

High quality draft genome sequence of *Olivibacter sitiensis* type strain (AW-6^T), a diphenol degrader with genes involved in the catechol pathway

Spyridon Ntougias¹, Alla Lapidus², James Han², Konstantinos Mavromatis², Amrita Pati², Amy Chen³, Hans-Peter Klenk⁴, Tanja Woyke², Constantinos Fasseas⁵, Nikos C. Kyrpides^{2,6}, Georgios I. Zervakis^{7,*}

¹Democritus University of Thrace, Department of Environmental Engineering, Laboratory of Wastewater Management and Treatment Technologies, Xanthi, Greece

²Department of Energy Joint Genome Institute, Genome Biology Program, Walnut Creek, CA, USA

³Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

⁴Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

⁵Agricultural University of Athens, Electron Microscopy Laboratory, Athens, Greece

⁶King Abdulaziz University, Jeddah, Saudi Arabia

⁷Agricultural University of Athens, Laboratory of General and Agricultural Microbiology, Athens, Greece

*Correspondence: Georgios I. Zervakis (zervakis@aua.gr)

Keywords: alkaline two-phase olive mill waste, *Bacteroidetes*, *Sphingobacteriaceae*, hemicellulose degradation, β -1,4-xylanase, β -1,4-xylosidase

Olivibacter sitiensis Ntougias et al. 2007 is a member of the family *Sphingobacteriaceae*, phylum *Bacteroidetes*. Members of the genus *Olivibacter* are phylogenetically diverse and of significant interest. They occur in diverse habitats, such as rhizosphere and contaminated soils, viscous wastes, composts, biofilter clean-up facilities on contaminated sites and cave environments, and they are involved in the degradation of complex and toxic compounds. Here we describe the features of *O. sitiensis* AW-6^T, together with the permanent-draft genome sequence and annotation. The organism was sequenced under the Genomic Encyclopedia for *Bacteria* and *Archaea* (GEBA) project at the DOE Joint Genome Institute and it is the first genome sequence of a species within the genus *Olivibacter*. The genome is 5,053,571 bp long and is comprised of 110 scaffolds with an average GC content of 44.61%. Of the 4,565 genes predicted, 4,501 were protein-coding genes and 64 were RNA genes. Most protein-coding genes (68.52%) were assigned to a putative function. The identification of 2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase-coding genes indicates involvement of this organism in the catechol catabolic pathway. In addition, genes encoding for β -1,4-xylanases and β -1,4-xylosidases reveal the xylanolytic action of *O. sitiensis*.

Introduction

The genus *Olivibacter* currently contains six species with validly published names, all of which are aerobic and heterotrophic, non-motile, rod-shaped Gram-negative bacteria [1-3]. Strain AW-6^T (= DSM 17696^T = CECT 7133^T = CIP 109529^T) is the type strain of *Olivibacter sitiensis* [1], which is the type species of the genus *Olivibacter*. The strain was isolated from alkaline alperujo, an olive mill sludge-like waste produced by two-phase centrifugal decanters located in the vicinity of

Toplou Monastery, Sitia, Greece [1]. The genus name derived from the Latin term *oliva* and the Neo-Latin *bacter*, meaning a rod-shaped bacterium living in olives/olive processing by-products [1]. The Neo-Latin species epithet *sitiensis* pertains to the region Sitia (Crete, Greece) where the olive mill is operating [1]. The other species of the genus are *O. soli*, *O. ginsengisoli*, *O. terrae*, *O. oleidegradans* and *O. jilunii* [2-4]. *O. soli* and *O. ginsengisoli* were isolated from soil of a ginseng field [2], *O. terrae* from a compost prepared of cow manure and rice straw [2], *O. oleidegradans* from a



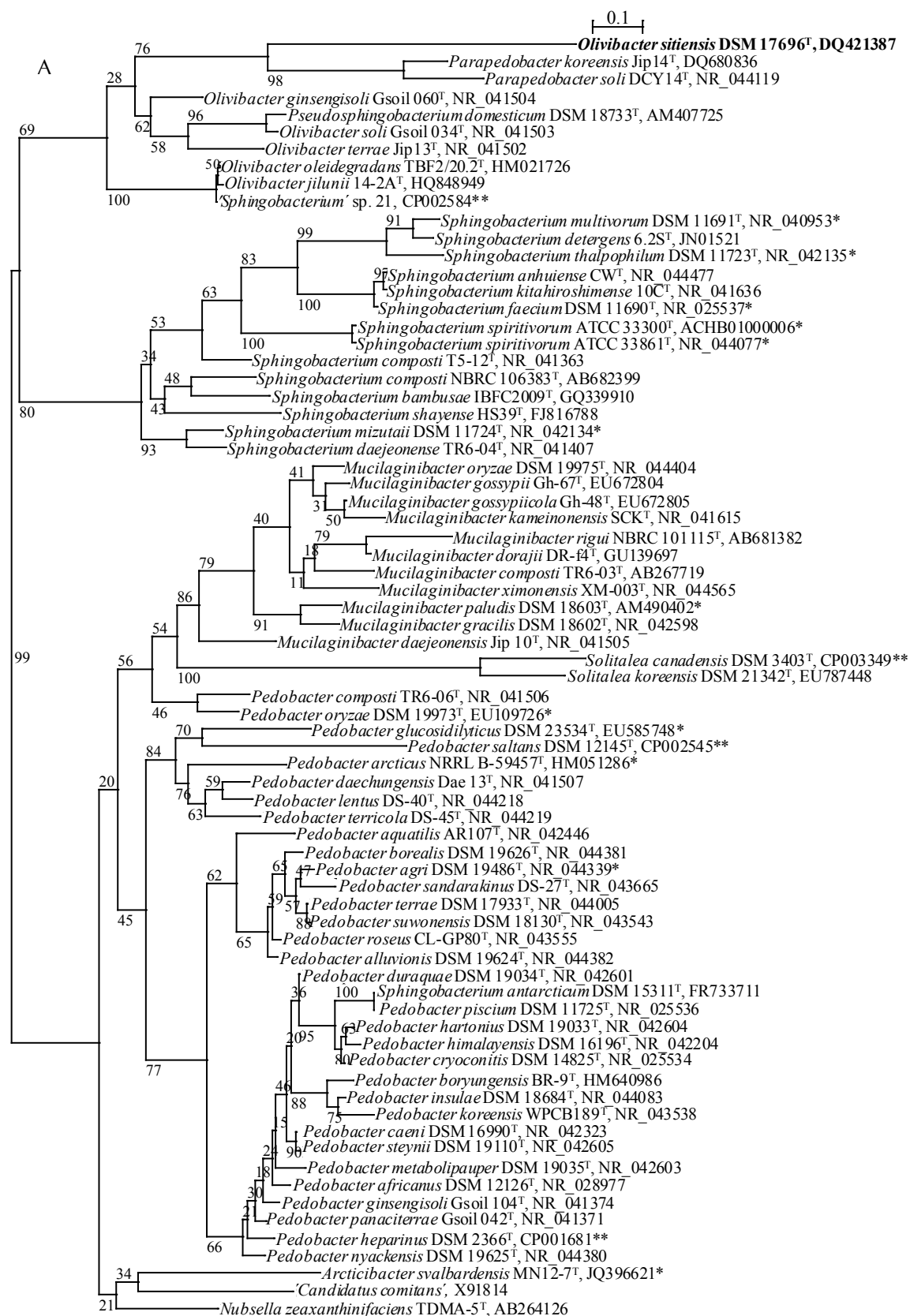
biofilter clean-up facility in a hydrocarbon-contaminated site [3] and *O. jilunii* from a DDT-contaminated soil [4]. *O. sitiensis* can be distinguished from *O. soli*, *O. ginsengisoli* and *O. terrae* on the basis of temperature and NaCl concentration ranges for growth, in its ability to assimilate N-acetyl-D-glucosamine, L-histidine, maltose and sorbitol, and for expression of naphthol-AS-BI-phosphohydrolase, in the presence/absence of iso-C_{15:1} F, C_{16:1} 2-OH, anteiso-C_{17:1} B and/or iso-C_{17:1} I, and in by its DNA G+C content [1,2,4]. Moreover, it differs from *O. soli* in terms of L-arabinose assimilation and valine arylamidase expression, from *O. ginsengisoli* in terms of inositol, mannitol and salicin assimilation and in oxidase reaction test, and from *O. terrae* in terms of L-arabinose and mannitol assimilation, and β -glucuronidase and valine arylamidase expression [1,2,4]. *O. sitiensis* can be differentiated from *O. oleidegradans* on the basis of DNA G+C content, pH upper limit for growth, in the ability for assimilation of D-adonitol, L-arabinose, N-acetyl-D-glucosamine, L-histidine, D-lyxose, maltose, melezitose, salicin and turanose, and for expression of esterase, β -galactosidase, α -mannosidase, urease and valine arylamidase as well as in the presence/absence of some minor fatty acid components of membrane lipids, menaquinone-6 (as minor respiratory quinone) and aminophospholipids (as cellular polar lipids) [1,3,4]. In addition, *O. sitiensis* can be distinguished from *O. jilunii* on the basis of DNA G+C content, pH, temperature and NaCl concentration upper limits for growth, lactose fermentation, in the ability for assimilation of acetate, L-arabinose, N-acetyl-D-glucosamine, L-histidine, malonate, maltose, D-mannose, salicin and L-serine, and for expression of α -mannosidase, oxidase and valine arylamidase as well as in the presence/absence of some minor fatty acid components of membrane lipids, menaquinone-8 (as minor respiratory quinone) and aminophospholipids (as cellular polar lipids) [1,4]. Here we present a summary classification and a set of features for *O. sitiensis* AW-6^T, together with the description of the permanent-draft genome sequencing and annotation.

Classification and features

The 16S rRNA gene sequence of *O. sitiensis* AW-6^T 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 [5] and the relative frequencies of taxa and keywords (reduced to their stem [6]) were determined and weighted by BLAST scores. The frequency of genera that belonged to the family *Sphingobacteriaceae* was 61.8%. The most fre-

quently occurring genera were in order *Sphingobacterium* (27.7%), *Pedobacter* (17.1%), *Flavobacterium* (8.5%), *Olivibacter* (6.4%), *Hymenobacter* (6.4%), *Mucilaginibacter* (4.3%), *Cytophaga* (4.3%), *Flectobacillus* (4.3%), *Parapedobacter* (2.1%), *Pseudosphingobacterium* (2.1%) and '*Hevizibacter*' (2.1%) (47 hits in total). The 16S rRNA gene sequence of *O. sitiensis* AW-6^T was the only hit on members of the species in INSDC (=EMBL/NCBI/DDBJ) under the accession number DQ421387 (=NR_043805). Among all other species, the two yielding the highest score were *Parapedobacter koreensis* Jip14^T (DQ680836) [7] and *Olivibacter ginsengisoli* Gsoil 060^T (AB267716) [2], showing similarity in 16S rRNA gene of 90.1% (both of them) and HSP coverages of 99.8% and 99.9% respectively. It is noteworthy 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 sequences was AM114441 ['Interactions U(VI) added natural dependence on various incubation conditions soil uranium mining waste pile clone JG35+U2A-AG9'], which showed identity of 90.3% with HSP coverage of 86.1%. The most frequently occurring keywords within the labels of all environmental samples that yielded hits were 'rumen' (23.1%), 'oil' (10.8%), 'water' (9.7%), 'soil' (9.7%), 'fluid' (9.1%) and 'gut' (9.1%) (186 hits in total). The most frequently occurring keywords within the labels of those environmental samples that yielded hits of a higher score than the highest scoring species were 'waste' (50.0%) and 'soil' (50.0%) (4 hits in total), which are keywords with biological meaning fitting the environment from which *O. sitiensis* AW-6^T was isolated.

Figure 1 shows the phylogenetic neighborhood of *O. sitiensis* in the 16S rRNA gene sequence-based trees constructed. Independently from the clustering method applied, all *Olivibacter* species together with *Pseudosphingobacterium domesticum* and '*Sphingobacterium*' sp. 21 fell into a distinct cluster, indicating the unique phylogenetic position of genus *Olivibacter* and the necessity for reconsidering the taxonomic status of the genus *Pseudosphingobacterium*. In addition, '*Sphingobacterium*' sp. 21 should be assigned to the genus *Olivibacter*, and not to the genus *Sphingobacterium*. In the ML tree, members of the genus *Parapedobacter* branched together with *O. sitiensis*, although the unique topology of the genus was established by applying a character-based (parsimony) method. As previously stated by Ntougias et al. [1], *S. antarcticum* should be reassigned to the genus *Pedobacter*.



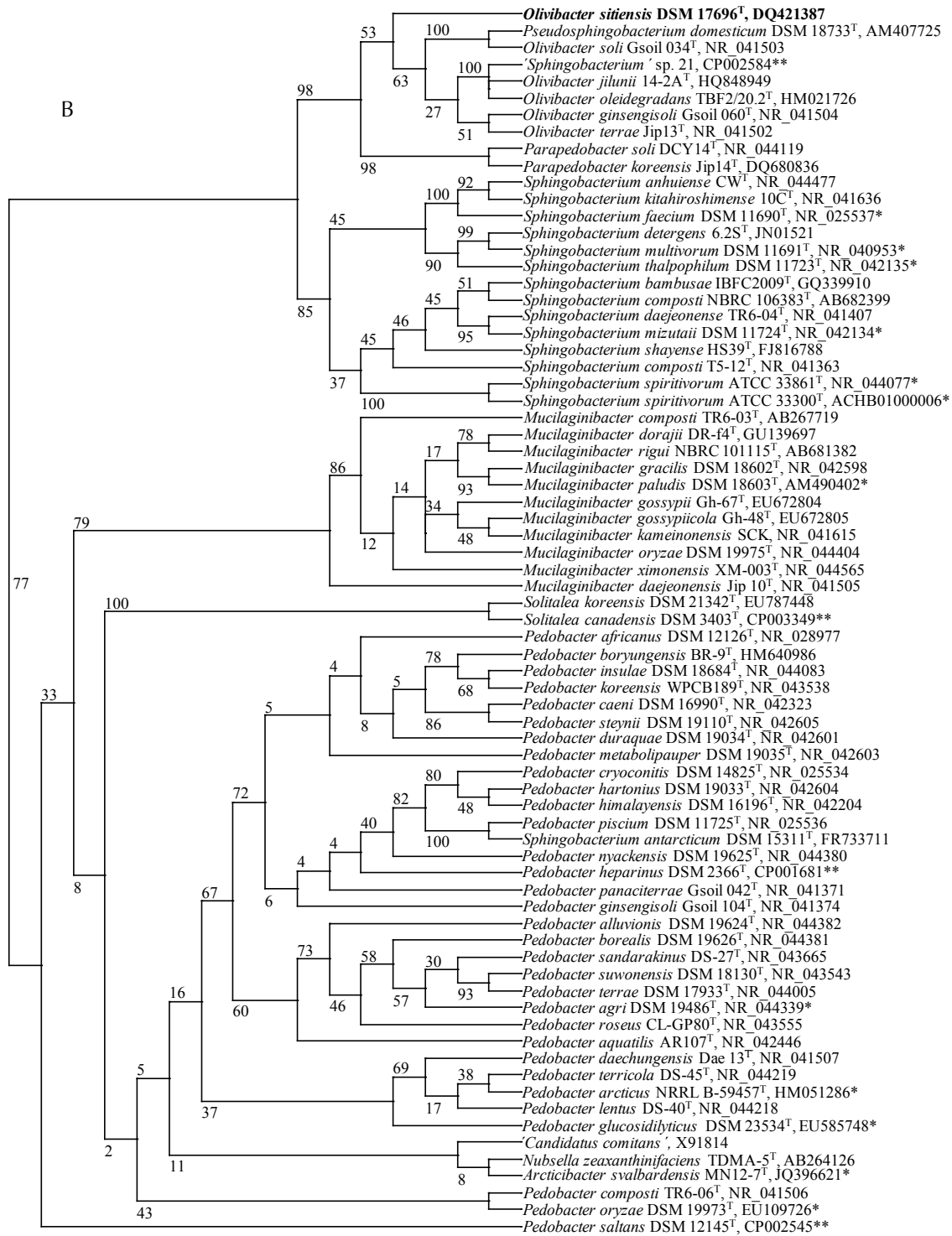


Figure 1. Phylogenetic trees highlighting the position of *O. sitiensis* relative to the type strains of the species within the family *Sphingobacteriaceae*. The tree was inferred from 1,288 aligned characters [8,9] of the 16S rRNA gene sequence under (A) [previous page] the maximum likelihood (ML) [10] and (B) [this page] the maximum-parsimony criterion. In ML tree, the branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches are support values from 100 ML bootstrap replicates (A) and from 1,000 maximum-parsimony bootstrap replicates (B) [11]. Lineages with strain genome sequencing projects registered in GOLD [12] are labeled with one asterisk, while those listed as 'Complete and Published' with two asterisks (e.g. *Pedobacter heparinus* [13] and *P. saltans* [14]).

Cells of *O. sitiensis* AW-6^T are Gram-negative non-motile rods [1] with a length of 1.0-1.3 μm and a width of 0.2-0.3 μm (Table 1 and Figure 2). The temperature range for growth is 5-45°C, with an optimum at 28-32°C [1]. *O. sitiensis* is neutrophilic, showing no growth at 30 g L⁻¹ NaCl [1]. The pH for growth ranges between 5 and 8, with pH values of 6-7 being the optimum [1]. *O. sitiensis* is strictly aerobic and chemo-organotrophic; it assimilates mostly D(+)-glucose,

protocatechuate and D(+)-xylose, while L-cysteine, D(-)-fructose, D(+)-galactose, L-histidine, lactose, sorbitol and sucrose are also utilized by strain AW-6^T [1]. *O. sitiensis* was found to be sensitive to ampicillin, bacitracin, chloramphenicol, penicillin, rifampicin, tetracycline and trimethoprim, and resistant to kanamycin, polymixin B and streptomycin (antibiotics' concentration of 50 $\mu\text{g ml}^{-1}$) [1].

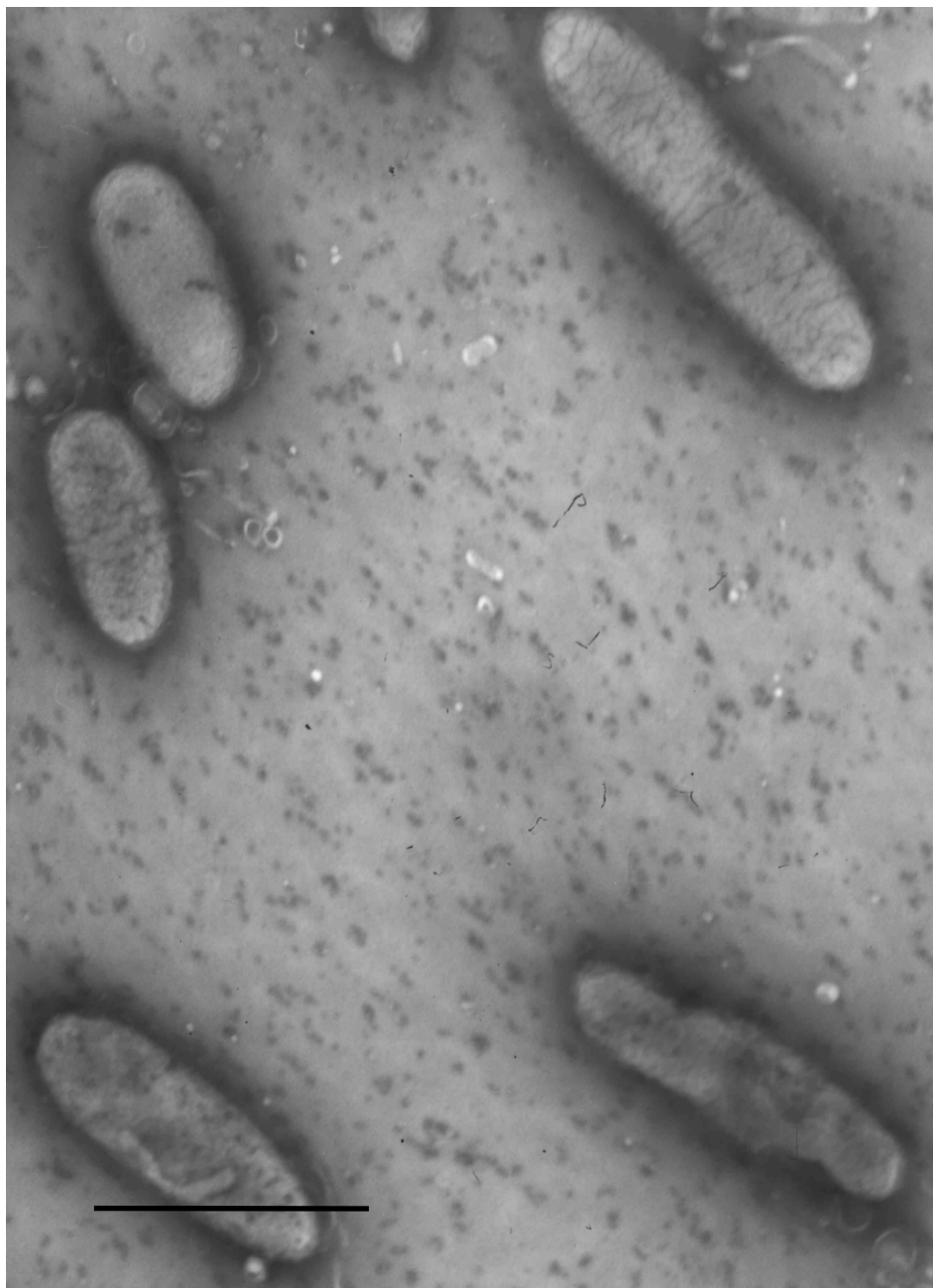


Figure 2. Electron micrograph of *O. sitiensis* AW-6^T negatively-stained cells. Bar represents 1 μm .

Table 1. Classification and general features of *O. sitiensis* AW-6^T, according to the MIGS recommendations [15].

MIGS ID	Property	Term	Evidence code ^a
		Domain <i>Bacteria</i>	TAS [16]
		Phylum <i>Bacteroidetes</i>	TAS [17,18]
		Class <i>Sphingobacteriia</i>	TAS [18,19]
	Current classification	Order <i>Sphingobacteriales</i>	TAS [18,20]
		Family <i>Sphingobacteriaceae</i>	TAS [21]
		Genus <i>Olivibacter</i>	TAS [1]
		Species <i>Olivibacter sitiensis</i>	TAS [1]
		Type-strain AW-6 ^T	TAS [1]
	Gram stain	negative	TAS [1]
	Cell shape	rod	TAS [1]
	Motility	non-motile	TAS [1]
	Sporulation	non-sporulating	TAS [1]
	Temperature range	mesophile, 5-45°C	TAS [1]
	Optimum temperature	28-32°C	TAS [1]
	Salinity	neutrophilic and non-halotolerant -no growth at 30 g l ⁻¹ NaCl	TAS [1]
MIGS-22	Oxygen requirement	strictly aerobic	TAS [1]
	Carbon source	carbohydrates and amino-acids, utilization of protocatechuate and sorbitol	TAS [1]
	Energy metabolism	chemo-organotroph	TAS [1]
MIGS-6	Habitat	olive mill waste	TAS [1]
MIGS-15	Biotic relationship	free living	TAS [1]
MIGS-14	Pathogenicity	none	NAS
	Biosafety level	1	TAS [22]
MIGS-23.1	Isolation	alkaline two-phase olive mill waste (alkaline alperujo)	TAS [1]
MIGS-4	Geographic location	Toplou Monastery, Sitia, Crete, Greece	TAS [1]
MIGS-5	Sample collection time	year 2003	NAS
MIGS-4.1	Latitude	35.220	TAS [1]
MIGS-4.2	Longitude	26.216	TAS [1]
MIGS-4.3	Depth	surface	NAS
MIGS-4.4	Altitude	161 m	NAS

^aEvidence codes - 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 the Gene Ontology project [23]. If the evidence code is IDA, then the property was directly observed for a living isolate by one of the authors or an expert mentioned in the acknowledgements.

Chemotaxonomy

The major polar lipids of *O. sitiensis* are phosphatidylethanolamine (PE), phosphatidylmonomethylethanolamine (PME), phosphatidylinositol mannoside (PIM), an unknown phospholipid (PL) and an unknown non-phosphorylated lipid (UL) [4]. Moreover, the main membrane fatty acids of *O. sitiensis* are C_{16:1}ω7c and/or iso-C_{15:0} 2-OH, iso-C_{15:0}, iso-C_{17:0} 3-OH and

C_{16:0} [1]. The only respiratory quinone found in *O. sitiensis* is menaquinone with seven isoprene subunits (MK-7) [1].

Genome sequencing and annotation

Genome project history

This microorganism was selected for sequencing on the basis of its phylogenetic position [24,25], and is part of the Genomic Encyclopedia of Type

Strains, Phase I: the one thousand microbial genomes (KMG) project [26] which aims in increasing the sequencing coverage of key reference microbial genomes. The genome project is deposited in the Genomes On Line Database [12] and the genome sequence is available from GenBank. Se-

quencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI) using state of the art sequencing technology [27]. A summary of the project information is presented in Table 2.

Table 2. Genome sequencing project information.

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-Quality Draft
MIGS-29	Sequencing platforms	Illumina
MIGS-31.2	Sequencing coverage	120×
MIGS-30	Assemblers	ALLPATHS v. r41043
MIGS-32	Gene calling method	Prodigal 2.5
	Genbank ID	ATZA00000000
	Genbank Date of Release	September 5, 2013
	GOLD ID	Gi11724
	NCBI project ID	165253
	Database: IMG	2515154027
	MIGS-13	Source material identifier
	Project relevance	GEBA-KMG, Tree of Life, Biodegradation

Growth conditions and DNA isolation

O. sitiensis strain AW-6^T was grown aerobically in DSMZ medium 92 (trypticase soy yeast extract medium) [28] 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 but applying a modified cell lysis procedure (1 hour incubation at 58°C with additional 50 µl proteinase K followed by overnight incubation on ice with additional 200 µl PPT-buffer). DNA is available via the DNA Bank Network [29].

Genome sequencing and assembly

The draft genome of *Olivibacter sitiensis* DSM 17696 was generated at the DOE Joint genome Institute (JGI) using the Illumina technology. An Illumina Standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform, which generated 13,155,872 reads totaling 1,973.4 Mbp. All general aspects of library construction and sequencing performed at the JGI can be found at the JGI website [30]. All raw Illumina sequence data were passed through DUK, a filtering program developed at JGI, which removes known Illumina sequencing and library preparation artifacts (Mingkun L, unpublished). The following steps were then performed for assembly: (i) filtered Illumina reads were assembled using Velvet (version 1.1.04) [31], (ii) 1–3 Kbp simulated paired end reads were created from

Velvet contigs using wgsim [32] (iii) Illumina reads were assembled with simulated read pairs using Allpaths-LG (version r41043) [33]. The final draft assembly contained 110 contigs in 110 scaffolds. The total size of the genome is 5.1 Mbp and the final assembly is based on 605.8 Mbp of Illumina data, which provides an average 120.0× coverage of the genome.

Genome annotation

Genes were identified using Prodigal [34] as part of the DOE-JGI Annotation pipeline [35]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analysis and functional annotation was performed within the Integrated Microbial Genomes (IMG-ER) [36].

Genome properties

The genome is 5,053,571 bp long and comprises 110 scaffolds with an average GC content of 44.61% (Table 3). Of the 4,565 genes predicted, 4,501 were protein-coding genes and 64 RNA genes. Most protein-coding genes (68.52%) were assigned to 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 ^a
Genome size (bp)	5,053,571	100.00%
DNA coding region (bp)	4,534,282	89.72%
DNA G+C content (bp)	2,254,441	44.61%
DNA scaffolds	110	
Total genes	4,565	
RNA genes	64	1.40%
tRNA genes	47	1.03%
Protein-coding genes	4,501	98.60%
Genes with function prediction (proteins)	3,128	68.52%
Genes in paralog clusters	1,777	38.93%
Genes assigned to COGs	3,062	67.08%
Genes assigned Pfam domains	3,471	76.04%
Genes with signal peptides	501	10.97%
Genes with transmembrane helices	1,124	24.62%
CRISPR repeats	0	

^aThe total is based on either the size of the genome in base pairs or the total number of protein coding genes in the annotated genome

Table 4. Number of genes associated with the 25 general COG functional categories

Code	Value	%age ^a	Description
J	159	4.7	Translation, ribosomal structure and biogenesis
A	1	0.0	RNA processing and modification
K	283	8.4	Transcription
L	190	5.7	Replication, recombination and repair
B	1	0.0	Chromatin structure and dynamics
D	22	0.6	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	99	2.9	Defense mechanisms
T	197	5.9	Signal transduction mechanisms
M	274	8.1	Cell wall/membrane biogenesis
N	7	0.2	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	63	1.9	Intracellular trafficking and secretion, and vesicular transport
O	120	3.6	Posttranslational modification, protein turnover, chaperones
C	168	5.0	Energy production and conversion
G	259	7.7	Carbohydrate transport and metabolism
E	211	6.3	Amino acid transport and metabolism
F	61	1.8	Nucleotide transport and metabolism
H	148	4.4	Coenzyme transport and metabolism
I	107	3.2	Lipid transport and metabolism
P	238	7.1	Inorganic ion transport and metabolism
Q	52	1.5	Secondary metabolites biosynthesis, transport and catabolism
R	419	12.5	General function prediction only
S	280	8.3	Function unknown
-	1,503	32.9	Not in COGs

^aThe total is based on the total number of protein coding genes in the annotated genome.

Based on genomic analysis of the metabolic features, *O. sitiensis* is an auxotroph for L-alanine, L-arginine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-phenylalanine, L-proline, L-serine, L-tyrosine, L-tryptophan and L-valine, and a prototroph for L-aspartate, L-glutamate and glycine. Selenocysteine and biotin cannot be synthesized by *O. sitiensis*. Strain AW-6^T can utilize L-arabinose and maltose (via orthophosphate activation), whereas no maltose hydrolysis is achieved [1].

Genome analysis revealed the genetic and molecular bases of the degradation of recalcitrant compounds by *O. sitiensis*. The ability of *O. sitiensis* to degrade phenolic compounds is verified by the distribution of genes encoding oxidoreductases that act on diphenols and related substances and by the 2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase-coding genes that are involved in the catechol pathway. Genes encoding β -1,4-xylanases and β -1,4-xylosidases were also identified in the genome of strain AW-6^T, indicating that *O. sitiensis* is a xylanolytic bacterium involved in the cleavage of β -1,4-xylosic bonds in hemicelluloses. The existence of protocatechuate 3,4-dioxygenase (dioxygenase_C)-coding genes are indicative of the ability of this bacterium to degrade benzoate and 2,4-dichlorobenzoate. Genes encoding carboxymethylenebutenolidase were distributed in the genome of *O. sitiensis*, indicating its potential for hexachlorocyclohexane and 1,4-dichlorobenzene degradation. Oxidoreductases related to aryl-alcohol dehydrogenases were predicted, showing that *O. sitiensis* may be also involved in biphenyl and toluene/xylene degradation. This is also strengthened by the identification of an uncharacterized protein, possibly involved in aromatic compounds catabolism. Moreover, putative multicopper oxidases with possible laccase-like activity were identified. Mercuric reductase and arsenate reductase-coding genes as well as organic solvent tolerance and chromate transport proteins encoded in the genome indicate possible resistance of *O. sitiensis* to the presence of heavy metals and organic solvents.

Acknowledgments

We would like to gratefully acknowledge the help of Brian J. Tindall and his team for growing *O. sitiensis* cultures, and Evelyne-Marie Brambilla for DNA extraction and quality control (all 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, as well as German Research Foundation (DFG) INST 599/1-2.

References

1. Ntougias S, Fasseas C, Zervakis GI. *Olivibacter sitiensis* gen. nov., sp. nov., isolated from alkaline olive-oil mill wastes in the region of Sitia, Crete. *Int J Syst Evol Microbiol* 2007; **57**:398-404. [PubMed http://dx.doi.org/10.1099/ijs.0.64561-0](http://dx.doi.org/10.1099/ijs.0.64561-0)
2. Wang L, Ten LN, Lee HG, Im WT, Lee ST. *Olivibacter soli* sp. nov., *Olivibacter ginsengisoli* sp. nov. and *Olivibacter terrae* sp. nov., from soil of a ginseng field and compost in South Korea. *Int J Syst Evol Microbiol* 2008; **58**:1123-1127. [PubMed http://dx.doi.org/10.1099/ijs.0.65299-0](http://dx.doi.org/10.1099/ijs.0.65299-0)
3. Szabó I, Szoboszlai S, Kriszt B, Háhn J, Harkai P, Baka E, Táncsics A, Kaszab E, Privler Z, Kukolya J. *Olivibacter oleidegradans* sp. nov., a hydrocarbon degrading bacterium isolated from a biofilter cleanup facility on a hydrocarbon-contaminated site. *Int J Syst Evol Microbiol* 2011; **61**:2861-2865. [PubMed http://dx.doi.org/10.1099/ijs.0.026641-0](http://dx.doi.org/10.1099/ijs.0.026641-0)
4. Chen K, Tang SK, Wang GL, Nie GX, Li QF, Zhang JD, Li WJ, Li SP. *Olivibacter jilunii* sp. nov., isolated from a DDT-contaminated soil. *Int J Syst Evol Microbiol* 2013; **63**:1083-1088. [PubMed http://dx.doi.org/10.1099/ijs.0.042416-0](http://dx.doi.org/10.1099/ijs.0.042416-0)
5. 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 http://dx.doi.org/10.1128/AEM.03006-05](http://dx.doi.org/10.1128/AEM.03006-05)
6. Porter MF. An algorithm for suffix stripping. Program: electronic library and information systems 1980; **14**:130-137.
7. Kim MK, Na JR, Cho DH, Soung NK, Yang DC. *Parapedobacter koreensis* gen. nov., sp. nov. *Int J Syst Evol Microbiol* 2007; **57**:1336-1341. [PubMed http://dx.doi.org/10.1099/ijs.0.64677-0](http://dx.doi.org/10.1099/ijs.0.64677-0)
8. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464. [PubMed http://dx.doi.org/10.1093/bioinformatics/18.3.452](http://dx.doi.org/10.1093/bioinformatics/18.3.452)
9. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic

- analysis. *Mol Biol Evol* 2000; **17**:540-552. [PubMed](#)
<http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334>
10. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* 2008; **57**:758-771. [PubMed](#)
<http://dx.doi.org/10.1080/10635150802429642>
 11. Gouy M, Guindon S, Gascuel O. SeaView Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010; **27**:221-224. [PubMed](#)
<http://dx.doi.org/10.1093/molbev/msp259>
 12. Pagani I, Liolios K, Jansson J, Chen IMA, Smirnova T, Nosrat B, Markowitz VM, Kyrpides NC. The Genomes OnLine Database (GOLD) v.4: Status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2012; **40**:D571-D579. [PubMed](#)
<http://dx.doi.org/10.1093/nar/gkr1100>
 13. Han C, Spring S, Lapidus A, Glavina Del Rio T, Tice H, Copeland A, Cheng JF, Lucas S, Chen F, Nolan M, et al. Complete genome sequence of *Pedobacter heparinus* type strain (HIM 762-3^T). *Stand Genomic Sci* 2009; **1**:54-62. [PubMed](#)
<http://dx.doi.org/10.4056/signs.22138>
 14. Liolios K, Sikorski J, Lu M, Nolan M, Lapidus A, Lucas S, Hammon N, Deshpande S, Cheng JF, Tapia R, et al. Complete genome sequence of the gliding, heparinolytic *Pedobacter saltans* type strain (113^T). *Stand Genomic Sci* 2011; **5**:30-40. [PubMed](#) <http://dx.doi.org/10.4056/signs.2154937>
 15. 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](#)
<http://dx.doi.org/10.1038/nbt1360>
 16. 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](#)
<http://dx.doi.org/10.1073/pnas.87.12.4576>
 17. Krieg NR, Ludwig W, Euzéby J, Whitman WB. Phylum XIV. *Bacteroidetes* phyl. nov. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 4, Springer-Verlag, New York, 2011, p. 25.
 18. Editor L. Validation List No. 143. *Int J Syst Evol Microbiol* 2012; **62**:1-4.
 19. Kampf P. Class III. *Sphingobacteriia* class. nov. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 4, Springer-Verlag, New York, 2011, p. 330.
 20. Kampf P. Order I. *Sphingobacteriales* ord. nov. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 4, Springer-Verlag, New York, 2011, p. 330.
 21. Steyn PL, Segers P, Vancanneyt M, Sandra P, Kersters K, Joubert JJ. Classification of heparinolytic bacteria into a new genus, *Pedobacter*, comprising four species: *Pedobacter heparinus* comb. nov., *Pedobacter piscium* comb. nov., *Pedobacter africanus* sp. nov. and *Pedobacter saltans* sp. nov. proposal of the family *Sphingobacteriaceae* fam. nov. *Int J Syst Bacteriol* 1998; **48**:165-177. [PubMed](#)
<http://dx.doi.org/10.1099/00207713-48-1-165>
 22. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA), Classification of prokaryotes (bacteria and archaea) into risk groups. Technical Rule for Biological Agents 466 (TRBA 466), Germany, 2010, p. 157.
 23. 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. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. [PubMed](#)
<http://dx.doi.org/10.1038/75556>
 24. Klenk HP, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst Appl Microbiol* 2010; **33**:175-182. [PubMed](#)
<http://dx.doi.org/10.1016/j.syapm.2010.03.003>
 25. 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](#)
<http://dx.doi.org/10.1038/nature08656>
 26. Kyrpides NC, Woyke T, Eisen JA, Garrity G, Lilburn TG, Beck BJ, Whitman WB, Hugenholtz P, Klenk HP. Genomic Encyclopedia of Type Strains, Phase I: the one thousand microbial genomes (KMG-I) project. *Stand Genomic Sci* 2013; **9**:628-634; <http://dx.doi.org/10.4056/signs.5068949>.
 27. Mavromatis K, Land ML, Brettin TS, Quest DJ, Copeland A, Clum A, Goodwin L, Woyke T, Lapidus A, Klenk HP, et al. The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. *PLoS ONE* 2012; **7**:e48837. [PubMed](#)
<http://dx.doi.org/10.1371/journal.pone.0048837>
 28. List of growth media used at DSMZ. <http://www.dsmz.de/catalogues/catalogue->

- microorganisms/culture-technology/list-of-media-for-microorganisms.html.
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. *Biopreserv Biobank* 2011; **9**:51-55.
<http://dx.doi.org/10.1089/bio.2010.0029>
 30. DOE Joint Genome Institute.
<http://www.jgi.doe.gov>
 31. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008; **18**:821-829. [PubMed](#)
<http://dx.doi.org/10.1101/gr.074492.107>
 32. wgsim. <https://github.com/lh3/wgsim>
 33. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA* 2011; **108**:1513-1518. [PubMed](#)
<http://dx.doi.org/10.1073/pnas.1017351108>
 34. 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](#) <http://dx.doi.org/10.1186/1471-2105-11-119>
 35. Mavromatis K, Ivanova NN, Chen IM, Szeto E, Markowitz VM, Kyrpides NC. The DOE-JGI Standard operating procedure for the annotations of microbial genomes. *Stand Genomic Sci* 2009; **1**:63-67. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.632>
 36. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [PubMed](#)
<http://dx.doi.org/10.1093/bioinformatics/btp393>