Title
Complete genome sequence of Syntrophobotulus glycolicus type strain (FIGlyRT)

Permalink
https://escholarship.org/uc/item/2w5118qp

Journal
Environmental Microbiome, 4(3)

ISSN
2524-6372

Authors
Han, Cliff
Mwirichia, Romano
Chertkov, Olga
et al.

Publication Date
2011-07-01

DOI
10.4056/sigs.2004684

Peer reviewed
Complete genome sequence of *Syntrophobotulus glycolicus* type strain (FLGlyRT)

Cliff Han1,2, Romano Mwirichia3, Olga Chertkov1,2, Brittany Held1,2, Alla Lapidus1, Matt Nolan1, Susan Lucas1, Nancy Hammon1, Shweta Deshpande1, Jan-Fang Cheng1, Roxanne Tapia1,2, Lynne Goodwin1,3, Sam Pitluck1, Marcel Huntemann1, Konstantinos Liolios1, Natalie Ivanova1, Ioanna Pagani1, Konstantinos Mavromatis1, Galina Ovchinikova1, Amirita Pati1, Amy Chen1, Krishna Palaniappan1, Miriam Land1, Loren Hauser1,5, Evelyne-Marie Brambilla6, Manfred Rohde7, Stefan Spring6, Johannes Sikorski6, Markus Göker6, Tanja Woyke1, James Bristow1, Jonathan A. Eisen1,8, Victor Markowitz4, Philip Hugenholtz1,9, Nikos C. Kyrpides1, Hans-Peter Klenk6*, and John C. Detter1,2

1 DOE Joint Genome Institute, Walnut Creek, California, USA
2 Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
3 Jomo Kenyatta University of Agriculture and Technology, Kenya
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 HZI – Helmholtz Centre for Infection Research, Braunschweig, 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: Hans-Peter Klenk*

**Keywords:** glycolate-oxidizing, Gram-negative with Gram-positive cell wall structure, strictly anaerobic, chemotrophic, mesophilic, non-motile, rod-shaped, spore-forming, *Peptococcaceae, Clostridiales*, GEBA

*Syntrophobotulus glycolicus* Friedrich et al. 1996 is currently the only member of the genus *Syntrophobotulus* within the family *Peptococcaceae*. The species is of interest because of its isolated phylogenetic location in the genome-sequenced fraction of tree of life. When grown in pure culture with glyoxylate as carbon source the organism utilizes glyoxylate through fermentative oxidation, whereas, when grown in syntrophic co-culture with homoacetogenic or methanogenic bacteria, it is able to oxidize glycolate to carbon dioxide and hydrogen. No other organic or inorganic carbon source is utilized by *S. glycolicus*. The subdivision of the family *Peptococcaceae* into genera does not reflect the natural relationships, particularly regarding the genera most closely related to *Syntrophobotulus*. Both *Desulfotomaculum* and *Pelotomaculum* are paraphyletic assemblages, and the taxonomic classification is in significant conflict with the 16S rRNA data. *S. glycolicus* is already the ninth member of the family *Peptococcaceae* with a completely sequenced and publicly available genome. The 3,406,739 bp long genome with its 3,370 protein-coding and 69 RNA genes is a part of the Genomic Encyclopedia of Bacteria and Archaea project.

**Introduction**

Strain FLGlyRT (DSM 8271) is the type strain of *Syntrophobotulus glycolicus* within the monotypic genus *Syntrophobotulus* [1], which is affiliated to the family *Peptococcaceae* within the order *Clostridiales* [2]. The genus name is derived from the latinized Greek *syntrophos* meaning *having grown up with one*, and the Latin *botulus*, sausage, a syntrophic sausage-like item [3]. The species epithet is derived from the Neo-Latin *acidan glycolicum* meaning 'glycolic acid', referring to the key substrate of this species, glycolic acid [3]. The major characteristic that differentiates this genus from other bacteria is the ability to oxidize glyoxylate under anaerobic conditions [3]. The major source of glycolate in n-
Syntrophobotulus glycolicus type strain (FlGlyRT)

Syntrophobotulus glycolicus is excretion by algae and other photoautotrophs and chemotrophic [4-7]. Strain FlGlyRT was isolated from anoxic sewage sludge in Konstanz, Germany [3], but was also mentioned in earlier reports [8]. No further isolates have been reported until now. Here we present a summary classification and a set of features for S. glycolicus FlGlyRT, together with the description of the complete genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of strain FlGlyRT was compared using NCBI BLAST [9] 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 Green-genes database [10] and the relative frequencies of taxa and keywords (reduced to their stem [11]) were determined, weighted by BLAST scores. The most frequently occurring genera were Desulfotabacterium (45.4%), Desulfovosorinus (19.3%), Dehalobacter (18.0%), Heliobacterium (13.8%) and Syntrophobotulus (2.6%) (85 hits in total). Regarding the single hit to sequences from members of the species, the average identity within HSPs was 99.7%, whereas the average coverage by HSPs was 99.7%. Among all other species, the one yielding the highest score was Dehalobacter restrictus (Y10164), which corresponded to an identity of 95.0% and an HSP coverage of 85.2%. (Note that the Greengenes database uses the INSDC (= EMBL/NCBI/DDBJ) annotation, which is not an authoritative source for nomenclature or classification.) The highest-scoring environmental sequence was AJ278164 (‘Dehalobacter sp. clone SHD-11’ [12]), which showed an identity of 95.3% and an HSP coverage of 86.8%. The most frequently occurring keywords within the labels of environmental samples which yielded hits were 'soil' (6.5%), 'microbi' (6.4%), 'respons' (4.8%), 'paddi, rice' (4.6%) and 'condit' (4.5%) (165 hits in total). The most frequently occurring keyword within the labels of environmental samples which yielded hits of a higher score than the highest scoring species was 'dehalobact' (100.0%) (1 hit in total). The BLAST analysis results concur with earlier reports on the ecology and the physiology of the isolate whereby it was isolated from a co-culture with other sulfate-reducing bacteria [3,8].

![Figure 1. Phylogenetic tree highlighting the position of S. glycolicus relative to the type strains of the most closely related genera within the family Peptococcaceae. The tree was inferred from 1,306 aligned characters [13,14] of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion [15] and rooted with the type species of the family. The branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches are support values from 700 ML bootstrap replicates [16] (left) and from 1,000 maximum parsimony bootstrap replicates [17] (right) if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [18] are labeled with one asterisk, those also listed as ‘Complete and Published’ with two asterisks [19-21].](image-url)
Figure 1 shows the phylogenetic neighborhood of *S. glycolicus* in a 16S rRNA based tree. The sequences of the four 16S rRNA gene copies in the genome differ from each other by up to eight nucleotides, and differ by up to 14 nucleotides from the previously published 16S rRNA sequence X99706, which contains two ambiguous base calls.

As two of the genera selected for Figure 1, *Desulfo- tomaculum* and *Pelotomaculum*, appeared as paraphyletic in the tree, we conducted both unconstrained heuristic searches for the best tree under the maximum likelihood (ML) [15] and maximum parsimony (MP) criterion [17] as well as searches constrained for the monophyly of all genera (for details of the data matrix see the figure caption). The best-known ML tree had a log likelihood of -12,054.61, whereas the best trees found under the constraint had a log likelihood of -12,209.39 and were significantly worse in the SH test as implemented in RAxML [15] (p < 0.01). The best-known MP trees had a score of 2,018 whereas the best trees found under the constraint had a score of 2,076 and were significantly worse in the KH test as implemented in PAUP* [17] (p < 0.0001). Accordingly, the current classification of the group is in significant conflict with the 16S rRNA data and apparently does not reflect its natural relationships. The classification could be improved if combinations of phenotypic character states were found which characterize a set of appropriately rearranged, then monophyletic genera. However, it might also be that the goal to 'define' each genus in terms of unique combinations of few, potentially arbitrarily selected character states is overambitious, if not misleading in this group of organisms. Apparently a taxonomic revision of the family appears to be necessary which focuses more strongly on the genealogy of the organisms than previous treatments.

Cells of strain *FlGlyR<sup>T</sup>* are Gram-positive, spore forming and slightly curved rods of 2.5-3.5 by 0.5 µm in size [3] (Figure 2). Though the organism is reported to be non-motile, numerous genes associated with flagellar motility are present in the genome (see below). Growth occurs between 15°C and 37°C with an optimum at 28°C, and in a pH range of 6.7 to 8.3, with an optimum at pH 7.3 [3] (Table 1). The reported habitat for this strain is sewage sludge and anoxic freshwater sediments [3]. Initial isolation condition was from defined co-cultures of fermenting bacteria with homoacetogenic or methanogenic bacteria which converted glycolate completely to CO<sub>2</sub> and H<sub>2</sub>, with concomitant reduction of CO<sub>2</sub> to either acetate or methane [3,8]. Later strain *FlGlyR<sup>T</sup>* was identified as the primary fermenting partner in these co-cultures and glyoxlate was the substrate [3]. Strain *FlGlyR<sup>T</sup>* grows optimally in freshwater medium although growth also occurred in brackishwater medium with 110 mM NaCl and 5 mM MgCl [3]. Strain *FlGlyR<sup>T</sup>* is strictly anaerobic, growing chemotrophically in pure culture by fermentative oxidation of glyoxylate [3]. In pure culture, glyoxyllic acid is fermented to carbon dioxide, hydrogen, and glycolic acid [3]. However, in syntrophic co-culture with, e.g., *Methanospirillum hungatei* or *Acetobacterium woodii* as a partner, glycolic acid is converted to carbon dioxide and hydrogen [3]. Glycolate oxidation to glyoxylate and vice versa is coupled to a membrane-bound electron transport system that catalyzes either a proton potential-driven reversed electron transport from glycolate to hydrogen or a hydrogen-dependent glyoxylate reduction coupled to ATP synthesis by electron transport phosphorylation [34,35]. Due to the oxygenase activity of the D-ribulose-1,5-bisphosphate carboxylase at low CO<sub>2</sub> and high O<sub>2</sub> concentrations, the phosphoglycolate formed in these organisms is subsequently dephosphorylated to glycolate [8]. It is reported that no other organic or inorganic substrates are used [3], even though a total of 78 carbohydrate transport and metabolism genes are found the genome of this organism (COGS table). Neither sulfate, sulfite, thiosulfate, elemental sulfur, nor nitrate are reduced [3].

![Figure 2. Scanning electron micrograph of *S. glycolicus FlGlyR<sup>T</sup>*](image_url)
Syntrophobotulus glycolicus type strain (FlGlyRT)

Chemotaxonomy
Strain FlGlyRT has no cytochromes and the cells contain menaquinone-7-10, with MK-9 as major fraction [3]. Although the cells stain Gram-negative, the ultrastructural analysis shows a Gram-positive cell wall architecture [3].

Genome sequencing and annotation
Genome project history
This organism was selected for sequencing on the basis of its phylogenetic position [36], and is part of the Genomic Encyclopedia of Bacteria and Archaea project [37]. The genome project is deposited in the Genome On Line Database [18] 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 1. Classification and general features of S. glycolicus FlGlyRT according to the MIGS recommendations [22] and the NamesforLife database [23].

| MIGS ID | Property               | Term                                      | Evidence code |
|---------|------------------------|-------------------------------------------|---------------|
|         | Current classification | Type strain FlGlyR                        | TAS [3]       |
|         | Gram stain             | negative                                  | TAS [3]       |
|         | Cell shape             | rod shaped, slightly curved               | TAS [3]       |
|         | Motility               | non-motile                                | TAS [3]       |
|         | Sporulation            | sporulating                               | TAS [3]       |
|         | Temperature range      | 15°C-37°C                                 | TAS [3]       |
|         | Optimum temperature    | 28°C                                     | TAS [3]       |
|         | Salinity               | tolerates ~6% NaCl                        | TAS [3]       |
| MIGS-22 | Oxygen requirement     | strictly anaerobic                        | TAS [3]       |
|         | Carbon source          | glyoxylate                                | TAS [3]       |
|         | Energy metabolism      | chemotrophic                              | TAS [3]       |
| MIGS-6  | Habitat                | marine, sludge, fresh water               | TAS [3]       |
| MIGS-15 | Biotic relationship    | free-living                               | TAS [3]       |
| MIGS-14 | Pathogenicity          | not reported                              |               |
|         | Biosafety level        | 1                                         | TAS [32]      |
|         | Isolation              | anoxic sludge from municipal sewage treatment plant | TAS [3,8] |
| MIGS-4  | Geographic location    | Konstanz, Germany                         | TAS [3,8]     |
| MIGS-5  | Sample collection time | 1991 or before                           | TAS [3,8]     |
| MIGS-4.1| Latitude               | 47.67                                     | NAS           |
| MIGS-4.2| Longitude              | 9.16                                      | NAS           |
| MIGS-4.3| Depth                  | unknown                                   |               |
| MIGS-4.4| Altitude               | about 420 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 [33]. If the evidence code is IDA, the property was directly observed by one of the authors or an expert mentioned in the acknowledgements.
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 (13 kb insert size), one Illumina library |
| MIGS-29 | Sequencing platforms      | Illumina GAii, 454 GS FLX Titanium                                     |
| MIGS-31.2| Sequencing coverage       | 167.0 × Illumina; 48.0 × pyrosequence                                  |
| MIGS-30 | Assemblers                | Newbler version 2.3 (Roche), Velvet 0.7.63, phrap SPS - 4.24          |
| MIGS-32 | Gene calling method       | Prodigal 1.4, GenePRIMP                                               |
| INSDC ID |                           | CP002547                                                             |
| Genbank Date of Release |                       | March 4, 2011                                                        |
| GOLD ID  |                           | Gc01670                                                              |
| NCBI project ID |                | 38111                                                               |
| Database: IMG-GEBA |                 | 2503707006                                                          |
| MIGS-13 | Source material identifier | DSM 8271                                                             |
| Project relevance |              | Tree of Life, GEBA                                                   |

Growth conditions and DNA isolation

*S. glycolicus* FlGlyR®, DSM 8271, was grown anaerobically in DSMZ medium 298b (FlGlyM-medium) [38] at 28°C. DNA was isolated from 0.5-1 g of cell paste using Jetflex Genomic DNA Purification kit (GENOMED 600100) following the standard protocol as recommended by the manufacturer, adding 10 µL proteinase K to the standard lysis solution for 50 minutes at 58°C. DNA is available through the DNA Bank Network [39].

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 [40]. Pyrosequencing reads were assembled using the Newbler assembler. The initial Newbler assembly consisting of 38 contigs in two scaffolds was converted into a phrap [41] assembly by making fake reads from the consensus, to collect the read pairs in the 454 paired end library. Illumina sequencing data (602.6 Mb) was assembled with Velvet [42] and the consensus sequences were shredded into 1.5 kb overlapped fake reads and assembled together with the 454 data. 454 draft assembly was based on 163.5 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/Phraps/Consed software package [41] 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 [40], Dupfinisher, or sequencing cloned bridging PCR fragments with subcloning or transposon bombing (Epicentre Biotechnologies, Madison, WI) [43]. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F. Chang, unpublished). A total of 331 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 [44]. 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 215.0 × coverage of the genome. The final assembly contained 327,738 pyrosequence and 15,336,223 Illumina reads.

Genome annotation

Genes were identified using Prodigal [45] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [46]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGR-Fam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [47].
**Syntrophobotulus glycolicus** type strain (FIGlyRT)

**Genome properties**
The genome consists of a 3,406,739bp long chromosome with a GC content of 46.4% (Figure 3 and Table 3). Of the 3,439 genes predicted, 3,370 were protein-coding genes, and 69 RNAs; 119 pseudogenes were also identified. The majority of the protein-coding genes (68.7%) 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 chromosome. 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)              | 3,406,739   | 100.00%    |
| DNA coding region (bp)        | 2,989,609   | 87.76%     |
| DNA G+C content (bp)          | 1,579,030   | 46.35%     |
| Number of replicons           | 1           |            |
| Extrachromosomal elements     | 0           |            |
| Total genes                   | 3,439       | 100.00%    |
| RNA genes                     | 69          | 2.01%      |
| rRNA operons                  | 4           |            |
| Protein-coding genes          | 3,370       | 97.99%     |
| Pseudo genes                  | 119         | 3.46%      |
| Genes with function prediction| 2364        | 68.74%     |
| Genes in paralog clusters     | 710         | 20.65%     |
| Genes assigned to COGs        | 2,399       | 69.76%     |
| Genes assigned Pfam domains   | 2,561       | 74.47%     |
| Genes with signal peptides    | 463         | 13.46%     |
| Genes with transmembrane helices| 848     | 24.66%     |
| CRISPR repeats                | 2           |            |

| Code | value | %age | Description                                      |
|------|-------|------|--------------------------------------------------|
| J    | 152   | 5.8  | Translation, ribosomal structure and biogenesis  |
| A    | 0     | 0.0  | RNA processing and modification                  |
| K    | 230   | 8.8  | Transcription                                    |
| L    | 156   | 6.0  | Replication, recombination and repair            |
| B    | 1     | 0.0  | Chromatin structure and dynamics                 |
| D    | 33    | 1.3  | Cell cycle control, cell division, chromosome partitioning |
| Y    | 0     | 0.0  | Nuclear structure                                |
| V    | 98    | 3.8  | Defense mechanisms                               |
| T    | 162   | 6.2  | Signal transduction mechanisms                   |
| M    | 170   | 6.5  | Cell wall/membrane/envelope biogenesis           |
| N    | 68    | 2.6  | Cell motility                                    |
| Z    | 0     | 0.0  | Cytoskeleton                                     |
| W    | 0     | 0.0  | Extracellular structures                         |
| U    | 43    | 1.7  | Intracellular trafficking, secretion, and vesicular transport |
| O    | 74    | 2.8  | Posttranslational modification, protein turnover, chaperones |
| C    | 156   | 6.0  | Energy production and conversion                 |
| G    | 78    | 3.0  | Carbohydrate transport and metabolism            |
| E    | 214   | 8.2  | Amino acid transport and metabolism              |
| F    | 65    | 2.5  | Nucleotide transport and metabolism              |
| H    | 140   | 5.4  | Coenzyme transport and metabolism               |
| I    | 49    | 1.9  | Lipid transport and metabolism                   |
| P    | 195   | 7.5  | Inorganic ion transport and metabolism           |
| Q    | 27    | 1.0  | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 282   | 10.8 | General function prediction only                 |
| S    | 220   | 8.4  | Function unknown                                 |
| -    | 1,040 | 30.2 | Not in COGs                                      |
Acknowledgements

We would like to gratefully acknowledge the help of Maren Schröder (DSMZ) for growing S. glycolicus cultures. 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-AC05-00R22725, as well as German Research Foundation (DFG) INST 599/1-2.

References

1. Euzéby JP. List of bacterial names with standing in nomenclature: A folder available on the internet. Int J Syst Bacteriol 1997; 47:590-592. PubMed doi:10.1099/00207713-47-2-590

2. Boone DR, Castenholz RW, Garrity GM, eds. 2001. Bergey’s Manual of Systematic Bacteriology, 2nd ed. Springer-Verlag. New York, NY.

3. Friedrich M, Springer N, Ludwig W, Schink B. Phylogenetic position of Desulfofustis glycolicus gen. nov. sp. nov. and Syntrophobotulus glycolicus gen. nov. sp. nov., two new strict anaerobes growing with glycolic acid. Int J Syst Bacteriol 1996; 46:1065-1069. PubMed doi:10.1099/00207713-46-4-1065

4. Whittingham CP, Pritchard GG. The production of glycolate during photosynthesis in Chlorella. Proc R Soc Lond 1963; 157:366-382. doi:10.1098/rspb.1963.0017

5. Codd GA, Smith BM. Glycolate formation and excretion by the purple photosynthetic bacterium Rhodospirillum rubrum. FEBS Lett 1974; 48:105-108. PubMed doi:10.1016/0014-5793(74)81073-2

6. Codd GA, Bowien B, Schlegel HG. Glycolate production and excretion by Alcaligenes eutrophus. Arch Microbiol 1976; 110:167-171. PubMed doi:10.1007/BF00690224

7. Beudeker RF, Kuenen JG, Codd GA. Glycolate metabolism in the obligate chemolithotrophic Thiobacillus neapolitanus grown in continuous culture. J Gen Microbiol 1981; 126:337-346.

8. Friedrich M, Laderer U, Schink B. Fermentative degradation of glycolic acid by defined syntrophic cocultures. Arch Microbiol 1991; 156:398-404. doi:10.1007/BF00248717

9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403-410. PubMed

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

12. Schlötelburg C, von Wintzingerode C, Hauck R, von Wintzingerode F, Hegemann W, Göbel UB. Microbial structure of an anaerobic bioreactor population that continuously dechlorinates 1,2-dichloropropene. FEMS Microb Ecol 2002; 39:229-237. PubMed doi:10.1016/S0168-6496(02)00177-0

13. 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

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. 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

16. 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

17. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0 b10. Sinauer Associates, Sunderland, 2002.

18. 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
19. Spring S, Lapidus A, Schröder M, Gleim D, Sims D, Meincke L, Gla
inga Del Rio T, Tice H, Copeland A, Cheng JF, et al. Complete genome se-
iquence of Desulfitomaculum acetoxidans type strain (5575T). *Stand Genomic Sci* 2009; 1:242-253. [PubMed doi:10.4056/sigs.39508]

20. Kosaka T, Kato S, Shimoyama T, Ishii S, Abe T, Watanabe K. The genome of *Pelotomaculum thermopropionicum* reveals niche-associated evolution in anaerobic microbota. *Genome Res* 2008; 18:442-448. [PubMed doi:10.1101/gr.7136508]

21. Nonaka H, Keresztes G, Shinoda Y, Ikenaga Y, Abe M, Naito K, Inatomi K, Furukawa K, Inui M, Yukawa H. Complete genome sequence of the dehalorespiring bacterium *Desulfitobacterium hafniense* Y51 and comparison with *Dehalococcoides ethenogenes* 195. *J Bacteriol* 2006; 188:2262-2274. [PubMed doi:10.1128/JB.188.6.2262-2274.2006]

22. 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]

23. Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; 37:9.

24. 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]

25. Gibbons NE, Murray RGE. Proposals Concerning the Higher Taxa of Bacteria. *Int J Syst Bacteriol* 1978; 28:1-6. [PubMed doi:10.1099/00207713-28-1-1]

26. 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.

27. List Editor. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; 60:469-472. [PubMed doi:10.1099/ijs.0.022855-0]

28. Rainey FA. Class II. *Clostridia* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds), Bergey’s Manual of Systematic Bacteriology, Second Edition, Volume 3, Springer-Verlag, New York, 2009, p. 736.

29. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; 30:225-420. [PubMed doi:10.1099/00207713-30-1-225]

30. Prévot AR. In: Hauderoy P, Ehringer G, Guillot G, Magrou. J.. Prévot AR, Rosset D, Urbain A (eds), Dictionnaire des Bactéries Pathogènes, Second Edition, Masson et Cie, Paris, 1953, p. 1-692.

31. Rogosa M. *Peptococccaeae*, a new family to include the Gram-positive, anaerobic cocci of the genera *Peptococcus*, *Peptostreptococcus* and *Ruminococcus*. *J Int J Bacteriol* 1971; 21:234-237. [PubMed doi:10.1099/00207713-21-3-234]

32. BAuA 2005. Classification of bacteria and archaea in risk groups. TRBA 466 p. 336; [http://www.baua.de](http://www.baua.de)

33. 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]

34. Friedrich M, Schink B. Hydrogen formation from glycolate driven by reversed electron transport in membrane vesicles of a syntrophic glycolate-oxidizing bacterium. *Eur J Biochem* 1993; 217:233-240. [PubMed doi:10.1111/j.1432-1033.1993.tb18238.x]

35. Friedrich M, Schink B. Electron transport phosphorylation driven by glyoxylate respiration with hydrogen as electron donor in membrane vesicles of a glyoxylate-fermenting bacterium. *Arch Microbiol* 1995; 163:268-275. [PubMed doi:10.1007/BF00393379]

36. 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]

37. 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]

38. List of growth media used at DSMZ: [http://www.dsmz.de/microorganisms/media_list.php](http://www.dsmz.de/microorganisms/media_list.php)

39. Gemeinhölzer 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
40. The DOE Joint Genome Institute. http://www.jgi.doe.gov
41. Phrap and Phred for Windows, MacOS, Linux, and Unix. http://www.phrap.com
42. 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
43. 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
44. 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.
45. Hyatt D, Chen GL, LoCescio PF, Land ML, Lai-mer FW, Hauser L. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010; 11:119. PubMed doi:10.1186/1471-2105-11-119
46. 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
47. 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