Complete genome sequence of *Oceanithermus profundus* type strain (506\(^\dagger\))

Amrita Pati\(^1\), Xiaojing Zhang\(^2\), Alla Lapidus\(^1\), Matt Nolan\(^1\), Susan Lucas\(^1\), Tijana Glavina Del Río\(^1\), Hope Tice\(^1\), Jan-Fang Cheng\(^1\), Roxane Tapia\(^1,2\), Cliff Han\(^1,2\), Lynne Goodwin\(^1,2\), Sam Pitluck\(^1\), Konstantinos Liolios\(^6\), Ioanna Pagani\(^1\), Natalia Ivanova\(^1\), Konstantinos Mavromatis\(^3\), Amy Chen\(^1\), Krishna Palaniappan\(^1\), Loren Hauser\(^1,4\), Cynthia D. Jeffries\(^1,4\), Evelyne-Marie Brambilla\(^1\), Alina Röhl\(^6\), Romano Mwiricha\(^1\), Manfred Rohde\(^6\), Brian J. Tindall\(^5\), Johannes Sikorski\(^2\), Reinhard Wirth\(^6\), Markus Gäker\(^3\), Tanja Woyke\(^1\), John C. Detter\(^1,2\), James Bristow\(^1\), Jonathan A. Eisen\(^1,9\), Victor Markowitz\(^3\), Philip Hugenholtz\(^1,10\), Nikos C. Kyrpides\(^1\), Hans-Peter Klenk\(^5\)*, and Miriam Land\(^1,4\)

\(^1\) DOE Joint Genome Institute, Walnut Creek, California, USA
\(^2\) Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
\(^3\) Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
\(^4\) Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
\(^5\) DSMZ – German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany
\(^6\) University of Regensburg, Microbiology – Archaeenzentrum, Regensburg, Germany
\(^7\) Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
\(^8\) HZI – Helmholtz Centre for Infection Research, Braunschweig, 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**: microaerophilic, non-motile, Gram-negative, nitrate-reducing, moderate thermophilic, neutrophilic, chemolithoheterotrophic, hydrothermal vent, *Thermaceae*, GEBA

*Oceanithermus profundus* Miroshnichenko et al. 2003 is the type species of the genus *Oceanithermus*, which belongs to the family *Thermaceae*. The genus currently comprises two species whose members are thermophilic and are able to reduce sulfur compounds and nitrite. The organism is adapted to the salinity of sea water, is able to utilize a broad range of carbohydrates, some proteinaceous substrates, organic acids and alcohols. This is the first completed genome sequence of a member of the genus *Oceanithermus* and the fourth sequence from the family *Thermaceae*. The 2,439,291 bp long genome with its 2,391 protein-coding and 54 RNA genes consists of one chromosome and a 135,351 bp long plasmid, and is a part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

**Introduction**

Strain 506\(^\dagger\) (DSM 14977 = NBRC 100410 = VKM B-2274) is the type strain of *Oceanithermus profundus*, which is the type species of the genus *Oceanithermus* [1] of the family *Thermaceae* [2]. Together with *O. desulfurans*, there are currently two species placed in the genus [1,3]. The generic name derives from the Latin noun oceanus, meaning ocean and the Neo-Latin masc. substantive (from Gr. adj. thermos) thermus which means hot. Therefore, the name *Oceanithermus* refers to warmth-loving organisms living in the ocean. The species epithet is derived from the Latin adjective *profundus* meaning deep, which means pertaining to the abyss, pertaining to the depths of the ocean [1]. Strain 506\(^\dagger\) was first isolated from samples of hydrothermal fluids and chimneys collected at the 13ºN hydrothermal vent field on the East Pacific Rise at a depth of 2600 m [1]. There are no further cultivated strains of this species known. The other member of the genus, *O. desulfurans*, is a thermophilic, sulfur-reducing bacterium isolated from a sulfide chimney in Suiyo Seamount, in the Western Pacific [3]. Here we present a summary classification and a set of features for *O. profundus* 506\(^\dagger\),
together with the description of the complete genomic sequencing and annotation.

**Classification and features**

A representative genomic 16S rRNA sequence of strain 506T was compared using NCBI BLAST 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 [4] and the relative frequencies, weighted by BLAST scores, of taxa and keywords (reduced to their stem) [5] were determined. The five most frequent genera were *Thermus* (52.0%), *Meiothermus* (37.0%), *Oceanithermus* (7.6%), *Marinithermus* (2.0%) and *Vulcanithermus* (1.4%) (156 hits in total). Regarding the four hits to sequences from members of the species, the average identity within HSPs was 99.6%, whereas the average coverage by HSPs was 94.8%. Regarding the two hits to sequences from other members of the genus, the average identity within HSPs was 99.3%, whereas the average coverage by HSPs was 91.0%. Among all other species, the one yielding the highest score was *O. desulfuricans*, which corresponded to an identity of 99.3% and an HSP coverage of 91.0%. The highest-scoring environmental sequence was EU555123 (*Microbial Sulfide Hydrothermal Vent Field Juan de Fuca Ridge Dudley hydothermal vent clone 4132B16*), which showed an identity of 99.1% and an HSP coverage of 98.0%. The five most frequent keywords within the labels of environmental samples which yielded hits were 'spring' (8.2%), 'hot' (6.2%), 'microbi' (4.5%), 'geochem, nation, park, yellowston' (2.8%) and 'hydrotherm/vent' (2.5%) (94 hits in total). The five most frequent keywords within the labels of environmental samples which yielded hits of a higher score than the highest scoring species were 'hydrotherm/vent' (12.2%), 'field, microbi, ridg' (6.1%), 'fluid' (5.9%), 'dudlei, fuca, juan, sulfid' (3.1%) and 'degre, east, north, ocean, pacif, rise' (3.0%) (3 hits in total). These 16S BLAST results are a confirmation of the kind of environment from which the living strain was isolated and therefore fits the description of the isolate.

Figure 1 shows the phylogenetic neighborhood of *O. profundus* in a 16S rRNA based tree. The sequences of the two identical 16S rRNA gene copies in the genome differ by one nucleotide from the previously published 16S rRNA sequence (AJ430586).

The cells of *O. profundus* are described as non-motile, rod-shaped, 0.5 – 0.7 µm in diameter and of various lengths (Figure 2). When grown on proteinaceous substrates, old cultures of *O. profundus* form filaments and large spheres resembling the ‘rotund bodies’ typical of aged cells of *Thermus* species [1,15]. The organism is Gram-negative and non spore-forming (Table 1).

*O. profundus* is microaerophilic, only being able to grow at oxygen concentrations below 6% [1]. No growth has been observed in an atmosphere of air, either in liquid medium or on plates. In an agar tube containing 5 ml of basal medium supplemented with 2 g sucrose and 1 g tryptone per liter with air in the headspace (10 ml), growth occurs in a zone located 20 mm below the agar/air interface [1]. Alternatively, the organism grows anaerobically using nitrate as the electron acceptor. *O. profundus* grows within a temperature range of 40-68°C, optimal growth being observed at 60°C. At 60°C, it grows between pH 5.5 and 8.4, with an optimum around 7.5 [1]. Strain 506T grows at NaCl concentrations ranging from 10 to 50 g/l, with an optimum at 30 g/l [1]. The organism is oxidase- and catalase positive and is able to utilize a wide spectrum of carbohydrates in the presence of either nitrate or oxygen [1]. The highest cell yield is observed in the presence of nitrate with fructose, maltose, sucrose, trehalose, galactose, rhamnose or xylose. Glucose, lactose and starch are utilized, but no growth has been reported with ribose, galactose, arabinose, dextrin or cellobiose [1]. Acetate and propionate are produced during growth with sucrose as a growth substrate and nitrate as the electron acceptor. Nitrite is the only product of denitrification [1]. *O. profundus* grows well with complex proteinaceous substrates such as beef extract, tryptone or papai digest of soybean (1-1.5 g/l). However, growth is strongly inhibited by higher concentrations of these substrates [1]. The isolate does not grow with Casamino acids or yeast extract as sole sources of carbon and energy, though 100 mg/l yeast extract is required for growth [1]. *O. profundus* is able to utilize acetate, pyruvate and propionate as growth substrates. It also grows with methanol, ethanol and mannitol, though the cell yield is lower [1]. *O. profundus* is able to grow lithoheterotrophically using molecular hydrogen as the energy source, yeast extract as the carbon source and nitrate as the electron acceptor. Other electron acceptors (sulfate, elemental sulfur, thiosulfate and nitrite) do not support growth, regardless of growth substrate [1]. Detailed studies on the metabolism of maltose, acetate, pyruvate, and hydrogen have been undertaken by Fedosov et al. [26].

http://standardsingenomics.org
Figure 1. Phylogenetic tree highlighting the position of *O. profundus* relative to the other type strains within the family *Thermaceae*. The tree was inferred from 1,420 aligned characters [6,7] of the 16S rRNA gene sequence under the maximum likelihood criterion [8]. Rooting was initially done using the midpoint method [9] and then checked for its accordance with the current taxonomy (see Table 1) and rooted accordingly. The branches are scaled in terms of the expected number of substitutions per site. Numbers to the right of bifurcations are support values from 1,000 bootstrap replicates [10] if larger than 60%. Lineages with type strain genome sequencing projects that are registered in GOLD [11] but remain unpublished are labeled with one asterisk, published genomes with two asterisks [12-14].

Figure 2. Scanning electron micrograph of *O. profundus* 506T
Table 1. Classification and general features of *O. profundus* 506\(^{T}\) according to the MIGS recommendations [16].

| MIGS ID | Property                  | Term                                         | Evidence code |
|---------|---------------------------|----------------------------------------------|---------------|
|         | **Current classification**| Domain *Bacteria*                            | TAS [17]      |
|         |                           | Phylum “*Deinococcus-Thermus*”                | TAS [18,19]   |
|         |                           | Class *Deinococci*                           | TAS [20,21]   |
|         |                           | Order *Thermales*                            | TAS [21,22]   |
|         |                           | Family *Thermaceae*                          | TAS [21,23]   |
|         |                           | Genus *Oceanithermus*                        | TAS [1]       |
|         |                           | Species *Oceanithermus profundus*            | TAS [1]       |
|         |                           | Type strain 506                             | TAS [1]       |
|         | Gram stain                | negative                                     | TAS [1]       |
|         | Cell shape                | rod-shaped                                   | TAS [1]       |
|         | Motility                  | non-motile                                   | TAS [1]       |
|         | Sporulation               | none                                         | TAS [1]       |
|         | Temperature range          | 40-68°C                                      | TAS [1]       |
|         | Optimum temperature       | 60°C                                         | TAS [1]       |
|         | Salinity                  | 1%-5%, optimum 3% NaCl                       | TAS [1]       |
| MIGS-22 | Oxygen requirement        | microaerophile                               | TAS [1]       |
|         | Carbon source             | carbohydrates                                | TAS [1]       |
|         | Energy metabolism         | chemoorganoheterotroph, lithoheterotroph, organotroph | TAS [1]       |
| MIGS-6  | Habitat                   | deep sea, hydrothermal vent, marine          | TAS [1]       |
| MIGS-15 | Biotic relationship       | free-living                                  | TAS [1]       |
| MIGS-14 | Pathogenicity             | none                                         | NAS [24]      |
|         | Biosafety level           | 1                                            | NAS [24]      |
|         | Isolation                 | deep-sea hot vent                            | TAS [1]       |
| MIGS-4  | Geographic location       | East Pacific Rise                            | TAS [1]       |
| MIGS-5  | Sample collection time    | 1999                                         | TAS [1]       |
| MIGS-4.1| Latitude                  | 12.8                                         | TAS [1]       |
| MIGS-4.2| Longitude                 | 103.93                                       | TAS [1]       |
| MIGS-4.3| Depth                    | 2,600 m                                      | TAS [1]       |
| MIGS-4.4| Altitude                 | -2,600 m                                     | NAS           |

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.

Chemotaxonomy

The polar lipid pattern of strain 506\(^{T}\) comprises three phospholipids, whereas glycolipids have not been detected [1]. This differentiates the organism from members of the genera *Vulcanithermus, Rhabdothermus, Thermus* and *Meiothermus*, where phospholipids and glycolipids have both been detected [27,28]. It should be noted that the major phospholipid detected in *O. profundus* has the same \(R_f\) and staining behavior as the \(2'\)-O-(1, 2-diacyl-sn-glycero-3-phospho)-3'-O-(\(\alpha\)-N-acetyl-glucosaminy1)-N-glyceroalkylamine reported to occur in members of the genera *Meiothermus* and *Thermus* [29]. On the basis of \(R_f\) value and staining behavior this lipid also appears to be present in

http://standardsingenomics.org
members of the genera *Vulcanithermus* and *Rhabdothermus*, which also synthesize glycolipids [30,31] Although members of the genus *Deinococcus* may also produce glycolipids in addition to a novel series of phosphoglycolipids [32,33] the latter are absent in members of the genera *Thermus* and *Meiothermus*. The absence of glycolipids was one of the arguments for Miroshnichenko et al. for placing strain 506T in a new genus [1].

Menaquinones are the sole respiratory lipoprenones detected, with MK-8 predominating (95%) and MK-9 being present in smaller proportions (5%) [1]. The predominance of MK-8 is consistent with reports of MK-8 in members of the genera *Thermus, Meiothermus* [34,35], *Marinithermus* [36], *Vulcanithermus, Rhabdothermus, Truerepa, Deinobacterium* and *Deinococcus* [30-33,37]. However, the presence of MK-9, albeit at only 5%, appears to be a unique feature of *O. profundus*.

The fatty acids comprise mainly iso- and anteiso-branched fatty acids though iso-unsaturated fatty acids are also present [1]. The major fatty acids are iso-C₁₅:₁ω7 (7.7%), iso-C₁₅:₀ (33.2%), iso-C₁₆:₁ω8 (2.6 iso-C₁₆:₀ (3.3%), iso-C₁₇:₁ω7c (18.8%), iso-C₁₇:₀ (12.3%), anteiso-C₁₅:₀ (5.1%) and anteiso-C₁₇:₀ (5.4%) [1]. The presence of iso- and anteiso-branched fatty acids is a feature of members of the genera *Deinococcus, Thermus, Meiothermus, Vulcanithermus, Rhabdothermus* and *Marinithermus* [27,28,30-34,37]. The presence of unsaturated branched-chain fatty acids is a distinctive feature of members of the genera *Oceanithermus, Vulcanithermus* and *Rhabdothermus* within the family *Thermaceae*. The unsaturated fatty acid content of the isolate is also higher (33-37%) as compared to the closest relative *O. desulfurans* (18%) [3].

**Genome sequencing and annotation**

**Genome project history**

This organism was selected for sequencing on the basis of its phylogenetic position [38] and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [39]. The genome project is deposited in the Genome On Line Database [11] 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**

*O. profundus* strain 506, DSM 14977, was grown anaerobically in DSMZ medium 975 (*Oceanithermus profundus* medium) [40] at 60°C. DNA was isolated from 0.5-1 g of cell paste using Jetflex Genomic DNA Purification Kit following the standard protocol as recommended by the manufacturer, but with an additional proteinase K (20 μl) digestion for 45 min at 58°C. DNA is available through the DNA Bank Network [41].

**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 [42]. Pyrosequencing reads were assembled using the Newbler assembler version 2.3-PreRelease-8-23-2009 (Roche). The initial Newbler assembly, consisting of nine contigs in four scaffolds, was converted into a phrap assembly by [43] making fake reads from the consensus, to collect the read pairs in the 454 paired end library. Illumina GAii sequencing data (208 Mb) was assembled with Velvet [44] 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 306.1 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 [43] 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 [42], Dupfinisher, or sequencing cloned bridging PCR fragments with subcloning or transposon bombing (Epícentre Biotechnologies, Madison, WI) [45]. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F.Chang, unpublished). A total of 177 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 [46]. 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 282.8 × coverage of the genome. The final assembly contained 1,258,374 pyrosequence and 5,792,221 Illumina reads.
### Table 2. Genome sequencing project information

| MIGS ID | Property                  | Term                                                                 |
|---------|---------------------------|----------------------------------------------------------------------|
| MIGS-31 | Finishing quality         | Finished                                                             |
| MIGS-28 | Libraries used            | Three genomic libraries: one 454 pyrosequence standard library, one 454 PE library (17 kb insert size), one Illumina library |
| MIGS-29 | Sequencing platforms      | Illumina GAII, 454 GS FLX Titanium                                    |
| MIGS-31.2 | Sequencing coverage    | 85.5 × Illumina; 197.3 × pyrosequence                                 |
| MIGS-30 | Assemblers                | Newbler version 2.3-PreRelease-8-23-2009, Velvet, phrap               |
| MIGS-32 | Gene calling method       | Prodigal 1.4, GenePRIMP                                             |
|         |                           | CP002361 chromosome                                                  |
|         |                           | CP002362 plasmid OCEPR01                                             |
| INSDC ID |                           |                                                                     |
| Genbank Date of Release | December 7, 2010         |                                                                     |
| GOLD ID  |                           | Gc01553                                                             |
| NCBI project ID | 40223                  |                                                                     |
| Database: IMG-GEBA | 2503508010              |                                                                     |
| MIGS-13 | Source material identifier | DSM 14977                                                           |
| Project relevance | Tree of Life, GEBA    |                                                                     |

### Genome annotation

Genes were identified using Prodigal [47] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [48]. 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) [49].

### Genome properties

The genome consists of a 2,303,940 bp long chromosome with a G+C content of 70% and a 135,351 bp plasmid with a G+C content of 66% (Table 3 and Figure 3). Of the 2,445 genes predicted, 2,391 were protein-coding genes, and 54 RNAs; 18 pseudogenes were also identified. The majority of the protein-coding genes (69.9%) 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)               | 2,439,291   | 100.00%    |
| DNA coding region (bp)         | 2,265,747   | 92.89%     |
| DNA G+C content (bp)           | 1,702,985   | 69.81%     |
| Number of replicons            | 2           |            |
| Extrachromosomal elements      | 1           |            |
| Total genes                    | 2,445       | 100.00%    |
| RNA genes                      | 54          | 2.21%      |
| rRNA operons                   | 2           |            |
| Protein-coding genes           | 2,391       | 97.79%     |
| Pseudo genes                   | 18          | 0.74%      |
| Genes with function prediction | 1,790       | 69.90%     |
| Genes in paralog clusters      | 25          | 1.02%      |
| Genes assigned to COGs         | 1,772       | 72.47%     |
| Genes assigned Pfam domains    | 1,842       | 75.34%     |
| Genes with signal peptides     | 615         | 25.15%     |
| Genes with transmembrane helices| 654         | 26.75%     |
| CRISPR repeats                 | 0           |            |
**Figure 3.** Graphical circular map of chromosome (map of plasmid not shown). 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 4. Number of genes associated with the general COG functional categories

| Code | value | %age  | Description                                           |
|------|-------|-------|-------------------------------------------------------|
| J    | 150   | 7.7%  | Translation, ribosomal structure and biogenesis       |
| A    | 1     | 0.0%  | RNA processing and modification                       |
| K    | 90    | 4.6%  | Transcription                                         |
| L    | 91    | 4.7%  | Replication, recombination and repair                 |
| B    | 1     | 0.0%  | Chromatin structure and dynamics                      |
| D    | 27    | 1.4%  | Cell cycle control, cell division, chromosome partitioning |
| Y    | 0     | 0.0%  | Nuclear structure                                     |
| V    | 31    | 1.6%  | Defense mechanisms                                    |
| T    | 80    | 4.1%  | Signal transduction mechanisms                        |
| M    | 79    | 4.1%  | Cell wall/membrane/envelope biogenesis                |
| N    | 23    | 1.2%  | Cell motility                                         |
| Z    | 0     | 0.0%  | Cytoskeleton                                          |
| W    | 0     | 0.0%  | Extracellular structures                              |
| U    | 47    | 2.4%  | Intracellular trafficking, secretion, and vesicular transport |
| O    | 82    | 4.2%  | Posttranslational modification, protein turnover, chaperones |
| C    | 154   | 7.9%  | Energy production and conversion                      |
| G    | 125   | 6.4%  | Carbohydrate transport and metabolism                 |
| E    | 203   | 10.4% | Amino acid transport and metabolism                   |
| F    | 72    | 3.7%  | Nucleotide transport and metabolism                   |
| H    | 93    | 4.8%  | Coenzyme transport and metabolism                     |
| I    | 66    | 3.4%  | Lipid transport and metabolism                        |
| P    | 100   | 5.1%  | Inorganic ion transport and metabolism                |
| Q    | 31    | 1.6%  | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 244   | 12.5% | General function prediction only                      |
| S    | 155   | 8.0%  | Function unknown                                      |
| -    | 673   | 27.6% | 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, as well as German Research Foundation (DFG) INST 599/1-2.

References
1. Miroshnichenko ML, L’haridon S, Jeanthon C, Antipov AN, Kostrikina NA, Tindall BJ, Schumann P, Spring S, Stackebrandt E, Bonch-Osmolovskaya EA. Oceanithermus profundus gen. nov., sp. nov., a thermophilic, microaerophilic, facultatively chemolithotrophic bacterium from a deep-sea hydrothermal vent. Int J Syst Evol Microbiol 2003; 53:747-752. PubMed doi:10.1099/ijs.0.02367-0
2. Da Costa MS, Rainey FA. Family I. Thermaceae fam. nov. In: Bergey’s Manual of Systematic Bacteriology 2001. Boone DR, Castenholz RW, Garrity GM (eds), 2nd edn, vol. 1, pp. 403-404. New York: Springer.
3. Mori K, Kakegawa T, Higashi Y, Nakamura K, Maruyama A, Hanada S. Oceanithermus desulfurans sp. nov., a novel thermophilic, sulfur-

http://standardsingenomics.org
reducing bacterium isolated from a sulfide chimney in Suiyo Seamount. *Int J Syst Evol Microbiol* 2004; **54**:1561-1566. PubMed doi:10.1099/ijs.0.02962-0

4. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brotie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL, Greenengenes, 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

5. Porter MF. An algorithm for suffix stripping. *Program: electronic library and information systems* 1980; **14**:130-137.

6. 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

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

8. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol* 2008; **57**:758-771. PubMed doi:10.1080/10635150802429780

9. Hess PN, De Moraes Russo CA. An empirical test of the midpoint rooting method. *Biol J Linn Soc Lond* 2007; **92**:669-674. doi:10.1111/j.1095-8312.2007.00864.x

10. 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

11. 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

12. Sikorski J, Tindall BJ, Lowry S, Lucas S, Nolan M, Copeland A, Rio TGD, Tice H, Cheng JF, Han C, et al. Complete genome sequence of *Meiothermus silvanus* type strain (VI-R2¹). *Stand Genomic Sci* 2010; **3**:37-46. PubMed doi:10.4056/sigs.1042812

13. Tindall BJ, Sikorski J, Lucas S, Goltsman E, Copeland A, Rio TGD, Nolan M, Tice H, Cheng JF, Han C, et al. Complete genome sequence of *Meiothermus ruber* type strain (21¹). *Stand Genomic Sci* 2010; **3**:26-36. PubMed doi:10.4056/sigs.1032748

14. Ivanova N, Rohde C, Munk C, Nolan M, Lucas S, Glavina Del Rio T, Tice H, Deshpande S, Cheng JF, Tapia R, et al. Complete genome sequence of *Truepera radiovictrix* type strain (RQ-24¹). *Stand Genomic Sci* 2011; **4**:91-99. PubMed doi:10.4056/sigs.1563919

15. Brock TD, Edwards MR. Fine structure of *Thermus aquaticus*, an extreme thermophile. *J Bacteriol* 1970; **104**:509-517. PubMed

16. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angioli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. PubMed doi:10.1038/nbt1360

17. 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**:5567-5573. PubMed doi:10.1073/pnas.87.12.5567

18. Garrity GM, Lilburn TG, Cole JR, Harrison SH, Ezubey J, Tindall BJ. Taxonomic outline of the *Bacteria* and *Archaea*, Release 7.7 March 6, 2007. Part 2 - The Bacteria: Phyla *'Aquificae', 'Thermotogae', 'Thermodesulfobacteria', 'Deinococcus-Thermus', 'Chrysiogenetes', 'Chloroflexi', 'Thermomicrobia', 'Nitrospira', 'Deferribacteres', 'Cyanobacteria*', and *'Chlorobi'*. http://www.taxonomicoutline.org/index.php/toba/article/view/187/211 2007.

19. Weisburg WG, Giovannoni SJ, Woese CR. The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. *Syst Appl Microbiol* 1989; **11**:128-134. PubMed

20. Garrity GM, Holt JG. Class I. *Deinococcus* class. *In*: Garrity GM, Boone DR, Castenholz RW (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, 2001, p. 395.

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. Rainey FA, da Costa MS. Order II. *Thermales* ord. *In*: Garrity GM, Boone DR, Castenholz RW (eds), Bergey's Manual of Systematic Bacteriolo-
23. da Costa MS, Rainey FA. Family I. Thermaceae fam. nov. In: Garrity GM, Boone DR, Castenholz RW (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 1, Springer, New York, 2001, p. 403.

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

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. Fedosov DV, Podkopaeva DA, Miroshnichenko ML, Bonch-Osmolovskaya EA, Lebedinsky AV, Grabovich MY. Metabolism of the thermophilic bacterium Oceanithermus profundus. Microbiology 2008; 77:159-165. PubMed doi:10.1134/S0026261708020069

27. Donato MM, Seleiro EA, Da Costa MS. Polar lipid and fatty acid composition of strains of the genus Thermus. Syst Appl Microbiol 1990; 13:234-239.

28. Donato MM, Seleiro EA, Da Costa MS. Polar lipid and fatty acid composition of strains of Thermus ruber. Syst Appl Microbiol 1991; 14:235-239.

29. Yang YL, Yang FL, Jao SC, Chen MY, Tsay SS, Zou W, Wu SH. Structural elucidation of phosphoglycolipids from strains of the bacterial thermophiles Thermus and Meiothermus. J Lipid Res 2006; 47:1823-1832. PubMed doi:10.1194/jlr.M600034-JLR200

30. Steinsbu BO, Tindall BJ, Torsvik VL, Ingunn H, Thorseth IH, Daae FL, Pedersen RB. Rhabdothermus arcticus gen. nov., sp. nov., a novel member of the family Thermaceae isolated from a hydrothermal vent chimney from Soria Moria vent field at the Arctic Mid-Ocean Ridge. Int J Syst Evol Microbiol 2010

31. Miroshnichenko ML, L’Haridon S, Nercuessian O, Antipov AN, Kostrikina NA, Tindall BJ, Schumann P, Spring S, Stackebrandt E, Bonch-Osmolovskaya EA, Jeannot C. Vulcanithermus mediatlanticus gen. nov., sp. nov., a novel member of the family Thermaceae from a deep-sea hot vent. Int J Syst Evol Microbiol 2003; 53:1143-1148. PubMed doi:10.1099/ijs.0.02579-0

32. Embley TM, O’Donnell AG, Wait R, Rostron J. Lipid and cell wall amino acid composition in the classification of members of the genus Deinococcus. Syst Appl Microbiol 1987; 10:20-27.

33. Ferreira AC, Nobre MF, Rainey FA, Silva MT, Wait R, Burghardt J, Chung AP, Da Costa MS. Deinococcus geothermalis sp. nov. and Deinococcus murrayi sp. nov., two extremely radiation-resistant and slightly thermophilic species from hot springs. Int J Syst Bacteriol 1997; 47:939-947. PubMed doi:10.1099/00207713-47-4-939

34. Hensel R, Demharter W, Kandler O, Kroppenstedt RM, Stackebrandt E. Chemotaxonomic and molecular-genetic studies of the genus Thermus: evidence for a phylogenetic relationship of Thermus aquaticus and Thermus ruber to the genus Deinococcus. Int J Syst Bacteriol 1986; 36:444-453. doi:10.1099/00207713-36-3-444

35. Chung AP, Rainey F, Nobre MF, Burghardt J, Da Costa MS. Meiothermus cerbereus sp. nov., a new slightly thermophilic species with high levels of 3-hydroxy fatty acids. Int J Syst Bacteriol 1997; 47:1225-1230. PubMed doi:10.1099/ijs.0.0207713-47-4-1225

36. Sako Y, Nakagawa S, Takai K, Horikoshi K. Marinithermus hydrothermalis gen. nov., sp. nov., a strictly aerobic, thermophilic bacterium from a deep-sea hydrothermal vent chimney. Int J Syst Evol Microbiol 2003; 53:59-65. PubMed doi:10.1099/ijs.0.02364-0

37. Ekman JV, Raulio M, Busse HJ, Fewer DP, Salkinoja-Salonen M. Deinobacterium chartae gen. nov., sp. nov., an extremely radiation-resistant, biofilm-forming bacterium isolated from a Finnish paper mill. Int J Syst Evol Microbiol 2011; 61:540-548. PubMed doi:10.1099/ijs.0.017970-0

38. 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

39. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Glöckner FO, Seedorf H, Yilmaz P, Castenholz RW, Rutgers van der Loeff MS, Woese CR, Qian X, Rohwer F, Hugenholtz P. The microbial world: a framework for the phylogenetic classification of non-model microorganisms. PLoS Biol 2007; 5:e74. PubMed doi:10.1371/journal.pbio.0050074

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

41. Gemeinholzer B, Dröge G, Zetzche 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

http://standardsingenomics.org
Oceanithermus profundus type strain (506T)

42. DOE Joint Genome Institute. http://www.jgi.doe.gov

43. Phrap and Phred for Windows. MacOS, Linux, and Unix. http://www.phrap.com

44. 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

45. 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. PubMed doi:10.4056/sigs.761

46. 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.

47. 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

48. 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 doi:10.1038/nmeth.1457

49. 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