Supplementary material

Quantitative metaproteomics highlight the metabolic contributions of uncultured phylotypes in a thermophilic anaerobic digester

Live H. Hagen, Jeremy A. Frank, Mirzaman Zamanzadeh, Vincent G.H. Eijsink, Phillip B. Pope, Svein J. Horn*, Magnus Ø. Arntzen

Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences (NMBU), P. O. Box 5003, N-1432 Ås, Norway

*Corresponding author. Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, P. O. Box 5003, N-1432 Ås, Norway. Tel.: + 47 67232488; Fax: + 47 64965901. E-mail address: svein.horn@nmbu.no

Supplementary Figures

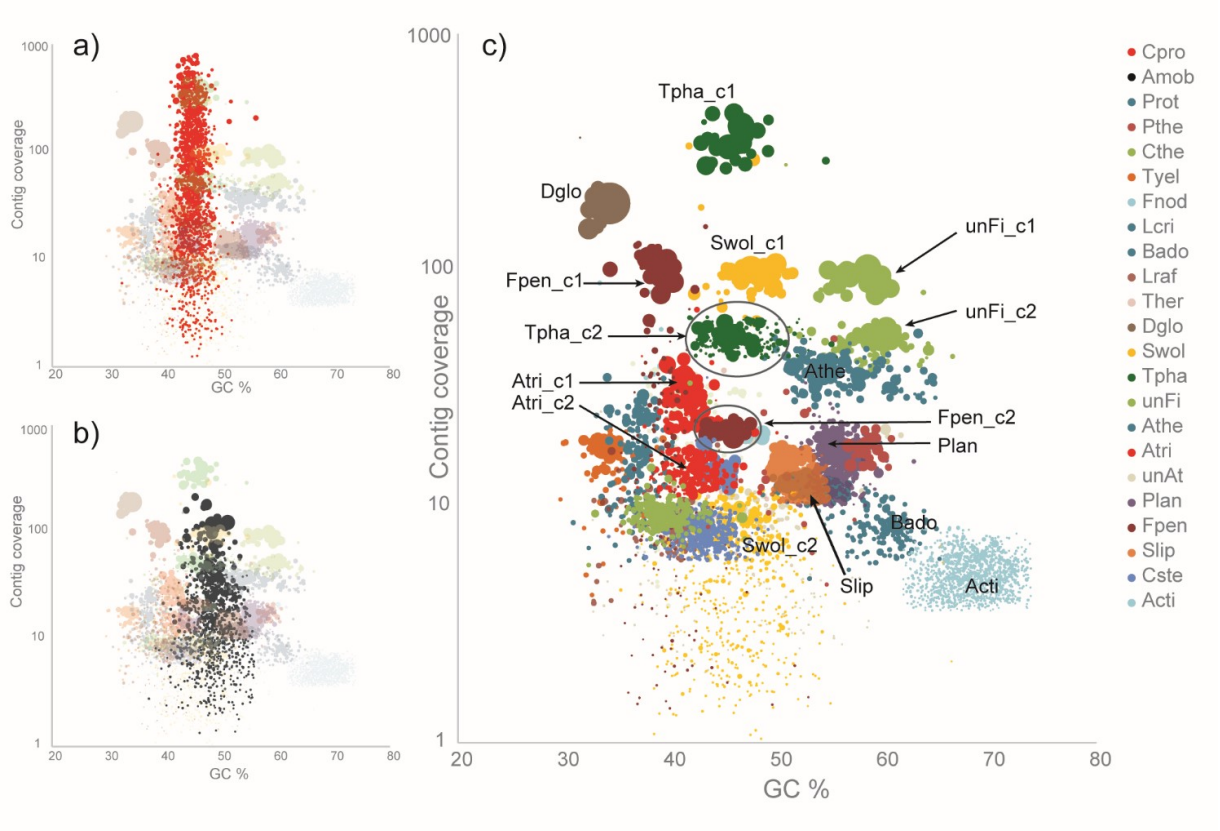


Figure S1: Visualization of PhylopytiaS+/manually refined metagenome binning results. Manual inspection and curation was based on visual clustering of metagenomics contigs in the GC% vs coverage plot, where the population bins are differentiated by color. Visual assessment of PhylopytiaS+-assigned contigs for *Coprothermobacter* spp. (**a**; highlighted in red) and *Anaerobaculum* spp. (**b;** highlighted in black) revealed no clustering based on coverage. These two bins were therefore grouped at a higher-ranked population level. The PhylopytiaS+/manual curation successfully recovered high quality reconstructed genomes for several populations, e.g. unFi_c1 and unFi_c2 (outlined, light green), Tpha_c1 and Tpha_c2 (outlined, dark green), Swol_c1 (outlined, yellow), Atri_c1 and Atri_c2 (outlined, red). For the purposes of clarity, only contigs greater than 5 kb are represented for *Coprothermobacter* spp. (a) and *Anaerobaculum* spp. (b). Contigs greater 1 kb are represented for the remaining population bins (c).

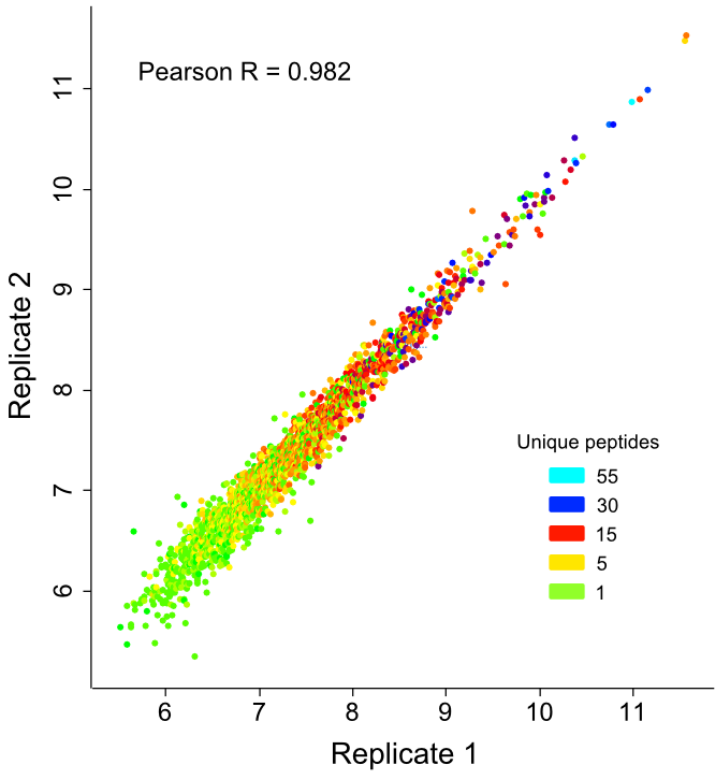


Figure S2: Quality of protein quantification. The figure shows the pairwise reproducibility of label-free quantification (LFQ) between the two replicates (Pearson correlation R=0.982). Proteins quantified in only one replicate were omitted from further analysis and are not shown in this figure.

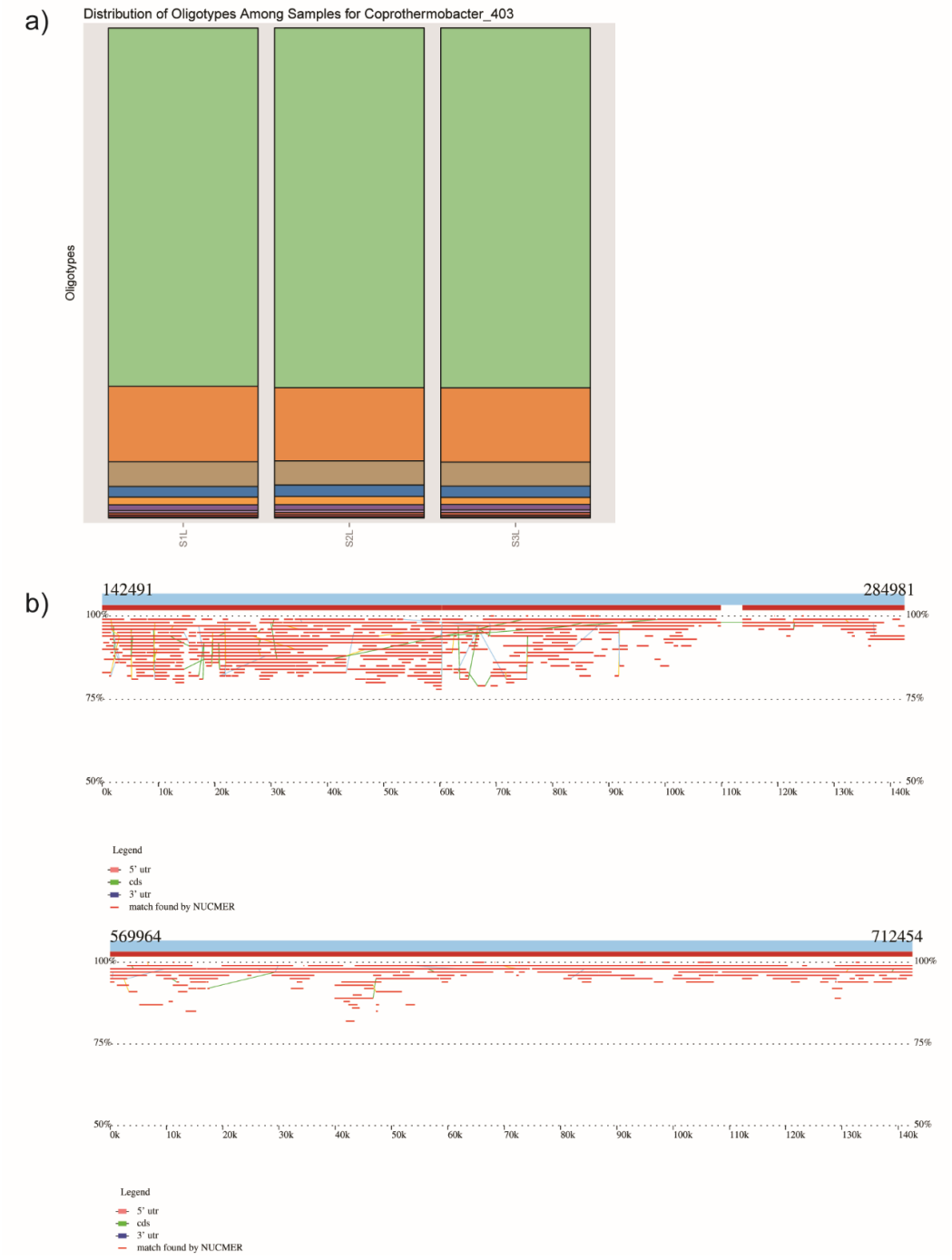
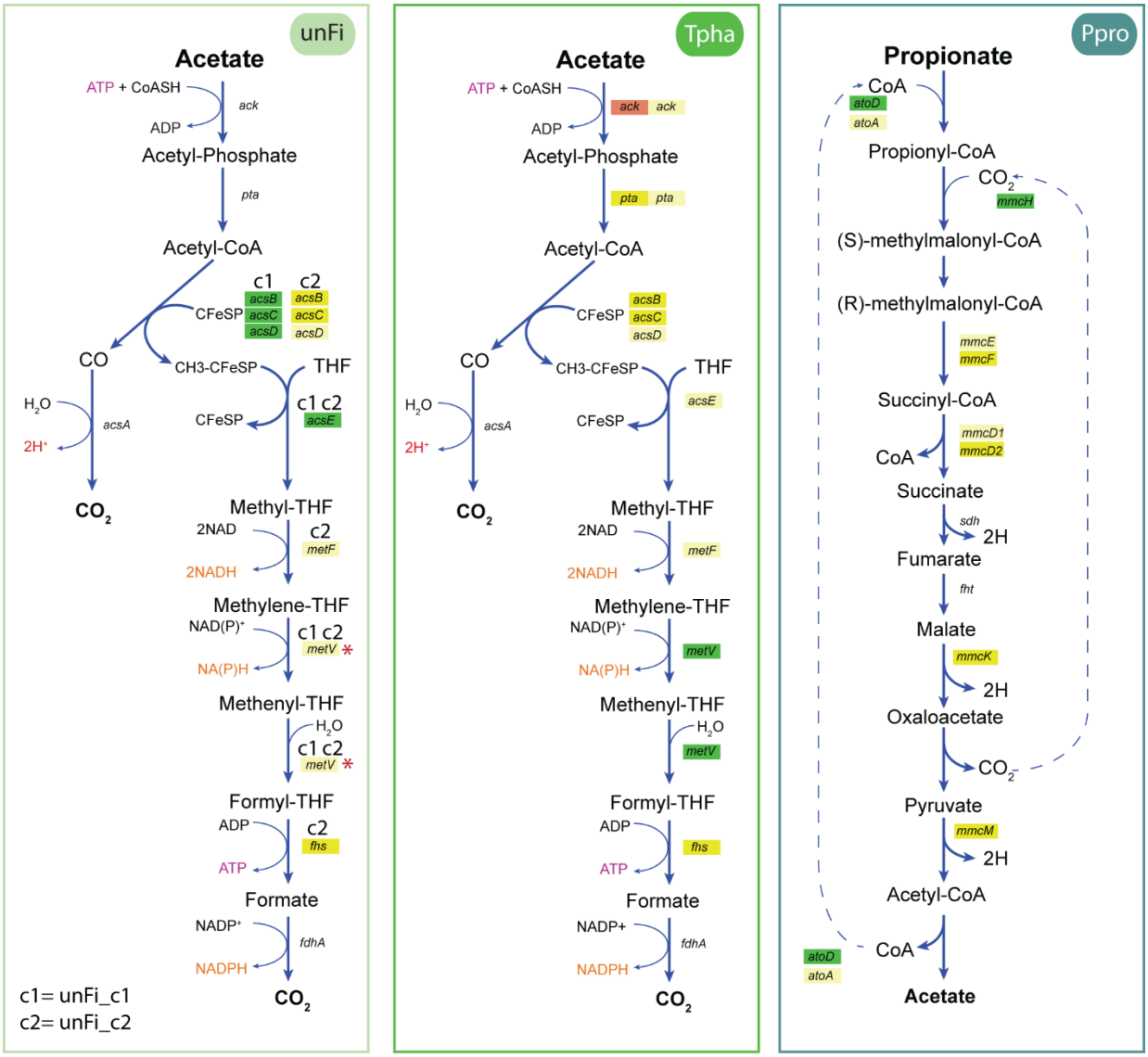


Figure S3 Indication of several strains of *Coprothermobacter proteolyticus* showed by (a) identification of several oligotypes within OTUs assigned to *Coprothermobacter* in the triplicated (S1L, S2L and S3L) 16S rRNA gene sequence data, and (b) alignment (NUCmer) of the draft genome against the representative sequenced genome, *Coprothermobacter proteolyticus* DSM5256.



(Figure continues on next page)

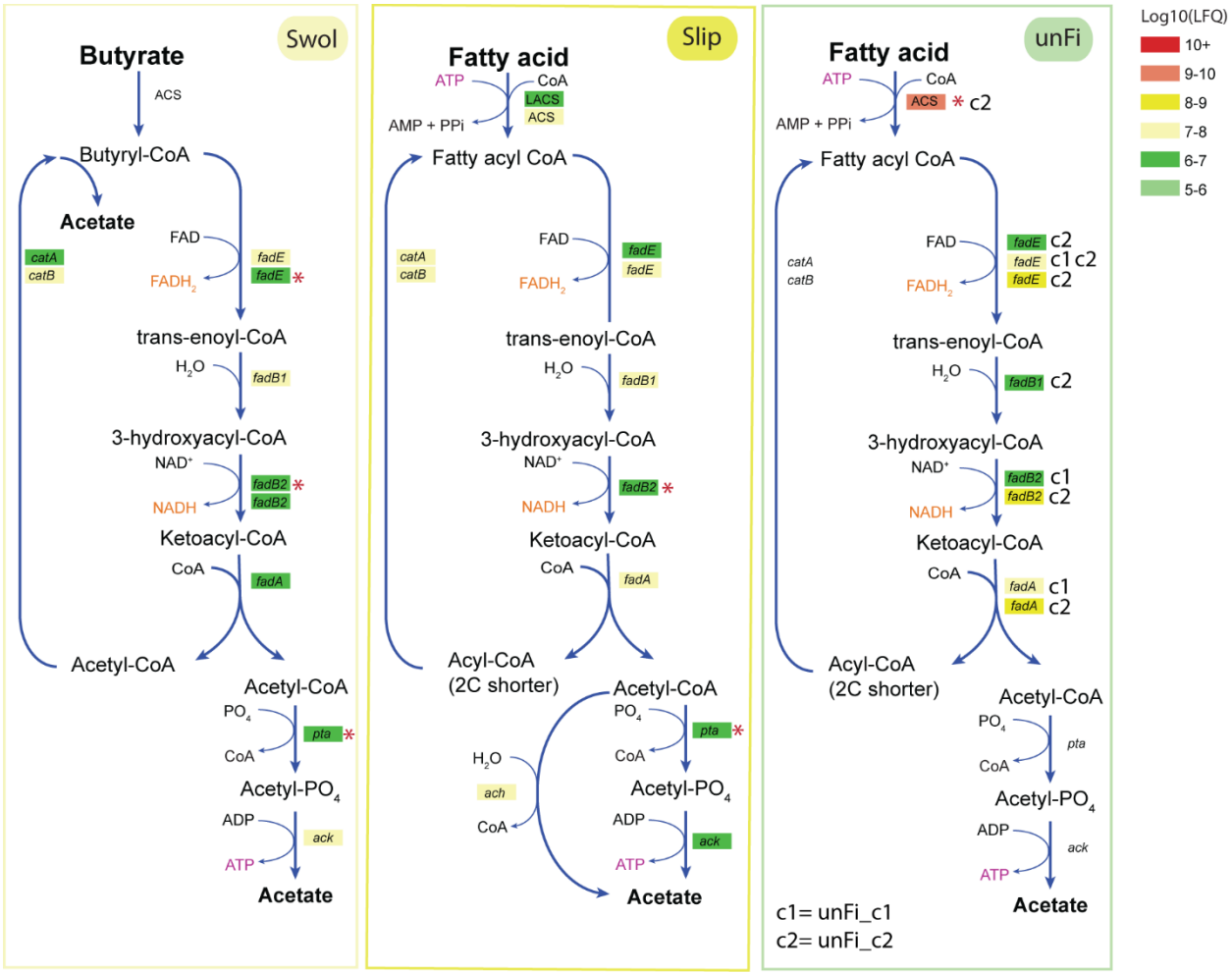


Figure S4 Hypothetical metabolic pathways for syntrophic oxidation of acetate via Wood Ljungdahl pathway (Tpha; *Thermacetogenium phaeum* cluster 1, and unFi; unFirm02_FrBGR cluster 1 and 2, as indicated), propionate via methylmalonyl-CoA pathway (Ppro; *Pelotomaculum thermopropionicum*) and butyrate and longer chain fatty acids via β-oxidation (Slip; *Syntrophothermus* spp, Swol*; Syntrophomonas wolfei* cluster 1 and unFi; unFirm02_FrBGR (cluster 1 and 2, as indicated). The pathways are proposed based on genome and proteome comparison, and protein abundances are indicated by color ranging from high abundant (red) to low abundant (green). Protein abbreviations used in this figure are given in supplementary Table S3. Instances where proteomics could not differentiate which organism the protein originated from are marked with an asterisk (*).

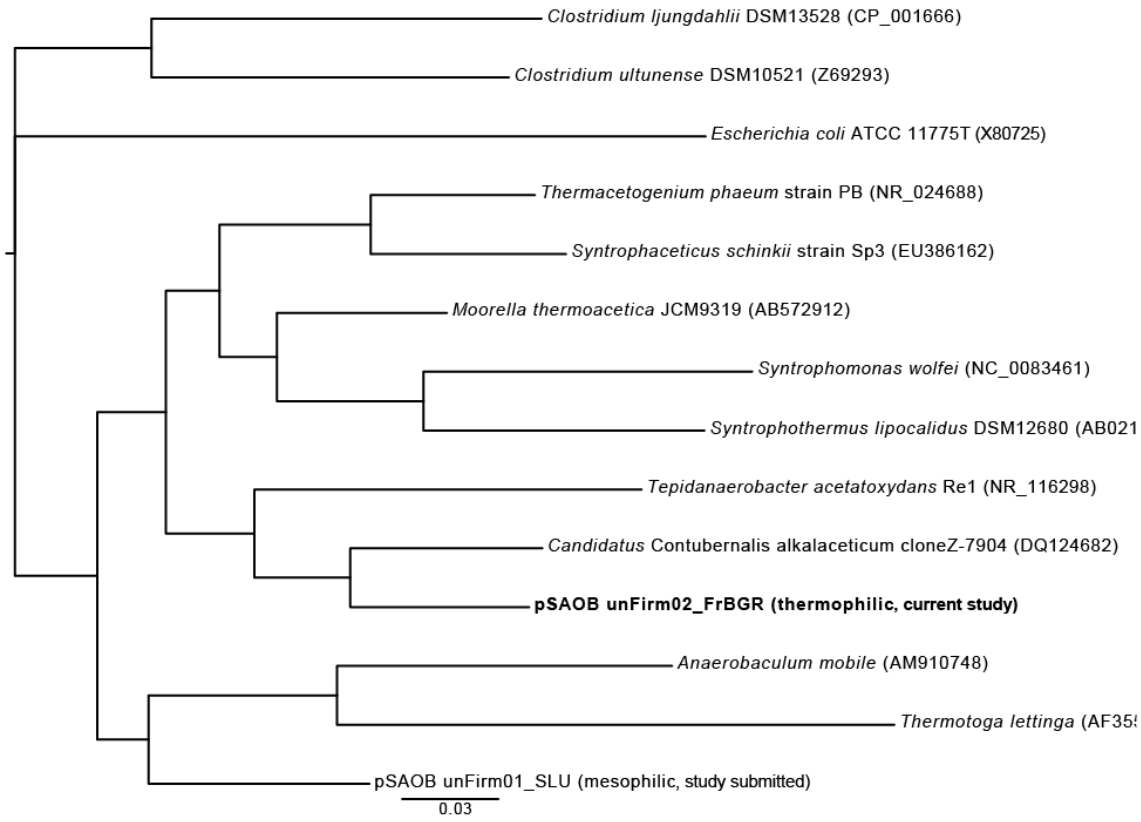


Figure S5 Phylogenetic tree based on 16S rRNA gene sequences highlighting the relationship of the putative novel SAOB unFirm02_FrBGR relative to known SAOBs, selected acetogens and one other potential novel SAOB recently assembled from a similar study of a mesophilic digester (*study submitted*). *Anaerobaculum mobile* and *Escherichia coli* was included as outgroups. The 16S rRNA-based alignment was carried out using MUSCLE, and the phylogenetic three was generated with FastTree.

Supplementary Tables

Table S1. List of reference genomes included in the protein mapping and the interpretation of the results. The reference genomes is only used if specified in the text, otherwise FrBGR genome bins were used.

Genomes	Accession number
Methanosaeta thermophila PT	CP000477.1
Methanothermobacter thermautotrophicus strain Delta H	NC_000916.1
Dictyoglomus thermophilum H-6-12	NC_011297
Pelotomaculum thermopropionicum SI	NC_009454
Tepidanaerobacter acetatoxydans Re1	NC_019954
Clostridium stercorarium subsp. stercorarium DSM 8532	NC_020134

Table S2. Sequencing statistics (yields after quality filtration as described in Materials and Methods)

	MiSeq 16S rRNA (n=3, high quality)	MiSeq Metagenomics (Q > 30)	8 SMRT cells Pac Bio RS (min. accuracy 99.0 %)
Reads (bp)	616 687	54 025 934	217 815
Average read length	300 PE	300 PE	1 253
Sequence information (Mb)	180	16 200	274
Contigs after assembly (n)	-	235 738	
Average contig size (nt)	-	2009	