Lawrence Berkeley National Laboratory
Recent Work

Title
Non-contiguous finished genome sequence of Aminomonas paucivorans type strain (GLU-3).

Permalink
https://escholarship.org/uc/item/06b517b3

Journal
Standards in genomic sciences, 3(3)

ISSN
1944-3277

Authors
Pitluck, Sam
Yasawong, Montri
Held, Brittany
et al.

Publication Date
2010-11-20

DOI
10.4056/sigs.1253298

Peer reviewed
Non-contiguous finished genome sequence of
Aminomonas paucivorans type strain (GLU-3T)

Sam Pitluck1, Montri Yasawong2, Brittany Held1,3, Alla Lapidus1, Matt Nolan1, Alex Copeland1, Susan Lucas1, Tijana Glavina Del Rio1, Hope Tice1, Jan-Fang Cheng1, Olga Chertkov1,3, Lynne Goodwin1,3, Roxane Tapia1,2, Cliff Han1,3, Konstantinos Liosis1, Natalia Ivanova1, Konstantinos Mavromatis1, Galina Ovchinnikova1, Amrita Pati1, Amy Chen1, Krishna Palaniappan1, Miriam Land1,3, Loren Hauser1,3, Yun-Juan Chang1,3, Cynthia D. Jeffries1,3, Rüdiger Pukall6, Stefan Spring6, Manfred Rohde2, Johannes Sikorski6, Markus Göker6, Tanja Woyke1, James Bristow1, Jonathan A. Eisen1,7, Victor Markowitz4, Philip Hugenholtz1, Nikos C. Kyrpides1, and Hans-Peter Klenk6

1 DOE Joint Genome Institute, Walnut Creek, California, USA
2 HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany
3 Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA
4 Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA
5 Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
6 DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany
7 University of California Davis Genome Center, Davis, California, USA

*Corresponding author: Hans-Peter Klenk

Keywords: strictly anaerobic, obligate amino-acid-degrading, Gram-negative, nonmotile, asaccharolytic, mesophilic, chemoorganotrophic, Synergistaceae, ‘Synergistetes’, GEBA

Introduction

Strain GLU-3T (= DSM 12260 = ATCC BAA-6) is the type strain of the species Aminomonas paucivorans, which in turn is the type and only species of the genus Aminomonas [1,2]. The generic name derives from the Latin word ‘aminum’ meaning ‘amine’ and the Greek word ‘monas’ meaning ‘a unit or monad’, referring to amine-degrading monads [2]. The species epithet is derived from the Latin word ‘paucus’ meaning ‘few or little’ and the Latin word ‘vorans’ meaning ‘digesting’, referring to digesting little [2]. Strain GLU-3T was isolated from anaerobic sludge of a dairy wastewater treatment plant in Santafé de Bogota, Colombia [2]. So far, no further isolates have been obtained for A. paucivorans. Here we present a summary classification and a set of features for A. paucivorans GLU-3T, together with the description of the non-contiguous finished genomic sequencing and annotation.

Classification and features

The 16S rRNA gene of A. paucivorans GLU-3T shares 96% sequence identity with that of the type strain of Thermanaerovibrio acidaminovorans, which was isolated from an upflow anaerobic sludge bed reactor of a sugar refinery, Breda, the Netherlands [3] (Figure 1), and 82.2-96.4% sequence identity with the type strains from the other members of the family Synergistaceae [11].
The sequences of four marine metagenomic clones in the env_nt database, 1096626071844 (AACY020063505), 1096626840052 (AACY020539193), 1096626748225 (AACY020105546) and 1096626774924 (AACY020274567) share 96% sequence identity with *A. paucivorans* GLU-3T (as of October 2010). A representative genomic 16S rRNA sequence of *A. paucivorans* was compared using NCBI BLAST under default values with the most recent release of the Greengenes database [12] and the relative frequencies of taxa and keywords, weighted by BLAST scores, were determined. The four most frequent genera were *Thermanaerovibrio* (65.5%), *Aminomonas* (18.0%), *Anaerobaculum* (9.0%) and *Aminiphilus* (7.6%). The species yielding the highest score was *T. acidaminovorans*. The five most frequent keywords within the labels of environmental samples which yielded hits were ‘anaerobic’ (7.2%), ‘sludge’ (6.9%), ‘wastewater’ (6.8%), ‘municipal’ (6.8%) and ‘digester’ (6.7%). These keywords corroborate the physiological and ecological features on strain GLU-3T as depicted in the original description [2]. The single most frequent keyword within the labels of environmental samples which yielded hits of a higher score than the highest scoring species was ‘harbor/sediment’ (50.0%).

Figure 1 shows the phylogenetic neighborhood of *A. paucivorans* GLU-3T in a 16S rRNA based tree. The sequences of the three 16S rRNA gene copies in the genome of *A. paucivorans* differ from each other by up to one nucleotide, and differ by up to eleven nucleotides from the previously published 16S rRNA sequence (AF072581), which contains 59 ambiguous base calls (ambiguous bases not count as differences).

**Genome sequencing and annotation**

**Genome project history**

This organism was selected for sequencing on the basis of its phylogenetic position [18], and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [19]. The genome project is deposited in the Genome OnLine Database [8,20] and the non-contiguous finished genome sequence has been deposited in DDBJ/EMBL/GenBank under the accession AEIV00000000. The version described in this paper is the first version, AEIV01000000. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2.

---

**Figure 1.** Phylogenetic tree highlighting the position of *A. paucivorans* GLU-3T relative to the other type strains within the family *Synergistaceae*. The tree was inferred from 1,347 aligned characters [4,5] of the 16S rRNA gene sequence under the maximum likelihood criterion [6] and rooted in accordance with the current taxonomy. The branches are scaled in terms of the expected number of substitutions per site. Numbers above branches are support values from 1,000 bootstrap replicates [7] if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [8] are shown in blue, published genomes in bold [3,9,10].
A. paucivorans GLU-3T is described as Gram-negative, slightly curved, rod-shaped bacterium (0.3 × 4.0-6.0 µm), which occurs singly or in pairs (Figure 2 and Table 1). Colonies of strain GLU-3T are round, smooth and white, with a diameter up to 1 mm [2]. Strain GLU-3T does not produce endospores [2]. The organism does not have flagella and motility is not observed [2], although plenty of motility genes are present in the genome. Strain GLU-3T is a strictly anaerobic, mesophilic, chemooorganotrophic and asaccharolytic bacterium [2]. The temperature range for growth is 20-40°C, with an optimum at 35°C [2]. The pH range for growth is 6.7-8.3, with an optimum at 7.5 [2]. The organism does not require NaCl for growth but tolerates up to 2.0% [2]. The optimum growth occurs in media with 0.05-0.5% of NaCl [2]. The species requires yeast extract for growth [2]. The organism is able to ferment arginine, histidine, glutamine, threonine, and glycine [2]. Arginine is fermented to acetate, formate and ornithine [2]. Histidine is fermented to acetate and formate [2]. Glutamate is fermented to acetate, formate and trace amounts of propionate [2]. Threonine and glycine are fermented to acetate [2]. Casamino acid, peptone and cysteine are only poorly used by the strain GLU-3T, and acetate is the end-product of the amino acid metabolism [2]. A mixed culture of strain GLU-3T and Methanobacterium formicicum does not extend the range of substrate utilization [2], as is observed for, e.g., Aminobacterium colombiense [9]. Methane is not detectable in mixed cultures, when grown in glycine and threonine [2], however, the end-product profiles are the same as in pure culture [2]. The major end-product is shifted from acetate to propionate, when strain GLU-3T was grown together with M. formicicum on arginine, histidine and glutamate [2]. Ornithine is not accumulated during arginine degradation in mixed culture [2]. Strain GLU-3T does not degrade alanine and branched-chain amino acids, valine, leucine and isoleucine either in pure culture or in syntrophic growth with M. formicicum [2]. Also, the range of amino acid utilization is not increased in co-culture with M. formicicum [2]. Strain GLU-3T does not grow on carbohydrates, gelatin, casein, pyruvate, succinate, malate, fumarate, α-ketoglutarate, mesaconate, β-methylaspartate, oxaloacetate, glycerol, ethanol, acetate, propionate, butyrate, lactate, citrate, leucine, lysine, alanine, valine, proline, serine, methionine, asparagines, phenylalanine and aspartate [2]. The organism does not utilize sulfate, thiosulfate, elemental sulfur, sulfite, nitrate and fumarate as electron acceptors [2].

**Chemotaxonomy**

No chemotaxonomic data are currently available for A. paucivorans or for other members of the genus Aminomonas.

Figure 2. Scanning electron micrograph of A. paucivorans GLU-3T

http://standardsingenomics.org
### Aminomonas paucivorans type strain (GLU-3T)

Table 1. Classification and general features of *A. paucivorans* GLU-3\(^T\) according to the MIGS recommendations [13].

| MIGS ID | Property           | Term                                           | Evidence code |
|---------|--------------------|------------------------------------------------|---------------|
|         | Current classification | Domain *Bacteria*                               | TAS [14]      |
|         |                     | Phylum "Synergistetes"                          |               |
|         |                     | Class *Synergista*                               |               |
|         |                     | Order *Synergistales*                            |               |
|         |                     | Family *Synergistaceae*                          |               |
|         |                     | Genus *Aminomonas*                               |               |
|         |                     | Species *Aminomonas paucivorans*                |               |
|         |                     | Type strain GLU-3                               |               |
|         | Gram stain          | negative                                       | TAS [2]       |
|         | Cell shape          | slightly curved rods occurring singly or in pairs |               |
|         | Motility            | none                                           | TAS [2]       |
|         | Sporulation         | none                                           | TAS [2]       |
|         | Temperature range   | 20°C-40°C                                      | TAS [2]       |
|         | Optimum temperature | 35°C                                           |               |
|         | Salinity            | 0-2% NaCl (optimum 0.05-0.50%)                 |               |
|         | Oxygen requirement  | strictly anaerobic                             |               |
|         | Carbon source       | amino acids                                    |               |
|         | Energy source       | chemoorganotroph                               |               |
|         | Habitat             | wastewater                                      |               |
|         | Biotic relationship | free-living                                    | NAS           |
|         | Pathogenicity       | none                                           | NAS           |
|         | Biosafety level     | 1                                              | TAS [16]      |
|         | Isolation           | anaerobic sludge of a dairy wastewater treatment plant |               |
|         | Geographic location | SantaFe de Bogota, Colombia                     |               |
|         | Sample collection time | 1996                                          | NAS           |
|         | Latitude            | 4.60                                           | NAS           |
|         | Longitude           | 74.08                                          | NAS           |
|         | Depth               | not reported                                    |               |
|         | Altitude            | 2620 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 [17]. If the evidence code is IDA, then the property was directly observed by one of the authors or an expert mentioned in the acknowledgements.

### Growth conditions and DNA isolation

*A. paucivorans* GLU-3\(^T\), DSM 12260, was grown anaerobically in DSMZ medium 846 (Anaerobic Serine/Arginine medium) [21] at 37°C. DNA was isolated from 0.5-1 g of cell paste using the MasterPure Gram-positive DNA purification kit (Epico centre MGP04100) following the standard protocol as recommended by the manufacturer, with modification st/LALM for cell lysis as described in Wu et al. [19].
Table 2. Genome sequencing project information

| MIGS ID | Property          | Term                                                                 |
|---------|-------------------|----------------------------------------------------------------------|
| MIGS-31 | Finishing quality | Non-contiguous finished                                              |
| MIGS-28 | Libraries used    | Three genomic libraries: one 454 pyrosequence standard library, 454 PE library (12 kb insert size), one Illumina standard library |
| MIGS-29 | Sequencing platforms | 454 GS FLX Titanium, Illumina GAii                                    |
| MIGS-31.2 | Sequencing coverage | 202.0 × Illumina; 72.4 × pyrosequence                                |
| MIGS-30 | Assemblers        | Newbler version 2.0.00.20-PostRelease-11-05-2008-gcc-3.4.6, phrap    |
| MIGS-32 | Gene calling method | Prodigal 1.4, GenePRIMP                                              |
| INSDC ID |                  | CM001022, AEIV00000000                                              |
| Genbank Date of Release |  | November 2, 2010                                                      |
| GOLD ID   |                  | Gi02542                                                              |
| NCBI project ID |  | 33371                                                               |
| Database: IMG-GEBA | | 2502790015                                                          |
| MIGS-13  | Source material identifier | DSM 12260                                                          |
| Project relevance |  | Tree of Life, GEBA                                                  |

Genome sequencing and assembly

The genome was sequenced using a combination of Illumina and 454 sequencing platforms. All general aspects of library construction and sequencing can be found at the JGI website. Pyrosequencing reads were assembled using the Newbler assembler version 2.0.00.20-PostRelease-11-05-2008-gcc-3.4.6 (Roche). The initial Newbler assembly consisted of 126 contigs in 103 scaffolds and was converted into a phrap assembly by making fake reads from the consensus for collecting the read pairs in the 454 paired end library. Illumina GAii sequencing data (525.3 Mb) was assembled with Velvet [22] 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 190.7 Mb 454 draft data and all of the 454 paired end data. Newbler parameters were -consed -a 50 -l 350 -g -m -ml 20.

The Phred/Phrap/Consed software package 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, Dupfinisher, or sequencing cloned bridging PCR fragments with subcloning or transposon bombing (Epicentre Biotechnologies, Madison, WI) [23]. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F.Chang, unpublished). A total of 259 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 [24]. 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 274.4× coverage of the genome. The final assembly contained 535,052 pyrosequences and 15,007,632 Illumina reads.

Genome annotation

Genes were identified using Prodigal [25] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [26]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGR-Fam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [27].

Genome properties

The genome consists of a 2,630,120 bp long chromosome with an overall GC content of 67.6% (Table 3 and Figure 3). Of the 2,494 genes predicted, 2,433 were protein-coding genes, and 61 RNAs; 34 pseudogenes were also identified. The majority of the protein-coding genes (77.2%) 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)                 | 2,630,120 | 100.00%    |
| DNA Coding region (bp)           | 2,411,389 | 91.68%     |
| DNA G+C content (bp)             | 1,777,554 | 67.59%     |
| Number of replicons              | 1         |            |
| Extrachromosomal elements        | 0         |            |
| Total genes                      | 2,494     | 100.00%    |
| RNA genes                        | 61        | 2.45%      |
| rRNA operons                     | 3         |            |
| Protein-coding genes             | 2,433     | 97.55%     |
| Pseudo genes                     | 34        | 1.36%      |
| Genes with function prediction   | 1,926     | 77.23%     |
| Genes in paralog clusters        | 338       | 13.55%     |
| Genes assigned to COGs           | 1,988     | 79.71%     |
| Genes assigned Pfam domains      | 2,047     | 82.08%     |
| Genes with signal peptides       | 446       | 17.88%     |
| Genes with transmembrane helices | 588       | 23.58%     |
| CRISPR repeats                   | 4         |            |

Figure 3. Graphical circular map of the genome. 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    | 158   | 7.2% | Translation, ribosomal structure and biogenesis |
| A    | 0     | 0.0% | RNA processing and modification |
| K    | 138   | 6.3% | Transcription |
| L    | 107   | 4.9% | Replication, recombination and repair |
| B    | 0     | 0.0% | Chromatin structure and dynamics |
| D    | 29    | 1.3% | Cell cycle control, cell division, chromosome partitioning |
| Y    | 0     | 0.0% | Nuclear structure |
| V    | 33    | 1.5% | Defense mechanisms |
| T    | 154   | 7.0% | Signal transduction mechanisms |
| M    | 123   | 5.6% | Cell wall/membrane/envelope biogenesis |
| N    | 90    | 4.1% | Cell motility |
| Z    | 0     | 0.0% | Cytoskeleton |
| W    | 0     | 0.0% | Extracellular structures |
| U    | 48    | 2.2% | Intracellular trafficking, secretion, and vesicular transport |
| O    | 64    | 2.9% | Posttranslational modification, protein turnover, chaperones |
| C    | 161   | 7.3% | Energy production and conversion |
| G    | 104   | 4.7% | Carbohydrate transport and metabolism |
| E    | 251   | 11.4% | Amino acid transport and metabolism |
| F    | 71    | 3.2% | Nucleotide transport and metabolism |
| H    | 111   | 5.0% | Coenzyme transport and metabolism |
| I    | 35    | 1.6% | Lipid transport and metabolism |
| P    | 102   | 4.6% | Inorganic ion transport and metabolism |
| Q    | 25    | 1.1% | Secondary metabolites biosynthesis, transport and catabolism |
| R    | 236   | 10.7% | General function prediction only |
| S    | 166   | 7.5% | Function unknown |
| -    | 506   | 20.3% | Not in COGs |

Acknowledgements
We would like to gratefully acknowledge the help of Katja Steenblock for growing *A. paucivorans* cultures and Susanne Schneider for DNA extraction and quality analysis (both at DSMZ). 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 and Thailand Research Fund Royal Golden Jubilee Ph.D. Program No. PHD/0019/2548 for MY.

References
1. Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; 37:9.
2. Baena S, Fardeau ML, Ollivier B, Labat M, Thomas P, Garcia JL, Patel BK. *Aminomonas paucivorans* gen. nov., sp. nov., a mesophilic, anaerobic, amino-acid-utilizing bacterium. *Int J Syst Bacteriol* 1999; 49:975-982. PubMed [doi:10.1099/00207713-49-3-975](https://doi.org/10.1099/00207713-49-3-975)
3. Chovatia M, Sikorski J, Schröder M, Lapidus A, Nolan M, Tice H, Glavina Del Rio T, Copeland A, Cheng JF, Lucas S, *et al*. Complete genome sequence of *Thermaanaerobirebrio acidaminovorans* type strain (Su883T). *Stand Genomic Sci* 2009; 1:254-261. [doi:10.4056/sigs.40645](https://doi.org/10.4056/sigs.40645)
4. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic
Aminomonas paucivorans type strain (GLU-3T)

analysis. Mol Biol Evol 2000; 17:540-552. PubMed

5. 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

6. 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

7. 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

8. 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 2009; 38:D346-D354. PubMed doi:10.1093/nar/gkp848

9. Chertkov O, Sikorski J, Brambilla E, Lapidus A, Copeland A, Glavina Del Rio T, Nolan M, Lucas S, Tice H, Cheng JF, et al. Complete genome sequence of Aminobacterium colombiense type strain (ALA-1'). Stand Genomic Sci 2010; 2:280-289. doi:10.4056/sigs.902116

10. LaButti K, Mayilraj S, Clum A, Lucas S, Glavina Del Rio T, Nolan M, Tice H, Cheng JF, Pitluck S, Liolios K, et al. Permanent draft genome sequence of Dethiosulfovibrio peptidovorans type strain (SEBR 4207T). Stand Genomic Sci 2010; 3:85-92. doi:10.4056/sigs.1092865

11. 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.0.064915-0

12. 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

13. 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

14. 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

15. Jumas-Bilak E, Roudiere L, Marchandin H. Description of 'Synergistetetes' phyI. nov. and emended description of the phylum 'Deferribacteres' and of the family Syntrophomadaceae, phyI. 'Firmicutes'. Int J Syst Evol Microbiol 2009; 59:1028-1035. PubMed doi:10.1099/ijs.0.006718-0

16. Biological Agents. Classification of Bacteria and Archaea in risk groups. www.baua.de TRBA 466

17. 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

18. Klenk HP, Goeker 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

19. 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

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

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

22. 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

23. 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 (541'). Stand Genomic Sci 2009; 1:12-20. doi:10.4056/sigs.761

24. Lapidus A, LaButti K, Foster B, Lowry S, Trong S, Goltsman E. POLISHER: An effective tool for us-
ing ultra short reads in microbial genome assembly and finishing. AGBT, Marco Island, FL, 2008.

25. Hyatt D, Chen GL, LoCascio PF, Land ML, Lari-mer 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

26. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. Gene-

27. 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