Complete genome sequence of *Calditerrivibrio nitroreducens* type strain (Yu37-1[T])

Sam Pitluck[1], Johannes Sikorski[2], Ahmet Zeytun[1,3], Alla Lapidus[1], Matt Nolan[1], Susan Lucas[1], Nancy Hammon[1], Shweta Deshpande[1], Jan-Fang Cheng[1], Roxane Tapia[1,3], Cliff Han[1,3], Lynne Goodwin[1,3], Konstantinos Liolios[1], Ioanna Pagani[1], Natalia Ivanova[1], Konstantinos Mavromatis[1], Amrita Pati[1], Amy Chen[4], Krishna Palaniappan[4], Loren Hauser[1,5], Yun-Juan Chang[1,5], Cynthia D. Jeffries[1,5], John C. Detter[1], Evelyne Brambilla[2], Oliver D. Ngatchou Djao[6], Manfred Rohde[6], Stefan Spring[2], Markus Göker[2], Tanja Woyke[1], James Bristow[1], Jonathan A. Eisen[1,7], Victor Markowitz[4], Philip Hugenholtz[1,8], Nikos C. Kyrpides[1], Hans-Peter Klenk[2*], and Miriam Land[1,5]

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

*Corresponding author: Hans-Peter Klenk

Calditerrivibrio nitroreducens Iino *et al*. 2008 is the type species of the genus *Calditerrivibrio*. The species is of interest because of its important role in the nitrate cycle as nitrate reducer and for its isolated phylogenetic position in the Tree of Life. Here we describe the features of this organism, together with the complete genome sequence and annotation. This is the third complete genome sequence of a member of the family *Deferribacteraceae*. The 2,216,552 bp long genome with its 2,128 protein-coding and 50 RNA genes is a part of the **G**enomic **En**cyclopedia of **B**acteria and **A**rchaea project.

Introduction

Strain Yu37-1[T] (= DSM 19672 = NBRC 101217) is the type strain of *Calditerrivibrio nitroreducens* which in turn is the type and sole species of the genus *Calditerrivibrio* [1]. The genus *Calditerrivibrio* is one out of six genera in the family *Deferribacteraceae* [2-6]. The genus name is derived from Latin adjective "*caldus*", hot, "*terra*", the earth, and "*vibrio*", a vibrio, referring to a vibroid shaped bacterium in a hot terrestrial environment. The species epithet *nitroreducens* derives from the Greek name "*nitron*", nitrite, nitrate, and "*reducens*", drawing backwards, referring to its nitrate-reducing physiology [1]. Strain Yu37-1[T] was iso-lated from hot-spring water from Yumata, Nagano, Japan. No further cultivated strains belonging to the species *C. nitroreducens* have been described so far. Here we present a summary classification and a set of features for *C. nitroreducens* strain Yu37-1[T], together with the description of the complete genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of strain Yu37-1[T] was compared using BLAST under default settings (e.g., considering only only the

high-scoring segment pairs (HSPs) from the best 250 hits) with the most recent release of the Greengenes database [7] and the relative frequencies of taxa and keywords (reduced to their stem [8]) were determined, weighted by BLAST scores. The most frequently occurring genera were *Deferribacter* (33.4%), *Alteromonas* (21.3%), *Magnetococcus* (9.4%), *Shuttleworthia* (7.5%) and *Geovibrio* (7.3%) (61 hits in total). Regarding the single hit to sequences from members of the species, the average identity within HSPs was 99.9%, whereas the average coverage by HSPs was 96.7%. Among all other species, the one yielding the highest score was *Deferribacter desulfuricans*, which corresponded to an identity of 88.1% and an HSP coverage of 86.0%. The highest-scoring environmental sequence was DQ424925 ('Enrichment and Thermophilic Mediator-Less Microbial Fuel Cell thermophilic microbial fuel cell enriched artificial wastewater clone 1B62') [9], which showed an identity of 99.7% and an HSP coverage of 98.2%. The most frequently occurring keywords within the labels of environmental samples were 'microbiota' (4.1%), 'microbi' (4.1%), 'intestin' (4.0%), 'mous' (3.8%) and 'compet, exploit, inflamm, salmonella, typhimurium' (3.7%) (183 hits in total). The most frequently occurring keywords within the labels of environmental samples which yielded hits of a higher score than the highest scoring species were 'microbi' (8.6%), 'thermophil' (6.7%), 'enrich' (5.7%), 'cell, fuel' (5.3%) and 'spring' (3.6%) (21 hits in total), which seem to fit to the features known for *C. nitroreducens*.

Figure 1 shows the phylogenetic neighborhood of *C. nitroreducens* Yu37-1^T in a 16S rRNA based tree. The two copies of the 16S rRNA gene in the genome differ by one nucleotide from each other any by up to one nucleotide from the previously published 16S rRNA sequence (AB364234).

Cells of the strain Yu37-1^T are vibrio-shaped, 0.4-0.5 x 1.4-2.0 μm in size, occur singly or in pairs and stain Gram-negative [1] (Table 1 and Figure 2). No spore formation was detected for Yu37-1^T [1]. No data is available on the generation time of strain Yu37-1^T. Nitrate is the only electron acceptor utilized, with ammonium as the end product [1]. Elemental sulfur, sulfate, sulfite, nitrite, iron (III) oxide, manganese (IV) oxide, selenate, selenite, arsenate, arsenite, fumarate and oxygen are not used as alternative electron acceptors [1]. Acetate, pyruvate, lactate, fumarate, succinate, malate, yeast extract, peptone and Casamino acids are utilized as electron donors with nitrate as the electron acceptor; fermentative growth has not been observed [1]. Strain Yu37-1^T is strictly anaerobic and catalase negative [1].

Chemotaxonomy

The predominant compounds in whole cell lipids of *C. nitroreducens* strain Yu37-1^T are saturated branched-chain fatty acids: iso-$C_{14:0}$ (26.3%), anteiso-$C_{15:0}$ (24.1%), iso-$C_{13:0}$ (7.7%), $C_{18:0}$ (7.2%), $C_{16:0}$ (6.2%), iso-$C_{16:0}$ (5.7%) and anteiso-$C_{13:0}$ (5.3%) [1]. Menaquinone MK-8 was identified as the major quinone [1].
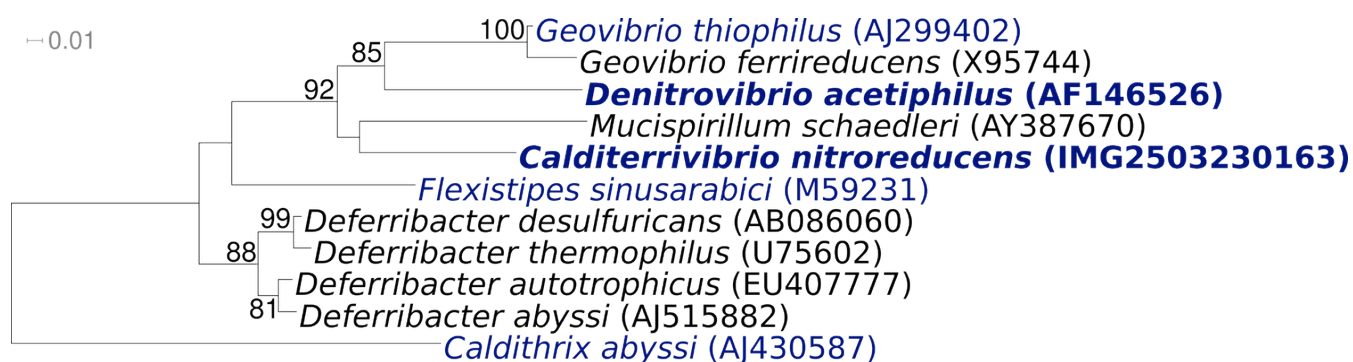
Figure 1. Phylogenetic tree highlighting the position of *C. nitroreducens* Yu37-1^T relative to the other type strains within the family *Deferribacteraceae*. The tree was inferred from 1,470 aligned characters [10,11] of the 16S rRNA gene sequence under the maximum likelihood criterion [12] and rooted 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 [13] if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [14] are shown in blue, published genomes in bold [15].
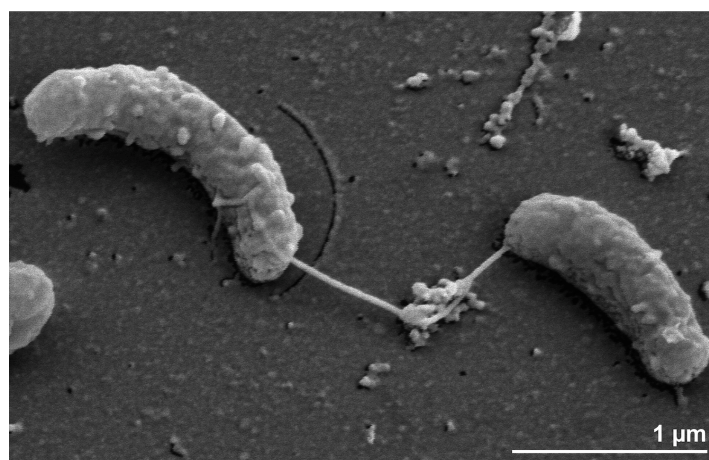
Figure 2. Scanning electron micrograph of *C. nitroreducens* Yu37-1[T]

Table 1. Classification and general features of *C. nitroreducens* Yu37-1[T] according to the MIGS recommendations [16]

MIGS ID	Property	Term	Evidence code
		Domain *Bacteria*	TAS [17]
		Phylum *Deferribacteres*	TAS [18-20]
		Class *Deferribacteres*	TAS [18,21]
	Current classification	Order *Deferribacterales*	TAS [18,22]
		Family *Deferribacteraceae*	TAS [18,23]
		Genus *Calditerrivibrio*	TAS [1]
		Species *Calditerrivibrio nitroreducens*	TAS [1]
		Type strain Yu37-1	TAS [1]
	Gram stain	negative	TAS [1]
	Cell shape	vibrio-shaped	TAS [1]
	Motility	motile, single polar flagellum	TAS [1]
	Sporulation	none	TAS [1]
	Temperature range	30°C–65°C	TAS [1]
	Optimum temperature	55°C	TAS [1]
	Salinity	< 0.5% NaCl	TAS [1]
MIGS-22	Oxygen requirement	strictly anaerobic	TAS [1]
	Carbon source	carbohydrates	TAS [1]
	Energy source	chemoorganoheterotrophic	TAS [1]
MIGS-6	Habitat	hot spring	TAS [1]
MIGS-15	Biotic relationship	not reported	
MIGS-14	Pathogenicity	not reported	
	Biosafety level	1	TAS [24]
	Isolation	hot spring water	TAS [1]
MIGS-4	Geographic location	Yumata, Nagano, Japan	TAS [1]
MIGS-5	Sample collection time	2008 or before	TAS [1]
MIGS-4.1	Latitude	36.83	TAS [1]
MIGS-4.2	Longitude	138.22	
MIGS-4.3	Depth	0 m, surface waters	TAS [1]
MIGS-4.4	Altitude	not reported	

Evidence codes - IDA: Inferred from Direct Assay (first time in publication); TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from of the Gene Ontology project [25]. If the evidence code is IDA, then the property was directly observed by one of the authors or an expert mentioned in the acknowledgements.

Genome annotation

Genes were identified using Prodigal [35] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [36]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes - Expert Review (IMG-ER) platform [37].

Genome properties

The genome consists of a 2,157,835 bp long chromosome with a 36% GC content and a 58,717 bp plasmid with 31% GC content (Figures 3a and 3b and Table 3). Of the 2,278 genes predicted, 2,128 were protein-coding genes, and 50 RNAs; 27 pseudogenes were identified. The majority of the protein-coding genes (76.5%) were assigned with a putative function while the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 4.
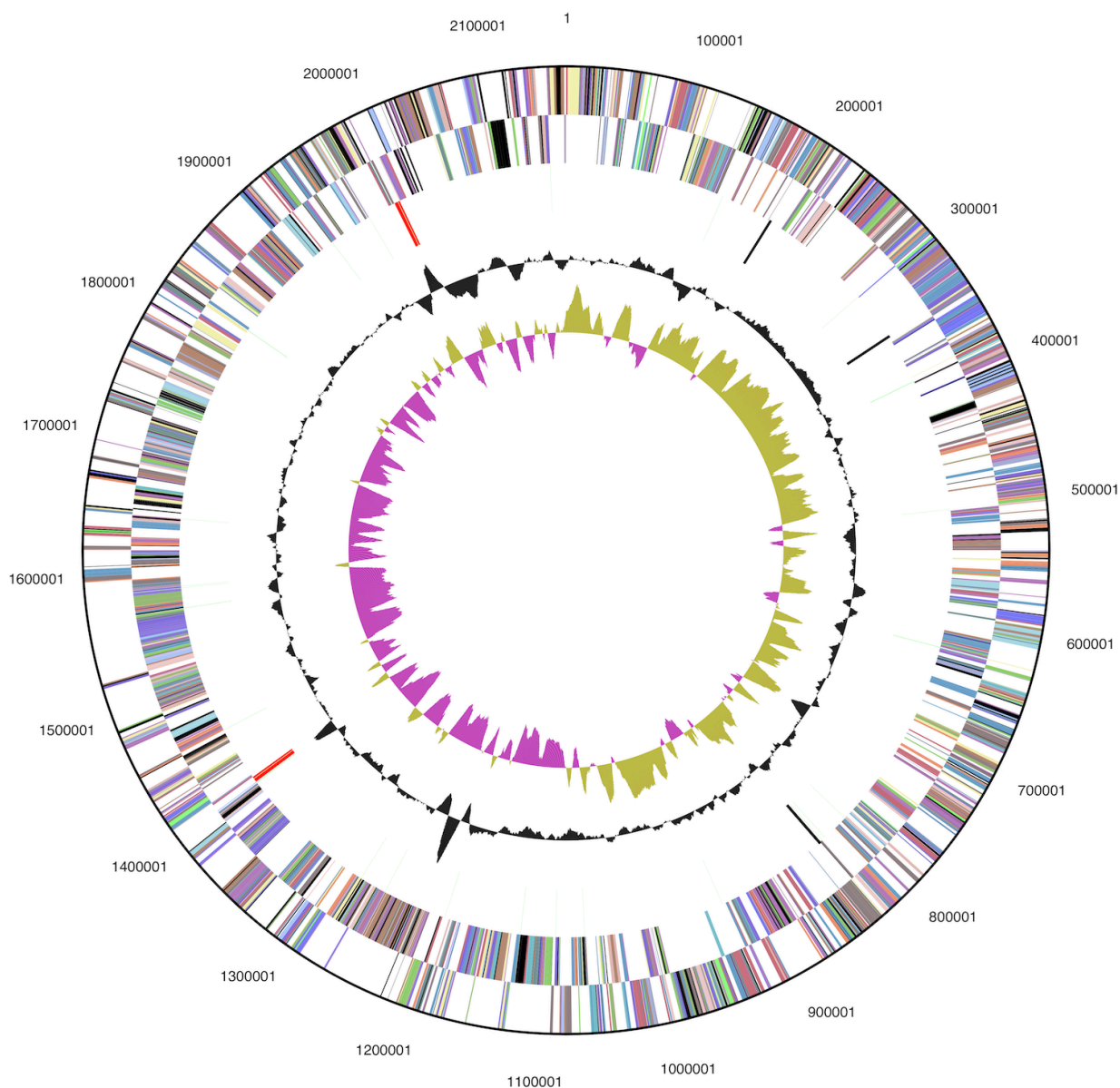
Figure 3a. Graphical circular map of the chromosome. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.
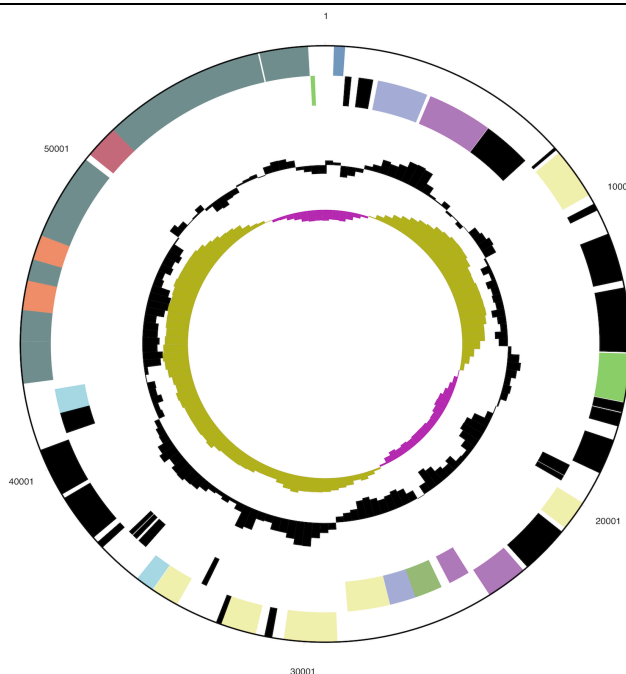
Figure 3b. Graphical circular map of the plasmid (not drown to scale with chromosome). From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	2,216,552	100.00%
DNA coding region (bp)	2,076,059	93.66%
DNA G+C content (bp)	789,723	35.63%
Number of replicons	2	
Extrachromosomal elements	1	
Total genes	2,178	100.00%
RNA genes	50	2.30%
rRNA operons	2	
Protein-coding genes	2,128	97.70%
Pseudo genes	27	1.24%
Genes with function prediction	1,666	76.49%
Genes in paralog clusters	190	8.72%
Genes assigned to COGs	1,731	79.48%
Genes assigned Pfam domains	1,800	82.64%
Genes with signal peptides	295	13.54%
Genes with transmembrane helices	520	23.88%
CRISPR repeats	3	

Table 4. Number of genes associated with the general COG functional categories

Code	value	%age	Description
J	141	7.4	Translation, ribosomal structure and biogenesis
A	0	0.0	RNA processing and modification
K	78	4.1	Transcription
L	96	5.0	Replication, recombination and repair
B	1	0.1	Chromatin structure and dynamics
D	22	1.2	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	34	1.8	Defense mechanisms
T	139	7.3	Signal transduction mechanisms
M	145	7.6	Cell wall/membrane/envelope biogenesis
N	81	4.3	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	68	3.6	Intracellular trafficking and secretion, and vesicular transport
O	84	4.4	Posttranslational modification, protein turnover, chaperones
C	145	7.6	Energy production and conversion
G	67	3.5	Carbohydrate transport and metabolism
E	142	7.5	Amino acid transport and metabolism
F	53	2.8	Nucleotide transport and metabolism
H	112	5.9	Coenzyme transport and metabolism
I	58	3.0	Lipid transport and metabolism
P	87	4.6	Inorganic ion transport and metabolism
Q	25	1.3	Secondary metabolites biosynthesis, transport and catabolism
R	214	11.2	General function prediction only
S	113	5.9	Function unknown
-	447	20.5	Not in COGs

Acknowledgements

References

1. Iino T, Nakagawa T, Mori K, Harayama S, Suzuki K. *Calditerrivibrio nitroreducens* gen. nov., sp. nov., a thermophilic, nitrate-reducing bacterium isolated from a terrestrial hot spring in Japan. *Int J Syst Evol Microbiol* 2008; **58**:1675-1679. PubMed doi:10.1099/ijs.0.65714-0

2. Fiala G, Woese CR, Langworthy TA, Stetter KO. *Flexistipes sinusarabici*, a novel genus and species of eubacteria occurring in the Atlantis II Deep brines of the Red Sea. *Arch Microbiol* 1990; **154**:120-126. doi:10.1007/BF00423320

3. Greene AC, Patel BKC, Sheehy AJ. *Deferribacter thermophilus* gen. nov., sp. nov., a novel thermophilic manganese and iron-reducing bacterium isolated from a petroleum reservoir. *Int J Syst Bacteriol* 1997; **47**:505-509. PubMed doi:10.1099/00207713-47-2-505

4. Myhr S, Torsvik T. *Denitrovibrio acetiphilus*, a novel genus and species of dissimilatory nitrate-

reducing bacterium isolated from an oil reservoir model column. *Int J Syst Evol Microbiol* 2000; **50**:1611-1619. PubMed

5. Caccavo F, Coates JD, Rossello-Mora RA, Ludwig W, Schleifer KH, Lovley DR, McInerney MJ. *Geovibrio ferrireducens*, a phylogenetically distinct dissimilatory Fe(III)-reducing bacterium. *Arch Microbiol* 1996; **165**:370-376. PubMed doi:10.1007/s002030050340

6. Robertson BR, O'Rourke JL, Neilan BA, Vandamme P, On SLW, Fox JG, Lee A. *Mucispirillum schaedleri* gen. nov., sp. nov., a spiral-shaped bacterium colonizing the mucus layer of the gastrointestinal tract of laboratory rodents. *Int J Syst Evol Microbiol* 2005; **55**:1199-1204. PubMed doi:10.1099/ijs.0.63472-0

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

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

9. Jong BC, Kim BH, Chang IS, Liew PW, Choo YF, Kang GS. Enrichment, performance, and microbial diversity of a thermophilic mediatorless microbial fuel cell. *Environ Sci Technol* 2006; **40**:6449-6454. PubMed doi:10.1021/es0613512

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

11. 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

12. 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

13. 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

14. 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

15. Kiss H, Lang E, Lapidus A, Copeland A, Nolan M, Rio TGD, Chen F, Lucas S, Tice H, Cheng JF, *et al*. Complete genome sequence of *Denitrovibrio acetiphilus* type strain (N2460[T]). *Stand Genomic Sci* 2010; **2**:270-279. PubMed doi:10.4056/sigs.892105

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

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**:4576-4579. PubMed doi:10.1073/pnas.87.12.4576

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

19. Garrity GM, Holt JG. Phylum BIX. *Deferribacteres* phy. nov. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 1 (*The Archaea and the Deeply Branching and Phototrophic Bacteria*), Edited by D. R. Boone, R. W. Castenholz and G. M. Garrity. *New York: Springer* 2001, pp.465-471.

20. Jumas-Bilak E, Roudière L, Marchandin H. Description of '*Synergistetes*' phyl. nov. and emended description of the phylum '*Deferribacteres*' and of the family *Syntrophomonadaceae*, phylum '*Firmicutes*'. *Int J Syst Evol Microbiol* 2009; **59**:1028-1035. PubMed doi:10.1099/ijs.0.006718-0

21. Huber H, Stetter KO. Class I. *Deferribacteres* class. nov. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol. 1 (*The Archaea and the Deeply Branching and Phototrophic Bacteria*), Edited by D. R. Boone, R. W. Castenholz and G. M. Garrity. *New York: Springer* 2001, pp. 465-471.

22. Huber H, Stetter KO. Order I. *Deferribacterales* ord. nov. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, vol.1 (*The Archaea and the Deeply Branching and Phototrophic Bacteria*)

Edited by D. R. Boone, R. W. Castenholz and G. M. Garrity. *New York: Springer* 2001:pp.465-471.

23. Huber H, Stetter KO. Family I. *Deferribacteraceae* fam. nov. In Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 1, Edited by D. R. Boone and R. W. Castenholz. *New York: Springer* 2002:pp. 465-466.

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

27. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al*. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 2009; **462**:1056-1060. PubMed doi:10.1038/nature08656

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

29. Gemeinholzer B, Dröge G, Zetzsche H, Haszprunar G, Klenk HP, Güntsch A, Berendsohn WG, Wägele JW. The DNA Bank Network: the start from a German initiative. *Biopreservation and Biobanking* (In press).

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

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

32. 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

33. Han C, Chain P. 2006. Finishing repeat regions automatically with Dupfinisher. in Proceeding of the 2006 international conference on bioinformatics & computational biology. Edited by Hamid R. Arabnia & Homayoun Valafar, CSREA Press. June 26-29, 2006: 141-146.

34. 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.

35. 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

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

37. 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