

Genome sequence of the homoacetogenic bacterium *Holophaga foetida* type strain (TMBS4^T)

Iain Anderson¹, Brittany Held^{1,2}, Alla Lapidus¹, Matt Nolan¹, Susan Lucas¹, Hope Tice¹, Tijana Glavina Del Rio¹, Jan-Fang Cheng¹, Cliff Han^{1,2}, Roxanne Tapia^{1,2}, Lynne Goodwin^{1,2}, Sam Pitluck¹, Konstantinos Liolios¹, Konstantinos Mavromatis¹, Ioanna Pagani¹, Natalia Ivanova¹, Natalia Mikhailova¹, Amrita Pati¹, Amy Chen³, Krishna Palaniappan³, Miriam Land^{1,4}, Evelyne-Marie Brambilla⁶, Manfred Rohde⁵, Stefan Spring⁶, Markus Göker⁶, John C. Detter^{1,2}, Tanja Woyke¹, James Bristow¹, Jonathan A. Eisen^{1,7}, Victor Markowitz³, Philip Hugenholtz^{1,8}, Hans-Peter Klenk^{6*}, Nikos C. Kyrpides^{1*}

¹ DOE Joint Genome Institute, Walnut Creek, California, USA

² Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

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

⁴ Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

⁵ HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany

⁶ Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures, 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 authors: Nikos C. Kyrpides and Hans-Peter Klenk

Keywords

anaerobic, motile, Gram-negative, mesophilic, chemoorganotrophic, sulfide-methylation, fresh water mud, *Acidobacteria*, *Holophagaceae*, GEBA

Abstract

Holophaga foetida Liesack *et al.* 1994 is a member to the genomically so far poorly characterized family *Holophagaceae* in the class *Holophagae* within the phylum *Acidobacteria*. *H. foetida* is of interest for its ability to anaerobically degrade aromatic compounds and for its production of volatile sulfur compounds through a unique pathway. The genome of *H. foetida* strain TMBS4^T is the first sequenced genome of a member of the class *Holophagae*. Here we describe the features of this organism, together with the complete genome sequence (improved high quality draft), and annotation. The 4,127,237 bp long chromosome with its 3,615 protein-coding and 57 RNA genes is a part of the *Genomic Encyclopedia of Bacteria and Archaea* project.

Introduction

Strain TMBS4^T (= DSM 6591) is the type strain of the species *Holophaga foetida* [1], which is the type species of the monotypic genus *Holophaga* [1,2]. The genus *Holophaga* is the type genus of the family *Holophagaceae* [3] in the order *Holophagales* [3] within the class *Holophagae* [3]. The genus name was derived from a combination of the Neo-Greek term *holos*, whole, and the Greek term *phagein*, to eat, meaning eating it all [1]; the species epithet was derived from the Latin adjective *foetidus*, smelling, stinking, referring to the production of ill-smelling methanethiol and dimethylsulfide [1]. Strain TMBS4^T was originally isolated from a black anoxic freshwater mud sample from a ditch near Konstanz, Germany [4]. It was found to transfer methyl groups from methoxylated aromatic compounds to sulfide, forming

methanethiol and dimethylsulfide [4]. Dimethylsulfide plays an important role in atmospheric chemistry, and is produced mainly by marine bacteria from dimethylsulfoniopropionate (reviewed in [5]). The production of dimethylsulfide from methoxylated aromatic compounds represents a unique pathway for production of this important compound. Strain TMBS4^T anaerobically degrades several aromatic compounds to acetate [1,4]. The only other cultured species within the order *Holophagales* is *Geothrix fermentans*, which is also an anaerobe but degrades small organic acids and fatty acids using Fe(III) as electron acceptor [6]. Here we present a summary classification and a set of features for *H. foetida* TMBS4^T, together with the description of the genomic sequencing and annotation.

Classification and features

A representative genomic 16S rRNA sequence of *H. foetida* TMBS4^T was compared using NCBI BLAST [7,8] 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 [9] and the relative frequencies of taxa and keywords (reduced to their stem [10]) were determined, weighted by BLAST scores. The most frequently occurring genera were *Holophaga* (52.9%), *Geothrix* (33.7%) and *Acidobacterium* (13.4%) (5 hits in total). Regarding the two hits to sequences from members of the species, the average identity within HSPs was 99.7%, whereas the average coverage by HSPs was 100.0%. Among all other species, the one yielding the highest score was *G. fermentans* (NR_036779), which corresponded to an identity of 91.6% and an HSP coverage of 97.8%. (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 DQ676369 ('Archaeal sediment and plankton freshwater pond suboxic freshwater-pond clone MVP-105'), which showed an identity of 97.6% and an HSP coverage of 94.9%. The most frequently occurring keywords within the labels of all environmental samples which yielded hits were 'lake' (6.2%), 'aquat' (4.6%), 'gatum, rank' (4.3%), 'soil' (3.4%) and 'microbi' (2.1%) (245 hits in total). The most frequently occurring keywords within the labels of those environmental samples which yielded hits of a higher score than the highest scoring species were 'situ' (3.3%), 'microbi' (3.0%), 'groundwat' (2.8%), 'activ' (2.5%) and 'aquif' (2.5%) (42 hits in total). Keywords with biological meaning fit to the environment from which strain TMBS4^T was isolated.

Figure 1 shows the phylogenetic neighborhood of *H. foetida* in a 16S rRNA based tree. The sequences of the two identical 16S rRNA gene copies in the genome differ by two nucleotides from the previously published 16S rRNA sequence (X77215), which contains one ambiguous base call.

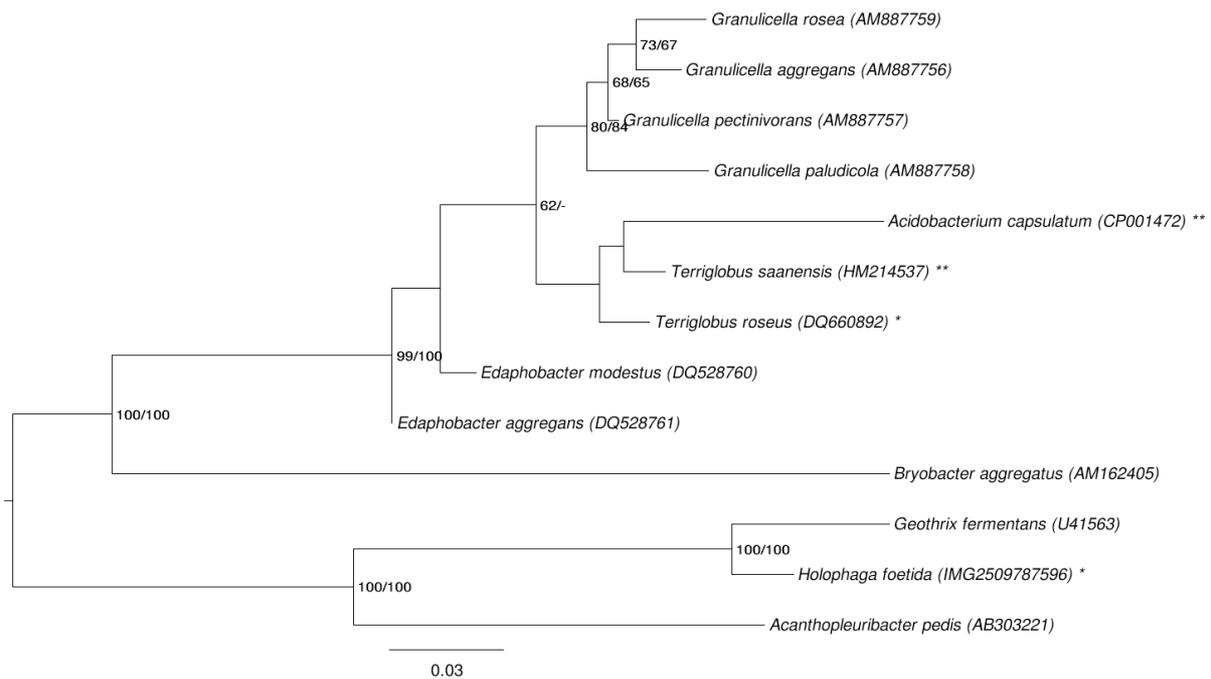


Figure 1. Phylogenetic tree highlighting the position of *H. foetida* relative to the type strains of the other species within the phylum 'Acidobacteria'. The tree was inferred from 1,395 aligned characters [11,12] of the 16S rRNA gene sequence under the maximum likelihood (ML) criterion [13]. Rooting was done initially using the midpoint method [14] and then checked for its agreement with the current classification (Table 1). The branches are scaled in terms of the expected number of substitutions per site. Numbers adjacent to the branches are support values from 400 ML bootstrap replicates [15] (left) and from 1,000 maximum-parsimony bootstrap replicates [16] (right) if larger than 60%. Lineages with type strain genome sequencing projects registered in GOLD [17] are labeled with one asterisk, those also listed as 'Complete and Published' with two asterisks [18] (see CP002467 for *Terriglobus saanensis*).

H. foetida TMBS4^T is Gram-negative, and an electron micrograph shows an inner and outer membrane [1]. Cells are rod-shaped, 1-3 μm long and 0.5-0.7 μm wide [1,4] (Figure 2). No motility was observed [1,4], although the genome is rich in genes classified under 'cell motility' (152 genes, see below in Tab 4). Growth was observed between 10°C and 35°C with an optimum at 28-32°C [1,4]. The pH range for growth was 5.5-8.0 with 6.8-7.5 as the optimum [1,4]. The salinity range for growth was 1-15 g/l NaCl [4]. Aromatic compounds utilized by TMBS4^T include 3,4,5-trimethoxybenzoate, syringate, 5-hydroxyvanillate, phloroglucinol monomethyl ether, sinapate, ferulate, caffeate, gallate, 2,4,6-trihydroxybenzoate, pyrogallol, and phloroglucinol [1,4]. The fastest growth occurred with syringate [4]. When sulfide was present in the medium, methyl groups from aromatic compounds were used to form methanethiol and dimethylsulfide [1,4]. Strain TMBS4^T could also grow with CO or CO₂ as methyl acceptors, and acetyl-CoA synthase activity was detected [19]. Growth was also observed on pyruvate [1,4].

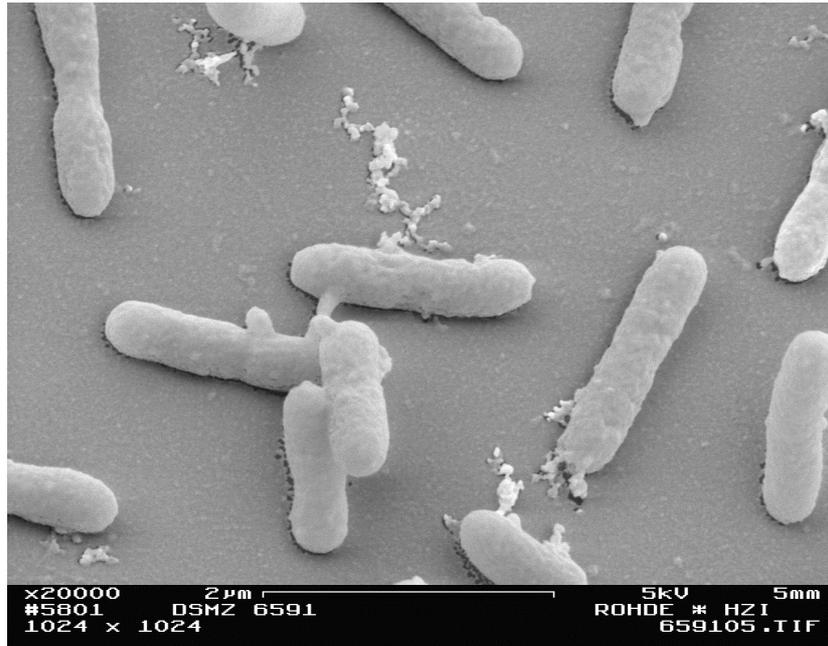


Figure 2. Scanning electron micrograph of *H. foetida* TMBS4^T

Chemotaxonomy

No data are available regarding chemotaxonomy of *H. foetida* TMBS4^T.

Table 1. Classification and general features of *H. foetida* TMBS4^T in accordance with the MIGS recommendations [20].

MIGS ID	Property	Term	Evidence code
	Current classification	Domain <i>Bacteria</i>	TAS [21]
		Phylum ' <i>Acidobacteria</i> '	TAS [22]
		Class <i>Holophagae</i>	TAS [3]
		Order <i>Holophagales</i>	TAS [3]
		Family <i>Holophagaceae</i>	TAS [3]
		Genus <i>Holophaga</i>	TAS [1,23]
		Species <i>Holophaga foetida</i>	TAS [1,23]
		Type-strain TMBS4	TAS [1]
	Gram stain	negative	TAS [1]
	Cell shape	rod-shaped	TAS [1,4]
	Motility	non-motile	TAS [1,4]
	Sporulation	non-sporulating	TAS [4]
	Temperature range	mesophile, 10-35°C	TAS [1,4]
	Optimum temperature	28-32°C	TAS [1,4]
	Salinity	1-15 g/l NaCl	TAS [4]
MIGS-22	Oxygen requirement	obligate anaerobe	TAS [1,4]
	Carbon source	methoxylated and hydroxylated aromatic compounds, pyruvate	TAS [1]
	Energy metabolism	chemoorganotroph	TAS [1]
MIGS-6	Habitat	freshwater mud	TAS [1]
MIGS-15	Biotic relationship	free living	TAS [1]
MIGS-14	Pathogenicity	none	NAS
	Biosafety level	1	TAS [24]
MIGS-23.1	Isolation	freshwater mud	TAS [1]
MIGS-4	Geographic location	near Konstanz, Germany	TAS [1]
MIGS-5	Sample collection time	not reported	

MIGS-4.1	Latitude – Longitude	47.663 – 9.175	TAS [1]
MIGS-4.2			
MIGS-4.3	Depth	not reported	
MIGS-4.4	Altitude	not reported	

Evidence 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 of the Gene Ontology project [25].

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its phylogenetic position [26], and is part of the *Genomic Encyclopedia of Bacteria and Archaea* project [27]. The genome project is deposited in the Genomes On Line Database [17] 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 2. Genome sequencing project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	Improved high quality draft
MIGS-28	Libraries used	Four genomic libraries: two 454 pyrosequence standard libraries, one 454 PE library (11.5 kb insert sizes), one Illumina library
MIGS-29	Sequencing platforms	Illumina GAii, 454 GS FLX Titanium
MIGS-31.2	Sequencing coverage	2,172.4 × Illumina; 20.0 × pyrosequence
MIGS-30	Assemblers	Newbler version 2.3, Velvet version 1.0.13, phrap version SPS - 4.24
MIGS-32	Gene calling method	Prodigal
	INSDC ID	AGSB000000000
	GenBank Date of Release	January 12, 2012
	GOLD ID	Gi05348
	NCBI project ID	53485
	Database: IMG	2509601028
MIGS-13	Source material identifier	DSM 6591
	Project relevance	Bioremediation and phylogenetic diversity

Growth conditions and DNA isolation

H. foetida strain TMBS4^T, DSM 6591, was grown anaerobically in DSMZ medium 559 (TMBS4 medium) [28] at 30°C. DNA was isolated from 0.5-1 g of cell paste using MasterPure Gram-positive DNA purification kit (Epicentre MGP04100) following the standard protocol as recommended by the manufacturer with modification st/DL for cell lysis as described in Wu *et al.* 2009 [27]. DNA is available through the DNA Bank Network [29].

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

[30]. Pyrosequencing reads were assembled using the Newbler assembler (Roche). The initial Newbler assembly consisting of 186 contigs in two scaffolds was converted into a phrap [31] assembly by making fake reads from the consensus, to collect the read pairs in the 454 paired end library. Illumina GAii sequencing data (9,124.2 Mb) was assembled with Velvet [32] 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 135.9 Mb of 454 draft data and all of the 454 paired end data. Newbler parameters are -consed -a 50 -l 350 -g -m -ml 20. The Phred/Phrap/Consed software package [31] 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 (C. Han, unpublished), Dupfinisher [33], or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR primer walks (J.-F. Chang, unpublished). A total of 192 additional reactions were necessary to close some gaps and to raise the quality of the assembled sequence. Illumina reads were also used to correct potential base errors and increase consensus quality using a software Polisher developed at JGI [34]. The error rate of the assembled genome sequence is less than 1 in 100,000. Together, the combination of the Illumina and 454 sequencing platforms provided $2,192.4 \times$ coverage of the genome. The final assembly contained 461,984 pyrosequence and 120,055,671 Illumina reads and consists of 39 contigs organised in three scaffolds.

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) non-redundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to assert a product description for each predicted protein. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE [37], RNAMMer [38], Rfam [39], TMHMM [40], and signalP [41].

Genome properties

The genome in its current assembly consists of three scaffolds of 3,443,192 bp, 677,300 bp and 6,745 bp length with a 63.0% G+C content (Table 3 and Figure 3). Of the 3,672 genes predicted, 3,615 were protein-coding genes, and 57 RNAs; 76 pseudogenes were also identified. The majority of the protein-coding genes (74.3%) 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.

Table 3. Genome Statistics

Attribute	Value	% of Total
Genome size (bp)	4,127,237	100.00%
DNA coding region (bp)	3,689,184	89.39%
DNA G+C content (bp)	2,595,845	62.95%
Number of scaffolds	3	
Total genes	3,672	100.00%
RNA genes	57	1.55%
rRNA operons	2	

tRNA genes	47	1.28%
Protein-coding genes	3,615	98.45%
Pseudo genes	76	2.07%
Genes with function prediction (proteins)	2,728	74.29%
Genes in paralog clusters	1,916	52.18%
Genes assigned to COGs	2,788	75.93%
Genes assigned Pfam domains	2,729	74.32%
Genes with signal peptides	746	20.32%
Genes with transmembrane helices	703	19.14%
CRISPR repeats	1	

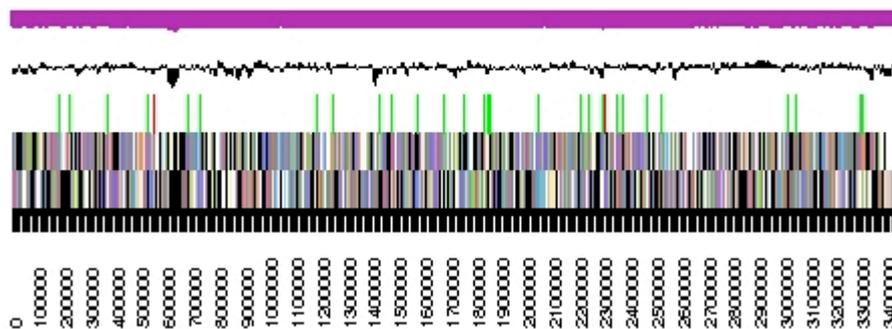


Figure 3. Graphical map of the largest scaffold. From bottom to the top: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, rRNAs red, other RNAs black), GC content, GC skew.

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

Code	value	%age	Description
J	159	5.1	Translation, ribosomal structure and biogenesis
A	2	0.1	RNA processing and modification
K	254	8.1	Transcription
L	191	6.1	Replication, recombination and repair
B	0	0.0	Chromatin structure and dynamics
D	35	1.1	Cell cycle control, cell division, chromosome partitioning
Y	0	0.0	Nuclear structure
V	54	1.7	Defense mechanisms
T	267	8.5	Signal transduction mechanisms
M	194	6.2	Cell wall/membrane biogenesis
N	152	4.8	Cell motility
Z	1	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	112	3.6	Intracellular trafficking and secretion, and vesicular transport
O	94	3.0	Posttranslational modification, protein turnover, chaperones
C	245	7.8	Energy production and conversion
G	120	3.8	Carbohydrate transport and metabolism

E	201	6.4	Amino acid transport and metabolism
F	69	2.2	Nucleotide transport and metabolism
H	183	5.8	Coenzyme transport and metabolism
I	87	2.8	Lipid transport and metabolism
P	135	4.3	Inorganic ion transport and metabolism
Q	32	1.0	Secondary metabolites biosynthesis, transport and catabolism
R	361	11.5	General function prediction only
S	201	6.4	Function unknown
-	884	24.1	Not in COGs

Insights into the genome sequence

H. foetida is known to utilize aromatic compounds through the phloroglucinol pathway, producing three molecules of acetate from the benzene ring. It is also capable of growing on methoxylated aromatic compounds, transferring methyl groups to sulfide or CO₂ to produce dimethylsulfide or acetyl-CoA. Genes have not been identified for the enzymes of the phloroglucinol pathway with one exception: the transhydroxylase that converts pyrogallol to phloroglucinol, which has been identified in *Pelobacter acidigallici* [42]. This enzyme has two subunits, and genes with high similarity to these subunits are found in *H. foetida*. HolfoDRAFT_0037 and HolfoDRAFT_0041 have 74% and 88% similarity to the large subunit, while HolfoDRAFT_0036, HolfoDRAFT_0040, and HolfoDRAFT_0058 have 63%, 73%, and 65% similarity to the small subunit.

H. foetida likely gains energy from the conversion of acetyl-CoA produced from aromatic compounds, pyruvate, and methyl groups from methoxylated aromatic compounds to acetate. Within the genome there are two phosphotransacetylase genes (HolfoDRAFT_0402, HolfoDRAFT_1130) and two acetate kinase genes (HolfoDRAFT_1418, HolfoDRAFT_3547). Several candidates for pyruvate:ferredoxin oxidoreductase were found in the genome. This enzyme would produce acetyl-CoA that can be used to produce ATP and acetate. *H. foetida* can combine methyl groups with CO or CO₂ to form acetyl-CoA, and acetyl-CoA synthase activity was detected [19]. Within the genome there are genes for a *Moorella* type CO dehydrogenase/acetyl-CoA synthase (HolfoDRAFT_1152 and HolfoDRAFT_1153) and the two subunits of the corrinoid Fe-S protein (HolfoDRAFT_1154 and HolfoDRAFT_1157).

H. foetida has potential symporters and ABC transporters for aromatic compounds. Four genes (HolfoDRAFT_0048, HolfoDRAFT_0224, HolfoDRAFT_0791, HolfoDRAFT_0858) belonging to the major facilitator superfamily have strong similarity to aromatic compound transporters of family 2.A.1.15. *H. foetida* has few ABC transporters for organic compounds, but it has 3 full transporters and 8 additional substrate binding proteins from family 4. Some members of this family are amino acid transporters, but one member has been found to transport protococatechuate [43].

Systems for demethylation of methoxyaromatic compounds have been identified in *Acetobacterium dehalogenans* [44] and *Moorella thermoacetica* [45]. Methyl groups are transferred first to a corrinoid protein, then to tetrahydrofolate, by two methyltransferases. The genes for two sets of enzymes from *A. dehalogenans* have been sequenced [46]. The corrinoid proteins belong to COG5012, the first methyltransferases belong to COG0407, and the second methyltransferases belong to COG1410. *H. foetida* likely uses the same type of process. It has six proteins belonging to COG5012 and 29 proteins belonging to COG0407. The only genome with more members of COG0407 is *Mahella australiensis* with 33. Some of the COG0407 proteins are found close to corrinoid proteins in the genome sequence. *H.*

foetida has two members of COG1410. One is adjacent to the CO dehydrogenase/acetyl-CoA synthase genes and has 61% identity to the *acsE* gene of *M. thermoacetica*. It probably transfers methyl groups from tetrahydrofolate to the corrinoid iron-sulfur protein of CO dehydrogenase/acetyl-CoA synthase. The other COG1410 gene is adjacent to a corrinoid protein and may transfer methyl groups from corrinoid proteins to tetrahydrofolate in the methoxyaromatic demethylation pathway.

Acknowledgements

We would like to gratefully acknowledge the help of Maren Schröder (DSMZ) for growing *H. foetida* cultures. This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396, UT-Battelle and Oak Ridge National Laboratory under contract DE-AC05-00OR22725, as well as German Research Foundation (DFG) INST 599/1-2.

References

1. Liesack W, Bak F, Kreft JU, Stackebrandt E. *Holophaga foetida* gen. nov., sp. nov., a new, homoacetogenic bacterium degrading methoxylated aromatic compounds. *Arch Microbiol* 1994; **162**:85-90.
2. Garrity G. NamesforLife. BrowserTool takes expertise out of the database and puts it right in the browser. *Microbiol Today* 2010; **37**:9.
3. Fukunaga Y, Kurahashi M, Yanagi K, Yokota A, Harayama S. *Acanthopleuribacter pedis* gen. nov., sp. nov., a marine bacterium isolated from a chiton, and description of *Acanthopleuribacteraceae* fam. nov., *Acanthopleuribacterales* ord. nov., *Holophagaceae* fam. nov., *Holophagales* ord. nov. and *Holophagae* classis nov. in the phylum 'Acidobacteria'. *Int J Syst Evol Microbiol* 2008; **58**:2597-2601.
4. Bak F, Finster K, Rothfuß F. Formation of dimethylsulfide and methanethiol from methoxylated aromatic compounds and inorganic sulfide by newly isolated anaerobic bacteria. *Arch Microbiol* 1992; **157**:529-534.
5. Reisch CR, Moran MA, Whitman WB. Bacterial catabolism of dimethylsulfoniopropionate (DMSP). *Front Microbiol* 2011; **2**:172.
6. Coates JD, Ellis DJ, Gaw CV, Lovley DR. *Geothrix fermentans* gen. nov., sp. nov., a novel Fe(III)-reducing bacterium from a hydrocarbon-contaminated aquifer. *Int J Syst Bacteriol* 1999; **49**:1615-1622.
7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**:403-410.
8. Korf I, Yandell M, Bedell J. BLAST, O'Reilly, Sebastopol, 2003.
9. 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 Env Microbiol* 2006; **72**:5069-5072.
10. Porter MF. An algorithm for suffix stripping. Program: *electronic library and information systems* 1980; **14**:130-137.

11. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics* 2002; **18**:452-464.
12. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000; **17**:540-552.
13. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* 2008; **57**:758-771.
14. Hess PN, De Moraes Russo CA. An empirical test of the midpoint rooting method. *Biol J Linn Soc* 2007; **92**:669-674.
15. 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.
16. Swofford DL. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0 b10. Sinauer Associates, Sunderland, 2002.
17. Pagani I, Liolios K, Jansson J, Chen IM, 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.
18. Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait H, Badger J. Three genomes from the phylum *Acidobacteria* provide insight into the lifestyle of these microorganisms in soil. *Appl Environ Microbiol* 2009; **75**:2046-2056.
19. Kreft JU, Schink B. Demethylation and degradation of phenylmethylethers by the sulfide-methylating homoacetogenic bacterium strain TMBS 4. *Arch Microbiol* 1993; **159**:308-315.
20. 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.
21. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms. Proposal for the domains *Archaea* and *Bacteria*. *Proc Nat Acad Sci USA* 1990; **87**:4576-4579.
22. 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.
23. List Editor. Validation List no. 52. Validation of the publication of new names and new combinations previously effectively published outside the IJSB. *Int J Syst Bacteriol* 1995; **45**:197-198.
24. BAuA 2010, Classification of bacteria and archaea in risk groups. <http://www.baua.de> TRBA 466, p. 105.
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. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29.
26. Klenk H, Göker M. En route to a genome-based classification of *Archaea* and *Bacteria*? *Syst Appl Microbiol* 2010; **33**:175-182
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.

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. [doi:10.1089/bio.2010.0029]
30. The DOE Joint Genome Institute. www.jgi.doe.gov
31. Phrap and Phred for Windows, MacOS, Linux, and Unix. 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.
33. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. *In: Proceeding of the 2006 international conference on bioinformatics & computational biology*. Arabnia HR, Valafar H (eds), 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 Dynamic Programming Genefinding Algorithm. *BMC Bioinformatics* 2010; **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.
37. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964.
38. Lagesen K, Hallin PF, Rødland E, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent annotation of rRNA genes in genomic sequences. *Nucleic Acids Res* 2007; **35**:3100-3108.
39. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res* 2003; **31**:439-441.
40. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 2001; **305**:567-580.
41. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**:783-795.
42. Baas D, Rétey J. Cloning, sequencing and heterologous expression of pyrogallol-phloroglucinol transhydroxylase from *Pelobacter acidigallici*. *Eur J Biochem* 1999; **265**:896-901.
43. Maclean AM, Haerty W, Golding GB, Finan TM. The LysR-type PcaQ protein regulates expression of a protocatechuate-inducible ABC-type transport system in *Sinorhizobium meliloti*. *Microbiology* 2011; **157**:2522-2533.
44. Kaufmann F, Wohlfarth G, Diekert G. O-demethylase from *Acetobacterium dehalogenans*—substrate specificity and function of the participating proteins. *Eur J Biochem* 1998; **253**:706-711.
45. Naidu D, Ragsdale SW. Characterization of a three-component vanillate O-demethylase from *Moorella thermoacetica*. *J Bacteriol* 2001; **183**:3276-3281.

46. Schilhabel A, Studenik S, Vödisch M, Kreher S, Schlott B, Pierik AJ, Diekert G. The ether-cleaving methyltransferase system of the strict anaerobe *Acetobacterium dehalogenans*: analysis and expression of the encoding genes. *J Bacteriol* 2009; **191**:588-599.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.