# Multiple Syntrophic Interactions in a Terephthalate-Degrading Methanogenic Consortium

*Athanasios Lykidis [1], Chia-Lung Chen [2], Susannah G. Tringe [1], Alice C. McHardy [3], Alex Copeland [1], Nikos C. Kyrpides [1], Philip Hugenholtz [1] and Wen-Tso Liu [2,4]*.*

[1] Joint Genome Institute, Lawrence National Berkeley Laboratory, Walnut Creek, CA 94598, USA
[2] Division of Environmental Science and Engineering, National University of Singapore, Singapore 117576
[3] Computational Genomics & Epidemiology, Max-Planck Institute for Informatics, 66123 Saarbrücken, Germany
[4] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, IL61801, USA

*Corresponding author: Wen-Tso Liu, University of Illinois at Urbana-Champaign

**August 5, 2010**

## ACKNOWLEDGMENT

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

# DISCLAIMER

# Multiple Syntrophic Interactions in a Terephthalate-Degrading Methanogenic Consortium

*Athanasios Lykidis[1], Chia-Lung Chen[2], Susannah G. Tringe[1], Alice C. McHardy[3], Alex Copeland[1], Nikos C. Kyrpides[1], Philip Hugenholtz[1] and Wen-Tso Liu[2,4*].*

[1]Joint Genome Institute, Lawrence National Berkeley Laboratory, Walnut Creek, CA 94598, USA

[2]Division of Environmental Science and Engineering, National University of Singapore, Singapore 117576

[3]Computational Genomics & Epidemiology, Max-Planck Institute for Informatics, 66123 Saarbrücken, Germany

[4]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, IL61801, USA

*Corresponding author: Wen-Tso Liu

Department of Civil and Environmental Engineering

University of Illinois at Urbana-Champaign

3207 Newmark Civil Engineering Laboratory, MC-250

205 North Mathews Avenue

Urbana, Illinois 61801, USA

Phone: (217) 333-8442, Fax: (217) 333-6968, Email: wtliu@illinois.edu

Running title: Metagenomics of a methanogenic consortium

**Abstract**

Terephthalate (TA) is one of the top 50 chemicals produced worldwide. Its production results in a TA-containing wastewater that is treated by anaerobic processes through a poorly understood methanogenic syntrophy. Using metagenomics, we characterized the methanogenic consortium inside a hyper-mesophilic (i.e., between mesophilic and thermophilic), TA-degrading bioreactor. We identified genes belonging to dominant *Pelotomaculum* species presumably involved in TA degradation through decarboxylation, dearomatization, and modified β-oxidation to $H_2/CO_2$ and acetate. These intermediates are converted to $CH_4/CO_2$ by three novel hyper-mesophilic methanogens. Additional secondary syntrophic interactions were predicted in *Thermotogae*, *Syntrophus* and candidate phyla OP5 and WWE1 populations. The OP5 encodes genes capable of anaerobic autotrophic butyrate production and *Thermotogae*, *Syntrophus* and WWE1 have the genetic potential to oxidize butyrate to $CO_2/H_2$ and acetate. These observations suggest that the TA-degrading consortium consists of additional syntrophic interactions beyond the standard $H_2$-producing syntroph – methanogen partnership that may serve to improve community stability.

## Introduction

Terephthalate (TA) is used as the raw material for the manufacture of numerous plastic products (e.g., polyethylene terephthalate bottles and textile fibers). During its production, TA-containing wastewater is discharged in large volumes (as high as 300 million $m^3$ per year) and high concentration (up to 20 kg COD $m^{-3}$) (Razo-Flores et al., 2006). This wastewater is generally treated by anaerobic biological processes under mesophilic conditions (~35$^o$C). However, anaerobic processes operated at hyper-mesophilic (46-50$^o$C) and thermophilic (~55$^o$C) temperatures may be preferable due to the ability to achieve higher loading rate (van Lier et al., 1997; Chen et al., 2004), which reduces the reactor volume. Moreover, TA wastewater is usually generated at 54–60 ºC, and does not require additional energy input for maintaining reactor temperature (Chen et al., 2004). The microbial biomass usually occurs in the form of granules or biofilms attaching on the surface of porous media. Under such environments, TA degradation has been hypothesized (Kleerebezem et al., 1999) to be based on a syntrophic microbial relationship whereby fermentative $H_2$-producing bacteria (syntrophs) convert TA via benzoate to acetate and $H_2/CO_2$, and acetoclastic and hydrogenotrophic methanogens further convert the intermediates to methane by physically positioning themselves close to the syntrophs to overcome the thermodynamic barrier (Stams, 1994; Conrad, 1999; Dolfing, 2001).

In practice the complexities of TA-degrading communities are not as well known. The communities require a long maturation phase (>200-300 days), are difficult to maintain, and do not always result in a successful syntrophic interaction. If the syntrophic interaction is disturbed and the treatment rendered ineffective, the resulting high-concentration effluent must be treated with a more energy-intensive down-stream aerobic biological process. These factors can

significantly increase the operational cost of the process and limit its application on a wider scale.

Several studies have investigated the microbial populations present in methanogenic TA degradation, often using laboratory-scale reactors operated at various temperatures (Kleerebezem et al., 1999; Wu et al., 2001; Chen et al., 2004). Using rRNA-based molecular methods, these studies have found that TA-degrading consortia in bioreactors are dominated by two to three bacterial populations and two types of methanogens (Wu et al., 2001; Chen et al., 2004). The methanogens are relatively straightforward to classify being mainly acetoclastic *Methanosaeta*-related species and a novel hydrogenotrophic methanogenic species in the family *Methanomicrobiales*. The syntrophic bacteria, however, are difficult to identify based on phylogenetic classification, and extremely difficult to obtain in pure culture. In the last decade, only three bacterial species that can degrade TA and its isomers have been successfully co-cultured with methanogens under mesophilic conditions (Qiu et al., 2006; Qiu et al., 2008), and these isolates are different from those found under thermophilic conditions (Chen et al., 2004). This greatly limits the effort to understand the microbial interaction and function in the TA-degrading consortia.

A metagenome analysis was chosen for this study since it has been proven as an effective method for retrieving nearly complete microbial genomes of dominant populations in relatively simple microbial ecosystems (Tyson et al., 2004; Martin et al., 2006). In particular, this study aims to elucidate the microbiology underpinning anaerobic TA-degrading processes, including improved knowledge of the diversity and physiology of participating syntrophs and methanogens, and the mechanism behind the establishment and maintenance of the partnership. This knowledge may lead to the generation of principles which could be applied to establish

different consortia for treating other chemicals discharged from industrial production lines, or to treat contaminants in other environments.

## Materials and Methods

The anaerobic microbial consortium that degrades terephthalate was selectively enriched using a 1-liter laboratory-scale hybrid bioreactor (Fig. 1a) as described previously (Angelidaki et al., 1990; Chen et al., 2004) (see supplementary text). Biomass was collected from only porous packing filters on Days 221 and 280, and from both filters and sludge bed on Days 346, and 430 for further analyses. These biomass samples were used for genomic DNA extraction, library construction, and sequencing according to standard protocols (http://www.jgi.doe.gov/sequencing/protocols/prots_production.html) (see supplementary text). Detailed metagenome analysis methods are described in the supplementary text. Data can be accessed through the IMG/M (http://img.jgi.doe.gov/m/) system.

## Results and Discussion

*Bioreactor operation and performance.*

An anaerobic hybrid reactor was successfully operated with TA as the only carbon and energy source for 480 days. This reactor was constructed with an upper section filled with ring-shape porous filters to support the growth of microbial biofilms, and a lower section for the development of anaerobic granular sludge (Fig. 1A). This reactor is unusual and novel in that it was the first methanogenic reactor operated in the hyper-mesophilic temperature zone (46-50$^{o}$C), whereas previously published studies of TA-degrading communities were at mesophilic (~35$^{o}$C) or thermophilic (~55$^{o}$C) temperatures (Kleerebezem et al., 1999; Wu et al., 2001; Chen et al.,

2004). After achieving a good TA removal efficiency (Fig. 1B), sludge samples were taken from the surface of the filter media at Days 221, 280, 346 and 430. Samples were also taken from the sludge bed at Days 346 and 430. These samples taken were used for 16S rRNA and metagenome analysis.

*16S rRNA-based community profiling.*

The rarefaction curve analysis indicates that the bacterial population diversity is much higher in terms of OTU number than the archaeal population diversity in any given sample taken from the TA reactor (Supplementary Fig. 1). The phylogenetic distribution of bacterial 16S rRNA clones (Fig. 2) indicates that *Peptococcaceaea* (mostly *Pelotomaculum*), *Thermotogae*, *Syntrophaceae*, and candidate phyla OP5 and WWE1 were the dominant bacterial lineages present. Between the biofilm samples (rings 1, 2, 3 and 5) and sludge bed samples (rings 4 and 6) taken, differences in the abundances of major phyla including *Firmicutes*, *Thermotogae*, *Proteobacteria*, and OP5 were observed. These differences are likely attributed to the differences in growth temperature ($46^{o}$C vs. $50^{o}$C) and form (biofilms vs. granules). Archaeal representatives were less diverse and consisted of two major types of methanogens belonging to the orders *Methanomicrobiales* and *Methanosarcinales* (Supplementary Fig. 2). Differences in growth temperature may further explain the variations observed in the microbial populations enriched in previous studies (Supplementary Fig. 3). Using the 16S rRNA gene and McrA gene as biomarkers, temperature-dependent variations were also observed with acetoclastic methanogen populations found in the order *Methanosarcinales* (Supplementary Fig. 4). The hydrogenotrophic methanogens identified here are closely related to methanogens found in mesophilic and thermophilic TA-degrading reactors (Wu et al., 2001; Chen et al., 2004), and together with *Methanolinea tarda* NOBI-1 recently isolated from anaerobic digestion processes (Imachi et al., 2008), form a novel cluster

separate from other known hydrogenotrophic methanogens. The comparison of McrA genes further suggests that the methanogens found in the new cluster are likely different from *M. tarda* NOBI-1 (Supplementary Fig. 4B).

*Shotgun sequencing.*

The assembled sequence data contained 37,818 singlets and 14,526 contiguous fragments of intermediate length (the largest fragment was approximately 240 kb (Supplementary Fig. 5) and 45 fragments between 24-167 kb). Gene prediction on the entire dataset using Genemark resulted in the prediction of 93,104 protein coding genes. A composition-based classifier, PhyloPythia (McHardy et al., 2007), was used to assign those contigs and singlets into major phylogenetic groups (Table 1), including *Pelotomaculum* species, candidate phylum OP5 species, *Methanolinea* species, and *Methanosaeta* species. The highest coverage of an isolate reference genome was observed for *M. thermophila* (~80%) followed by *Pelotomaculum thermopropionicum* (~60%) (Supplementary Fig. 6). However, the *Methanolinea* population appears to be the best covered one as the average read depth is 5.3X with many contigs having 10X read depth (Supplementary Fig. 5). In the case of OP5, where a closely related microbial genome was not present in the database, the occurrence was calculated with the phylogenetic marker COGs in the OP5 bin (1.41Mb; G+C content, 28%). Approximately 50% of the OP5 genome is estimated to be covered by the metagenomic data (Table 1).

*Pelotomaculum.*

As a known catabolic degrading organism abundant in the reactor, *Pelotomaculum* is assumed to be largely responsible for catabolic degradation of TA to $CO_2$, $H_2$, and acetate. With an average read depth of 3.2X, 1083 contigs were assigned to the *Pelotomaculum* population, comprising 4.3 Mb in total (Table 1). We first searched for genes with known decarboxylase functions that

are responsible for the first decarboxylation step of TA degradation (Supplementary Fig. 7). Two gene sets (tadcc27178-79-80 and tadcc16349) from the *Pelotomaculum* bin were identified to have high sequence similarity and a subunit complement with a known 4-hydroxybenzoate decarboxylase, EC 4.1.1.61, from *Sedimentibacter hydroxybenzoicum* that consists of three subunits (AAD50377, AAY67850, and AAY67851)and belongs to the UbiD family of proteins (Lupa et al., 2005). Two mechanisms have been described for the subsequent fermentation of benzoate to acetate and $CO_2$: the well known benzoyl-CoA reductase (BCR, EC 1.3.99.15) route used by *Thauera aromatica* (Boll and Fuchs, 1995) and the less understood BCR independent mechanism for reductive dearomatization (Wischgoll et al., 2005). Examining the TA decarboxylase dataset did not show any clear homologs of the *Thauera* benzoyl-CoA reductases. Instead, homologs are found in the alternative BCR independent mechanism within a set of 44 genes that have been postulated to operate in *Geobacter metallireducens* and "Syntrophus aciditrophicus" (Butler et al., 2007; McInerney et al., 2007; Peters et al., 2007). The key BCR enzyme in *G. metallireducens* was successfully characterized *in-vitro* (Kung et al., 2009). The metagenome analysis further predicted the pathways that are used for conversion of benzoate to hydroxypimelyl-CoA and subsequent conversion of 3-hydroxypimelyl-CoA via β-oxidation to acetyl-CoA, which in turn gives rise to acetate via substrate-level phosphorylation (Supplementary Fig. 7; Fig. 3). The *Pelotomaculum* bin also contains genes and pathways for the production of butyrate (Supplementary Fig. 7). These observations indicate that TA fermentation by *Pelotomaculum* may lead to the formation of butyrate in addition to acetate. Two genes assigned to *Pelotomaculum* (tadcc25255 and tadcc12813) belong to the Fe-only hydrogenase protein family and are potentially involved in hydrogen generation.

*Methanogens.*

Three major groups or bins of methanogens belonging to the genera *Methanosaeta* and *Methanolinea* were identified and are known to be syntrophic partners of *Pelotomaculum* (Table 1). Complete pathways for both acetoclastic and hydrogenotrophic methanogenesis were identified (Supplementary Table 1). The first step in acetoclastic methanogenesis is the formation of acetyl-CoA from acetate. It has been proposed (Smith and Ingram-Smith, 2007) that acetoclastic methanogenesis in *Methanosaeta* proceeds with a modified version of the pathway compared to *Methanosarcina*, which utilize the acetate kinase/phosphotransacetylase pathway to convert acetate to acetyl-CoA. In contrast, the *M. thermophila* genome does not include a readily identifiable acetate kinase and it has been proposed that this species utilizes an acetate transporter coupled with acetyl-CoA synthetases to convert acetate to acetyl-CoA (Smith and Ingram-Smith, 2007). Analysis of the TA dataset indicates the presence of acetate transporters (tadcc8417) and acetyl-CoA synthetases, E.C. 6.2.1.1, (tadcc27524, tadcc27520, tadcc27522, tadcc21744, tadcc21743) in the *Methanosaeta* bin. A complete set of the five acetyl-CoA decarbonylase subunits (E.C. 1.2.99.2) was identified as well as genes for the remaining steps of methanogenesis (Supplementary Table 1).

Hydrogenotrophic methanogenesis in the TA community is performed by the *Methanolinea* group. A complete set of Eha hydrogenase enzyme subunits is found in the *Methanolinea* bin. This set is adjacent to formylmethanofuran dehydrogenase which reduces $CO_2$ to formyl-methanofuran (tadcc39592-39604), suggesting that it may be the enzyme reducing the ferredoxin used by the dehydrogenase (Anderson et al., 2008). In addition, complete sets of *ech* (tadcc3040-tadcc3045) and *mbh* (tadcc17854-tadcc17865) hydrogenases can be found in the *Methanolinea* bin. These hydrogenases are proposed to provide $H_2$ for the reduction of heterodisulfide (CoM-S-S-CoB) by the heterodisulfide reductase in the absence of

MvhADG hydrogenase (Anderson et al., 2008; Thauer et al., 2008). In this way they link the regeneration of CoM to the reduction of ferredoxin. No homologs to the MvhADG hydrogenase were identified in the *Methanolinea* bin, suggesting that this organism couples ferredoxin and CoB-S-S-CoM reduction to hydrogen (Fig. 4).

*OP5.*

Analysis of the gene content in the OP5 bin (Table 1) revealed the existence of a gene fragment (tadcc9232) related to the *Archaeoglobus* type III RuBisCO. This fragment contains 181 amino acids and exhibits 60% identity to the N-terminus of the large subunit of the *Archaeoglobus* ribulose 1,5-bisphosphate carboxylase raising, EC 4.1.1.39, a link between OP5 and autotrophic $CO_2$ fixation via the Calvin–Benson–Bassham cycle. Previous work has established that type III RuBisCOs are functional enzymes *in vitro* and also complement RuBisCO deletion in photosynthetic organisms indicating their functionality *in vivo* (Tabita et al., 2007). However, other experiments have shown that type III RuBisCO enzymes are involved in AMP metabolism (Sato et al., 2007). Thus, future experiments are required to validate whether OP5 speices can use type III RuBisCO enzymes for autotrophic $CO_2$ fixation. Although organisms that contain type III RuBisCOs usually lack recognizable phosphoribulokinases (as is the case for *Archaea*), the OP5 bin contains a gene (tadcc30466) that belongs to the phosphoribulokinase protein family (Pfam domain 00485), which provides the second substrate for the RuBisCO reaction, ribulose 1,5-bisphosphate. These are the two unique enzymatic activities required for $CO_2$ assimilation. The OP5 bin also contains genes encoding phosphoglycerate kinase, E.C. 2.7.2.3, (tadcc33464), glyceraldehyde-3-phosphate dehydrogenase, E.C. 2.7.1.12, (tadcc33465), and phosphoglycerate mutase, E.C. 5.4.2.1, (tadcc16672) although no representatives of the remaining Calvin–Benson–Bassham cycle genes are readily recognizable in the OP5 bin.

OP5 also contains phosphate butyryltransferase, E.C. 2.3.1.19, (tadcc17546) and two copies of butyrate kinase, E.C. 2.7.2.7, (tadcc17547 and tadcc17544) indicating its ability to produce butyrate and gain energy by substrate level phosphorylation. No acetate kinases or ADP-forming acyl-CoA synthetases were detected in the OP5 genes binned by PhyloPythia. However, inspecting unassigned contigs with GC content lower than 31% identified an acetate kinase (E.C. 2.7.2.1) gene (tadcc15543 on contig taComm3_C5047) that may originate from OP5. Based on these observations, it is proposed that the OP5 populations within the TA community participate in the syntrophic interactions by removing $CO_2$ and $H_2$ and producing butyrate and potentially acetate. An operon on contig C11376 binned in OP5 was found to contain a system of Ni-hydrogenases (tadcc33916 and tadcc3391) potentially involved in hydrogen utilization.

*Syntrophaceae, Thermotogae, and WWE1.*

*Syntrophaceae* are members of syntrophic communities and are a minor component of the TA-degrading community (Fig 2). To our knowledge, no known *Syntrophaceae* isolates have been reported to degrade TA and most of the isolates utilize propionate, long-chain fatty acids and benzoate. The *Thermotogae* and WWE1 groups were estimated to constitute a significant proportion of the community based on the 16S rRNA analysis. Both for WWE1 and *Thermotogae*, the respective sample populations could not be modeled directly in composition-based binning, due to a lack of sample-specific training data, for WWE1 there was also not sufficient data to directly model the clade (Table 1). A protein-similarity comparison with sequenced members of the phylum *Thermotogae* (utilizing the distribution of BLAST matches for protein-coding genes in the dataset) resulted in 1066 genes with a BLAST matches greater than 60%, with additional 1646 genes having BLAST matches greater than 30%. Among these,

acetate kinase (tadcc6136) (Supplementary Fig. 8) and a phosphotransacetylase (tadcc64919) were identified, suggesting that members of the *Thermotogae* in the TA community may participate in the syntrophic interactions by producing acetate from an intermediate molecule. This intermediate molecule may be the butyrate produced by the OP5 population. A fragment encoding butyryl-CoA dehydrogenase, E.C. 1.3.99.2, (tadcc28367) further suggests the existence of the butyrate utilizing pathway, and a contig encoding a Fe-only hydrogenase (tadcc1433) suggests the ability of this population to generate $H_2$. Based on these observations, we hypothesize that the *Thermotogae* species may oxidize butyrate to acetate and $H_2$.

The sequence similarity-based phylogenetic profiler tool of IMG identified a set of genes from the TA degrading community with high similarity to *Candidatus* Cloacamonas acidaminovorans, that presents the only sequenced bacterial genome of the WWE1 candidate phylum through genome sequence reconstruction and is predicted as a syntrophic bacterium in anaerobic digesters (Pelletier et al., 2008). Comparing the common genes between the TA community dataset and the *C. acidaminovorans* genome identified 1607 and 3228 genes with sequence identity greater than 80% or 60%, respectively. These genes are likely to originate from the WWE1 population in the TA-degrading community. Among them, acetate kinases (tadcc38857) and acetyl phosphotransferases, E.C. 2.3.1.8, (tadcc25853, tadcc25854) were identified, suggesting an oxidative pathway generating acetate and energy via substrate-level phosphorylation and Fe-only hydrogenases (tadcc1522, tadcc13376, and tadcc38378), which presumably produce hydrogen. The substrate for this oxidative pathway may be butyrate since members of the butyrate-oxidizing pathway can be identified in the dataset.

*Methanogenic syntrophy*

Methanogenic syntrophy plays a critical role in the complete degradation of TA to methane (Fig. 4). Thermodynamic considerations suggest a low and narrow-range hydrogen concentration as the essential regulator to establish the syntrophic association between the $H_2$-producing bacteria and the $H_2$-consuming methanogens (Schink, 1997; Conrad, 1999). This is because the first reaction from TA to acetate and $CO_2/H_2$ (TA + $8H_2O \rightarrow$ 3 Acetate + $3H^+ + 2HCO_3^- + 3H_2$, $\Delta G^{o'}$ = 43.2 kJ/mol) can occur only at a low $pH_2$ (<5 Pa, 1 atm = 101325 Pa) by coupled with methanogenesis ($4TA + 35\ H_2O \rightarrow 17HCO_3^- + 9H^+ + 15CH_4$, $\Delta G^{o'} < -151.9$) (Schink, 1997). Also, a minimal threshold $pH_2$ is required for the $H_2$-dependent methanogenesis step to produce the minimum amount of energy required for cell maintenance (Conrad, 1999; Dolfing, 2001). When $H_2$ concentration is higher than this threshold level, $H_2$-dependent methanogenesis ceases. Such a low, narrow-range $pH_2$ is thought to be maintained by "interspecies hydrogen transfer" (Stams, 1994), in which $H_2$-producing syntrophs and $H_2$-consuming methanogens cooperate intimately by arranging themselves in close physical proximity in flocs or in a biofilm with short diffusion distances to facilitate hydrogen transfer. TA community metagenomic data revealed a set of hydrogenases, which generate hydrogen in *Pelotomaculum* and consume hydrogen in methanogens.

Using fluorescence *in-situ* hybridization analysis (Supplementary Fig. 9), we observed not only close physical proximity among methanogens and *Pelotomaculum* but also the presence of other microbes (i.e., OP5, WWE1, *Thermotogae*, and *Syntrophus*), as illustrated in Figure 3, associating with syntrophs and methanogens. Metagenomic analysis suggests that these populations can actively participate in the syntrophic interactions to tightly regulate $pH_2$. OP5 are likely to consume $CO_2$ and $H_2$ that are produced by *Pelotomaculum* through the degradation of TA, and produce butyrate. This concept is supported by the $\Delta G^{o'}$ value (-198.05 KJ/mol) for

the conversion of $CO_2$+$H_2$ to butyrate ($10H_2 + 4CO_2 \rightarrow C_4H_8O_2 + 6H_20$), which is even more favorable under hypermesophilic conditions than mesophilic conditions. OP5 and methanogens can compete for $H_2$ but the competition is likely to be $pH_2$-dependent.

Populations of *Syntrophus*, *Thermotogae*, and WWE1 may be involved in utilizing and recycling butyrate produced by OP5 probably through the secondary β-oxidation pathway. The presence of hydrogenases indicates that both *Thermotogae* and WWE1 gain energy through substrate-level phosphorylation. Although there is no clear evidence for the carbon source that these populations utilize, butyrate (produced by OP5) may serve as a key carbon source. This would suggest that, like *Syntrophus*, some members of *Themotogae* and WWE1 are possibly syntrophs. However, they are not persistently dominant populations and their abundance varies throughout the reactor operation (Fig. 2). TA metagenomics data indicates the presence of butyrate kinases and phosphotransacetylases in the *Pelotomaculum* bin, suggesting that this population may ferment TA not only to acetate but also butyrate. Our previous study also observed a detectable level of butyrate by using 2-bromoethanesulfonate to inhibit the methanogenesis step in a mesophilic TA-degrading consortium (Wu et al., 2001). It is possible that this type of fermentation results in the production of a second end product (butyrate in addition to acetate) and triggers a secondary syntrophic interaction involving butyrate-oxidizing organisms.

Several studies (Chan, 2000; Qiu et al., 2006; Imachi et al., 2008) have observed the existence of multiple bacterial populations in highly enriched methanogenic cultures degrading carbon substrates like formate, acetate, propionate and phthalate isomers. These observations were demonstrated through a defined mixed culture (Dolfing et al., 2008), suggesting that syntrophic interactions in methanogenic enrichments are more complex than simple pairwise

syntroph-methanogen relationships. Rather, they include other members that maintain and regulate the interspecies hydrogen transfer, which is the cornerstone of syntrophy. In conclusion, our overall observations imply that degradation of organic carbon is not simply a syntrophic interaction between $H_2$-producing syntrophs and methanogenic archaea. They further support the hypothesis that additional secondary interactions take place in order to maintain the stability of the TA degrading community.

ACKNOWLEDGMENTS

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.

**References**

Anderson, I., Rodriguez, J., Susanti, D., Porat, I., Reich, C., Ulrich, L.E. et al. (2008) Genome sequence of Thermofilum pendens reveals an exceptional loss of biosynthetic pathways without genome reduction. *Journal Of Bacteriology* **190**: 2957-2965.

Angelidaki, I., Petersen, S.P., and Ahring, B.K. (1990) Effects of lipids on thermophilic anaerobic-digestion and reduction of lipid inhibition upon addition of bentonite. *Applied Microbiology and Biotechnology* **33**: 469-472.

Boll, M., and Fuchs, G. (1995) Benzoyl-coenzyme A reductase (dearomatizing), a key enzyme of anaerobic aromatic metabolism - ATP dependence of the reaction, purification and some

properties of the enzyme from Thauera aromatica strain K172. *European Journal of Biochemistry* **234**: 921-933.

Butler, J.E., He, Q., Nevin, K.P., He, Z.L., Zhou, J.Z., and Lovley, D.R. (2007) Genomic and microarray analysis of aromatics degradation in Geobacter metallireducens and comparison to a Geobacter isolate from a contaminated field site. *BMC Genomics* **8**.

Chan, O.C. (2000) Characterization of microbial consortia in anaerobic granular sludge - a ribosomal RNA-based molecular approach. In *Civil and Environmental Engineering*. Hong Kong: University of Hong Kong, p. 221.

Chen, C.L., Macarie, H., Ramirez, I., Olmos, A., Ong, S.L., Monroy, O., and Liu, W.T. (2004) Microbial community structure in a thermophilic anaerobic hybrid reactor degrading terephthalate. *Microbiology* **150**: 3429-3440.

Conrad, R. (1999) Contribution of hydrogen to methane production and control of hydrogen concentrations in methanogenic soils and sediments. *FEMS Microbiology Ecology* **28**: 193-202.

Dolfing, J. (2001) The microbial logic behind the prevalence of incomplete oxidation of organic compounds by acetogenic bacteria in methanogenic environments. *Microbial Ecology* **41**: 83-89.

Dolfing, J., Jiang, B., Henstra, A.M., Stams, A.J.M., and Plugge, C.M. (2008) Syntrophic growth on formate: a new microbial niche in anoxic environments. *Applied and Environmental Microbiology* **74**: 6126-6131.

Imachi, H., Sakai, S., Sekiguchi, Y., Hanada, S., Kamagata, Y., Ohashi, A., and Harada, H. (2008) Methanolinea tarda gen. nov., sp nov., a methane-producing archaeon isolated from a methanogenic digester sludge. *International Journal of Systematic and Evolutionary Microbiology* **58**: 294-301.

Kleerebezem, R., Pol, L.W.H., and Lettinga, G. (1999) Anaerobic degradation of phthalate isomers by methanogenic consortia. *Appl Environ Microbiol* **65**: 1152-1160.

Kung, J.W., Loffler, C., Dorner, K., Heintz, D., Gallien, S., Van Dorsselaer, A. et al. (2009) Identification and characterization of the tungsten-containing class of benzoyl-coenzyme A reductases. *Proc Natl Acad Sci U S A* **106**: 17687-17692.

Lupa, B., Lyon, D., Gibbs, M.D., Reeves, R.A., and Wiegel, J. (2005) Distribution of genes encoding the microbial non-oxidative reversible hydroxyarylic acid decarboxylases/phenol carboxylases. *Genomics* **86**: 342-351.

Martin, H.G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy, A.C. et al. (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature Biotechnology* **24**: 1263-1269.

McHardy, A.C., Martin, H.G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007) Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods* **4**: 63-72.

McInerney, M.J., Rohlin, L., Mouttaki, H., Kim, U., Krupp, R.S., Rios-Hernandez, L. et al. (2007) The genome of Syntrophus aciditrophicus: Life at the thermodynamic limit of microbial growth. *Proc Natl Acad Sci U S A* **104**: 7600-7605.

Pelletier, E., Kreimeyer, A., Bocs, S., Rouy, Z., Gyapay, G., Chouari, R. et al. (2008) "Candidatus Cloacamonas acidaminovorans": Genome sequence reconstruction provides a first glimpse of a new bacterial division. *Journal Of Bacteriology* **190**: 2572-2579.

Peters, F., Shinoda, Y., McInerney, M.J., and Boll, M. (2007) Cyclohexa-1,5-diene-1-carbonyl-coenzyme A (CoA) hydratases of Geobacter metallireducens and Syntrophys aciditrophicus: Evidence for a common benzoyl-CoA degradation pathway in facultative and strict anaerobes. *Journal Of Bacteriology* **189**: 1055-1060.

Qiu, Y.L., Hanada, S., Ohashi, A., Harada, H., Kamagata, Y., and Sekiguchi, Y. (2008) Syntrophorhabdus aromaticivorans gen. nov., sp nov., the first cultured anaerobe capable of degrading phenol to acetate in obligate syntrophic associations with a hydrogenotrophic methanogen. *Applied and Environmental Microbiology* **74**: 2051-2058.

Qiu, Y.L., Sekiguchi, Y., Hanada, S., Imachi, H., Tseng, I.C., Cheng, S.S. et al. (2006) Pelotomaculum terephthalicum sp. nov. and Pelotomaculum isophthalicum sp. nov.: two anaerobic bacteria that degrade phthalate isomers in syntrophic association with hydrogenotrophic methanogens. *Arch Microbiol* **185**: 172-182.

Razo-Flores, E., Macarie, H., and Morier, F. (2006) Application of biological treatment systems for chemical and petrochemical wastewaters. In *Advanced Biological Treatment Processes for Industrial Wastewaters*. Cervantes, F.J., Pavlostathis, S.G., and van Haandel, A.C. (eds). London, UK: IWA Publishing., pp. 267-297.

Sato, T., Atomi, H., and Imanaka, T. (2007) Archaeal type IIIRuBisCOs function in a pathway for AMP metabolism. *Science* **315**: 1003-1006.

Schink, B. (1997) Energetics of syntrophic cooperation in methanogenic degradation. *Microbiology And Molecular Biology Reviews* **61**: 262-&.

Smith, K.S., and Ingram-Smith, C. (2007) Methanosaeta, the forgotten methanogen? *Trends in Microbiology* **15**: 150-155.

Stams, A.J.M. (1994) Metabolic interactions between anaerobic bacteria in methanogenic environments. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **66**: 271-294.

Tabita, F.R., Hanson, T.E., Li, H.Y., Satagopan, S., Singh, J., and Chan, S. (2007) Function, structure, and evolution of the RuBisCO-like proteins and their RubisCO homologs. *Microbiology and Molecular Biology Reviews* **71**: 576-599.

Thauer, R.K., Kaster, A.K., Seedorf, H., Buckel, W., and Hedderich, R. (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature Reviews Microbiology* **6**: 579-591.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M. et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37-43.

van Lier, J.B., Rebac, S., and Lettinga, G. (1997) High-rate anaerobic wastewater treatment under psychrophilic and thermophilic conditions. *Water Sci Technol* **35**: 199-206.

Wischgoll, S., Heintz, D., Peters, F., Erxleben, A., Sarnighausen, E., Reski, R. et al. (2005) Gene clusters involved in anaerobic benzoate degradation of Geobacter metallireducens. *Molecular Microbiology* **58**: 1238-1252.

Wu, J.H., Liu, W.T., Tseng, I.C., and Cheng, S.S. (2001) Characterization of microbial consortia in a terephthalate-degrading anaerobic granular sludge system. *Microbiology* **147**: 373-382.

**Table 1**. Phylogenetic affiliations of major bins in the TA dataset identified with the composition-based classifier, PhyloPythia (McHardy et al., 2007) [*].

| Phylogenetic affiliation | No of DNA contigs | Total sequence (Mb) | Average read depth | Expected genome size (Mb) [**] |
|---|---|---|---|---|
| *Bacteroidetes* (class) | 70 | 0.155 | 2.3±1.1 | - |
| *Bacteroidales* | 120 | 0.235 | 2.2±0.9 | - |
| *Betaproteobacteria* | 73 | 0.096 | 1.7±0.6 | - |
| *Deltaproteobacteria* | 160 | 0.343 | 2.1±0.8 | - |
| Uncultured *Syntrophus* | 196 | 0.724 | 2.9±1.3 | 1.9 |
| *Geobacter* | 264 | 0.578 | 2.1±0.7 | - |
| *Firmicutes* | 430 | 0.628 | 2.1±0.7 | - |
| *Clostridia* | 66 | 0.372 | 3.3±1.6 | - |
| Uncultured *Pelotomaculum* spp[#] | 1083 | 4.256 | 3.2±1.5 | 3.6 |
| OP5 | 228 | 1.411 | 3.8±1.9 | 2.8 |
| *Spirochaetes* (class) | 81 | 0.177 | 2.3±0.9 | - |
| *Euryarchaeota* | 1560 | 2.648 | 2.1±0.9 | - |
| *Thermoplasmata* | 71 | 0.148 | 1.9±0.8 | - |
| *Methanomicrobiales* | 36 | 0.098 | 2.4±1.8 | - |
| Uncultured *Methanolinea* spp[#] | 78 | 2.162 | 5.3±3.9 | 3.7 |
| *Methanosarcinales* | 15 | 0.095 | 3.7±1.9 | - |
| Uncultured *Methanosaeta* | 1180 | 2.613 | 2.6±1.1 | 3.1 |
| Uncultured *Methanosaeta* | 351 | 2.361 | 4.2±1.3 | 2.8 |

[*] No DNA contigs were binned to WWE1 related to *C. acidaminovorans* due to insufficient training data for PhyloPythia. 16S rRNA clone library indicted that132 sequences were affiliated with WWE1 and grouped into two different clusters. One of the clusters (37/132) was closely related to *C. acidaminovorans* (similarity = 96-98.8%).

[**] Expected genome size was calculated based on the percent coverage of the corresponding isolate genomes. For example, there are 1735 genes in *Pelotomaculum thermopropionicum* that are best-BLAST matches to genes from the metagenome dataset. Given that *P. thermopropionicum* contains 2920 genes, we estimate the genome size of the uncultured *Pelotomaculum* sp. was 7.16 Mb (4.256 *2920/1735). However, there are at least two strains of *Pelotomaculum* present in the sample. Therefore, the individual genome size for each strain is estimated to be around 3.6 Mb. For estimating the *Methanolinea* genome size we used as a reference genome *Methanoculleus marisnigri*. For *Methanosaeta* genome size, 2.6 Mb of sequences give hits to 1438 proteins in *M. thermophila* genome has 1730 CDS predicted so the expected genome size would be (2.6x1730)/1438=3.1Mb. In the case of OP5, expected genome size was calculated based on the occurrence of phylogenetic marker COGs that are defined as COGs having one or mostly one member in the genomes that are present and are available in IMG/M. The OP5 bin contained 91 out of 180 phylogenetic marker COGs.

[#], At least two species/strains were observed in each bin. With SNP frequencies of at least 0.03% to 0.07% (data not shown), we concluded that these species/strains are not clonal.

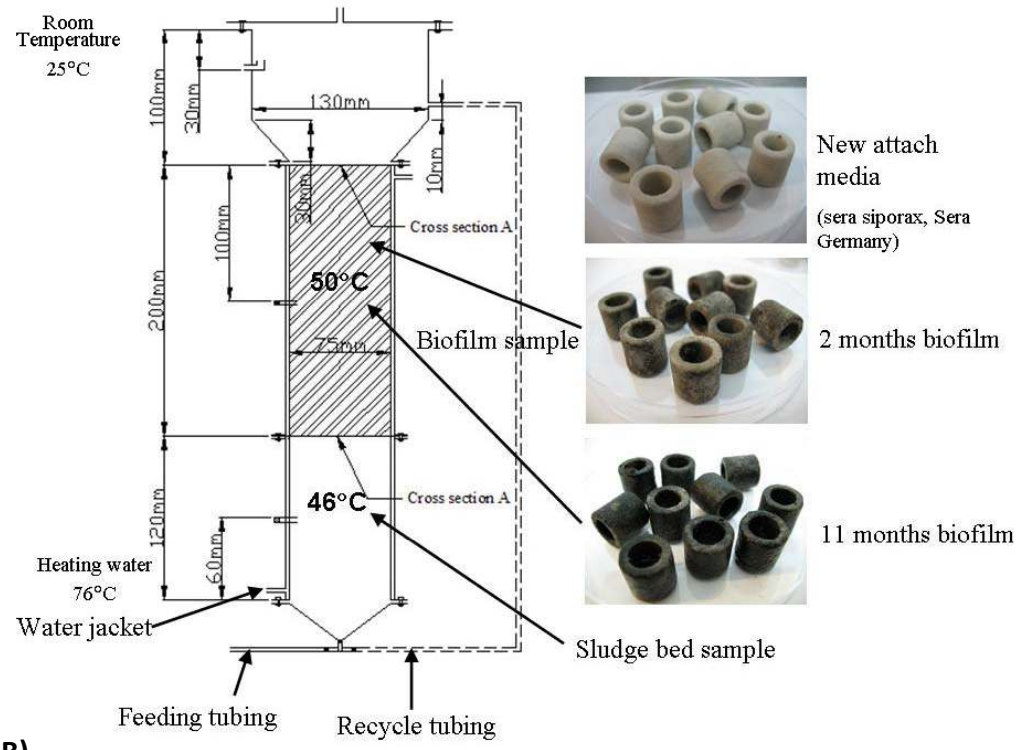
**Figure legends**

**Fig. 1.** Terephthalate (TA)-degrading laboratory-scale anaerobic hybrid bioreactor. **(A)** Schematic of the laboratory-scale anaerobic terephthalate-degrading hybrid reactor operated with a temperature gradient from ~ 46°C at the bottom to 50°C in the upper zone. Inserts illustrate freshly grown biofilm biomass on the surface of the media after two and 11months of enrichment; and **(B)** Performance of the reactor over 480-day operation. Under the initial operational conditions (i.e., TA-loading rate of 0.70-0.78 gTA/d.l, and hydraulic retention time (HRT) of 4 days), TA removal efficiency was gradually improved to 72.7% by day 124. By shortening the HRT (3 d on Day 127, 2 d on Day 168 and then to 1.5 d on Day 182) and increasing the TA loading concentration (to 3.2 g on Day 364), the TA loading rate was increased to 2.13 gTA/d·l by Day 364. Concurrently, the TA removal efficiency increased over the operation period reaching a 99% removal efficiency by Day 308. During the entire operation no sulfate reduction activity was detected. Samples were removed at the indicated time points (arrows) and the genomic DNA was extracted for 16S rRNA clone library construction and metagenomics analysis.

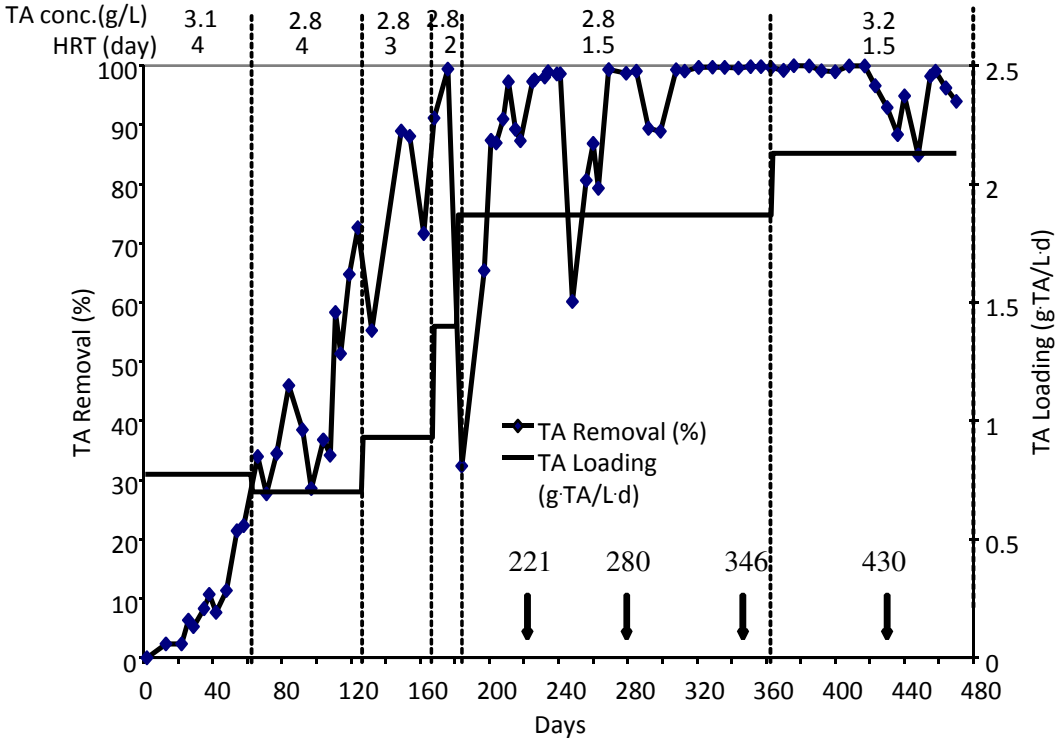
**Fig. 2**. Bacterial population dynamics of the TA-degrading bioreactor as revealed by 16S rRNA clone library. Samples (number of 16S rRNA sequences) from inner to outer of the ring chart were Day 221 biofilms (287), Day 280 biofilms (254), Day 346 biofilms (337), Day 346 sludge bed (289), Day 430 biofilms (352), and Day 430 sludge bed (287).

**Fig. 3.** Metabolic reconstruction of the TA-degrading syntrophic community. The metagenomic data revealed pathways for the degradation of the aromatic (terephthalate) compound, the recycling of the intermediates ($H_2$/$CO_2$, acetate, and butyrate) and the subsequent syntrophic methanogenesis.

**(A)**


Room Temperature 25°C
100mm
30mm
130mm
10mm
30mm
Cross section A
50°C
100mm
200mm
75mm
Biofilm sample
46°C
Cross section A
120mm
60mm
Heating water 76°C
Water jacket
Sludge bed sample
Feeding tubing
Recycle tubing
New attach media
(sera siporax, Sera Germany)
2 months biofilm
11 months biofilm

**(B)**


TA conc.(g/L) 3.1 2.8 2.8 2.8 2.8 3.2
HRT (day) 4 4 3 2 1.5 1.5
TA Removal (%)
TA Loading (g·TA/L·d)
TA Removal (%)
TA Loading (g·TA/L·d)
221
280
346
430
Days
0
40
80
120
160
200
240
280
320
360
400
440
480
0
10
20
30
40
50
60
70
80
90
100
0
0.5
1
1.5
2
2.5

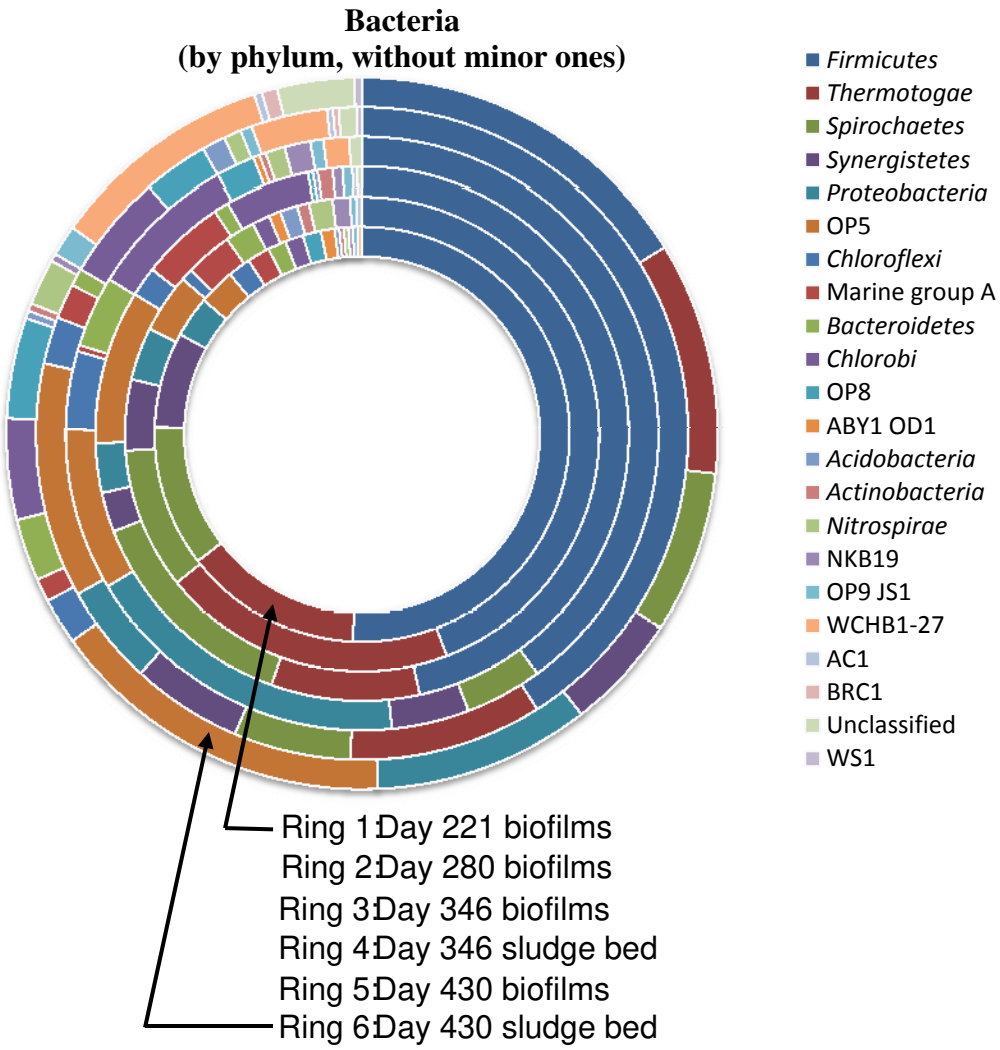
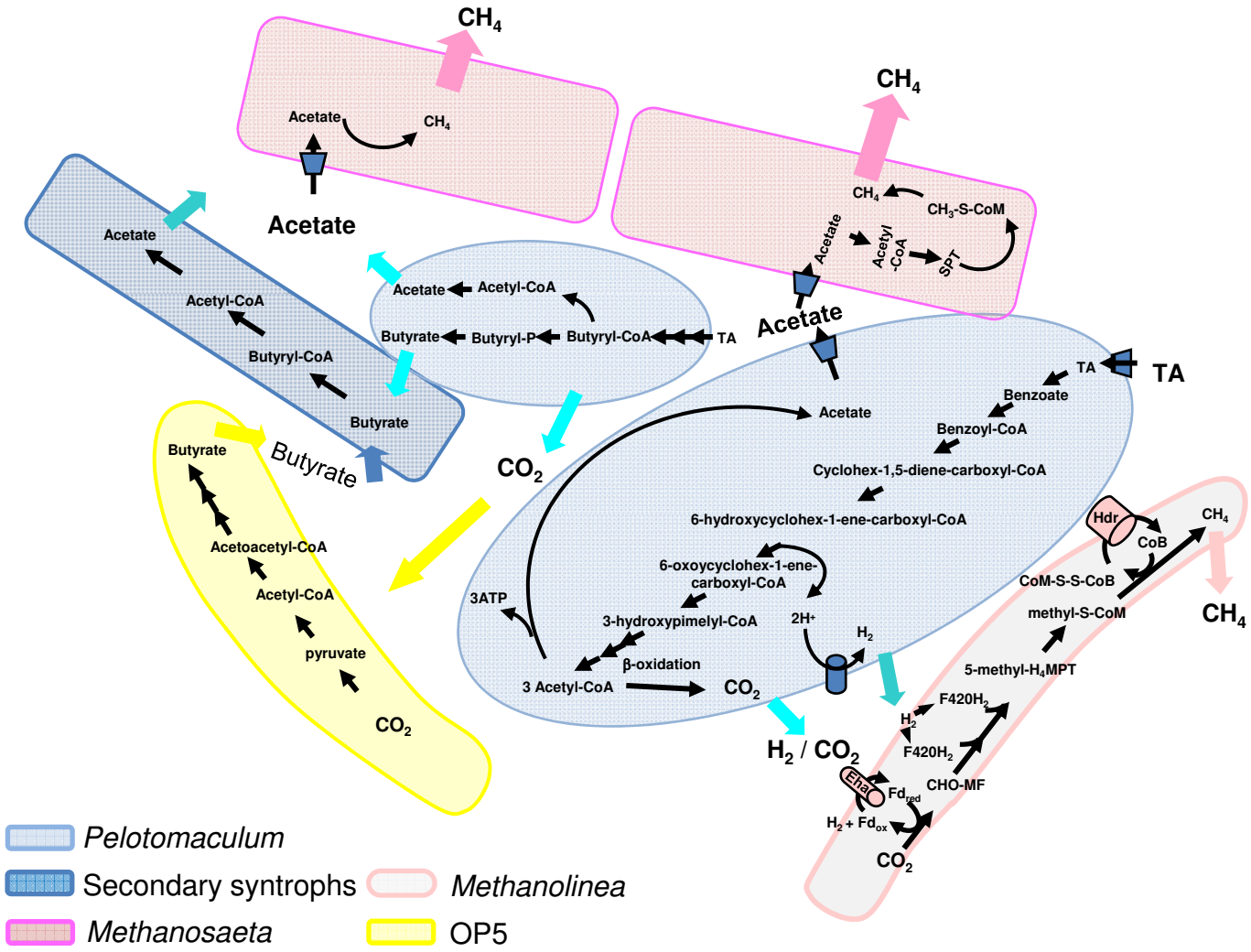



**Fig. 2**. Bacterial population dynamics of the TA-degrading bioreactor as revealed by 16S rRNA clone library. Samples (number of 16S rRNA sequences) from inner to outer of the ring chart were Day 221 biofilms (287), Day 280 biofilms (254), Day 346 biofilms (337), Day 346 sludge bed (289), Day 430 biofilms (352), and Day 430 sludge bed (287).


CH4
Acetate
CH4
Acetate
CH4
CH4
CH3-S-CoM
Acetate
Acetyl-CoA
SPT
Acetate
Acetyl-CoA
Butyryl-CoA
Butyrate
Acetate
Acetyl-CoA
Butyrate
Butyryl-P
Butyryl-CoA
TA
Acetate
TA
TA
Benzoate
Benzoyl-CoA
Cyclohex-1,5-diene-carboxyl-CoA
6-hydroxycyclohex-1-ene-carboxyl-CoA
6-oxoycyclohex-1-ene-carboxyl-CoA
3ATP
3-hydroxypimelyl-CoA
β-oxidation
3 Acetyl-CoA
CO2
2H+
H2
H2 / CO2
Butyrate
Butyrate
Acetoacetyl-CoA
Acetyl-CoA
pyruvate
CO2
CO2
Hdr
CoB
CH4
CoM-S-S-CoB
methyl-S-CoM
CH4
5-methyl-H4MPT
F420H2
H2
F420H2
Eha
Fdred
CHO-MF
H2 + Fdox
CO2
Pelotomaculum
Secondary syntrophs
Methanolinea
Methanosaeta
OP5

**Supplementary Figure Legend**

**Fig. 1.** The rarefaction curve generated for the bacterial populations and archaeal populations in those six samples taken from the TA reactor. Sequence dissimilarities of 3 and 5% were both used to produce the rarefaction curves of the domain *Bacteria* and domain *Archaea*. Higher microbial diversity was observed with the domain *Bacteria* than the domain *Archaea* in any given sample taken from the TA reactor. Within the domain *Archaea*, all samples had a similar rarefaction curve except for the sample taken at day 430 from the sludge bed.

**Fig. 2** Archaeal population dynamics of the TA-degrading bioreactor as revealed by 16S rRNA clone library. Samples (number of 16S rRNA sequences) from inner to outer of the ring chart were Day 221 biofilms (359), Day 280 biofilms (362), Day 346 biofilms (343), Day 346 sludge bed (356), Day 430 biofilms (278), and Day 430 sludge bed (211).

**Fig. 3** 16S rRNA neighbor joining phylogeny tree of *Pelotomaculum* species from the family *Desulfotomaculum* (number of bootstrapping, 500). The dominant bacterial 16S rRNA sequences obtained from hypermesophilic and thermophilic TA-degrading bioreactors were shown. The *Pelotomaculum*-related sequences obtained under hyper-mesophilic conditions (46-50$^{o}$C) are different from those *Pelotomaculum*-related syntrophs obtained under thermophilic conditions (~55$^{o}$C).

**Fig. 4** Methanogen diversity in TA-degrading reactor as determined by 16S rRNA and McrA. (A) 16S rRNA neighbor joining phylogeny tree (number of bootstrapping, 500). Within the *Methanosarcinales*, the methanogen sequences obtained from mesophilic and thermophilic TA-

degrading consortia were closely related to *Methanosaeta concilii* and *Methanosaeta thermophila*, respectively. However the methanogen sequences obtained from the hyper-mesophilic TA-degrading consortium formed two novel clusters, suggesting these methanogens are different from those predominating in mesophilic and thermophilic conditions. Within the order *Methanomicrobiales*, the methanogen sequences obtained from mesophilic, hyper-mesophilic and thermophilic TA-degrading consortia were all clustered with a novel methanogen, *Methanolinea tarda*, recently isolated from a propionate-degrading anaerobic reactor. *M. tarda* is a hydrogen-utilizing methanogen. (B) McrA-based neighbor joining phylogeny tree (bootstrap = 500). Sequences obtained from the TA metagenomics were denoted with 'tadcc'. tadcc5364 and tadcc11668 likely represented one or two different clusters within the genus *Methanosaeta* and are different from *Methanosaeta thermophila* primarily found in thermophilic conditions (~55$^{o}$C). Likewise, tadcc40165 and tadcc3600 represented a cluster that is different from the cluster containing tadcc70096 and *M. tarda* within the *Methanomicrobiales*.

**Fig. 5** Size distribution and read depth of DNA contigs obtained by shotgun sequencing and assembly. In total, 52342 DNA contigs were obtained. The largest contig is approximately 240 kb in size and contains 271 genes. Additional 45 fragments are 24 to 167 kb long.

**Fig. 6** Genome coverage of sequenced isolate bacterial and archaeal genomes. Genome coverage is calculated based on the distribution of best BLAST matches of protein-coding genes in the metagenome dataset as described in the supplementary text.

**Fig. 7** Reconstruction of the TA degradation pathway operating in *Pelotomaculum* sp. All genes necessary for terephthalate degradation have been identified in the *Pelotomaculum* bin. To identify specific TA decarboxylases, searches for genes with known decarboxylase functions revealed two gene sets (tadcc27178-79-80 and tadcc16349) that have sequence similarity and a

subunit complement with a known 4-hydroxybenzoate decarboxylase (EC 4.1.1.61) from *Clostridium hydroxybenzoicum* (1). This decarboxylase consists of three subunits and belongs to the UbiD family of proteins. Only one of these genes, tadcc16349, resides on a contig assigned to the *Pelotomaculum* bin. The contig containing the tadcc27178 gene set (taComm3_C8975) is binned to a higher-level classification of *Clostridia*, but contains two genes with a high degree of similarity to *Pelotomaculum* genes (a dnaE gene (tadcc27166) and an endonuclease (tadcc27168)), and includes three fosmid end-reads whose mate pairs map to contigs in the *Pelotomaculum* bin. Thus we conclude that the tadcc27178 gene set likely derives from a *Pelotomaculum* population. All genes necessary for terephthalate degradation have been identified in the *Pelotomaculum* bin except the phosphotransacetylase gene. Multiple phosphate butyryltransferases (tadcc9380, tadcc37666, tadcc28170) and butyrate kinases, EC 2.7.2.7, (tadcc9381, tadcc37665, tadcc28171) are present in the *Pelotomaculum* dataset.

**Fig. 8** Protein neighbor joining phylogenetic tree of acetate (Ack) and butyrate kinases (Btk) (number of bootstrapping, 500). Locus tags correspond to the genes from the following organisms: CPR, *Clostridium perfringens* SM101; CPE, *Clostridium perfringens* 13; CTC, *Clostridium tetani* E88; CAC, *Clostridium acetobutylicum* ATCC 824; CD, *Clostridium difficile* 630; Amet, *Alkaliphilus metalliredigens* QYMF; TTE, *Thermoanaerobacter tengcongensis* MB4; Gmet, *Geobacter metallireducens* GS-15; Geob, *Geobacter* sp. FRC-32; TRQ2, *Thermotoga* sp. RQ2; TM, *Thermotoga maritima* MSB8; CTN, *Thermotoga neapolitana* DSM 4359; Tmel, *Thermosipho melanesiensis* BI429; THA, *Thermosipho africanus* TCF52B; TLET, *Thermotoga lettingae* TMO; Hore, *Halothermothrix orenii* H 168; BLI, *Bacillus licheniformis* ATCC 14580; GK, *Geobacillus kaustophilus* HTA426; Bcer98, *Bacillus cereus*; BT9727,

*Bacillus thuringiensis*; CAO, *Candidatus* Cloacamonas acidaminovorans; Cbei, *Clostridium beijerinckii* NCIMB 8052.

**Fig. 9** Biofilm samples of TA-degrading methanogenic consortium as revealed by FISH. (A) The *Desulfotomaculum*-related cells (yellow color) as targeted by probe Ih820 accounted for 85.2 $\pm$6.4 % of the total bacterial populations detected by probe EUB338mix (in Green). (B) Bacterial and archaeal cells targeted by probes EUBmix (green color) and ARC915 (red color), respectively, appeared in equal abundance in the biofilm sample. Size bar = 10 µm.

Fig. S1

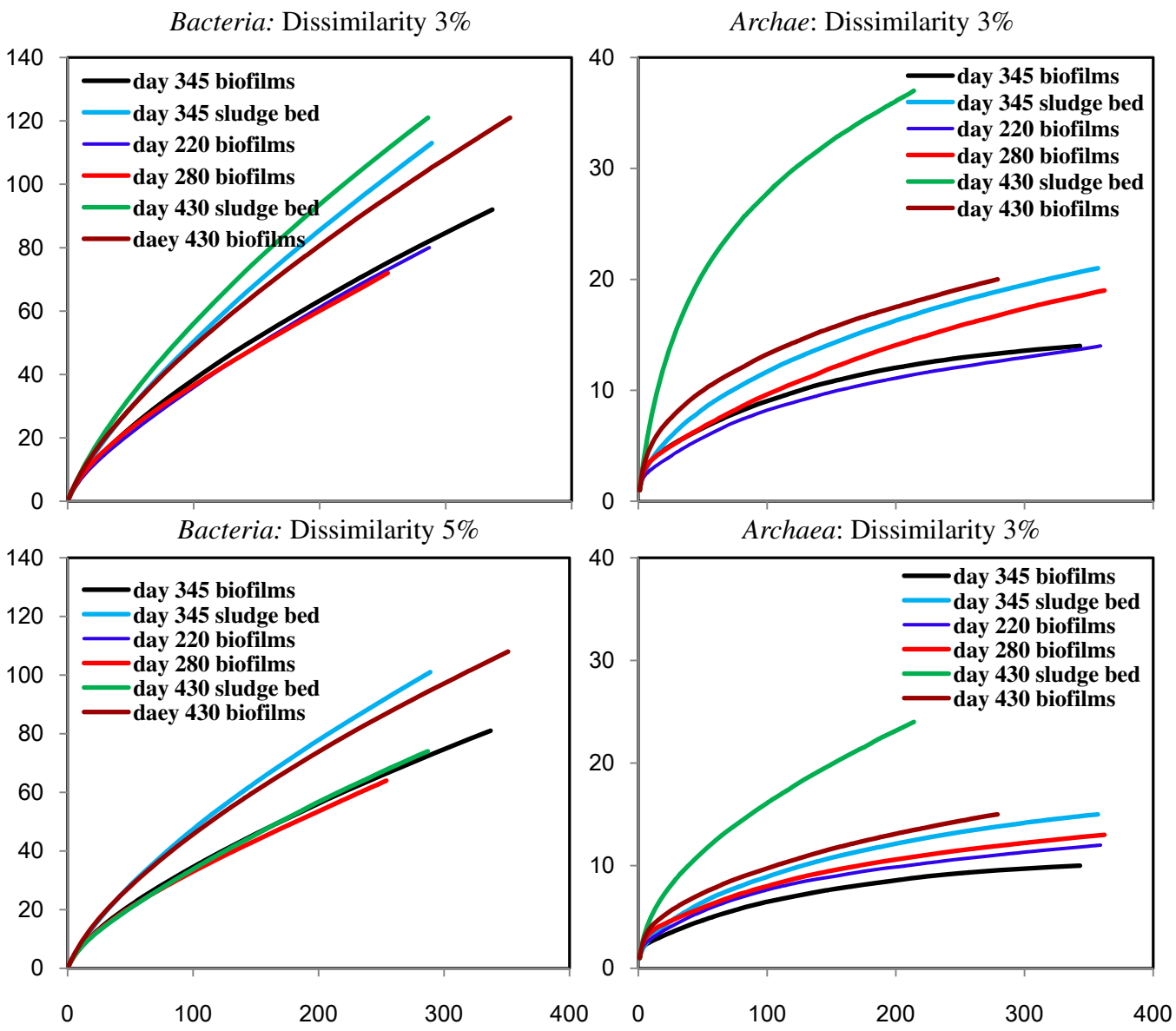

*Bacteria:* Dissimilarity 3%
day 345 biofilms
day 345 sludge bed
day 220 biofilms
day 280 biofilms
day 430 sludge bed
daey 430 biofilms
*Archae*: Dissimilarity 3%
day 345 biofilms
day 345 sludge bed
day 220 biofilms
day 280 biofilms
day 430 sludge bed
day 430 biofilms
*Bacteria:* Dissimilarity 5%
day 345 biofilms
day 345 sludge bed
day 220 biofilms
day 280 biofilms
day 430 sludge bed
daey 430 biofilms
*Archaea*: Dissimilarity 3%
day 345 biofilms
day 345 sludge bed
day 220 biofilms
day 280 biofilms
day 430 sludge bed
day 430 biofilms
0
100
200
300
400

Fig. S2

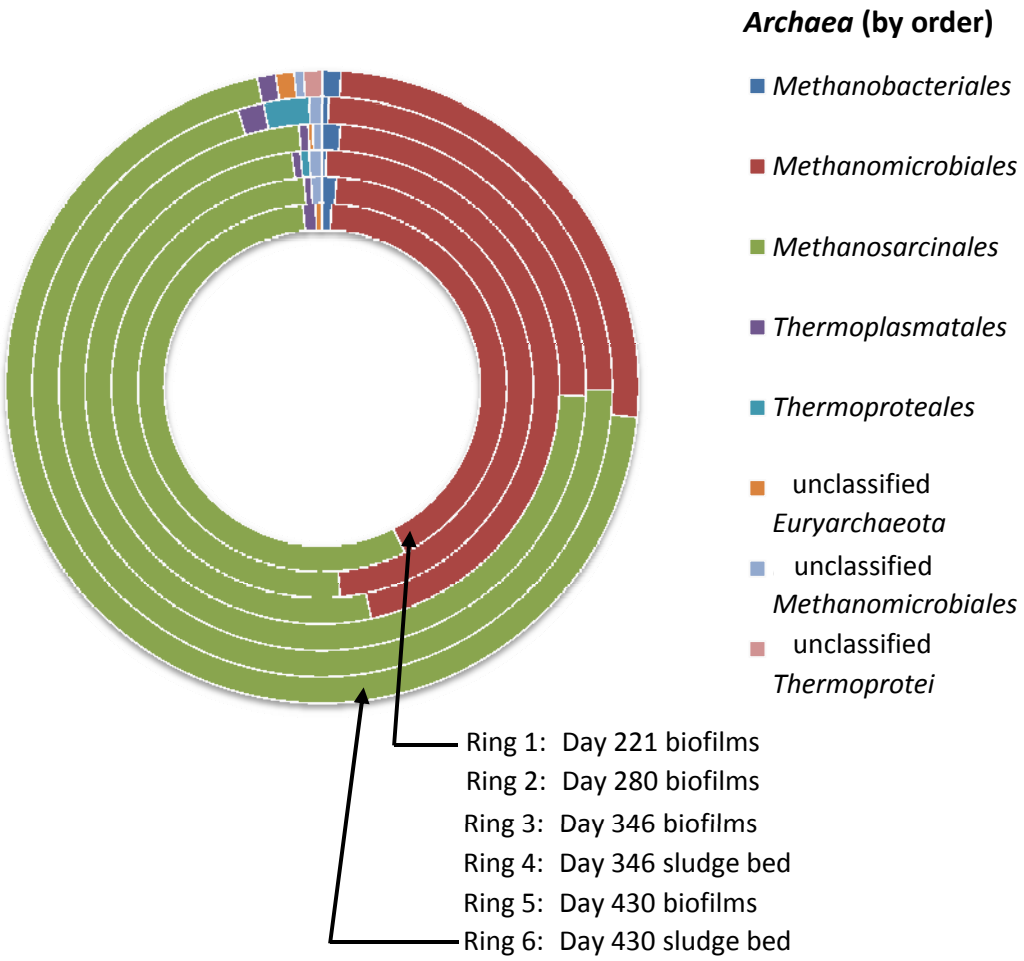

***Archaea* (by order)**
*Methanobacteriales*
*Methanomicrobiales*
*Methanosarcinales*
*Thermoplasmatales*
*Thermoproteales*
unclassified *Euryarchaeota*
unclassified *Methanomicrobiales*
unclassified *Thermoprotei*
Ring 1: Day 221 biofilms
Ring 2: Day 280 biofilms
Ring 3: Day 346 biofilms
Ring 4: Day 346 sludge bed
Ring 5: Day 430 biofilms
Ring 6: Day 430 sludge bed

Fig. S3

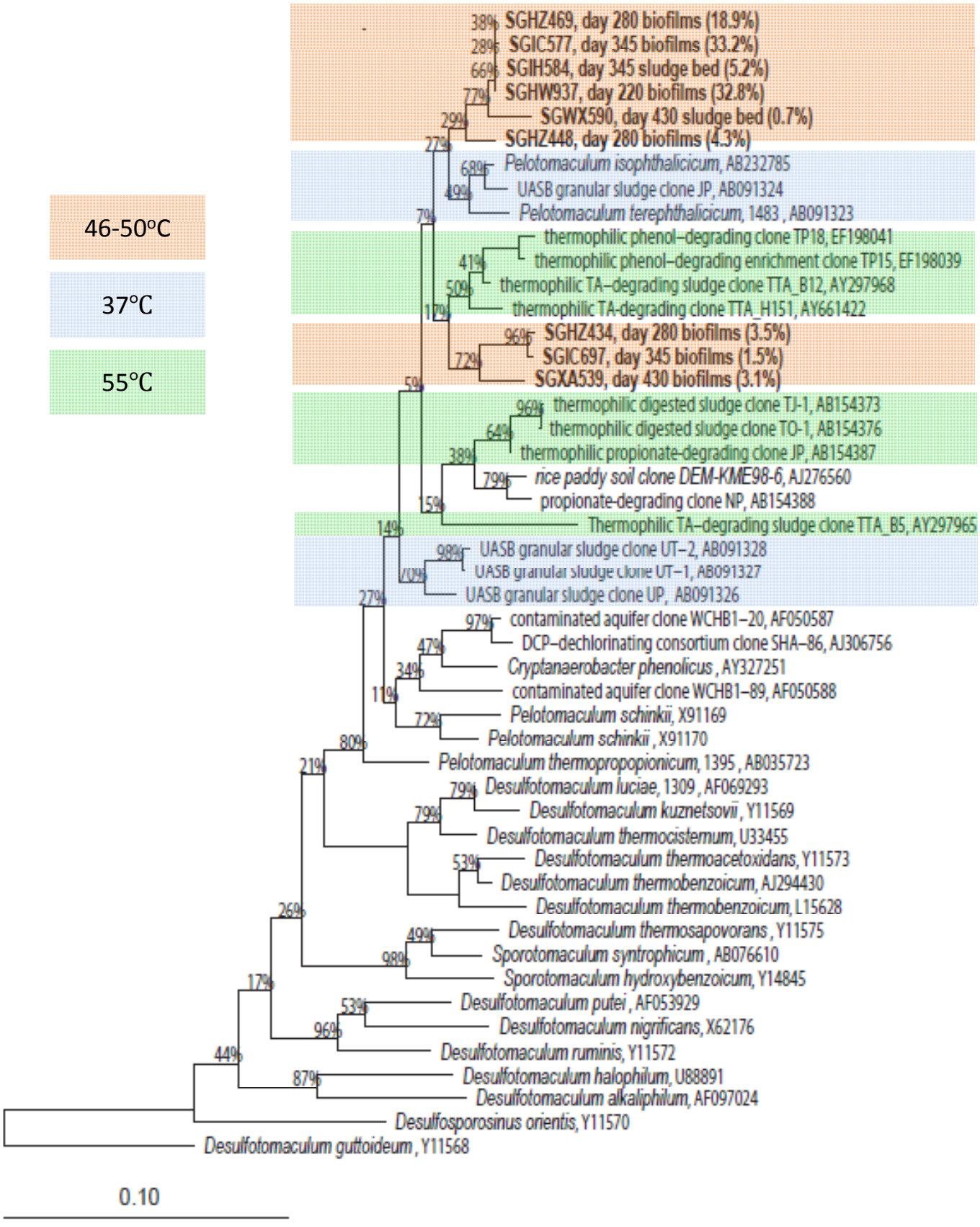

46-50°C
37°C
55°C
38% SGHZ469, day 280 biofilms (18.9%)
28% SGIC577, day 345 biofilms (33.2%)
66% SGIH584, day 345 sludge bed (5.2%)
77% SGHW937, day 220 biofilms (32.8%)
SGWX590, day 430 sludge bed (0.7%)
29% SGHZ448, day 280 biofilms (4.3%)
27% Pelotomaculum isophthalicicum, AB232785
68% UASB granular sludge clone JP, AB091324
49% Pelotomaculum terephthalicicum, 1483 , AB091323
7% thermophilic phenol–degrading clone TP18, EF198041
thermophilic phenol–degrading enrichment clone TP15, EF198039
41% thermophilic TA–degrading sludge clone TTA_B12, AY297968
50% thermophilic TA-degrading clone TTA_H151, AY661422
17% 96% SGHZ434, day 280 biofilms (3.5%)
72% SGIC697, day 345 biofilms (1.5%)
SGXA539, day 430 biofilms (3.1%)
5% 96% thermophilic digested sludge clone TJ-1, AB154373
64% thermophilic digested sludge clone TO-1, AB154376
38% thermophilic propionate-degrading clone JP, AB154387
79% rice paddy soil clone DEM-KME98-6, AJ276560
propionate-degrading clone NP, AB154388
15% Thermophilic TA–degrading sludge clone TTA_B5, AY297965
14% 98% UASB granular sludge clone UT–2, AB091328
70% UASB granular sludge clone UT–1, AB091327
UASB granular sludge clone UP, AB091326
27% 97% contaminated aquifer clone WCHB1–20, AF050587
47% DCP–dechlorinating consortium clone SHA–86, AJ306756
34% Cryptanaerobacter phenolicus , AY327251
11% contaminated aquifer clone WCHB1–89, AF050588
72% Pelotomaculum schinkii, X91169
Pelotomaculum schinkii , X91170
80% Pelotomaculum thermopropopionicum, 1395 , AB035723
21% 79% Desulfotomaculum luciae, 1309 , AF069293
79% Desulfotomaculum kuznetsovii , Y11569
Desulfotomaculum thermocisternum, U33455
53% Desulfotomaculum thermoacetoxidans, Y11573
Desulfotomaculum thermobenzoicum, AJ294430
Desulfotomaculum thermobenzoicum, L15628
26% 49% Desulfotomaculum thermosapovorans , Y11575
98% Sporotomaculum syntrophicum , AB076610
Sporotomaculum hydroxybenzoicum, Y14845
17% 53% Desulfotomaculum putei , AF053929
96% Desulfotomaculum nigrificans, X62176
Desulfotomaculum ruminis, Y11572
44% 87% Desulfotomaculum halophilum, U88891
Desulfotomaculum alkaliphilum, AF097024
Desulfosporosinus orientis, Y11570
Desulfotomaculum guttoideum , Y11568
0.10

Fig. S4 (A)

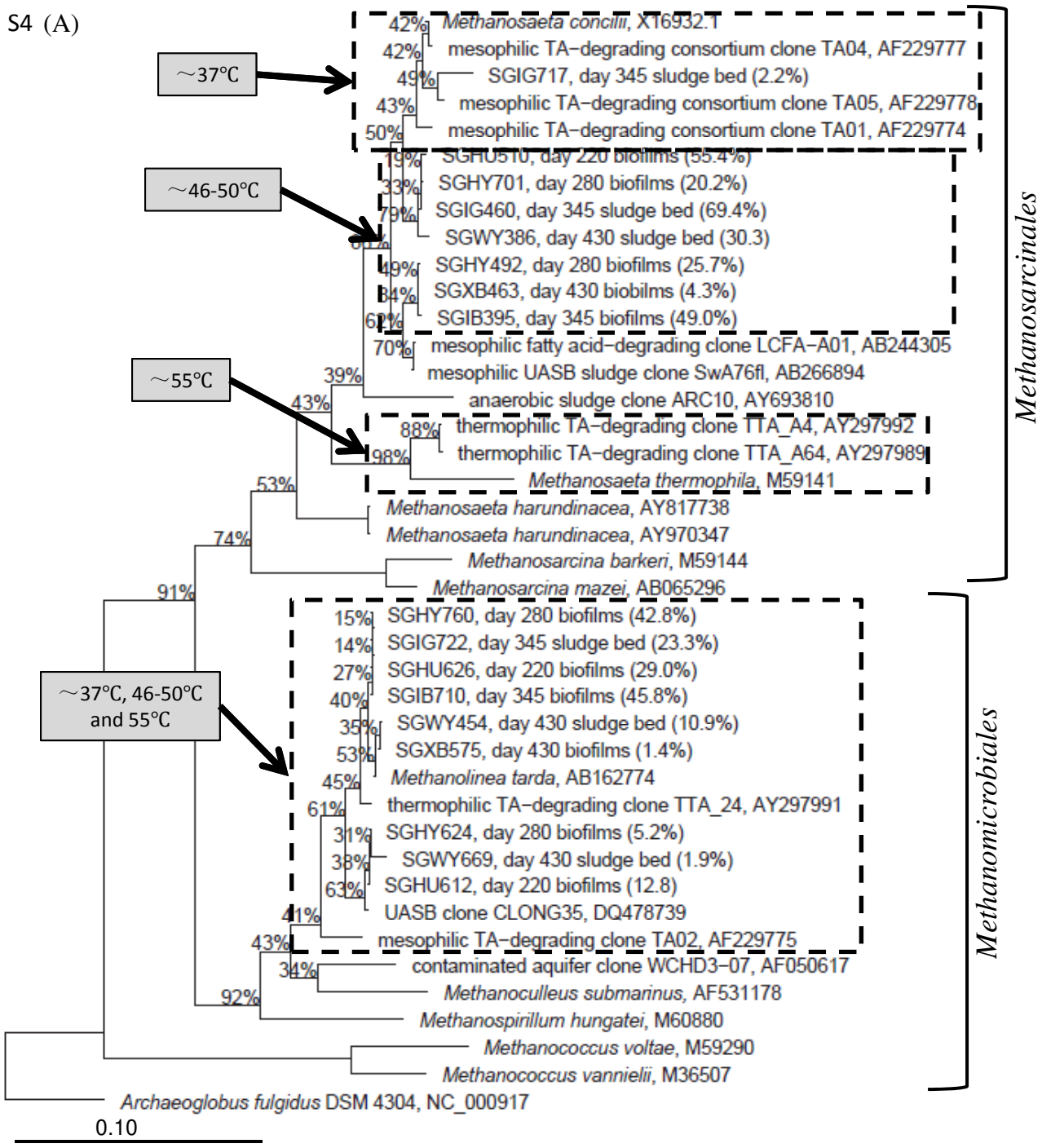

~37°C
~46-50°C
~55°C
~37°C, 46-50°C and 55°C
Methanosarcinales
Methanomicrobiales
42% Methanosaeta concilii, X16932.1
42% mesophilic TA-degrading consortium clone TA04, AF229777
49% SGIG717, day 345 sludge bed (2.2%)
43% mesophilic TA-degrading consortium clone TA05, AF229778
50% mesophilic TA-degrading consortium clone TA01, AF229774
79% SGHU510, day 220 biofilms (55.4%)
33% SGHY701, day 280 biofilms (20.2%)
79% SGIG460, day 345 sludge bed (69.4%)
SGWY386, day 430 sludge bed (30.3)
49% SGHY492, day 280 biofilms (25.7%)
34% SGXB463, day 430 biobilms (4.3%)
62% SGIB395, day 345 biofilms (49.0%)
70% mesophilic fatty acid-degrading clone LCFA-A01, AB244305
39% mesophilic UASB sludge clone SwA76fl, AB266894
43% anaerobic sludge clone ARC10, AY693810
88% thermophilic TA-degrading clone TTA_A4, AY297992
98% thermophilic TA-degrading clone TTA_A64, AY297989
Methanosaeta thermophila, M59141
53% Methanosaeta harundinacea, AY817738
Methanosaeta harundinacea, AY970347
74% Methanosarcina barkeri, M59144
91% Methanosarcina mazei, AB065296
15% SGHY760, day 280 biofilms (42.8%)
14% SGIG722, day 345 sludge bed (23.3%)
27% SGHU626, day 220 biofilms (29.0%)
40% SGIB710, day 345 biofilms (45.8%)
35% SGWY454, day 430 sludge bed (10.9%)
53% SGXB575, day 430 biofilms (1.4%)
45% Methanolinea tarda, AB162774
61% thermophilic TA-degrading clone TTA_24, AY297991
31% SGHY624, day 280 biofilms (5.2%)
38% SGWY669, day 430 sludge bed (1.9%)
63% SGHU612, day 220 biofilms (12.8)
41% UASB clone CLONG35, DQ478739
43% mesophilic TA-degrading clone TA02, AF229775
34% contaminated aquifer clone WCHD3-07, AF050617
92% Methanoculleus submarinus, AF531178
Methanospirillum hungatei, M60880
Methanococcus voltae, M59290
Methanococcus vannielii, M36507
Archaeoglobus fulgidus DSM 4304, NC_000917
0.10

Fig. S4 (B)

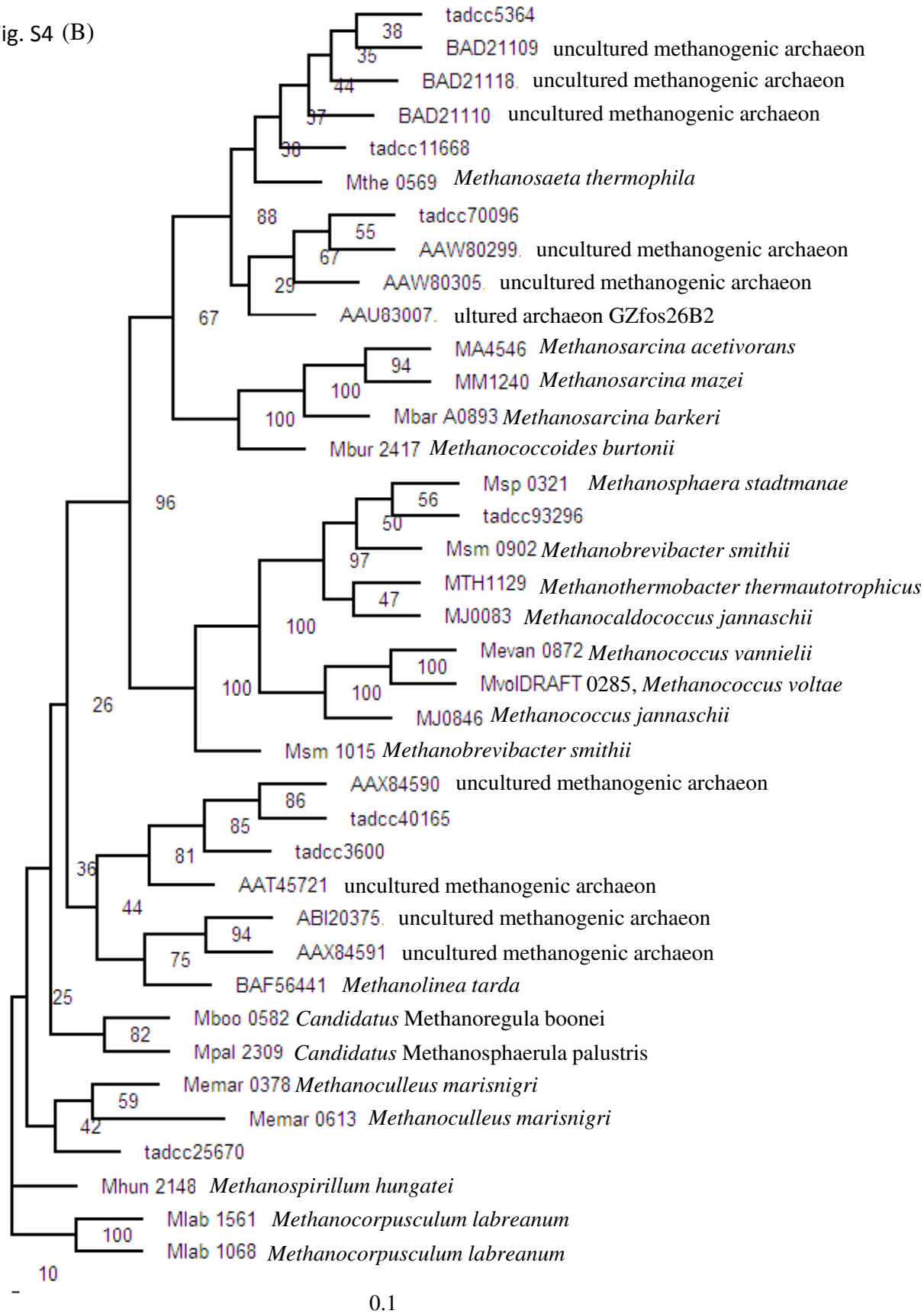

tadcc5364
BAD21109 uncultured methanogenic archaeon
BAD21118. uncultured methanogenic archaeon
BAD21110. uncultured methanogenic archaeon
tadcc11668
Mthe 0569 Methanosaeta thermophila
tadcc70096
AAW80299. uncultured methanogenic archaeon
AAW80305. uncultured methanogenic archaeon
AAU83007. ultured archaeon GZfos26B2
MA4546 Methanosarcina acetivorans
MM1240 Methanosarcina mazei
Mbar A0893 Methanosarcina barkeri
Mbur 2417 Methanococcoides burtonii
Msp 0321 Methanosphaera stadtmanae
tadcc93296
Msm 0902 Methanobrevibacter smithii
MTH1129 Methanothermobacter thermautotrophicus
MJ0083 Methanocaldococcus jannaschii
Mevan 0872 Methanococcus vannielii
MvolDRAFT 0285, Methanococcus voltae
MJ0846 Methanococcus jannaschii
Msm 1015 Methanobrevibacter smithii
AAX84590 uncultured methanogenic archaeon
tadcc40165
tadcc3600
AAT45721 uncultured methanogenic archaeon
ABI20375. uncultured methanogenic archaeon
AAX84591 uncultured methanogenic archaeon
BAF56441 Methanolinea tarda
Mboo 0582 Candidatus Methanoregula boonei
Mpal 2309 Candidatus Methanosphaerula palustris
Memar 0378 Methanoculleus marisnigri
Memar 0613 Methanoculleus marisnigri
tadcc25670
Mhun 2148 Methanospirillum hungatei
Mlab 1561 Methanocorpusculum labreanum
Mlab 1068 Methanocorpusculum labreanum
0.1

Fig. S5

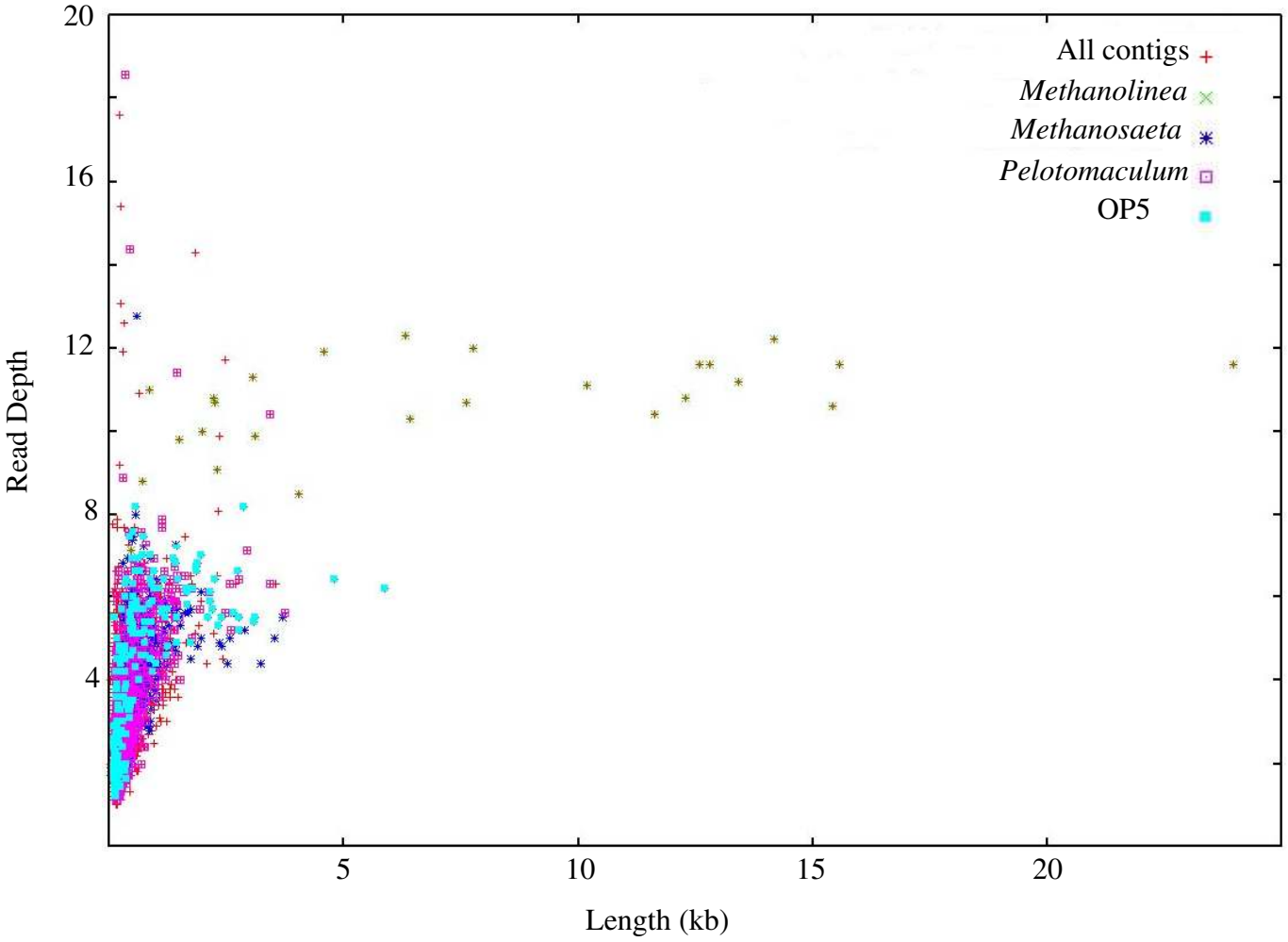

20
16
12
8
4
Read Depth
5
10
15
20
Length (kb)
All contigs
Methanolinea
Methanosaeta
Pelotomaculum
OP5

Fig. S6

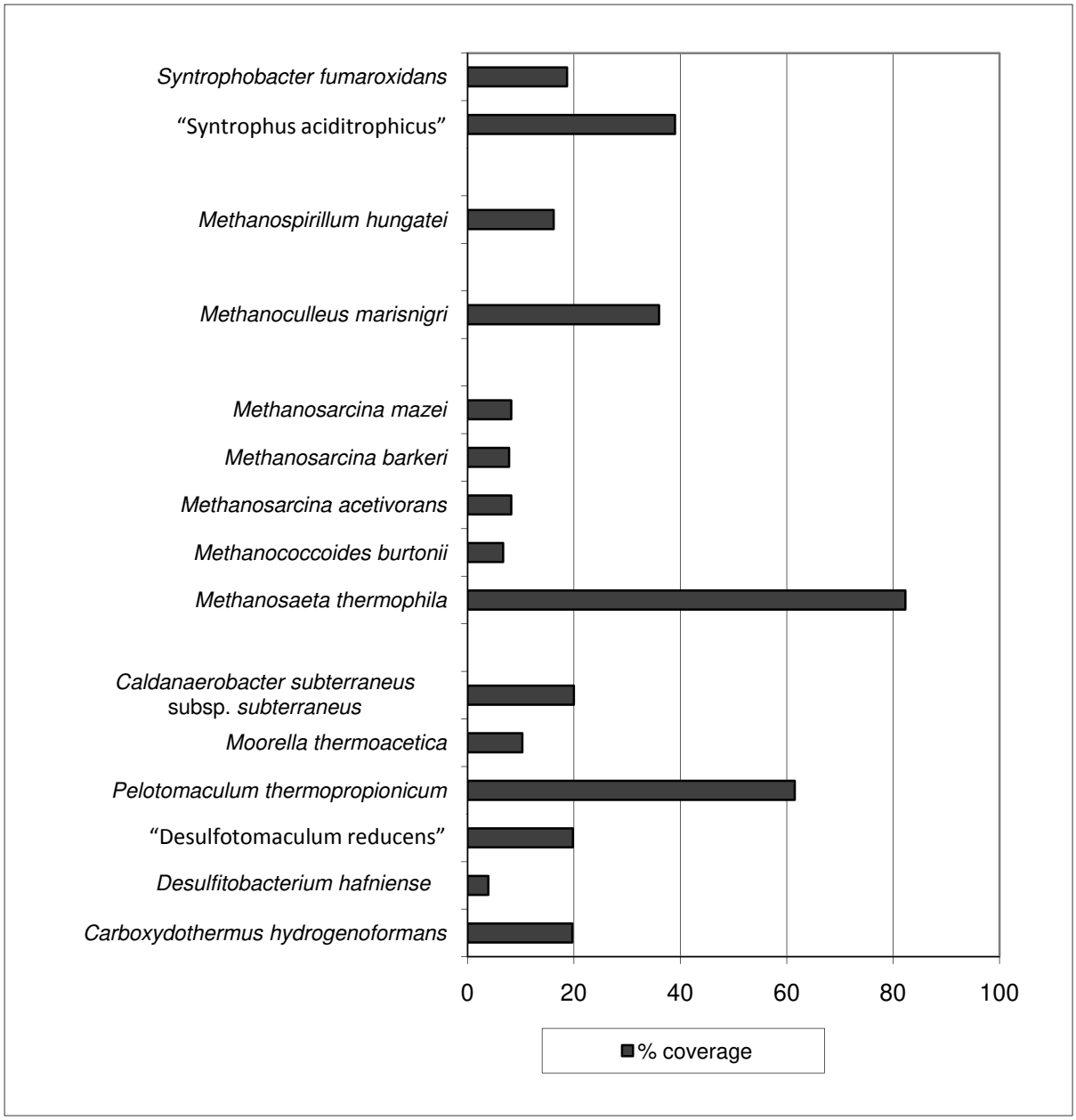

Syntrophobacter fumaroxidans
"Syntrophus aciditrophicus"
Methanospirillum hungatei
Methanoculleus marisnigri
Methanosarcina mazei
Methanosarcina barkeri
Methanosarcina acetivorans
Methanococcoides burtonii
Methanosaeta thermophila
Caldanaerobacter subterraneus subsp. subterraneus
Moorella thermoacetica
Pelotomaculum thermopropionicum
"Desulfotomaculum reducens"
Desulfitobacterium hafniense
Carboxydothermus hydrogenoformans
0
20
40
60
80
100
% coverage

Fig. S7

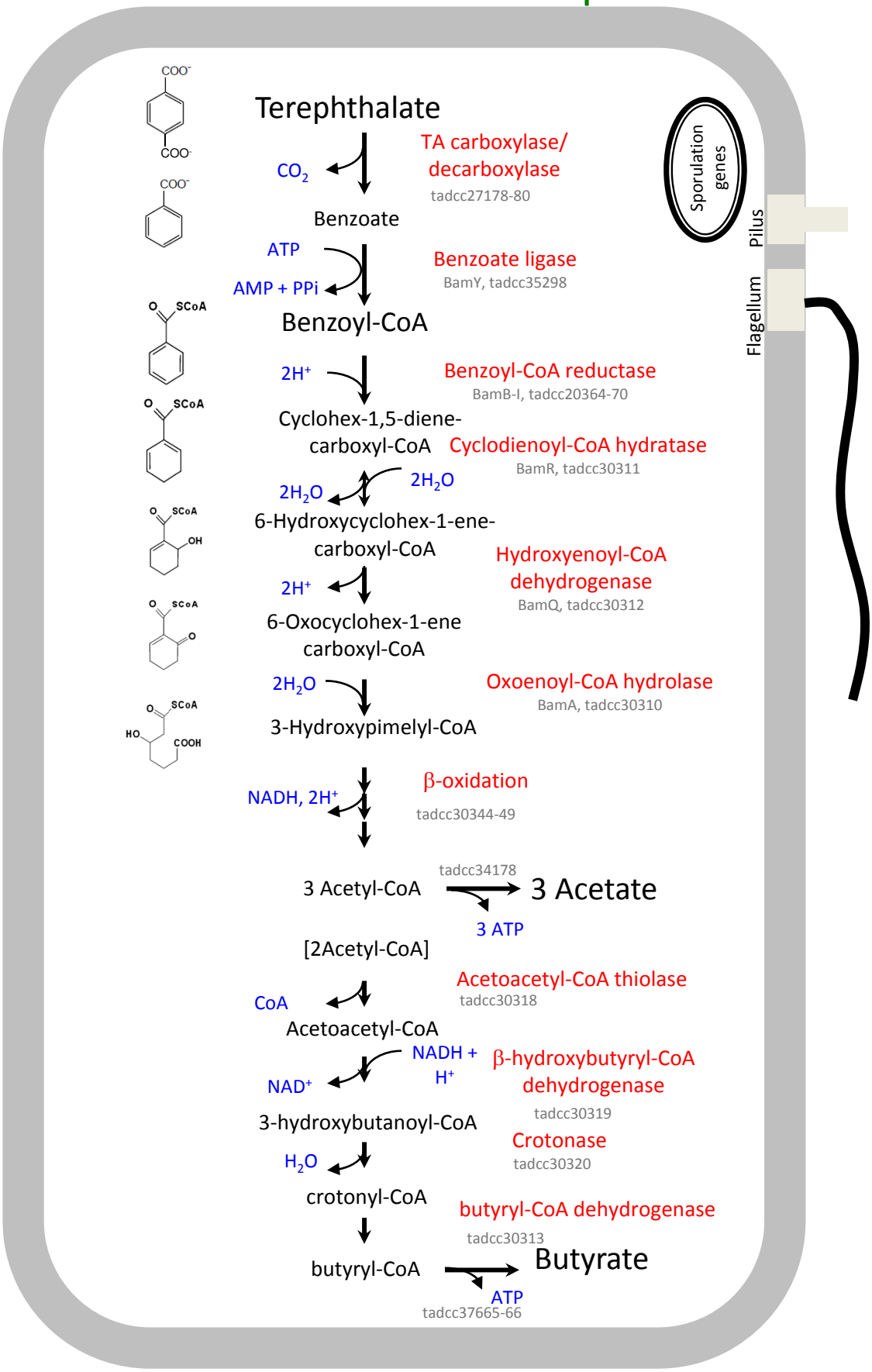

Pelotomaculum sp.
Terephthalate
CO2
TA carboxylase/ decarboxylase
tadcc27178-80
Benzoate
ATP
AMP + PPi
Benzoate ligase
BamY, tadcc35298
Benzoyl-CoA
2H+
Benzoyl-CoA reductase
BamB-I, tadcc20364-70
Cyclohex-1,5-diene-carboxyl-CoA
Cyclodienoyl-CoA hydratase
BamR, tadcc30311
2H2O
2H2O
6-Hydroxycyclohex-1-ene-carboxyl-CoA
Hydroxyenoyl-CoA dehydrogenase
BamQ, tadcc30312
2H+
6-Oxocyclohex-1-ene carboxyl-CoA
2H2O
Oxoenoyl-CoA hydrolase
BamA, tadcc30310
3-Hydroxypimelyl-CoA
NADH, 2H+
β-oxidation
tadcc30344-49
3 Acetyl-CoA
tadcc34178
3 Acetate
3 ATP
[2Acetyl-CoA]
CoA
Acetoacetyl-CoA thiolase
tadcc30318
Acetoacetyl-CoA
NADH + H+
NAD+
β-hydroxybutyryl-CoA dehydrogenase
tadcc30319
3-hydroxybutanoyl-CoA
H2O
Crotonase
tadcc30320
crotonyl-CoA
butyryl-CoA dehydrogenase
tadcc30313
butyryl-CoA
Butyrate
ATP
tadcc37665-66
Sporulation genes
Pilus
Flagellum

Fig. S8

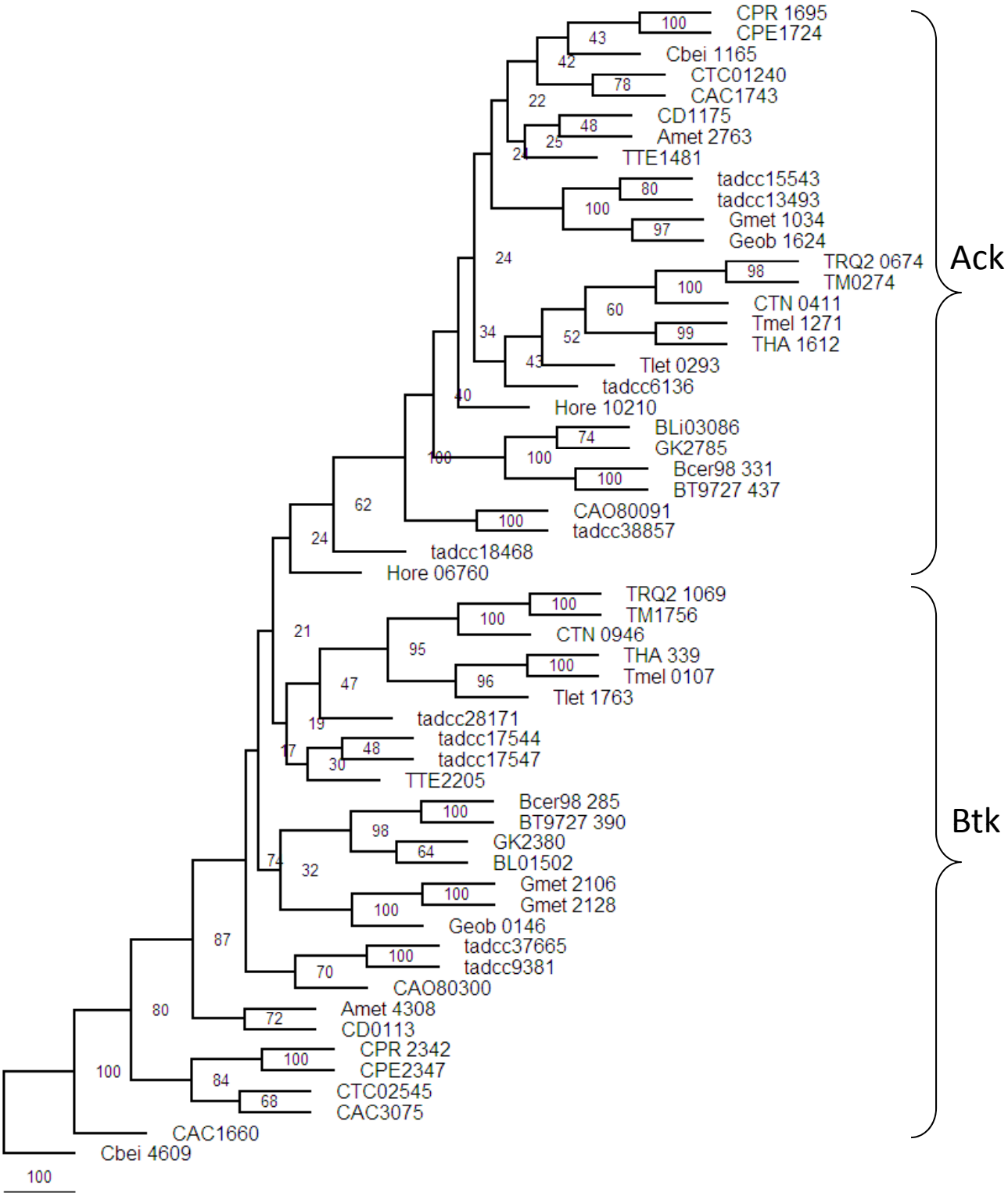
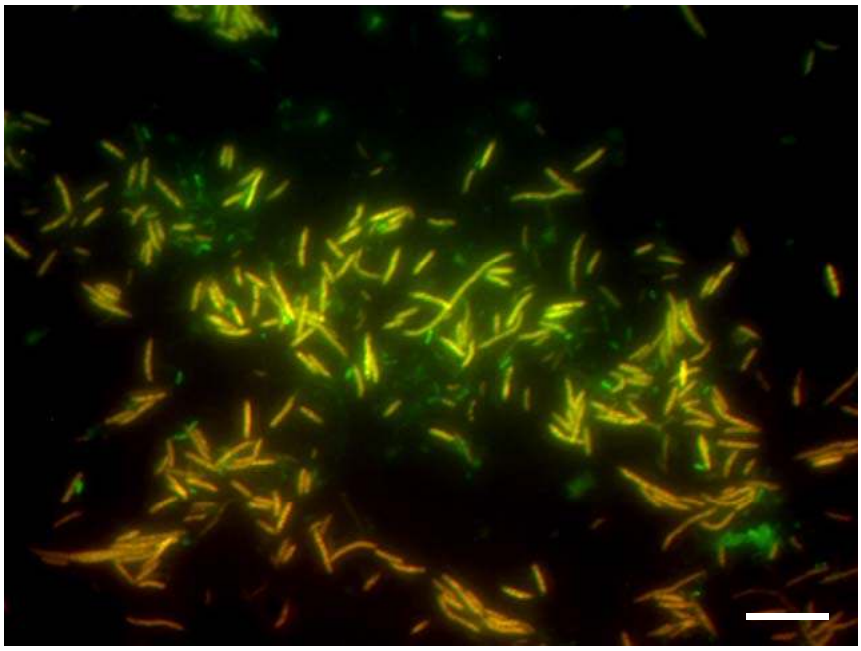
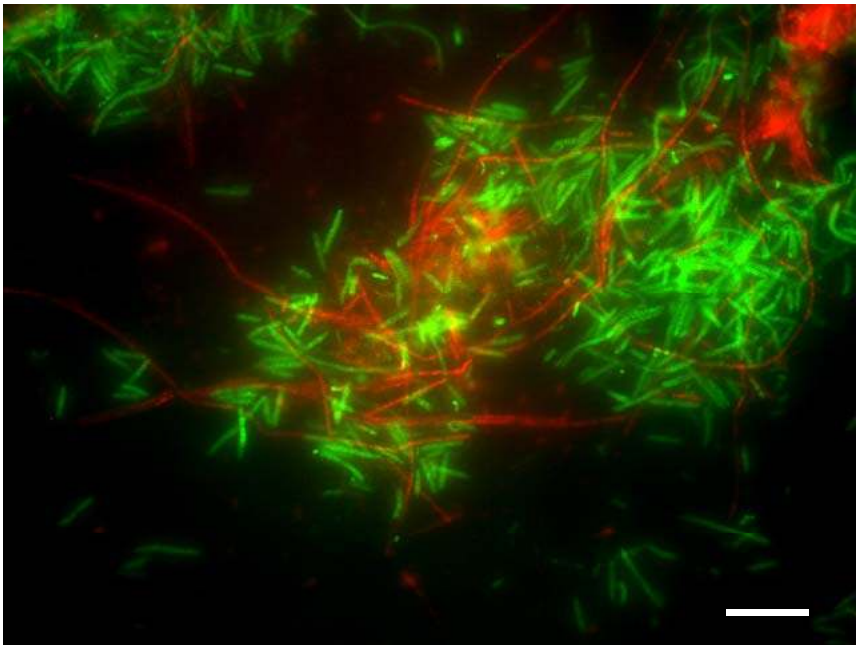

CPR 1695
CPE1724
Cbei 1165
CTC01240
CAC1743
CD1175
Amet 2763
TTE1481
tadcc15543
tadcc13493
Gmet 1034
Geob 1624
TRQ2 0674
TM0274
CTN 0411
Tmel 1271
THA 1612
Tlet 0293
tadcc6136
Hore 10210
BLi03086
GK2785
Bcer98 331
BT9727 437
CAO80091
tadcc38857
tadcc18468
Hore 06760
TRQ2 1069
TM1756
CTN 0946
THA 339
Tmel 0107
Tlet 1763
tadcc28171
tadcc17544
tadcc17547
TTE2205
Bcer98 285
BT9727 390
GK2380
BL01502
Gmet 2106
Gmet 2128
Geob 0146
tadcc37665
tadcc9381
CAO80300
Amet 4308
CD0113
CPR 2342
CPE2347
CTC02545
CAC3075
CAC1660
Cbei 4609
100
Ack
Btk

Fig. S9

**(A) Biofilms (Day 430)**



**(B) Biofilms (Day 430)**



Supplementary Table 1. Genes involved in methanogenesis pathways.

| | | *Methanolinea* | *Methanosaeta* |
|---|---|---|---|
| **formylmethanofuran dehydrogenase** | | | |
| | subunit A | tadcc3594, tadcc39608 | tadcc27005, tadcc35500, tadcc16765 |
| | subunit B | tadcc3593, tadcc39607 | tadcc27002, tadcc25501 |
| | subunit C | tadcc3595, tadcc39609 | tadcc27004, tadcc35499 |
| | subunit D | tadcc3592, tadcc39606 | tadcc27003, tadcc35502 |
| | subunit E | tadcc2822 | tadcc27006, tadcc8453 |
| | subunit F | tadcc3587 | tadcc16469 |
| | subunit G | tadcc3586 | |
| | subunit H | | |
| **formylmethanofuran-tetrahydromethanopterin formyltransferase** | | tadcc3585, tadcc35387 | tadcc6376, tadcc33681 |
| **methenyltetrahydromethanopterin cyclohydrolase** | | tadcc3557, tadcc37277 | tadcc4083, tadcc2370 |
| **methylenetetrahydromethanopterin dehydrogenase** | | tadcc3651 | tadcc14525, tadcc2531 |
| **methylenetetrahydromethanopterin reductase** | | tadcc2844, tadcc3653 | tadcc29239, tadcc4464 |
| **tetrahydromethanopterin S-methyltransferase** | | | |
| | subunit A | tadcc3607 | tadcc28098 |
| | subunit B | tadcc3606 | tadcc28097 |
| | subunit C | tadcc3605 | tadcc28096 |
| | subunit D | tadcc3604 | tadcc14510 |
| | subunit E | tadcc3603 | tadcc14509, tadcc12676 |
| | subunit F | tadcc3608 | tadcc28099 |
| | subunit G | | tadcc28100 |
| | subunit H | | tadcc28101 |
| **methyl-coenzyme M reductase** | | | |
| | α subunit | tadcc3600 | tadcc5364, tadcc6053 |
| | β subunit | tadcc3596 | tadcc5367, tadcc6056 |
| | γ subunit | tadcc3599 | tadcc5365, tadcc6054 |
| | δ subunit | tadcc3597 | tadcc5366, tadcc6055 |
| **acetyl-CoA decarbonylase/synthase** | | | |

| | | | |
|---|---|---|---|
| | α subunit | | tadcc8401, tadcc17916 |
| | β subunit | | tadcc8403, tadcc17919 |
| | γ subunit | | tadcc6386, tadcc12273, tadcc26855 |
| | δ subunit | | tadcc35692 |
| | ε subunit | | tadcc8402, tadcc17918 |
| **ech hydrogenase** | | | |
| | subunit A | tadcc3040 | |
| | subunit B | tadcc3041 | |
| | subunit C | tadcc3042 | |
| | subunit D | tadcc3043 | |
| | subunit E | tadcc3044 | |
| | subunit F | tadcc3045 | |
| **coenzyme F420-reducing hydrogenase** | | | |
| | α subunit | tadcc3698 | |
| | β subunit | tadcc3701 | |
| | γ subunit | tadcc3700 | |
| | δ subunit | tadcc3699 | |
| **Formate dehydrogenase** | | | |
| | α subunit | tadcc20799, tadcc3416 | tadcc34985, tadcc814 |
| | β subunit | tadcc20800, tadcc3417 | tadcc813 |
| **CoB--CoM heterodisulfide reductase** | | | |
| | subunit A | tadcc3590 | tadcc16468 |
| | subunit B | tadcc3589 | |
| | subunit C | tadcc3588 | |
| **CoB-CoM heterodisulfide reductase HdrDE** | | | |
| | subunit D | | tadcc23065 |
| | subunit E | | tadcc23066 |
| **Membrane bound hydrogenase** | | | |
| | subunit mbhA | tadcc17854 | |
| | subunit mbhB | tadcc17855 | |

| | | |
|---|---|---|
| | subunit mbhC | tadcc17856 |
| | subunit mbhD | tadcc17857 |
| | subunit mbhE | tadcc17858 |
| | subunit mbhF | tadcc17859 |
| | subunit mbhG | tadcc17860 |
| | subunit mbhH | tadcc17861 |
| | subunit mbhI | tadcc17862 |
| | subunit mbhJ | tadcc17863 |
| | subunit mbhK | tadcc17864 |
| | subunit mbhL | tadcc17865 |

Supplementary Table 2. Minimum Information about a Genome Sequence

| MIGS ID | Metadata Information Type | Metadata Value |
|---|---|---|
| MIGS 2 | Investigation type | Metagenome |
| MIGS 3 | Project Name | Wastewater Terephthalate-degrading communities from Bioreactor |
| | GOLD Stamp ID | Gm00012, http://genomesonline.org/GOLD_CARDS/Gm00012.html |
| MIGS 1.2 | Trace Archive | - |
| | Habitat Category | Synthetic-Bioreactor |
| | Number of samples | 4 |
| | Project Relevance | Biotechnological, Plastic production, Wastewater treatment |
| MIGS 32 | Relevant SOP | - |
| MIGS 33 | Database | IMG/M, http://img.jgi.doe.gov/m |
| | Contact Name | Wen-Tso Liu |
| | Contact Email | wtliu@ad.uiuc.edu |
| MIGS 29 | Sequencing Method | Sanger |
| MIGS 30 | Assembly Method | PGA |
| MIGS 28.2 | Number of Reads | - |
| | Gene Calling Method | Genemark |
| MIGS 4 | Geographic Location | National University of Singapore |
| MIGS 4.1 | Latitude | 1.29973 |
| MIGS 4.2 | Longitude | 103.771791 |
| MIGS 5 | Isolation Country | Singapore |
| MIGS 5 | Collection Date | ? |

**Supplementary text**

*Reactor operation*

In order to enrich the anaerobic microbial consortium that degrades terephthalate, 1-liter laboratory-scale hybrid bioreactor was filled with approximately 78 packing materials (sera siporax, Sera Germany) in the upper part of the reactor and seeded with sludge obtained from mesophilic upflow anaerobic sludge bed reactors treating dimethylterephthalate (Fig. 1a). The reactor temperature was controlled under 46-50$^{o}$C from Day 1 by recirculating heated water through a double-jacket column. Terephthalate (TA) was used as the sole carbon and energy source in the synthetic wastewater (Angelidaki et al., 1990; Chen et al., 2004). The overall TA loading per day was gradually increased from 0.7 to 3.6 (g TA $l^{-1}$ • $day^{-1}$) by increasing the TA concentration in the influent or decreasing the hydraulic retention time (HRT). During the reactor operation, reactor performance was constantly monitored based on the TA removal efficiency. Terephthalate was measured by a high-performance liquid chromatography (model FCV-10AL, Shimadzu, Japan) equipped with an Eclipse XDB-C18 column (Agilent, USA) and an SPD-M10A UV-detector. In addition, the TA concentration was also measured using spectrophotometer under a UV wavelength of 239 nm. Biomass was collected only from the packing materials on Days 221 and 280, and from both the packing materials and sludge bed on Days 346, and 430 for further microbial analyses. During each sampling point, 10 filters with mature biofilms were retrieved from the reactor and replaced with 10 new filters. This ensured that freshly grown biofilms during a period of 60 to 80 days could be recovered from the last three sampling points.

*DNA extraction and 16S rRNA clone library*

Genomic DNA from those samples taken at different time points was extracted using a protocol described previously (Liu et al., 1997). 16S rRNA gene libraries were generated for each sample taken based on a protocol described previously (Chen et al., 2004). The following primer sets, 27F (5'-AGAGTTTGATCCTGGCTCAG) and 1391R (1391R primer (5'-GACGGGCRGTGWGTRCA)) for bacterial 16S rRNA and 4aF (5'-TCCGGTTGATCCTGCCRG) and 1391R for archaeal 16S rRNA gene amplification (Hugenholtz and Goebel, 2001 ), were used. All libraries were sequenced from both ends and the data were analyzed using ARB software (Ludwig et al., 2004).

*Rarefaction curve*

16S rRNA gene sequences were aligned against the Greengenes database using the NAST alignment tool (DeSantis et al., 2006). The aligned sequences were imported into ARB software. A distance matrix was constructed and this served as the input to DOTUR (Schloss and Handelsman, 2005) for the clustering analysis of sequences into OTUs at defined sequence identity. Based on the OTU analysis, the rarefaction curves were generated.

*Fluorescence in-situ hybridization (FISH)*

FISH were performed on paraformaldehyde-fixed samples (Amann et al., 1995) according to the procedures described previously (Chen et al., 2004). The oligonucleotide probes used included EUBmix (i.e., EUB338, EUB338-II, EUB338-III) that targets most of the *Bacteria* (Amann et al., 1995; Daims et al., 1999), ARCH915 that targets most of *Archaea* (Amann et al., 1995), and specific probes that target different phylogenetic groups, for example, the *Desulfotomaculum* group (Loy et al., 2002; Imachi et al., 2006).

*Metagenome sequencing*

DNA samples taken from the surface of the filters at Days 221, 280, and 430, as well as the sludge bed sample at Day 430, were used. Three whole genome shotgun libraries, containing inserts of ~3, 8 and 40kb, were created for each of the four DNA samples (http://www.jgi.doe.gov/sequencing/protocols/prots_production.html). Initially 10 Mb of sequence from a 3 kb library was generated for each time point. For the Day 430 biofilm sample, an additional 68 Mb were sequenced from the short-insert library in addition to a small amount from 8 kb and 40 kb (fosmid) libraries (5 Mb and 3 Mb, respectively). As the microbial community composition among biofilms samples taken at Day 221, 280 and 430 are similar (Fig. 1a), the sequence data were pooled together with that obtained from the sludge bed sample at Day 430, and assembled using the PGA assembler.

*Metagenome analysis*

The identification of COG families was based on the IMG/M database, using a cut-off of 20% identity and e-value 0.01. Phylogenetic marker COGs were generated manually based on the occurrence of each COG family in finished isolate genomes. Phylogenetic marker COGs constitute a set of 101 COGs that are found once or twice in at least 90% of the complete isolate genomes. We used this COG set to estimate the size of the OP5 genome since there is no any closely related isolate genomes.

For the remaining populations we estimated genome size based on the Phylogenetic Distribution of Genes tool in IMG/M (Markowitz et al., 2008). This tool allows the assessment of the composition of a metagenome based on the distribution of the best BLAST hits on isolate genomes. For example, there are 1735 genes in *Pelotomaculum thermopropionicum* that are best-BLAST hits to genes from the metagenome dataset. Given that *P. thermopropionicum* contains 2920 protein coding genes we estimate that we have covered (1735/2920) x 100 = 59.4% of the

genome. This was a conservative estimate given the expected variation in gene content among members of a genus. However, there was no other better way.

For sequence-composition based binning, a sequence-composition based taxonomic classifier for the phylogenetic binning of the sequence sample was created from sample-specific sequence data and available isolate genome sequences as described before (McHardy et al., 2007). Sample-specific training data for the sample populations were identified based on phylogenetic marker genes, comprising data for an uncultured *Methanomicrobiales* species (2 contigs, 171,430 bp), a *Methanosaeta* species (8 contigs, 73,335 bp) and a species of the candidate phylum OP5 (7 contigs, 48,419 bp). At the ranks of species, genus and order, models were trained with sample-specific sequences and sequenced genomes and combined with higher level models for clades at the ranks of class, phylum and domain, created from available isolate genome sequences. At the species level, the model contains classes for the three sample-specific populations trained with sample-specific data, as well as classes for *Pelotomaculum thermoproprionicum*, "Syntrophus aciditrophicus", and *Candidatus* Cloacamonas acidaminovorans, trained with sequence data from one genome each. At the genus level, the model includes additional classes for *Methanosaeta* and *Geobacter*, trained with data from sequenced genomes and sample-specific data for the uncultured *Methanosaeta* population. At the order level, the model contains classes for the *Methanomicrobiales* and *Methanosarcinales*, trained with data from sequenced genomes and the sample-specific data for the respective populations. Thermotogae was not included as a clade in the model. We had no sample-specific training data for the populations from WWE1 and thus could not model these directly, nor WWE1 directly (which requires data from 2 or more different species, so e.g. one sequenced genome and sample-specific data). We mainly checked for chimeras in the sequences of the

genes discussed in the Results section. We used the SNP BLAST tool of IMG/m that aligns reads to assembled nucleotide sequences to check for misassembled contigs.

The generation of McrA and Ack amino acid sequence alignments was performed using Clustalw (Thomson et al., 1994). The phylogenetic tree of McrA was constructed using Clustalw and the Ack phylogenetic tree was constructed using the Phylip package (http://evolution.genetics.washington.edu/phylip.html).

The present work complies with the minimum information about a genome sequence (MIGS) (Table S2) (Field et al., 2008).

**References**

Amann, R.I., Ludwig, W., and Schleifer, K.-H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews* **59**: 143-169.

Angelidaki, I., Petersen, S.P., and Ahring, B.K. (1990) Effects of lipids on thermophilic anaerobic-digestion and reduction of lipid inhibition upon addition of bentonite. *Applied Microbiology and Biotechnology* **33**: 469-472.

Chen, C.L., Macarie, H., Ramirez, I., Olmos, A., Ong, S.L., Monroy, O., and Liu, W.T. (2004) Microbial community structure in a thermophilic anaerobic hybrid reactor degrading terephthalate. *Microbiology* **150**: 3429-3440.

Daims, H., Bruhl, A., Amann, R., Schleifer, K.H., and Wagner, M. (1999) The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**: 434-444.

DeSantis, T.Z., Hugenholtz, P., Keller, K., Brodie, E.L., Larsen, N., Piceno, Y.M. et al. (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Research* **34**: W394-W399.

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P. et al. (2008) The minimum information about a genome sequence (MIGS) specification. *Nature Biotechnology* **26**: 541-547.

Hugenholtz, P., and Goebel, B.M. (2001 ) The polymerase chain reaction as a tool to investigate microbial diversity in environmental samples. In *Environmental Molecular Microbiology: Protocols and Applications*. Rochelle, P.A. (ed). Norfolk, UK.: Horizon Scientific Press, pp. 31-42.

Imachi, H., Sekiguchi, Y., Kamagata, Y., Loy, A., Qiu, Y.L., Hugenholtz, P. et al. (2006) Non-sulfate-reducing, syntrophic bacteria affiliated with desulfotomaculum cluster I are widely distributed in methanogenic environments. *Appl Environ Microbiol* **72**: 2080-2091.

Liu, W.-T., Marsh, T.L., Cheng, H., and Forney, L.J. (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63**: 4516-4522.

Loy, A., Lehner, A., Lee, N., Adamczyk, J., Meier, H., Ernst, J. et al. (2002) Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl Environ Microbiol* **68**: 5064-5081.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Research* **32**: 1363-1371.

Markowitz, V.M., Ivanova, N.N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D. et al. (2008) IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Research* **36**: D534-D538.

McHardy, A.C., Martin, H.G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007) Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods* **4**: 63-72.

Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology* **71**: 1501-1506.

Thomson, J.D., Higgins, D.G., and Gibson, T.G. (1994) CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.