the \textit{in vitro} pH optimum of strain YNP1\textsuperscript{T} (pH 7) is much higher than that of the soil from which it was isolated (pH 4-5) [1]. In pure culture, strain YNP1\textsuperscript{T} failed to grow at such low pH values, suggesting that the thermal soil habitat is not optimal for the strain [1].

**Chemotaxonomy**

Murein is present in large amounts, which is consistent with the observed thick (approximately 34 nm) cell walls with a muramic acid content similar to that of \textit{Bacillus subtilis} [1]. The muramic acid content of strain YNP1\textsuperscript{T} was roughly one quarter of that measured for \textit{B. subtilis} but almost 40-fold greater than in \textit{E. coli} [1]. Lipopolysaccharide (LPS) was not detected [1]. Major fatty acids were dominated by straight and branched chain saturated acids: \( \text{C}_{18:0} \) (27.0%); iso-\( \text{C}_{17:0} \) (11.6%); iso-\( \text{C}_{19:0} \) (12.9%); anteiso-\( \text{C}_{18:0} \) (12.5%); \( \text{C}_{20:0} \) (16.5%) and \( \text{C}_{19:0} \) (6.6%). The pink pigment associated with strain YNP1\textsuperscript{T} exhibited a significant absorption at wavelengths 267, 326, 399, 483, 511, and 549 nm [1].

\[ \text{Figure 2. Transmission electron micrograph of ‘}T. \text{terrenum}’ \text{ strain YNP1}\textsuperscript{T}, scale bar 0.1 \, \mu\text{m} \]
Table 1. Classification and general features of ‘T. terrenum’ strain YNP1T according to the MIGS recommendations [18].

| MIGS ID | Property                | Term                                      | Evidence code |
|---------|-------------------------|-------------------------------------------|---------------|
|         | **Domain** Bacteria     | TAS [19]                                  |               |
|         | **Phylum** Chloroflexi  | TAS [20-23]                               |               |
|         | **Class** Incertae sedis | NAS                                      |               |
| Current classification | **Order** Incertae sedis | NAS                                      |               |
|         | **Family** Incertae sedis | NAS                                      |               |
|         | **Genus** ‘Thermobaculum’ | TAS [1]                                  |               |
|         | **Species** ‘Thermobaculum terrenum’ | TAS [1] |               |
|         | **Type strain** YNP1    | TAS [1]                                  |               |
|         | **Gram stain** positive | TAS [1]                                  |               |
|         | **Cell shape** short rods (1-1.5 × 2-3 µm) | TAS [1] |               |
|         | **Motility** non-motile | TAS [1]                                  |               |
|         | **Sporulation** none    | TAS [1]                                  |               |
|         | **Temperature range** 65°C-92°C | TAS [1] |               |
|         | **Optimum temperature** 67°C | TAS [1] |               |
|         | **Salinity** 0.5% NaCl | TAS [1]                                  |               |
| MIGS-22 | **Oxygen requirement** | obligate aerobic | TAS [1] |
|         | **Carbon source** glucose, fructose, ribose, sorbitol, sucrose, xylose, xylitol | TAS [1] | |
|         | **Energy source** heterotrophic | TAS [1] | |
| MIGS-6  | **Habitat**             | thermal soil                             | TAS [1]       |
| MIGS-15 | **Biotic relationship** | free living                             | NAS           |
| MIGS-14 | **Pathogenicity**       | none                                     | NAS           |
|         | **Biosafety level**     | 1                                        | TAS [24]      |
|         | **Isolation** 65°C hot extreme thermal soil in small depression site | TAS [1] | |
| MIGS-4  | **Geographic location** | Near Rabbit Creek and Firehole River in Yellowstone National Park, Wyoming, USA | TAS [1] |
| MIGS-5  | **Sample collection time** | June 1998                           | NAS           |
| MIGS-4.1| **Latitude**            | 44.394                                   | NAS           |
| MIGS-4.2| **Longitude**           | -110.568                                 | NAS           |
| MIGS-4.3| **Depth**               | 0 m, soil surface                        | TAS [1]       |
| 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 [25]. 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 [26], and is part of the Genomic Encyclopedia of Bacteria and Archaea project [27]. The genome project is deposited in the Genome OnLine Database [15] 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                                                             |
|         | Two genomic libraries:    |                                                                     |
|         | one Sanger 8 kb pMCL200 library, |                                                      |
|         | one 454 pyrosequence standard library |                                                        |
| MIGS-28 | Libraries used            | Two genomic libraries:                                                |
|         | one Sanger 8 kb pMCL200 library, |                                                     |
|         | one 454 pyrosequence standard library |                                                        |
| MIGS-29 | Sequencing platforms      | ABI3730, Illumina GAii, 454 GS FLX                                   |
| MIGS-31.2| Sequencing coverage       | 9.5 × Sanger; 31.8 x pyrosequence                                   |
| MIGS-30 | Assemblers                | Newbler version 1.1.02.15, phrap                                     |
| MIGS-32 | Gene calling method       | Prodigal 1.4, GenePRIMP                                              |
|         | INSDC ID                  | CP0001825 (chromosome 1)                                             |
|         | Genbank Date of Release   | November 23 and 25, 2009                                             |
|         | GOLD ID                   | Gc01150                                                              |
|         | NCBI project ID           | 29523                                                                |
|         | Database: IMG-GEBA        | 2501533217                                                           |
| MIGS-13 | Source material identifier| ATCC BAA-798                                                          |
|         | Project relevance         | Tree of Life, GEBA                                                   |

**Growth conditions and DNA isolation**

*T. terrenum* strain YNP17, ATCC BAA-798, was grown in ATCC medium 1981 (M-R2A medium) [28] at 60°C. The culture used to prepare genomic DNA (gDNA) for sequencing was only two transfers from the original deposit. The purity of the culture was determined by growth on general maintenance media under both aerobic and anaerobic conditions. Cells where harvested after 24 hours by centrifugation and gDNA was extracted from lysozyme-treated cells using CTAB and phenol-chloroform. The purity, quality and size of the bulk gDNA preparation was assessed according to DOE-JGI guidelines. Amplification and partial sequencing of the 16S rRNA gene confirmed the isolate as ‘*T. terrenum*’. The quantity of the DNA was determined on a 1% agarose gel using mass markers of known concentration supplied by JGI. The average fragment size of the purified gDNA determined to be ~43kb by pulsed-field gel electrophoresis.

**Genome sequencing and assembly**

The genome was sequenced using a combination of Sanger and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the [JGI website](http://standardsingenomics.org). Pyrosequencing reads were assembled using the Newbler assembler version 1.1.02.15 (Roche). Large Newbler contigs were broken into 3,926 overlapping fragments of 1,000 bp and entered into assembly as pseudo-reads. The sequences were assigned quality scores based on Newbler consensus q-scores with modifications to account for overlap redundancy and adjust inflated q-scores. A hybrid 454/Sanger assembly was made using the parallel phrap assembler (High Performance Software, LLC). Possible misassemblies were corrected with Dupfinisher or transposon bombing of bridging clones [29]. A total of 432 Sanger finishing reads were produced to close gaps, to resolve repetitive regions, and to raise the quality of the finished sequence. Illumina reads were used to improve the final consensus quality using an in-house developed tool (the Polisher [30]). The error rate of the completed genome sequence is less than 1 in 100,000. Together, the combination of the Sanger and 454 sequencing platforms provided 10.0× coverage of the genome. The final assembly contains 32,920 Sanger reads.

**Genome annotation**

Genes were identified using Prodigal [31] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [32]. 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 [33].
Genome properties
The genome consists of two chromosomes: the low G+C (48%) 2,026,947 bp long chromosome 1, and the high G+C (64%) 1,074,634 bp long chromosome 2 (Table 3, Figure 3, Figure 4). Of the 2,930 genes predicted (1,935 on chromosome 1 and 995 on chromosome 2), 2,872 were protein-coding genes, and 58 RNAs; forty one pseudogenes were also identified. The majority of the protein-coding genes (73.4%) 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.

Figure 3. Graphical circular map of the 2Mb low G+C chromosome 1. From outside to the center: 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 3. Genome Statistics

| Attribute                          | Value    | % of Total |
|------------------------------------|----------|------------|
| Genome size (bp)                   | 3,101,581| 100.00%    |
| DNA coding region (bp)             | 2,825,726| 91.11%     |
| DNA G+C content (bp)               | 1,659,573| 53.51%     |
| Number of replicons                | 2        |            |
| Extrachromosomal elements          | 0        |            |
| Total genes                        | 2,930    | 100.00%    |
| RNA genes                          | 58       | 1.98%      |
| rRNA operons                       | 2        |            |
| Protein-coding genes               | 2,872    | 98.02%     |
| Pseudo genes                       | 41       | 1.40%      |
| Genes with function prediction     | 2,151    | 73.41%     |
| Genes in paralog clusters          | 439      | 14.98%     |
| Genes assigned to COGs             | 2,223    | 75.78%     |
| Genes assigned Pfam domains        | 2,308    | 78.77%     |
| Genes with signal peptides         | 573      | 19.56%     |
| Genes with transmembrane helices   | 777      | 26.52%     |
| CRISPR repeats                     | 6        |            |

### Figure 4. Graphical circular map of the 1 Mb high-G+C chromosome 2. From outside to the center: 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.
‘Thermobaculum terrenum’ type strain (YNP1T)

| Code | Value | %age | Description |
|------|-------|------|-------------|
| J    | 139   | 5.6  | Translation, ribosomal structure and biogenesis |
| A    | 0     | 0.0  | RNA processing and modification |
| K    | 158   | 6.4  | Transcription |
| L    | 110   | 4.5  | Replication, recombination and repair |
| B    | 1     | 0.0  | Chromatin structure and dynamics |
| D    | 18    | 0.7  | Cell cycle control, cell division, chromosome partitioning |
| Y    | 0     | 0.0  | Nuclear structure |
| V    | 56    | 2.3  | Defense mechanisms |
| T    | 106   | 4.3  | Signal transduction mechanisms |
| M    | 131   | 5.3  | Cell wall/membrane/envelope biogenesis |
| N    | 2     | 0.1  | Cell motility |
| Z    | 0     | 0.0  | Cytoskeleton |
| W    | 0     | 0.0  | Extracellular structures |
| U    | 35    | 1.4  | Intracellular trafficking and secretion, and vesicular transport |
| O    | 105   | 4.3  | Posttranslational modification, protein turnover, chaperones |
| C    | 160   | 6.5  | Energy production and conversion |
| G    | 325   | 13.2 | Carbohydrate transport and metabolism |
| E    | 206   | 8.4  | Amino acid transport and metabolism |
| F    | 60    | 2.4  | Nucleotide transport and metabolism |
| H    | 128   | 5.2  | Coenzyme transport and metabolism |
| I    | 67    | 2.7  | Lipid transport and metabolism |
| P    | 134   | 5.4  | Inorganic ion transport and metabolism |
| Q    | 44    | 1.8  | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 316   | 12.8 | General function prediction only |
| S    | 162   | 6.6  | Function unknown |
| -    | 707   | 24.1 | Not in COGs |

Acknowledgements

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.

References

1. Botero LM, Brown KB, Brumefield S, Burr M, Castenholz RW, Young M, McDermott TR. Thermobaculum terrenum gen. nov., sp. nov.; a non-phototrophic gram-positive thermophile representing an environmental clone group related to the Chloroflexi (green non-sulfur bacteria) and Thermomicrobia. Arch Microbiol 2004; 181:269-277. PubMed doi:10.1007/s00203-004-0647-7

2. Lapage SP, Sneath PHA, L essel EF, Skerman VBD, Seeliger HPR, Clark WA. International Code of Nomenclature of Bacteria (1990) Revision. American Society for Microbiology, Washington, DC, 1992.

3. Demharter W, Hensel R, Smida J, Stackebrandt E. Sphaerobacter thermophilus gen. nov., sp. nov. A deeply rooting member of the actinomycetes subdivision isolated from thermophilically treated sewage sludge. Syst Appl Microbiol 1989; 11:261-266.

4. Zarilla KA, Perry JJ. Thermoleophilum album gen. nov. and sp. nov., a bacterium obligately for
thermophily and n-alkane substrates. *Arch Microbiol* 1984; 137:286-290. doi:10.1007/BF00410723

5. Zarilla KA, Perry JJ. Deoxyribonucleic acid homology and other comparisons among obligately thermophilic hydrocarbonoclastic bacteria, with a proposal for *Thermoleophilum minutum* sp. nov. *Int J Syst Bacteriol* 1986; 36:13-16. doi:10.1099/00270737-13-1-13

6. List editor. Validation list No. 20. *Int J Syst Evol Microbiol* 1986; 36:354-356.

7. Reddy GSN, Garcia-Pichel F. Description of *Patulibacter americanus* sp. nov., isolated from biological soil crusts, emended description of the genus *Patulibacter Takashahi* et al. 2006 and proposal of *Solirubrobacterales* ord. nov. and *Thermoleophilales* ord. nov. *Int J Syst Evol Microbiol* 2009; 59:87-94. PubMed doi:10.1099/ijs.0.64185-0

8. Chun J, Lee JH, Jung Y, Kim M, Kim S, Kim BK, Lim YW. EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int J Syst Evol Microbiol* 2007; 57:2259-2261. PubMed doi:10.1099/ijs.0.64915-0

9. Kuhlman KR, Fusco WG, La Duc MT, Allenbach LB, Ball CL, Kuhlman GM, Anderson RC, Erickson IK, Stuecker T, Benardini J, et al. Diversity of microorganisms within rock varnish in the Whipple Mountains, California. *Appl Environ Microbiol* 2006; 72:1708-1715. PubMed doi:10.1128/AEM.72.2.1708-1715.2006

10. 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 doi:10.1128/AEM.03006-05

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

12. 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

13. 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

14. 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

15. 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 doi:10.1093/nar/gkp848

16. Wu D, Raymond J, Wu M, Chatterji S, Ren Q, Graham JE, Bryant DA, Robb F, Colman A, Tallon LJ, et al. Complete genome sequence of the aerobic CO-oxidizing thermophile *Thermomicrobiurn roseum*. *PLoS ONE* 2009; 4:e4207. PubMed doi:10.1371/journal.pone.0004207

17. Pati A, LaButti K, Pukall R, Nolan M, Glavina del Rio T, Tice H, Cheng JF, Lucas S, Chen F, Lucas S, et al. Complete genome sequence of *Sphaerobacter thermophilus* type strain (S 6022\(^{+}\)). *Stand Genomic Sci* 2010; 2:49-56. doi:10.4056/sigs.601105

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 doi:10.1038/nbt1360

19. 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

20. Garrity GM, Holt JG. Phylum BVI. *Chlorolexi* phy. nov. In: DR Boone, RW Castenholz, GM Garrity (eds): Bergey’s Manual of Systematic Bacteriology, second edition, vol. 1 (The *Archaea* and the deeply branching and phototrophic *Bacteria*), Springer-Verlag, New York, 2001, p. 427-446.

21. List Editor. Validation List no. 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

22. 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.
23. Hugenholtz P, Stackebrandt E. Reclassification of *Sphaerobacter thermophilus* from the subclass *Sphaerobacteridae* in the phylum *Actinobacteria* to the class *Thermomicrobia* (emended description) in the phylum *Chloroflexi* (emended description). *Int J Syst Evol Microbiol* 2004; 54:2049-2051. PubMed doi:10.1099/ijs.0.03028-0
24. LGC Advanced Catalogue Search .
25. 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
26. 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
27. 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 doi:10.1038/nature08656
28. http://www.lgcstandards-atcc.org/Attachments/3145.pdf.
29. Sims D, Brettin T, Detter JC, 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. doi:10.4056/sigs.761
30. Lapidus A, LaButti K, Foster B, Lowry S, Trong S, Golsman E. POLISHER: An effective tool for using ultra short reads in microbial genome assembly and finishing. AGBT, Marco Island, FL, 2008.
31. Hyatt D, Chen GL, Locascio PF, Land ML, Laplante 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
32. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. Gene-PRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; 7:455-457. PubMed doi:10.1038/nmeth.1457
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 doi:10.1093/bioinformatics/btp393