Supplementary material

Metabolic and trophic interactions modulate the methane production by Arctic peat microbiota in response to warming

Alexander T. Tveit, Tim Urich, Peter Frenzel, Mette M. Svenning

Contents

Sequence analyses. Metatranscriptomic and metagenomic paired end sequence reads were first assembled using Pandaseq (1), with minimum overlap of 10 bp between the sequences and otherwise default settings. Metagenomes were then submitted to the MG-Rast server (2), where all further preprocessing was carried out (recommended filtering: dereplication for the removal of artificially replicated sequences where the first 50 bp are identical, and removal of sequences with >5 ambiguous bases). RNA was preprocessed using Prinseq (3); poly-A/T tails longer than 15 bp were trimmed away, sequences with >5 ambiguous bases were removed, and all but one sequence in pools of exactly identical sequences (same length, exact sequence identity, and exact reverse complement) were removed (dereplication).

Although one is expected to find exact replicates in metatranscriptomes, in contrast to metagenomes, we performed analyses that led us to remove them from the datasets. In total 11.5–16.9 % of the rRNA and 13.7–17.8 % of mRNA were exactly identical sequences. The analysis showed that >70–85 % of all taxonomically assigned mRNA replicates were assigned to only 10 different mRNA, 8 of which encoded hypothetical proteins. The same mRNAs were the most replicated in all metatranscriptomes. Thus, the majority of replicates did not originate from highly expressed genes of dominant community members, but from a few transcripts present at very high numbers. These replicates accounted for up to 99 % of the mRNA reads assigned to their respective genomes. This suggested an artificial origin of a majority of the replicate reads, justifying the removal of exact replicates. Further, since only small numbers of reads were removed from most transcript pools (transcript assigned to the same function and taxa), the removal did not create or remove any temperature related patterns within the data. In any case, most removed reads encoded hypothetical proteins thus not affecting the functional analysis of mRNA. The temperature related patterns in the SSU rRNA was the same after de-replication. We suspect that low input RNA (12.5 ng) in the RNA linear amplification protocol may have caused the observed biases, but we do not know the exact mechanism. The G + C % was variable in the highly replicated sequence reads and the reads did not have low-GC content. High proportion of artificial replicates in low input DNA (10 ng) Illumina metagenomes has been observed (4). However, the mechanism was not identified, and might differ from what we have observed in our metatranscriptomes. Finally, since the same wetlab and bioinformatic protocols were applied to all samples and sequence data, we are confident that the comparative analysis as applied was justified and that the derived conclusions are robust.

Protein-coding transcripts and genes, and rRNA and its genes were separated by comparing the metatranscriptomes/metagenomes against a combined database of small and large subunits of rRNA (SSU and LSU rRNA) using BLASTN (5). Sequences with a bit score <50 were assigned as putative protein-coding sequences; those with a higher bit score were assigned as rRNA or its gene in metatranscriptomes and metagenomes, respectively.

SSU ribo-tags (due to the large size of the datasets, subsets of 500,000 were analyzed further) and SSU rRNA gene tags were taxonomically assigned by MEGAN analysis of a BLASTN file against an SSU rRNA reference database (parameters: minimum bit score 150, minimum support 1, top 2%; 50 best blast hits) (6, 7). The metagenome sequences and the mRNA fraction of the metatranscriptome were functionally annotated for screening using the SEED subsystem classification of the metagenomics (MG)-RAST server with maximum e-value for a significant match set to 1e-4 (2). The putative mRNA of the metatranscriptomes and a subset of the putative protein-coding sequences of the metagenomes (3,000,000) were also taxonomically binned by MEGAN analysis (parameters: minimum bit score 50, minimum support 1, top %: best blast hit only) of BLASTX files against the RefSeq protein database (e-value < 1e-1). Reads taxonomically binned to taxa within the order *Methylococcales* were extracted and analyzed in detail. Specific protein-coding genes and transcripts were screened using custom reference sequence collections from the fungene database (http://fungene.cme.msu.edu/) or generated from the UniProtKB/Swiss-Prot or UniProtKB/TrEMBL databases (June, 2013), using E.C. numbers and gene names as queries for selecting sequence representatives for each function. TBLASTN searches with the sequence collections were carried out with an e-value threshold of 1e-4. All sequences within the threshold were selected as query sequences for a BLASTX search with the same parameters against the RefSeq-protein database. The functional assignments were screened, and only those sequences assigned to the same function as intended in the first round were considered positive. These were extracted from the dataset, and compared again to the Refseq-protein database using an e-value threshold value of 1e-04. The results were uploaded to MEGAN (parameters: minimum bit score 70, minimum support 1, top 2%; 50 best blast hits). The genes and transcripts encoding carbohydrate-active enzymes, including polysaccharide hydrolases, and peptidases were annotated by searching for the respective protein family domains in the sequences. The metagenomic and metatranscriptomic reads were translated into all six frames, each frame into separate ORFs, avoiding any '*' characters marking stop codons in a resulting ORF. All ORFs of 40 amino acids or larger were screened for assignable conserved protein domains. All ORFs were inspected by reference HMMs (Hidden Markov Models) using HMMER tools (hmmsearch) (http://hmmer.janelia.org/) with the Pfam database HMMs (Pfam release 25, http://pfam.janelia.org). All database hits with e-values below a threshold of 10^{-4} were counted. Screening was run on the HPC computer STALLO at the University of Tromsø

(http://docs.notur.no/uit). A selection of sequences was also used as query sequences in BLASTP searches against the RefSeq-protein database of the NCBI. Within the chosen threshold of the HMMER, identified carbohydrate-active enzyme-coding genes and transcripts gave either a corresponding hit or no hit in the RefSeq database (BLASTP, e-value threshold 1e-4). For the taxonomic assignment of gene and transcript sequences, the BLASTP outputs were uploaded in MEGAN (parameters: minimum bit score 50, minimum support 1, top 2%; 50 best blast hits).

1. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD (2012) PANDAseq: PAired-eND Assembler for Illumina sequences. *BMC Bioinform* 13. doi: 10.1186/1471-2105-13-31.
2. Meyer F*,* et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform* 9:8. doi: 10.1186/1471-2105-9-386.
3. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863-864.
4. Solonenko S*, et al.* (2013) Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genomics* 14(1):320.
5. Altschul SF*,* et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17):3389-3402.
6. Urich T*,* et al. (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *Plos One* 3(6):13. doi: 10.1371/journal.pone.0002527.
7. Lanzen A*,* et al. (2011) Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiol Ecol* 77(3):577-589.

SI appendix, section 2

Statistical and quantitative data analyses. Pearson correlation coefficients (r) were calculated from normalized values. Plotting (package:ggplot2), statistical modeling (packages: mgcv and lmPerm), analysis of variance (function: aov), and evaluation of significance (threshold applied: 0.05) and multivariate analyses were performed using the R package (79). Linear and polynomial regression was performed with the lm function. The Ratkowsky model is the square root of CH_4 accumulation rates plotted against temperature (32); the Arrhenius function is the natural logarithm of CH_4 accumulation rate plotted against the reciprocal of the absolute temperature. Generalized additive modeling was used to inspect the plot of the Arrhenius function (function: gam). An effective degree of freedom of 1 is equivalent to a simple linear relationship; larger values indicate non-linearity. Thus, the number of effective degrees of freedom associated with the model fit was used as a criterion for assumptions of linearity or non-linearity. The difference in total nucleic acids (µg) per gram of soil between temperature windows was evaluated by analysis of variance (function: anova). Correspondence analysis (CA) was carried out as described in (80), and contribution biplots were generated as described in (81) (taxonomic level applied in analyses: phylum; SEED subsystem level applied in analyses: level 2). We treated the SSU ribosomal RNA as a measure of living biomass, as discussed in Urich and Schleper, 2011 (1), and not as a measure of activity. The relationships between rRNA, growth and activity are complex and often contradictory, as discussed in Blazewicz et al., 2013 (2). Therefore, we refrain from generally using normalization of rRNA to rDNA for the distinction of abundance and activity. Similarly we did not normalize transcript abundances to the corresponding genes since we were not trying to study the transcriptional regulation of metabolisms in specific populations, but the increase and decreases in relative abundances of transcripts as an indicator of the changes in importance of taxon-assigned metabolisms. CA was applied because it grants a larger impact of low-abundance variables in the analysis than alternative methods. Also, it weights the samples based on the number of reads to ensure that the ordination is not biased by the low variance that is characteristic for small samples. Canonical (constrained) CA (CCA) was carried out for in-depth analysis of temperature responsive taxa (taxonomic level: order; SEED subsystems: functions). The dimensions of the initial analysis were regressed onto the temperature gradient, forcing all inertia explained by temperature into the first dimension of the CA. In the CA space, all taxa/functions for the rRNA, rRNA gene, and mRNA were sorted by the load exerted on the first dimension, from the largest increase with increasing temperature to the largest decrease. For rRNA and its genes, the load for the most responsive taxa on the taxonomic level order was plotted in a XY plot, which identified whether the increase or decrease in relative abundance was in the same direction for both molecules. For the transcripts, the number of hits assigned to each taxa for every metabolic function identified in the key enzyme screening (see Sequence analysis) was divided by the total number of

sequences homologous to any sequence in Refseq-protein database (BLASTx: 1e-1, MEGAN:parameters: minimum bit score 50, minimum support 1, top %: best blast hit only). Circle plots were generated by linear regression of the relative abundance of transcripts across temperatures for selected transcripts assigned to taxa. The three data points within each temperature window (3–5 °C, 14–16 °C, and 24–26 °C) were treated as (pseudo)replicates for the median temperatures (4, 15, and 25 °C). The slope of each regression line served as the template for a linear color-coded scheme, ranging from dark blue (largest observed decrease in relative abundance with increasing temperature) to dark red (largest observed increase). The circle sizes corresponded to the sum of relative abundance across temperatures.

1. R Development Core Team (2009) R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria).
2. Greenacre M (2007) Computation of correspondence analysis. *Correspondence Analysis in Practice,* Interdisciplinary statistics, eds Keiding N, Morgan B, Speed T, van der Heijden P (Chapman and Hall/CRC), 2nd edn, pp 213-259.
3. Greenacre M (2010) Correspondence analysis biplots. *Biplots in practice*, (Fundación BBVA), pp 79-88. http://www.multivariatestatistics.org/biplots.html.
4. Urich T, Schleper C (2011). The "double RNA" approach to simultaneously assess the structure and function of environmental microbial communities by meta-transcriptomics. *Handbook of Molecular Microbial Ecology,* eds de Bruijn F. J. (Wiley-Blackwell:Hoboken, NJ, USA).
5. Blazewicz, et al. (2013) Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses *ISME J* 7, 2061–2068.

Table S1. Number of assembled sequence reads generated from paired-end sequencing of metatranscriptomes and metagenomes. An Illumina HiSeq2000 was used (see Materials and methods for details).

Sample ID	Raw seq	After preprocessing	Average length after preprocessing	Putative protein coding	Functionally annotated (mg-rast $1e^{-4}$)	Subset PFAM analysis	Assigned pfam domain ($1e^{-4}$)	Subset refseq_ protein analysis	Homologous to refseq_protein (bit score 50)	rRNA_gene /molecule	Subset for SSU rRNA analysis	Homologous to SSU rRNA ($1e^{-4}$)	Assigned to taxa SSU rRNA (order level)
3dna	31717455	27748218	160	27687713	10875672	5000000	306419	3000000	1314435	60505	-	14999	10962
4dna	28511584	24728184	157	24677970	8448523	5000000	293934	3000000	1277414	50214	-	13103	9436
5dna	28813611	24681609	160	24630800	7324063	5000000	296038	3000000	1293356	50809	-	13309	9593
14dna	32413933	28376096	161	28320755	5547215	5000000	294680	3000000	1265960	55341	-	14550	10183
15dna	30182258	25376709	157	25327692	8831828	5000000	291653	3000000	1248531	49017	-	12870	8875
16dna	27980209	22770666	167	22725003	6810095	5000000	313208	3000000	1264013	45663	-	11846	7871
24dna	27975349	23775702	156	23728016	8004559	5000000	297777	3000000	1250209	47686	-	12459	8820
25dna	25089093	17555270	165	17519498	6263962	5000000	304853	3000000	1243851	35772	-	9259	6454
26dna	39495971	31269091	160	31207239	8534771	5000000	300949	3000000	1239396	61852	-	15866	11189
3rna	34321242	20683475	163	2695294	191395	-	105574	-	282835	17988181	500000	126183	91027
4rna	30295107	18732709	161	2629463	186225	-	101496	-	278914	16103246	500000	120470	87459
5rna	27689678	16850559	160	3128360	185145	-	96934	-	288313	13722199	500000	90834	66226
14rna	27934836	17113452	160	3639188	192955	-	95649	-	289466	13474264	500000	92049	65245
15rna	25142399	15550682	161	2410387	144192	-	74108	-	216348	13140295	500000	103703	73788
16rna	26359953	16501460	160	2850473	104232	-	88974	-	246293	13650987	500000	93723	65604
24rna	32152575	19515426	159	3526864	186453	-	101112	-	291329	15988562	500000	104921	76715
25rna	17489033	7603945	151	1592199	79985	-	34452	-	101364	6011746	500000	111241	78437
26rna	26863216	11234064	153	2276112	122487	-	52500	-	152409	8957952	500000	96610	66891

Table S2. Metatranscriptome sequence reads assigned to the genome of *Methylobacter tundripaludum*. From the RefSeq_protein database, those sequences that had the best score with those in the annotated genome of *M. tundripaludum* were sorted by assigned name and counted (threshold: bit score 50). Numbers indicate the sum of reads across the temperature windows (e.g., 575 methane monooxygenase subunit A transcripts in the temperature window 3–5° is the sum of reads in libraries from 3, 4 and 5°C) Parentheses: for "Sum of reads assigned to *Methylobacter tundripaludum"*, the fraction in parenthesis is the fraction of all reads homologous to the RefSeq_protein database that was assigned to *M. tundripaludum*; for specific functions: the fraction in parenthesis indicate the fraction of all reads assigned to *M. tundripaludum* that was assigned to that specific function.

Methane/methanol/formaldehyde/formate oxidation	3-5°C	14-16°C	24-26°C
Methane monooxygenase/ammonia monooxygenase, subunit A	575 (1.1e-1)	379 (1.2e-1)	92 (8.9e-02)
Methane monooxygenase/ammonia monooxygenase, subunit B	667 (1.3e-1)	382 (1.2e-1)	129 (1.3e-1)
Methane monooxygenase/ammonia monooxygenase, subunit C	1614 (3.2e-1)	1012 (3.3e-1)	252 (2.4e-1)
PQQ-dependent dehydrogenase, methanol/ethanol family	14 (2.8e-3)	7 (2.3e-3)	3 (2.9e-3)
Aldehyde dehydrogenase (pyrroloquinoline-quinone)	0 (0)	1 (3.2e-4)	2 (1.9e-3)
Cytochrome *c* class I	1 (2.0e-4)	0 (0)	0 (0)
Formaldehyde-activating enzyme	5 (9.9e-4)	1 (3.2e-4)	1 (9.7e-4)
Formate dehydrogenase	10 (2.0e-3)	5 (1.6e-3)	1 (9.7e-4)
Formate dehydrogenase, alpha subunit	11 (2.2e-3)	6 (1.9e-3)	2 (1.9e-3)
Formate dehydrogenase, delta subunit	0 (0)	1 (3.2e-4)	0 (0)
Sum of reads assigned to *Methylobacter tundripaludum*	5055 (5.9e-3)	3099 (4.1e-3)	1032 (1.9e-3)
Sum total of all reads assigned to microbial taxa	850062	752107	545102

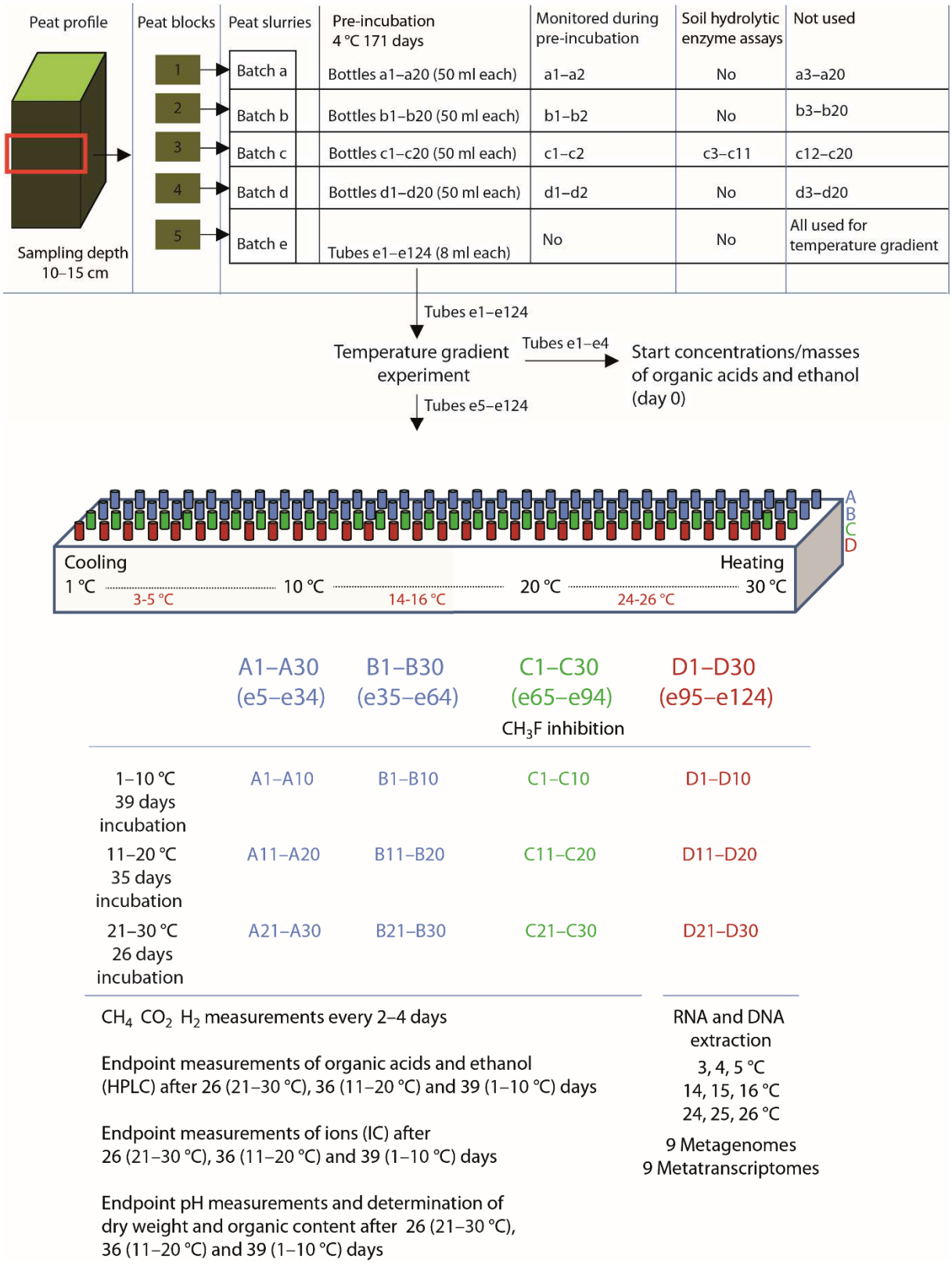

Fig. S1. Experimental set-up. Note: all peat blocks were sampled at 10–15 cm and in close proximity.

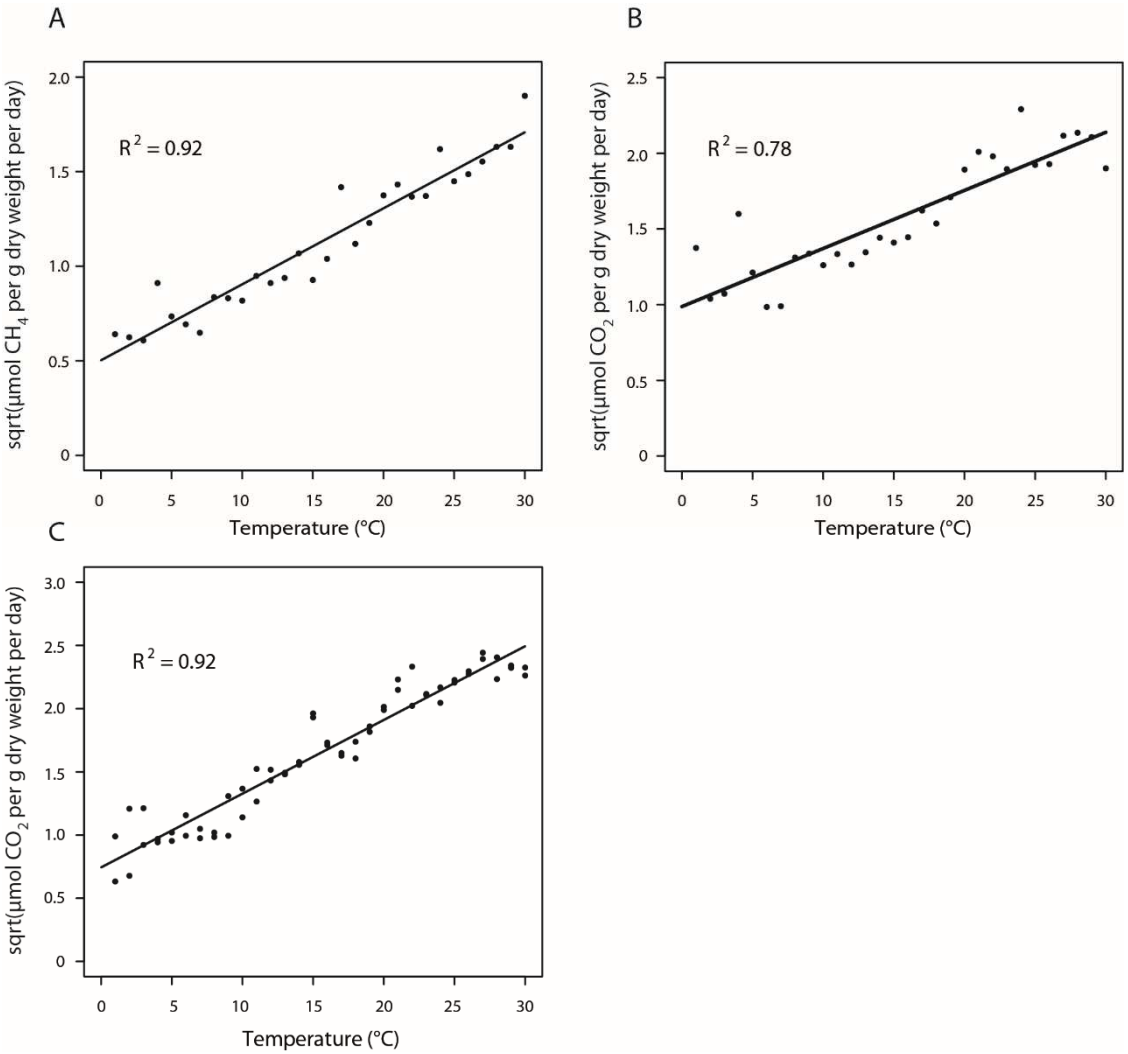

Fig. S2. Linear regression of the plotted Ratkowsky function (square root of the CH_4 or CO_2 production rate vs. temperature). CH_4 and CO_2 production under inhibition with CH_3F, and CO_2 production without inhibition. (A) The square root of CH_4 production under CH_3F inhibition. (B) Square root of CO_2 production under CH_3F inhibition. (C) Square root of CO_2 production without inhibition for two replicate temperature gradients.

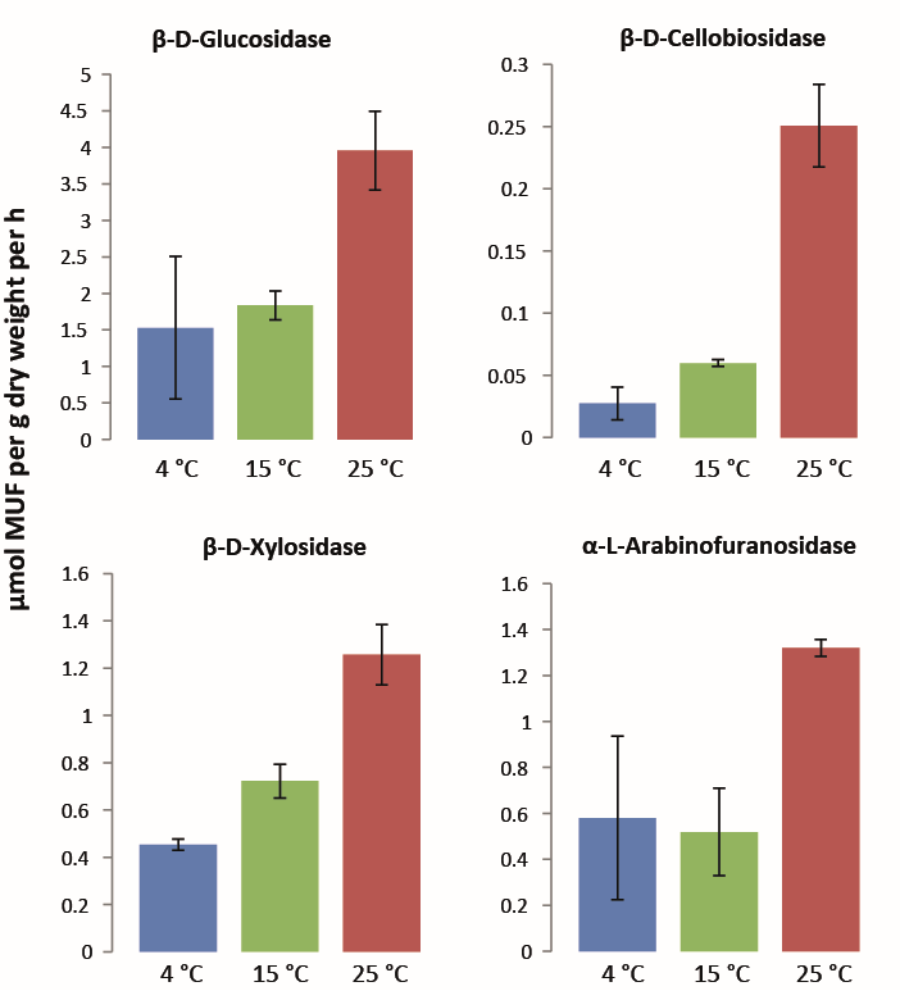

Fig. S3. Potential activity (V_{max}) at 4, 15, and 25 °C of enzymes involved in polysaccharide hydrolysis. The enzymes catalyze the hydrolysis of methylumbelliferone (MUF)-labeled β-D-glucopyranoside, β-D-cellobioside, β-D-xylopyranoside, and α-L-arabinofuranoside. The enzymes catalyzing these reactions are β-D-glucosidase, β-D-cellobiosidase, β-D-xylosidase, and α-L-arabinofuranosidase, respectively. These are involved in the degradation of cellulose (first two), hemicellulose, and hemicellulose branches, respectively. Error bars show standard deviation of biological duplicates.

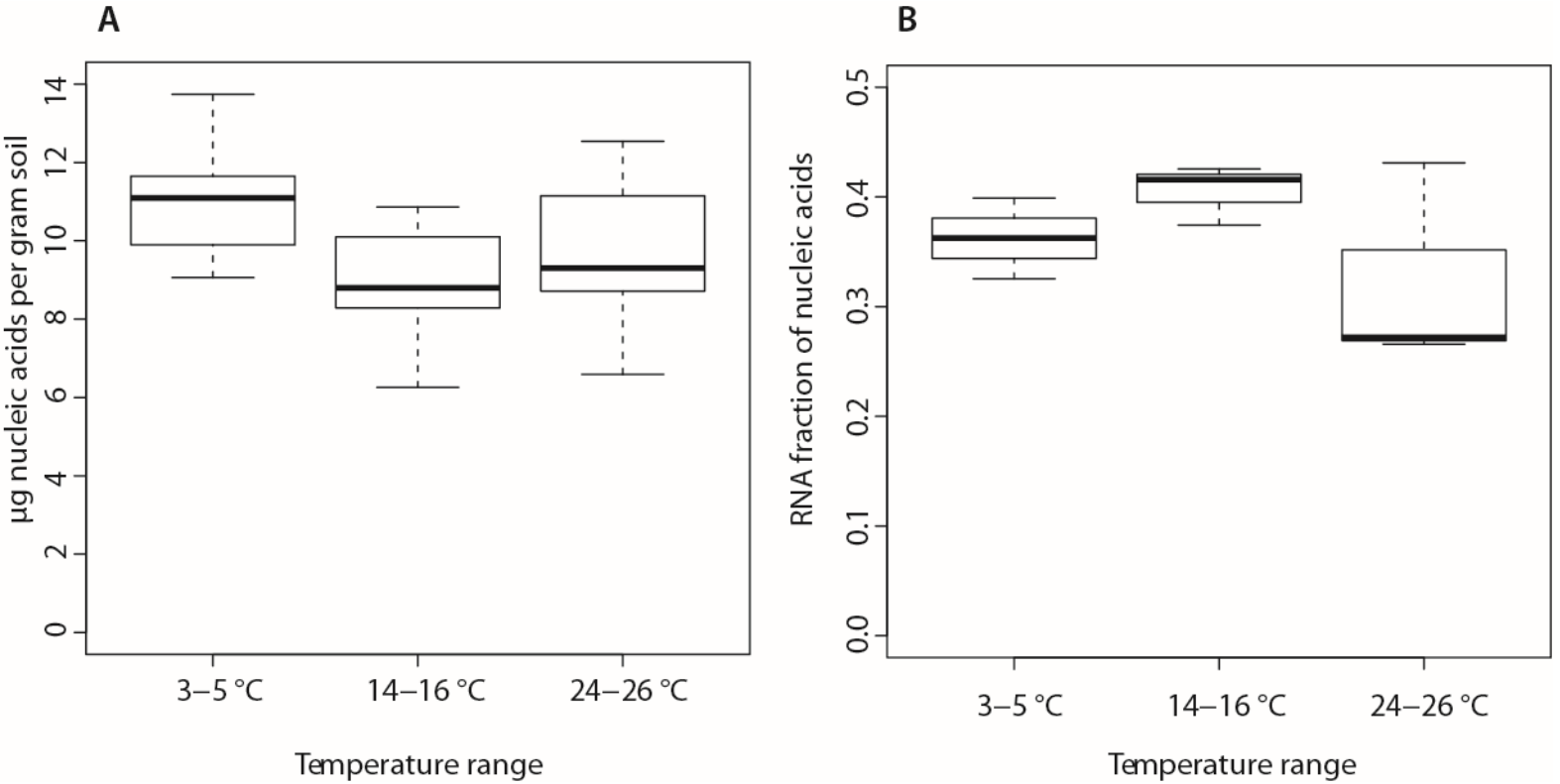

Fig. S4.Temperature dependence of (A) soil nucleic acids content and (B) RNA fraction of nucleic acids. (A) Nucleic acid yield per gram of soil and nucleic acids at the temperature windows of 3–5 °C (n = 19), 14–16 °C (n = 15), and 24–26 °C (n = 12). (B) RNA fraction of nucleic acids at the temperature windows of 3-5 °C (n =3), 14–16 °C (n = 3), and 24–26 °C (n = 3). All values within each temperature window were pooled and plotted to visualize the variation in extraction yield and RNA fraction along the temperature gradient. Analysis of variance for (A) showed significantly (p-value = 0.0009) more variation between temperature windows than within temperature windows, with the difference being between temperature window 3–5 °C and the two other windows. Analysis of variance for (B) indicated no significant differences.

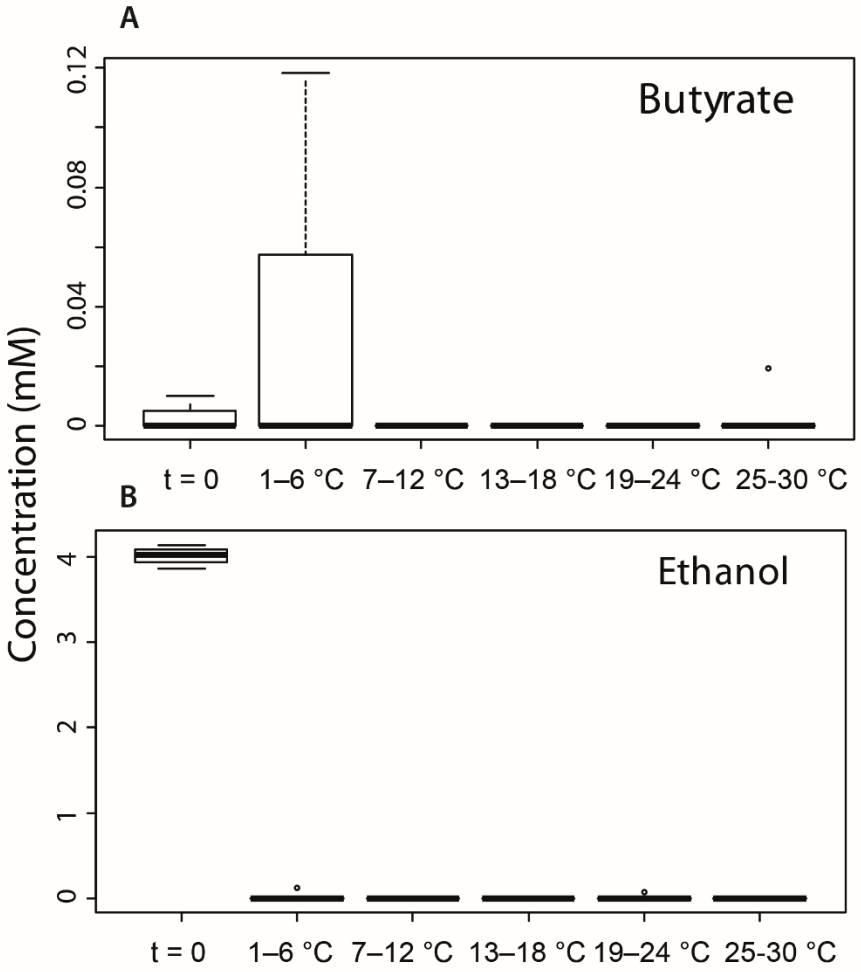

Fig. S5. Concentration of butyrate and ethanol in temperature-gradient experiments. Boxplots of (A) butyrate and (B) ethanol concentrations (mM) in soils incubated at temperatures ranging from 1 to 30 °C. The day zero concentrations of the experiment are indicated by t = 0. All other measurements are end-point. Measurements from the temperature gradient series A and B are shown. Whiskers indicate the most extreme values within 1.5 multiplied by the interquartile range; values outside that range are represented by circles. Box: 25% quartile; median, 75% quartile.

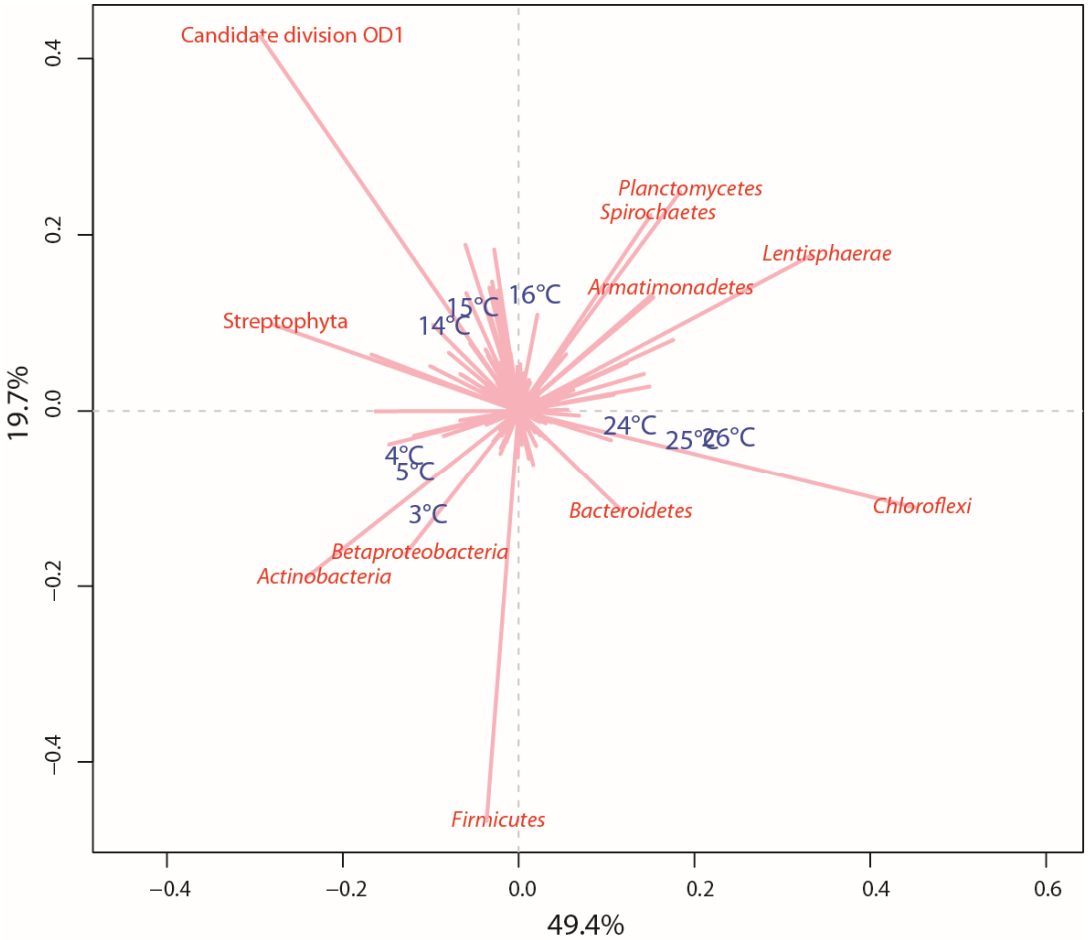

Fig. S6. Correspondence analysis of taxa abundances based on SSU rRNA genes. Sample profiles are shown in in blue; taxa are shown in red. The x-axis represents the first dimension, which explained 49.4% of the inertia; the y-axis represents the second dimension, which explained 19.7% of the inertia. The distance between the sample profiles corresponds to the difference in taxonomic composition. The length of the line connecting the taxon to the center of the plot is equivalent to the weight of this taxon profile in the final solution of the samples, given the inertia explained by the dimensions (i.e., the longer the line, the larger part of the inertia it explains). The direction of the lines indicates the sample orientation of its weight (e.g., if the line points towards higher temperatures, the relative abundance of that taxon is highest at high temperatures).

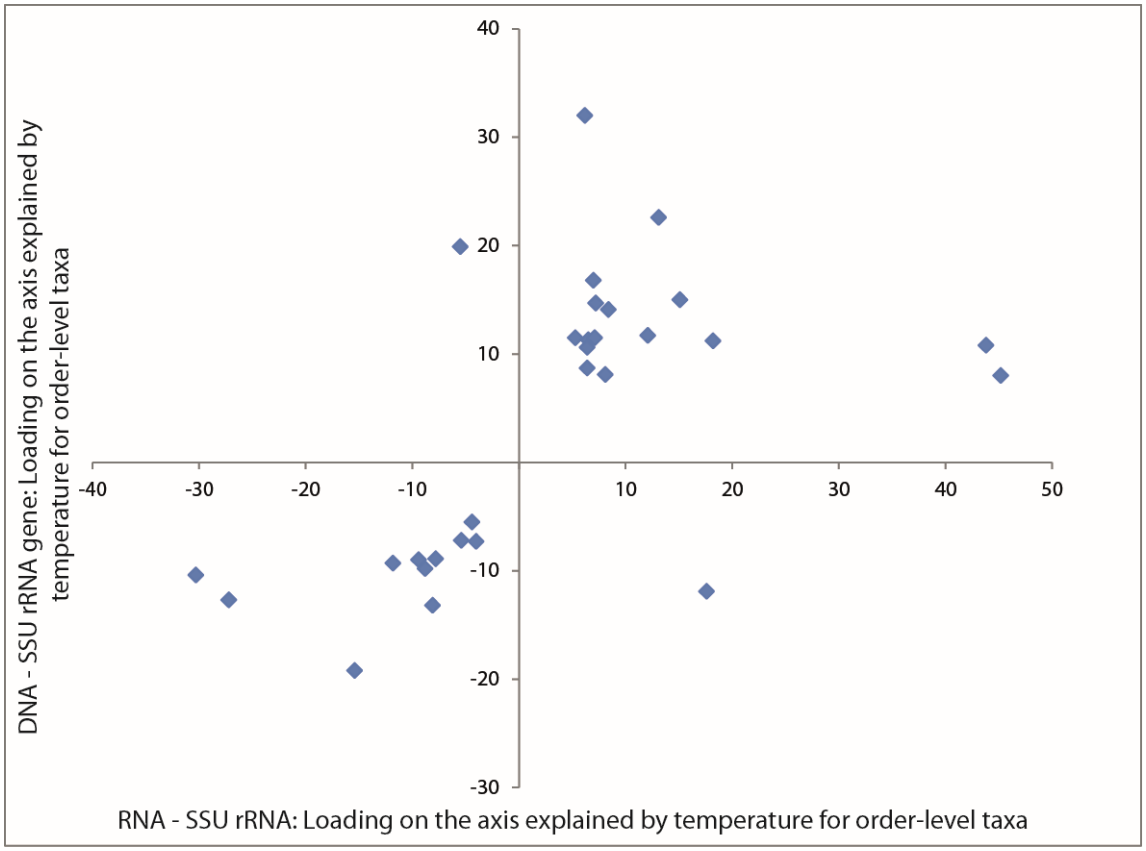

Fig. S7. Corresponding changes with temperature in relative abundance of SSU rRNA and its gene of microbial orders. XY plot of the loading on the correspondence analysis (CA) axis constrained to temperature for DNA (x-axis) and RNA (y-axis). Positive loading means that the relative abundance for that order increases with temperature; negative loading means that the relative abundance for that order decreases with temperature. Since these are two different datasets, and the actual loading is not comparable, correlation analysis has not been performed. Thus, the important information is that for all but two orders, the change in relative abundance with temperature is in the same direction for both the SSU rRNA and its gene.

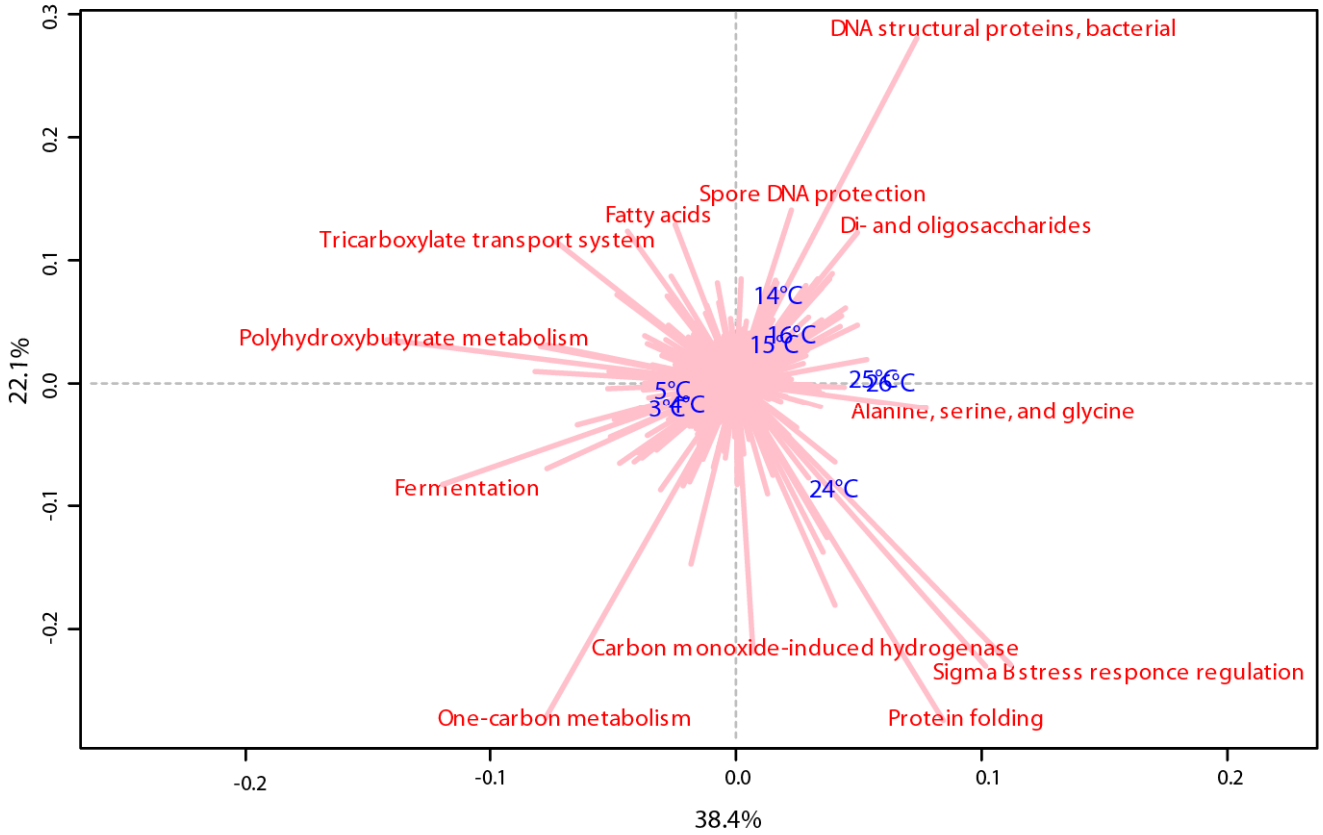

Fig. S8. Correspondence analysis of functional category transcript abundances (SEED; subsystem level 2). The abundances correspond to the number of transcript sequences assigned to each functional category. Sample profiles are shown in blue; functional categories are shown in red. The x-axis represents the first dimension, which explained 38.4% of the inertia; the y-axis represents the second dimension, which explained 22.1% of the inertia. The distance between the sample profiles indicates the difference in the relative abundance of transcripts within functional categories. The length of the line connecting each taxon to the center of the plot is equivalent to the weight of this taxon profile in the final solution of the samples, given the inertia explained by the dimensions (i.e., the longer the line, the larger part of the inertia it explains). The direction of the lines indicates the sample orientation of its weight (e.g., if a line points towards higher temperature, the relative abundance of that taxon is highest at high temperature).

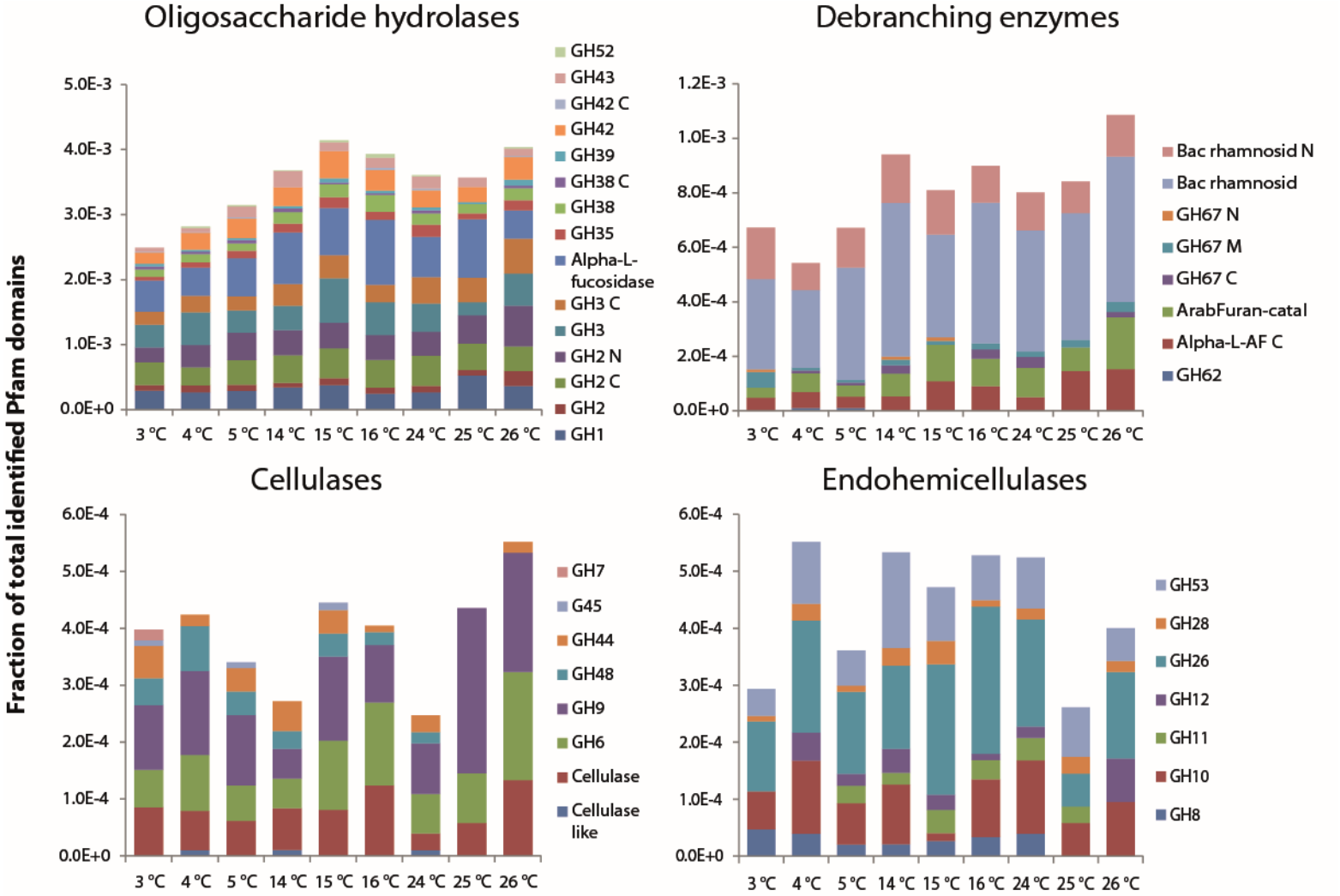

Fig. S9. Relative abundance of transcripts encoding oligosaccharide hydrolases, cellulases, endohemicellulases, and debranching enzymes. Protein family domains were identified by searches in the protein family database Pfam (see Materials and methods). Each name given for the color code represents a Pfam domain. Specific enzymatic functions associated with glycoside hydrolases (GH) and other Pfam domains can be found in the Pfam (http://pfam.sanger.ac.uk/) and CAZy (carbohydrate active enzymes) (http://www.cazy.org/) databases.

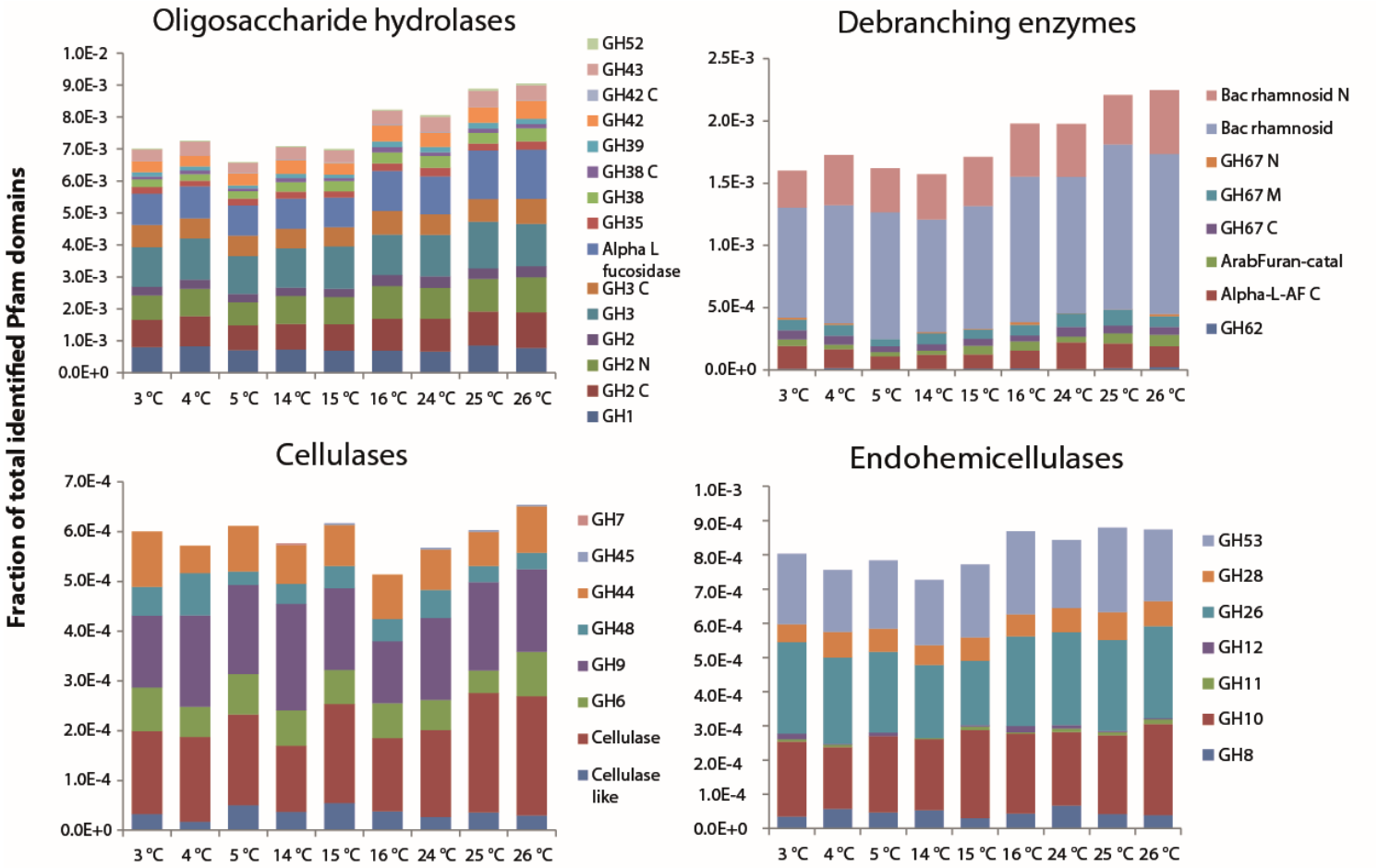

Fig. S10. Relative abundance of genes encoding oligosaccharide hydrolases, cellulases, endohemicellulases, and debranching enzymes. The respective protein family domains were identified by searches in the protein family database Pfam (see Materials and methods). Each name given for the color code represents a Pfam domain. Specific enzymatic functions associated with glycoside hydrolases (GH) and other Pfam domains can be found in the Pfam (http://pfam.sanger.ac.uk/) and CAZy (carbohydrate active enzymes) (http://www.cazy.org/) databases.

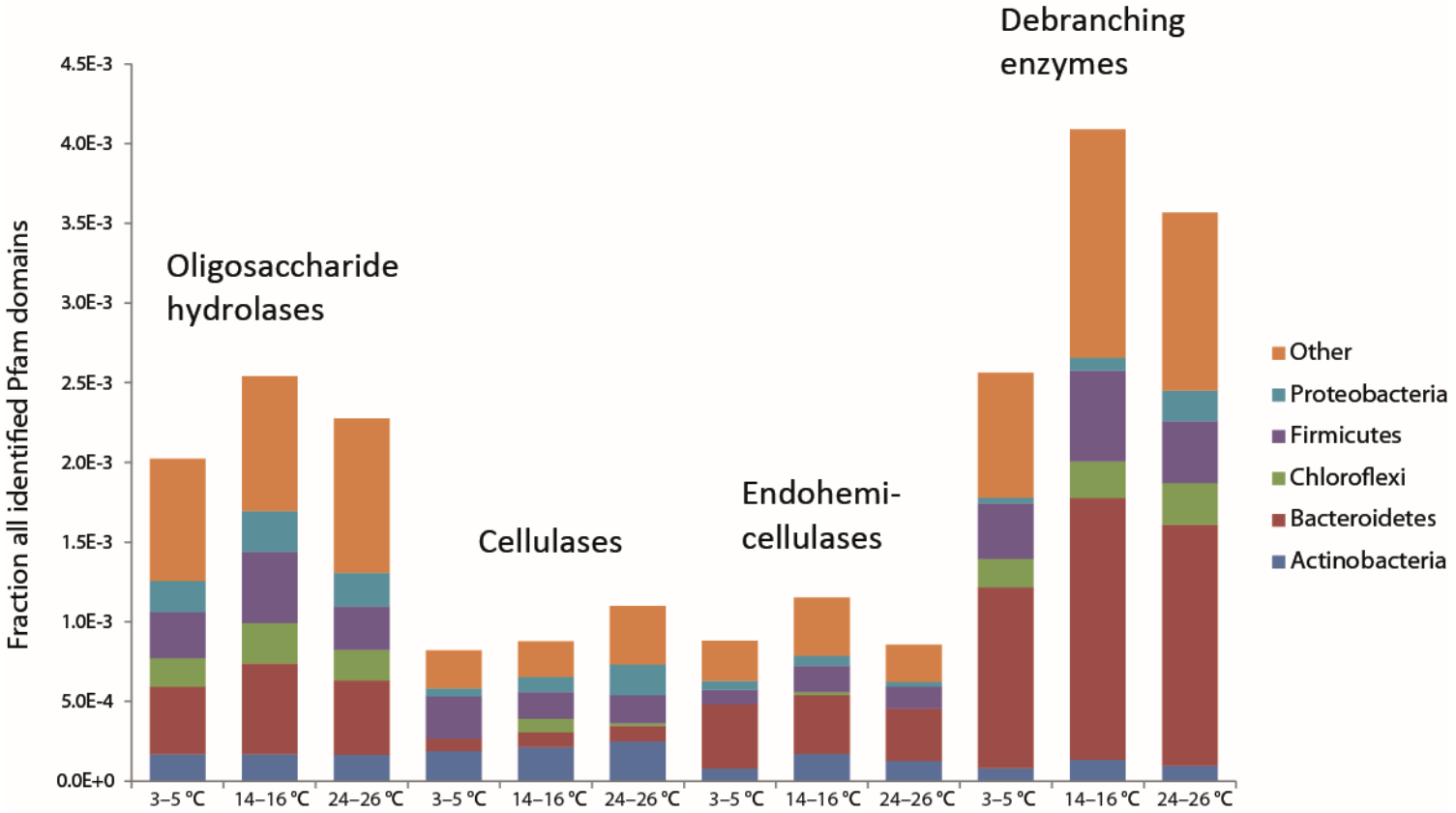

Fig. S11. Taxonomic assignment of transcripts encoding oligosaccharide hydrolases, cellulases, endohemicellulases, and debranching enzymes. For each temperature window, the average relative abundance was calculated at each temperature range. Protein family domains were identified by searches in the Pfam database. Sequences containing identified domains were taxonomically assigned using BLASTP (see Materials and methods).

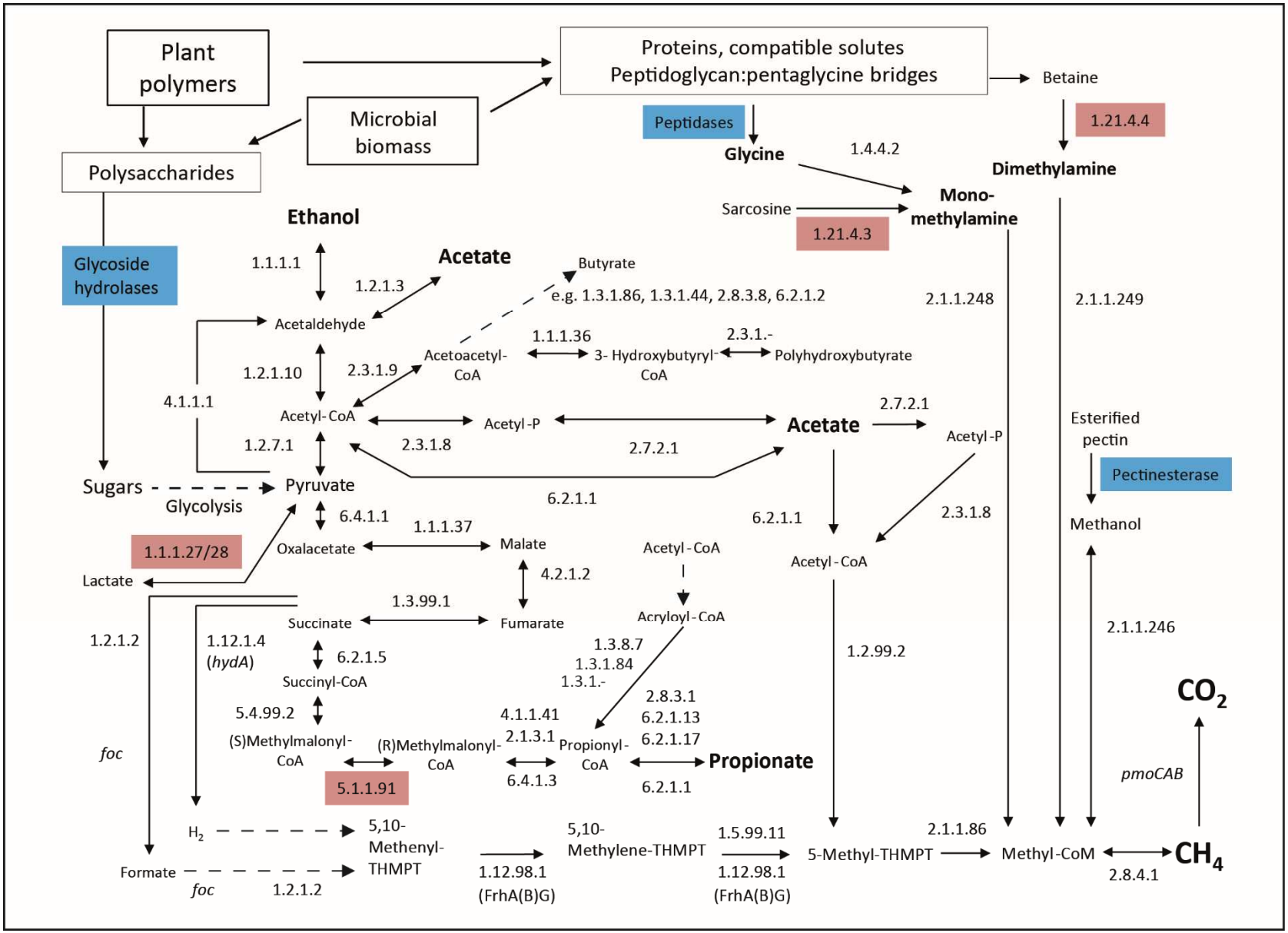
Plant polymers
Microbial biomass
Polysaccharides
Proteins, compatible solutes
Peptidoglycan:pentaglycine bridges
Peptidases
Betaine
1.21.4.4
Glycine
1.4.4.2
Dimethylamine
Sarcosine
1.21.4.3
Mono-methylamine
Glycoside hydrolases
Ethanol
1.1.1.1
Acetate
1.2.1.3
Butyrate
e.g. 1.3.1.86, 1.3.1.44, 2.8.3.8, 6.2.1.2
2.1.1.248
2.1.1.249
Acetaldehyde
1.1.1.36
2.3.1.-
Acetoacetyl-CoA
3- Hydroxybutyryl-CoA
Polyhydroxybutyrate
1.2.1.10
2.3.1.9
4.1.1.1
Acetyl-CoA
Acetyl-P
Acetate
2.7.2.1
Esterified pectin
1.2.7.1
2.3.1.8
2.7.2.1
Acetyl-P
Sugars
Glycolysis
Pyruvate
6.2.1.1
Pectinesterase
6.4.1.1
1.1.1.37
6.2.1.1
2.3.1.8
Methanol
1.1.1.27/28
Oxalacetate
Malate
Acetyl-CoA
Acetyl-CoA
Lactate
4.2.1.2
1.3.99.1
Succinate
Fumarate
Acryloyl-CoA
2.1.1.246
1.2.1.2
1.12.1.4 (hydA)
6.2.1.5
1.3.8.7
1.3.1.84
1.3.1.-
1.2.99.2
Succinyl-CoA
CO_2
5.4.99.2
2.8.3.1
6.2.1.13
6.2.1.17
4.1.1.41
2.1.3.1
foc
(S)Methylmalonyl-CoA
(R)Methylmalonyl-CoA
Propionyl-CoA
Propionate
pmoCAB
5.1.1.91
6.4.1.3
6.2.1.1
H_2
5,10-Methenyl-THMPT
5,10-Methylene-THMPT
1.5.99.11
2.1.1.86
Formate
foc
1.2.1.2
1.12.98.1 (FrhA(B)G)
1.12.98.1 (FrhA(B)G)
5-Methyl-THMPT
Methyl-CoM
CH_4
2.8.4.1

Fig. S12. Major functions in SOC decomposition analyzed using metagenomic and metatranscriptomic data. Formyltetrahydrofolate synthetase (fhs) is not included. Abbreviations: THMPT: tetrahydromethanopterin (THS(sarcina)PT within Methanosarcinales); CoA: coenzyme A. Italics: gene names; blue boxes: analyzed with hidden Markov models and the Pfam database; red boxes: only analyzed from the relative abundance of SEED functions; no box or color: SEED and custom key enzyme search; dashed arrows: pathway steps omitted; *foc:* formate channel; *pmoCAB*: particulate methane monoxygenase subunits c, a, and b. Numbers indicate enzyme categories (E.C.): 1.2.1.2: formate dehydrogenase; 1.12.1.4 (*hydA*): [FeFe] hydrogenase subunit a; 2.3.1.8: phosphate acetyltransferase; 2.7.2.1: acetate kinase; 2.8.3.1: propionate-CoA transferase; 6.2.1.17: propionate-CoA ligase; 6.2.1.13: acetate-CoA ligase (ADP forming); 6.2.1.1: acetate-CoA ligase (AMP forming); 6.4.1.3: propionyl-CoA carboxylase; 2.1.3.1: methylmalonyl-CoA carboxytransferase; 4.1.1.41: methylmalonyl CoA decarboxylase; 5.4.99.2: methylmalonyl-CoA mutase; 6.2.1.5: succinate-CoA ligase (ADP forming); 1.3.99.1: succinate dehydrogenase; 4.2.1.2: fumarate hydratase; malate dehydrogenase; 6.4.1.1: pyruvate carboxylase; 1.2.7.1: pyruvate synthase; 1.2.1.3: aldehyde dehydrogenase (NAD^+); 1.2.1.10: acetaldehyde dehydrogenase (acetylating); 1.1.1.1: alcohol dehydrogenase; 1.4.4.2: glycine dehydrogenase (aminomethyl transferring); 2.3.1.-: polyhydroxybutyrate polymerase; 2.3.1.9: acetyl-CoA acetyltransferase; 1.1.1.36: acetoacetyl-CoA reductase; 1.2.99.2: carbon monoxide dehydrogenase; 1.5.99.11: 5,10-methylenetetrahydromethanopterin reductase; 1.12.98.1: coenzyme F420 hydrogenase alpha and gamma subunits; 2.1.1.249: dimethylamine methyltransferase; 2.1.1.248: monomethylamine methyltransferase; 2.1.1.246: methanol methyltransferase; 1.21.4.3: sarcosine reductase; 1.21.4.4: betaine reductase; 2.1.1.86: tetrahydromethanopterin S-methyltransferase; 2.8.4.1: methyl-CoM reductase.

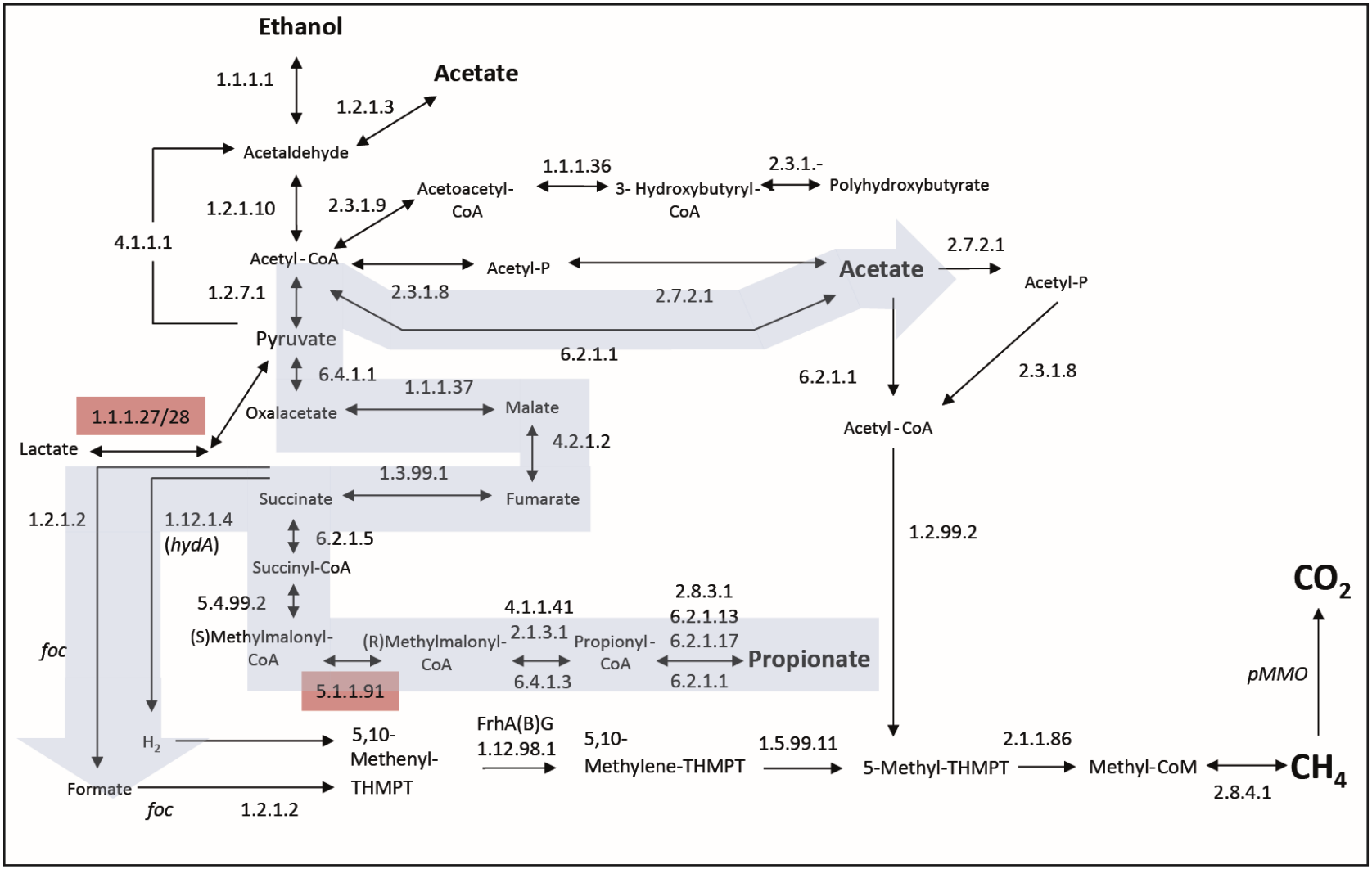

Fig. S13. Proposed pathway for propionate fermentation to acetate, formate, and H_2, carried out by *Firmicutes*. See Fig. S12 for details.

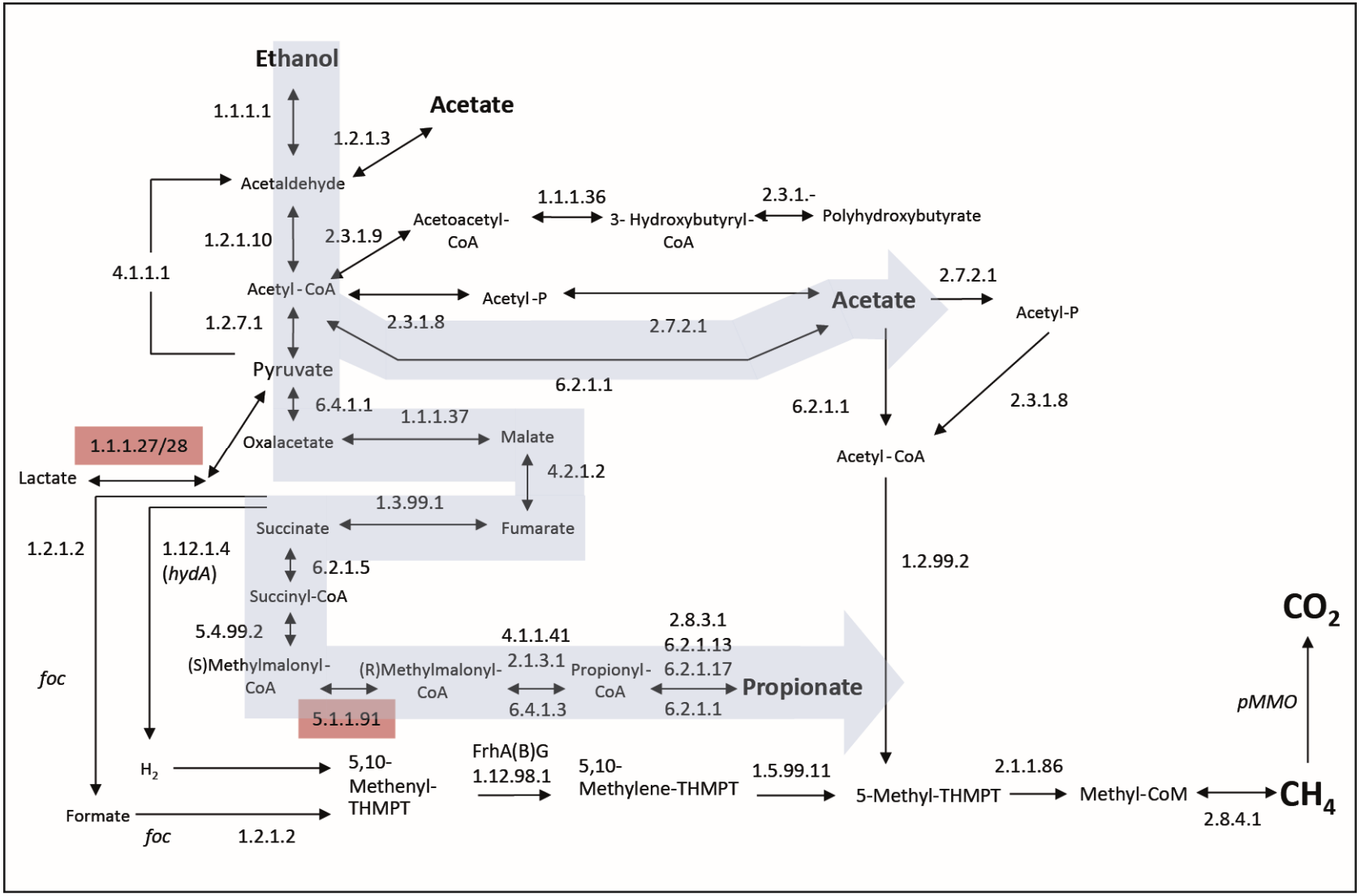

Fig. S14. Proposed pathway for ethanol fermentation to propionate and acetate carried out by *Actinobacteria*. See Fig. S12 for details.

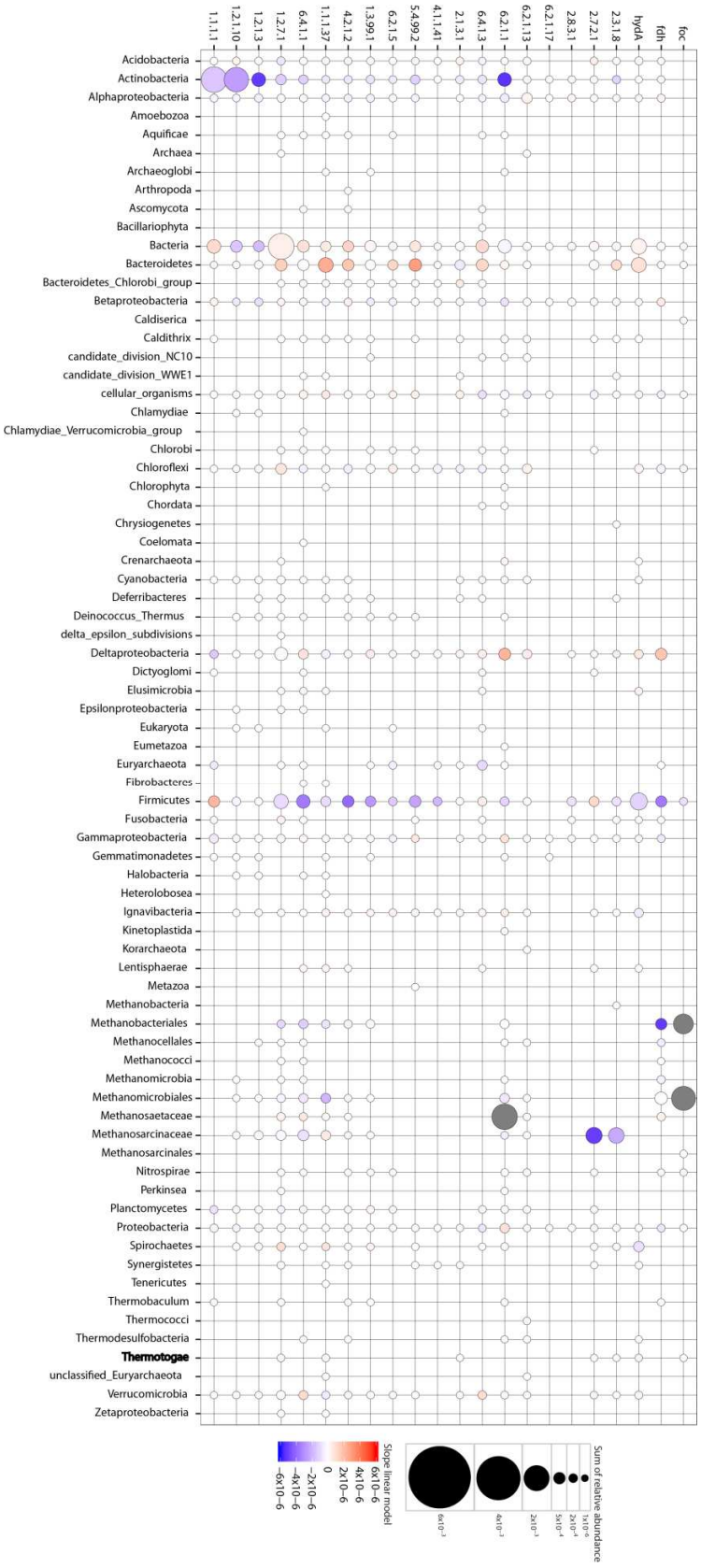

Fig. S15A: Taxonomic assignment of all transcripts encoding enzymes of fermentative pathways from figure 3A. Grey circles are outside the color scale here, but shown in a fitting scale in figure S15C.

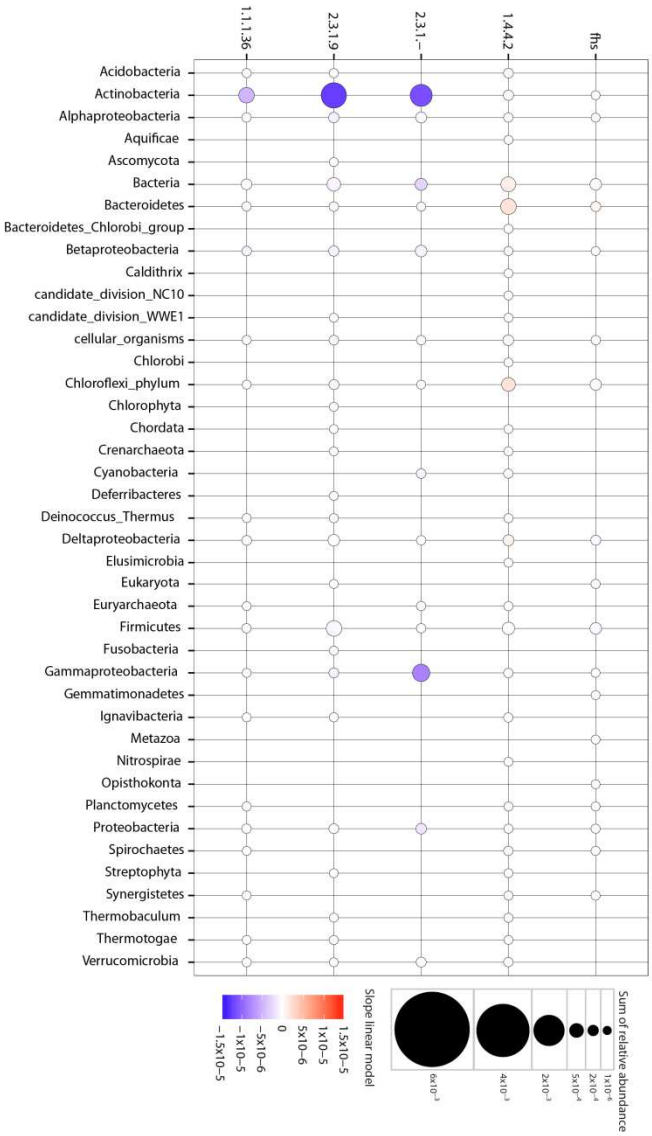

Fig. S15B: Taxonomic assignment of all transcripts encoding enzymes of fermentative pathways from figure 3B

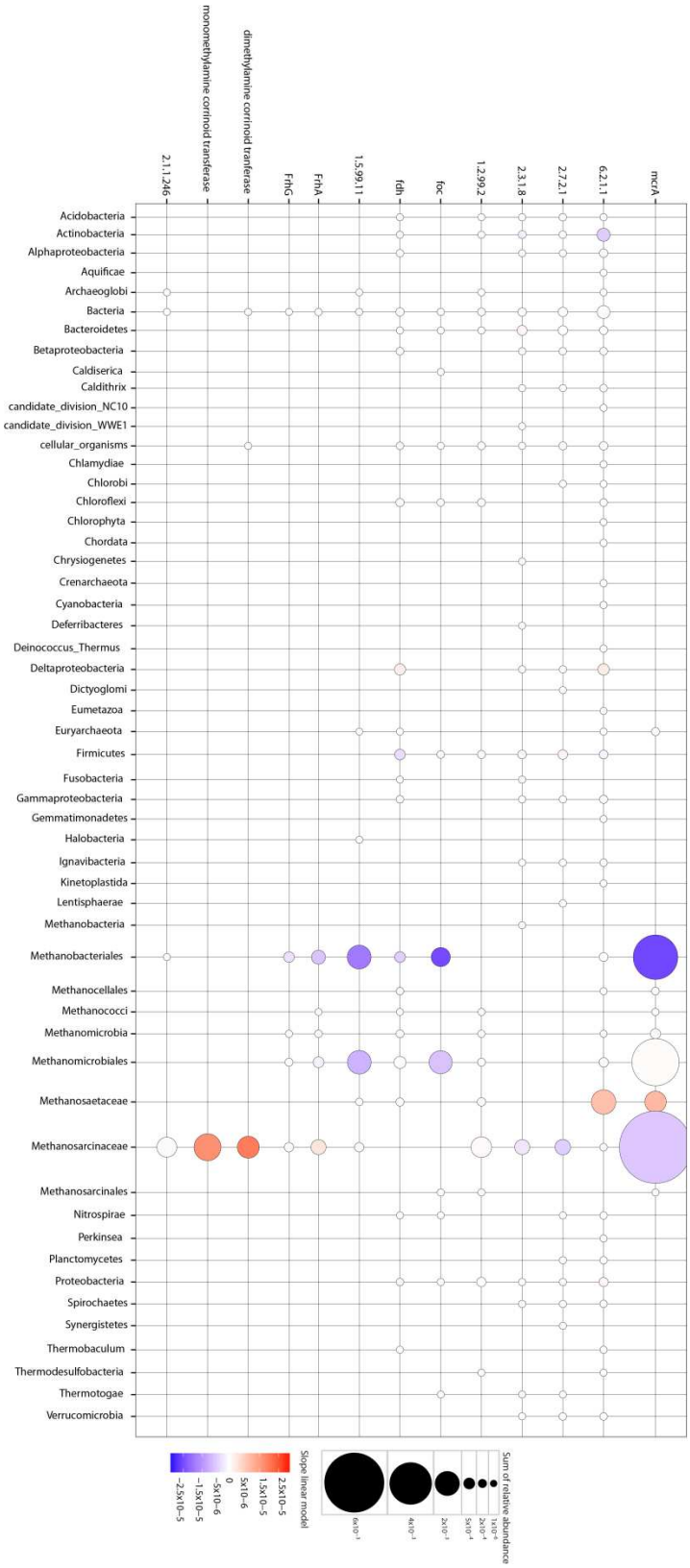

Fig. S15C: Taxonomic assignment of all transcripts encoding enzymes of methanogenic pathways from figure 3C

Fig. S16. Relative abundance of SSU rRNA, SSU rRNA gene, mRNA, and protein-coding DNA assigned to *Methylococcales*.

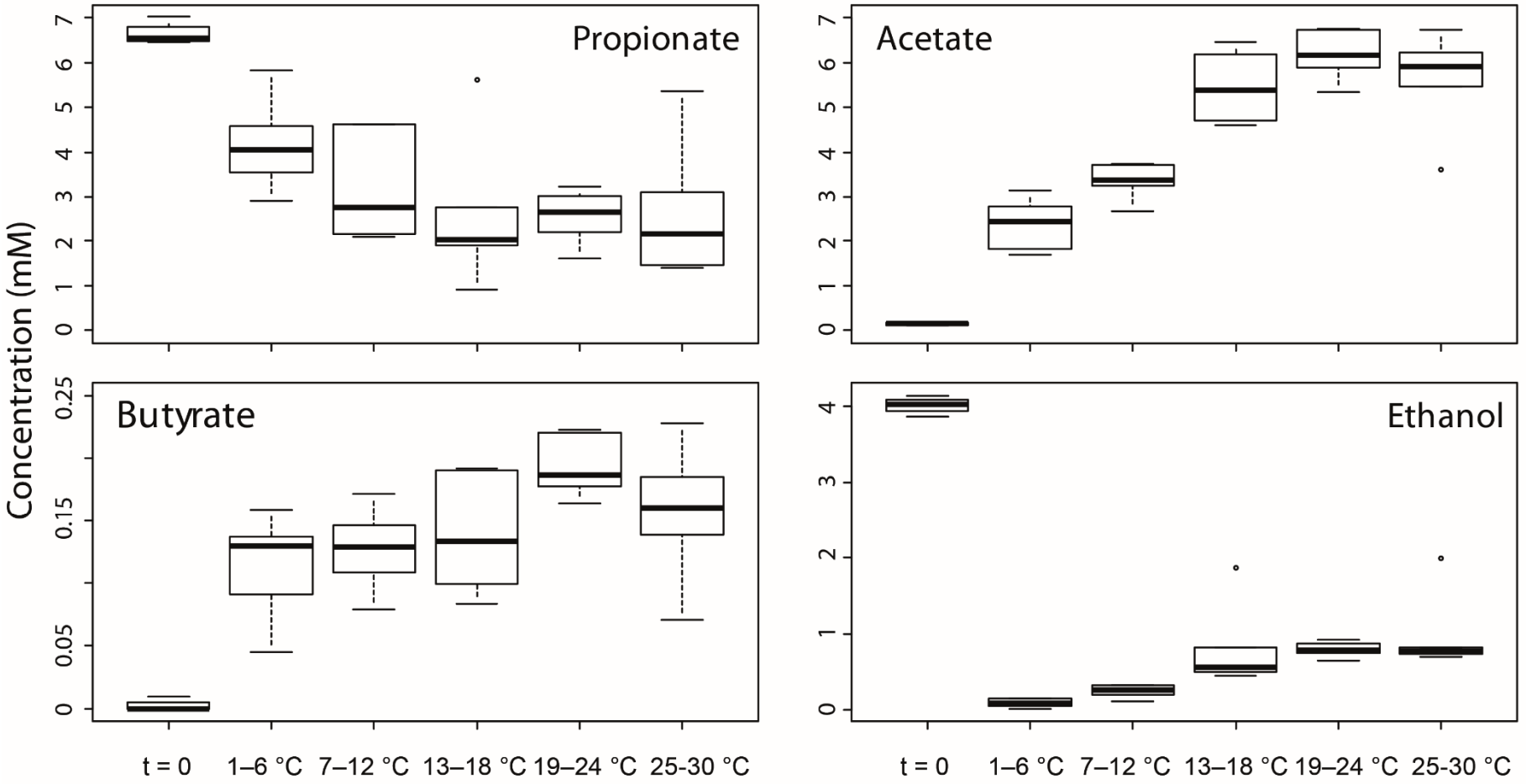

Fig. S17. Concentrations of organic acids and ethanol during methylfluoride inhibition. Boxplots of propionate, acetate, butyrate, and ethanol concentrations (mM) for temperature windows of 6 °C ranging from 1 to 30 °C. The measurements are from one temperature gradient amended with methylfluoride. The day zero concentrations of the experiment are indicated by t = 0. All other measurements are end-point. Whiskers indicate the most extreme value within 1.5 multiplied by the interquartile range; values outside that range represented by circles. Box: 25% quartile; median, 75% quartile.

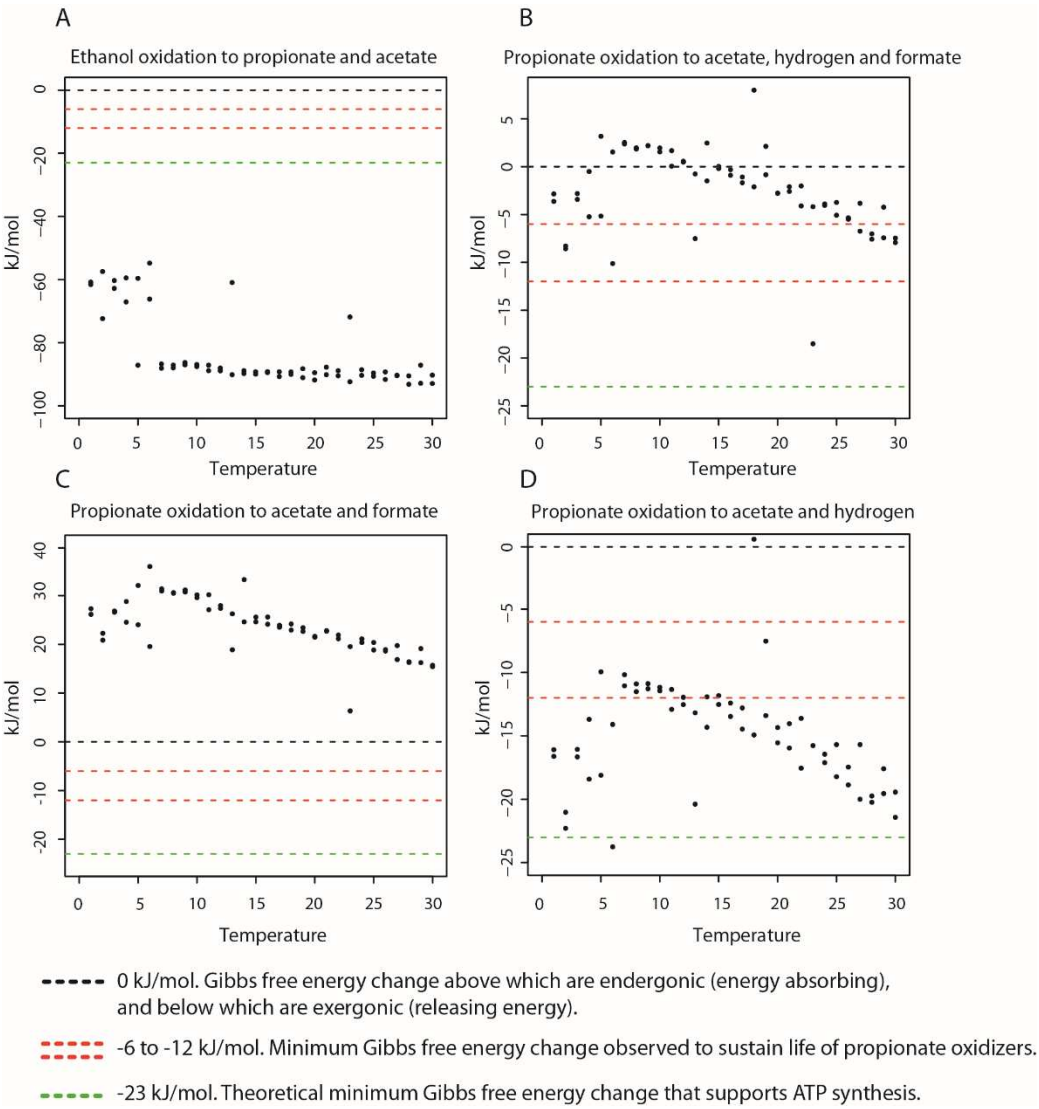

Fig. S18. Gibbs free energy change of fermentation reactions. (A) $3(CH_3CH_2OH) + 2HCO_3^- = CH_3COO^- + H^+ + 2(CH_3CH_2COO^-) + 3H_2O$. (B) $CH_3CH_2COO^- + 2H_2O = CH_3COO^- + HCOO^- + 2H_2 + H^+$. (C) $CH_3CH_2COO^- + 2H_2O + 2CO_2 = CH_3COO^- + 3HCOO^- + 3H^+$. (D) $CH_3CH_2COO^- + 2H_2O = CH_3COO^- + 3H_2 + CO_2$.

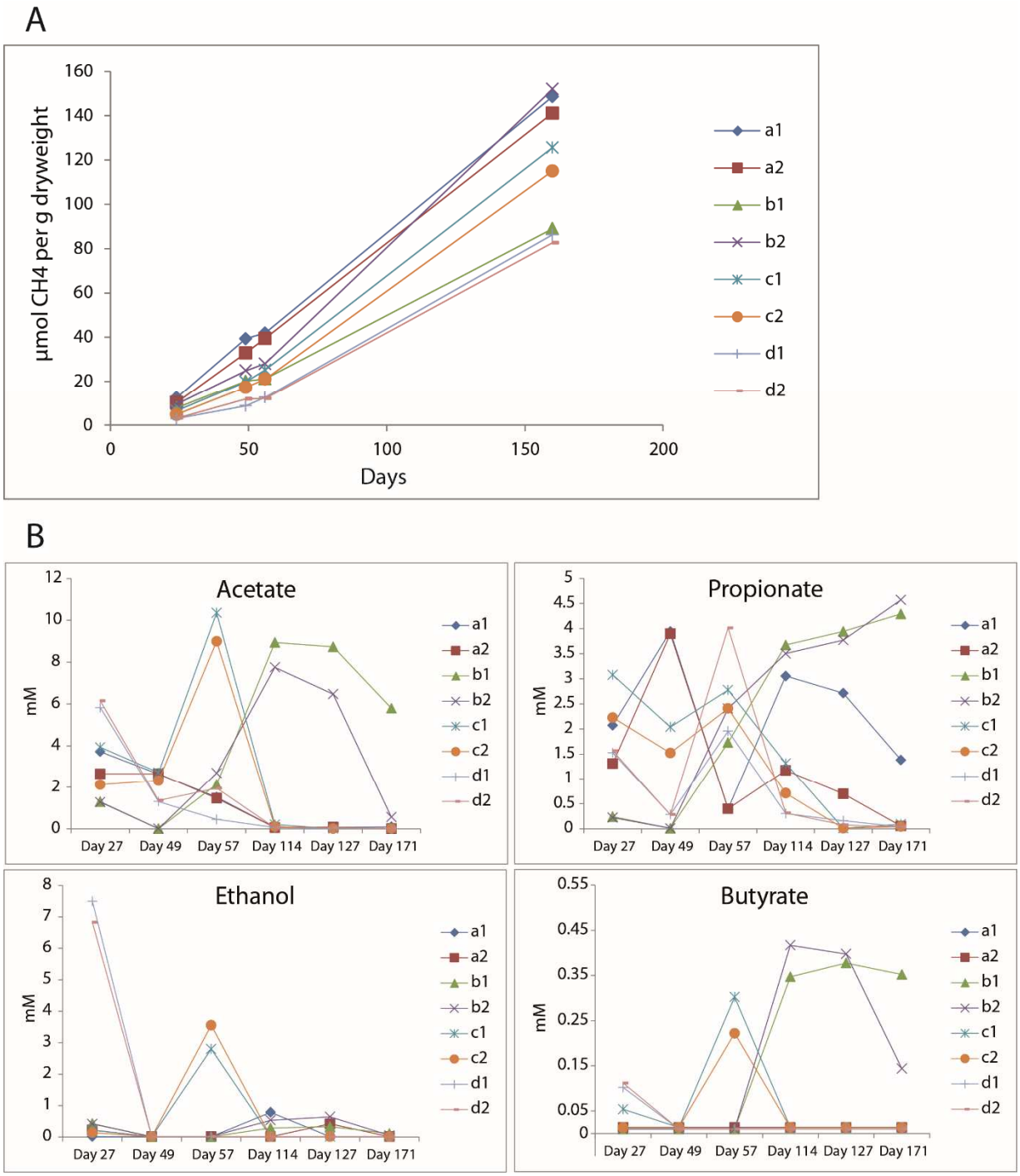

Fig. S19. CH_4 production and concentrations of organic acids and ethanol during the pre-incubation period. Five (a–e) different batches of peat soil slurries were prepared. Four of these batches (a–d) were monitored during the pre-incubation. (A) CH_4 production during pre-incubation of slurry batches a–d. (B) Concentration of organic acids and ethanol during pre-incubation of batches a–d. Detection limit for organic acids and ethanol was 10 µM.