Supplement of Biogeosciences, 19, 831–843, 2022
https://doi.org/10.5194/bg-19-831-2022-supplement
© Author(s) 2022. CC BY 4.0 License.

Biogeosciences
Open Access

Supplement of

The biogeographic pattern of microbial communities inhabiting terrestrial mud volcanoes across the Eurasian continent

Tzu-Hsuan Tu et al.

Correspondence to: Pei-Ling Wang (plwang@ntu.edu.tw)

The copyright of individual parts of the supplement might differ from the article licence.

Materials and methods

Sites and sampling procedures

Muddy fluids from bubbling pools and a total of 13 sediment cores from the adjacent mud platform were retrieved from 13 MVs across the Eurasian continent during 2011 to 2013 (Fig. 1; Table S1) for geochemical and molecular analyses. In brief, bubbling fluids and cores were collected using sterilized cups and liners, respectively. The lengths of the cores ranged between 21 and 67 cm. Samples were transported to the nearby laboratory or accommodation within 5 hours after retrieval. The cores were immediately sectioned at an interval of 1.5 to 3 cm (Table S1) with the average depth of individual sectioned intervals as the representative depth. For gas geochemistry, 6 mL of sediments were preserved in a 36-mL serum bottle with 10 mL of 1 M NaOH and sealed with a butyl rubber stopper capped with an aluminum ring. Following the gas sampling, 3 mL of sediments were collected in a 15-mL centrifuge tube for the determination of water content, and subject to freeze drying in the laboratory. The weight difference was used to calculate the water weight content or porosity assuming that the density of dry sediment was 2.5 g cm^{-3} and the pore space was completely saturated with pore water. For aqueous geochemistry, the remaining sediments were placed in a 50-mL centrifuge tube and centrifuged at 8,200 x g for 15 minutes to collect pore water. The obtained pore water was collected and 0.22-µm-filtered using syringe filters for ion chromatographic analyses of anion abundances. For molecular analyses, sediments were placed in a 50-mL centrifuge tube and kept frozen. All samples were stored in a cooler filled with blue ice during transportation. Upon arriving at the laboratory, anion and DNA samples were stored in a 4°C refrigerator and a −80°C freezer, respectively, until further analysis. Data obtained in this study were merged with companion geochemical data for 4 MVs in Italy (AR01, COM01, PA01, and PA02; Chiu, 2015), and geochemical and molecular data for 2 MVs in Taiwan (LGH03 and SYNH02; Tu et al., 2017; Lin et al., 2018) to generate a total of 136 sample sets for 16 cores from 15 MVs.

Geochemical analyses

Concentrations of gaseous hydrocarbon compounds in head space were analyzed using a 6890N gas chromatograph (GC; Agilent Technologies, USA) equipped with a Porapak Q packed column (3 m) in line with a flame ionization detector and a thermal conductivity detector. The measured partial pressure of methane was used to calculate the equilibrium dissolved concentration with the Henry's law constant (Wiesenburg and Guinasso, 1979). The total moles in headspace and dissolved phase were summed up and normalized to the volume of pore water in order to obtain the dissolved concentration.

Carbon isotope compositions of methane were measured using a MAT253 isotope ratio mass spectrometer connected with a GC Isolink (Thermo Fisher Scientific, USA). The isotopic compositions were reported as the δ notation (δ value = $(R_{sample} / R_{standard} - 1) \times 1000‰$), where R is the ratio of ^{13}C to ^{12}C, and the standard is Pee Dee Belemnite (PDB).

Two anions in pore water, chloride, and sulfate were analyzed using an ICS-3000 ion chromatograph (Thermo Fisher Scientific, USA). Concentrations of particulate total organic carbon (TOC), total inorganic carbon (TIC), total nitrogen (TN), and total sulfur (TS) were determined by an elemental analyzer (MICROcube, Elementar, Germany). The uncertainties for aqueous and gas geochemistry, elemental abundance, and $\delta^{13}C$ value are ±2%, ±5%, ±2%, and

±0.3‰, respectively. The detectable limits for anions with the consideration of dilution were 10 ppm.

Microbial community compositions

DNA extraction and amplification of 16S rRNA gene

Crude DNA for 16S rRNA gene analyses was extracted from 2 to 5 g of fluids/sediments using the PowerSoil DNA Isolation Kit (Qiagen, Germany). Bubbling fluids (if available) and sediments distributed across the geochemical transition were selected for DNA extraction. These samples are representative of communities inhabiting the subsurface source region (represented by bubbling fluids) or subjected to the redox gradient developed after the sediment deposition (represented by cored sediments in the adjacent mud platform). Obtained DNA extracts were stored at −80 °C for subsequent analyses. Polymerase chain reaction (PCR) was applied to amplify the V4 hypervariable region of 16S rRNA genes using the primers F515 (5′–GTG CCA GCM GCC GCG GTA A–3′) and R806 (5′–CCC GTC AAT TCM TTT RAG T–3′) that target both bacterial and archaeal communities (Kozich et al., 2013). Sample specific barcodes and Illumina-specific adapters were appended with both forward and reverse primers. The ingredients of each PCR mixture contained 1.1–1.5 ng of purified genomic DNA, 1 U of ExTaq polymerase (TaKaRa Bio, Japan), 0.2 mM of dNTPs, 0.2μM of each primer, and 2.5μl of 10 × PCR buffer in a total volume of 25 μL. The program of thermal cycling involved a denaturation step at 94°C for 3 minutes followed by 30 cycles of denaturation at 94°C for 45 seconds, annealing at 55°C for 45 seconds, extension at 72°C for 90 seconds, and a final extension step at 72°C for 10 minutes. The products of three PCR runs for individual samples were pooled, analyzed by gel electrophoresis for size verification (~350 bp), and purified using the DNA Clean and Concentrator Kit (Zymo Research, United States). Amplicons from different samples were pooled in equal quantities sufficient for sequencing on an Illumina MiSeq platform (Illumina, United States).

Sequence processing

Sequences of 16S rRNA gene amplicons obtained in this study were pooled with those for LGH03 (Tu et al., 2017) and SYNH02 (Lin et al., 2018) and analyzed using the mothur and QIIME2 (Schloss et al., 2009; Bolyen et al., 2019). Specifically, sequences for individual samples were binned in accordance with the barcode sequences. To minimize the effects of random-sequencing errors, reads that had two or more mismatches to the barcode sequences were discarded. The split raw FASTQ data were processed with the DADA2 (Callahan et al., 2016) implemented in the QIIME2 (version 2018.8; http://qiime2.org/) (Bolyen et al., 2018; Caporaso et al., 2010) to calculate the amplicon sequence variants (ASVs) in each sample. After removing the sequencing adapters, the first 31 nucleotides of primer sequences were trimmed off. Due to the decrease of quality at the end of each read, forward and reverse sequences were truncated to a length of 220 and 200 base pairs, respectively, to obtain individual sequences with a quality score greater than 20. Denoised reads were assembled to full sequences, aligned, and taxonomically assigned against the Silva v.132 reference set using mothur. Sequences identified as chloroplasts and mitochondria were removed. The obtained sequences were deposited in GenBank with accession number PRJNA560274.

Statistics

Microbial community analyses

Sequence data were rarefied to 9,413 sequences per sample through 100 sequence random re-sampling (without replacement) of the original amplicon sequence variant (ASV) table to account for the difference in sequencing depth for the calculation of alpha diversity indices, such as observed ASV richness, Chao1 and Shannon indices (Hill, 1973; Chao et al., 1984). For the beta diversity, the entire ASV table was used and normalized using the function cumNorm from the R package metagenomeSeq (Paulson et al., 2013) A cumulative-sum scaling method was used to calculate the scaling factors equal to the sum of counts up to a particular quantile in order to normalize the read counts with uneven sequencing depth (Paulson et al., 2013) The dissimilarity matrix between samples was computed using the Bray-Curtis method (Bray and Curtis, 1957) and visualized through the ordination of non-metric multidimensional scaling (NMDS). Constrained correspondence analysis (CCA) was performed to elucidate the relationship between microbial community compositions and geochemical variables. The significance of environmental variables relative to the CCA ordinations was computed using "envfit" and 999 permutations. All statistical analyses were performed in R using the packages *vegan, ggplot2,* and *phyloseq.*

Habitat similarities

Habitat similarities were calculated from the Euclidean distances between paired samples with the available concentrations of chloride, sulfate, methane, TN, TS, TIC, and TOC. To reduce the effects of large concentration scales, environmental factors were normalized to their minimum and maximum values to scale the data to a fixed range between 0 and 1. The transformed dataset was used to evaluate habitat similarity using the following formula (Ranjard et al., 2013; Powell et al., 2015):

$$E_d = \left(1 - {\mathrm{Euc}_d}/{\mathrm{Euc}_{max}}\right) \qquad \text{Eq.1}$$

where E_d is the habitat similiarity, Euc_d is the Euclidean distance, and Euc_{max} is the maximum distance between sites in the matrix. To test whether community similarities were significantly correlated with a variety of spatial components, non-parametric Mantel tests based on the Pearson correlation coefficient were applied with significance assessed on the basis of 1000 Monte Carlo permutations. All statistical analyses were performed in R using the package *vegan.*

Distance decay relationships (DDR)

To assess the DDR, pairwise community similarities between samples were calculated using the Sørensen-Dice index (Dice, 1945). The pairwise similarity was transformed in a logarithmic space to enhance the linear fitting using the following equation:

$$\log_{10}(S_{com}) = \log_{10}(a) + \beta \log_{10}(D) \qquad \text{Eq.2}$$

where S_{com} is the pairwise similarity in community composition, D *is* the distance between two samples, and β is the slope. Null values in the similarity/distance matrices were assumed to be 0.001 prior to the log-transformation. The distance between samples was aggregated from two categories for samples in separate cores or within the individual cores. For samples in separate cores, the distance represents the geographic distance between MVs and was calculated using the function *geodist* in the R package 'gmt'. For samples within the individual cores, the distance represents the depth difference between samples. Samples collected from the bubbling pools were regarded as the surface material (0 cm) of each sediment core. The DDR relationships were assessed for data encompassing all samples or either categories. The significance of β was tested by 1000 Monte Carlo permutations of the residuals under the full regression (Legendre and Legendre, 1998). The β was found to be significant for each sample surveyed ($P < 0.001$).

Results

Physical and geochemical characteristics

Geochemical profiles of pore water showed various characteristics related to abiotic and microbial processes. Chloride concentrations varied highly among MVs (ranging between 82 mM at SI02 in Myanmar and 4890 mM at GG01 in Iran) and generally decreased with increasing depth in individual cores (Fig. S1). Exceptions occurred for PA02, SH01, SI02, and LGH03, with substantial fluctuations in the middle or bottom part of the cores. Sulfate concentrations ranged from below the detectable level at SM22, AK03, GJ01, TA, PA01, PA02, and LGH03 to 288 mM at GG01, with most data clustering between 0.5 and 2 mM. Variations in sulfate concentration for cores with detectable sulfate were further categorized into three patterns, including depth-dependent decrease (DSZ01 and SYNH02C4) and increase (GG01, COM01, and SH01), and substantial fluctuation along the depth (AR01 and SI02) (Fig. S1). Methane concentrations ranged between 0.006 mM (PA02) and 3.98 mM (SYMH02C4), with most data clustering between 0.2 and 1 mM (Fig. S2). Methane concentrations either increased (DSZ01, SM22, GJ01, TA, SYNH02C4, and LGH03) or decreased (GG01, AR01, COM01, SH01, and SI02) with increasing depth. The $\delta^{13}C$ values of methane clustered between -58‰ and -35‰ and exhibited a trend opposite to that of methane concentration. The molar ratios of methane over ethane and propane (C1 (methane) / (C2 (ethane) + C3 (propane))) were variable and ranged from 22 (SI02) to approximately 1200 (AR01 and COM01; Fig. S3).

Community structures and compositions

Analyses of all available 16S rRNA genes yielded a total of 4,562,760 sequences. The numbers of observed ASVs for individual samples ranged between 58 and 1,462 with an average value of 449 ± 250 when singletons (presence of one sequence for an ASV at only one depth) were included. The numbers of observed ASVs for individual MVs ranged between 204 (SI02) and 4,203 (AR01). Accumulation curves at the coarse taxonomic resolution (i.e., phylum to family) revealed the sufficient sequencing effort. At the level of ASV, the accumulation curve showed a continuously increasing trend, indicating that the diversity of the entire MV community was not fully captured (Figure S5). The trends in diversity index all exhibited a similar pattern (Figure S4) with the lowest values of alpha diversity indices at SI02 and SH01 in Myanmar and the highest values at AR01 in Italy.

168

169 ***pNST results***

170 The pNST values varied from 65% for bubbling fluid, 73% for surface sediment, to between
171 27% and 92% for within-MV sediment (Table S1). For within-MV sediment, the pNST values for
172 7 out of 15 MVs were less than 50% (Table S1).

173

174 **References**

175 Bolyen, E., Rideout, J.R., Dillon, M.R. et al. Reproducible, interactive, scalable and extensible
176 microbiome data science using QIIME 2. Nat. Biotechnol. 37, 852–857,
177 https://doi.org/10.1038/s41587-019-0209-9, 2019.

178 Bray, J. R. and Curtis, J. T.: An ordination of the upland forest communities of southern Wisconsin,
179 Ecol. Monogr., 27, 325–349, 1957.

180 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P.:
181 DADA2: High-resolution sample inference from Illumina amplicon data., Nat. Methods, 13, 581–
182 583, https://doi.org/10.1038/nmeth.3869, 2016.

183 Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer,
184 N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig,
185 J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky,
186 J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight,
187 R.: QIIME allows analysis of high-throughput community sequencing data., Nat. Methods, 7, 335–
188 336, https://doi.org/10.1038/nmeth.f.303, 2010.

189 Chao, S., Sederoff, R., and Levings, C. S.: Nucleotide sequence and evolution of the 18S ribosomal
190 RNA gene in maize mitochondria, Nucleic Acids Res., 12, 6629–6644, 1984.

191 Chiu, Y.-P.: Microbial community structure and methane cycling in mud volcanoes of Sicily, Italy,
192 M.S. thesis, National Taiwan University, Taiwan, 137 pp., 2015.

193 Dice, L. R.: Measures of the amount of ecologic association between species, Ecology, 26, 297–
194 302, https://doi.org/10.2307/1932409, 1945.

195 Hill, M. O.: Diversity and evenness: a unifying notation and its consequences, Ecology, 54, 427–
196 432, 1973.

197 Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D.: Development
198 of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data
199 on the MiSeq Illumina sequencing platform, Appl. Environ. Microbiol., 79, 5112–5120,
200 https://doi.org/10.1128/aem.01043-13, 2013.

201 Legendre, P. and Legendre, L.: Numerical Ecology, Volume 24, Second Edition (Developments
202 in Environmental Modelling), https://doi.org/10.1234/12345678, 1998.

Lin, Y.-T., Tu, T.-H., Wei, C.-L., Rumble, D., Lin, L.-H., and Wang, P.-L.: Steep redox gradient and biogeochemical cycling driven by deeply sourced fluids and gases in a terrestrial mud volcano., Fems. Microbiol. Ecol., 94, 3796, https://doi.org/10.1093/femsec/fiy171, 2018.

Paulson, J. N., Stine, O. C., Bravo, H. C., and Pop, M.: Differential abundance analysis for microbial marker-gene surveys., Nat. Methods, 10, 1200–2, https://doi.org/10.1038/nmeth.2658, 2013.

Powell, J. R., Karunaratne, S., Campbell, C. D., Yao, H., Robinson, L., and Singh, B. K.: Deterministic processes vary during community assembly for ecologically dissimilar taxa., Nat. Commun., 6, 8444, https://doi.org/10.1038/ncomms9444, 2015.

Ranjard, L., Dequiedt, S., Prévost-Bouré, N. C., Thioulouse, J., Saby, N. P. A., Lelievre, M., Maron, P. A., Morin, F. E. R., Bispo, A., Jolivet, C., Arrouays, D., and Lemanceau, P.: Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity., Nat. Commun., 4, 1434, https://doi.org/10.1038/ncomms2431, 2013.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., and Robinson, C. J.: Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities, Appl. Environ. Microbiol., 75, 7537-- 7541, 2009.

Tu, T.-H., Wu, L.-W., Lin, Y.-S., Imachi, H., Lin, L.-H., and Wang, P.-L.: Microbial community composition and functional capacity in a terrestrial ferruginous, sulfate-depleted mud volcano, Front. Microbiol., 8, 2137, https://doi.org/10.3389/fmicb.2017.02137, 2017.

Wiesenburg, D. A. and Guinasso, N. L.: Equilibrium solubilities of methane, carbon monoxide, and hydrogen in water and sea water, J. Chem. Eng. Data., 24, 356-- 360, https://doi.org/10.1021/je60083a006, 1979.

Supplementary Tables

Table S1. Site, sample, and sequence information for investigated MVs.

Core name	Country	Core length (cm)	Section interval (cm)	Longitude	Latitude	Number of ASVs	NST_{bray} (%)	$NST_{Jaccard}$ (%)	pNST (%)
AR01C1[1]	Italy	67	3	13.60000	37.37667	4,203	36	36	39
COM01C1[1]	Italy	45	2.5	13.65194	37.44306	3,684	60	60	47
PA01C1[1]	Italy	55	3	14.91972	37.54472	3,272	62	61	65
PA02C1[1]	Italy	41	2.75	14.89028	37.57278	2,125	34	38	92
AK03C1	Georgia	49	2.5	45.91322	41.41953	3,072	32	31	58
GJ01C1	Georgia	44	2.5	45.79261	41.74531	2,101	35	35	27
QK01C1	Georgia	22	1.5	45.80564	41.28905	2,495	63	57	75
GG01C1	Iran	46	2.5	54.39608	37.11856	2,075	36	37	78
TA01C1	Iran	47	3	59.93306	25.46697	871	6	5	40
SM22C1	China	33	2	84.38722	44.18269	1,364	19	18	27
DSZ01C1	China	21	2	84.84636	44.30517	1,761	42	40	65
SH01C1	Myanmar	38	2	93.57119	19.36975	332	1	1	51
SI02C1	Myanmar	48	2	93.59169	19.39778	204	1	1	44
LGH03C4[2]	Taiwan	160	5	121.20940	22.98306	3,176	24	25	42
SYNH02C4[3]	Taiwan	52	2.5	120.40948	22.80313	2,044	44	43	53
SYNH02C11[3]	Taiwan	20	2	120.40948	22.80313	2,116			

[1]Sample retrieval and geochemical data were adopted from Chiu (2015).
[2]Sample retrieval, geochemical and raw sequence data were adopted from Tu et al. (2017).
[3]Sample retrieval, geochemical and raw sequence data were adopted from Lin et al. (2018).

Table S2. Distribution and sequence information for two cosmopolitan ASVs and 10 most abundant ASVs in any individual MVs.

Name of ASV	Phylum	Genus	Site	Proportion (%)
		The most widespread ASVs		
WS_1	Proteobacteria	Unclassified genus in Desulfuromonadaceae	SM22, GJ01, QK01, GG01, AR01, COM01, PA01, PA02, SYNH02	0.019, 0.182, 0.094, 0.028, 0.411, 0.250, 0.208, 0.027, 0.082
WS_2	Proteobacteria	*Desulfotignum*	DSZ01, AK03, GJ01, QK01, AR01, COM01, PA01, LGH03, SYNH02	0.017, 0.051, 0.107, 0.009, 0.031, 0.016, 0.042, 0.012, 0.033
		The 10 most abundant ASVs at individual sites		
AR01_ASV_1	Bacteroidetes	Unclassified genus in Lentimicrobiaceae	AR01	0.459
AR01_ASV_2	Cyanobacteria	Unclassified genus in Chloroplast	AR01	0.203
AR01_ASV_3	Cyanobacteria	Unclassified genus in Chloroplast	AR01	0.174
AR01_ASV_4	Proteobacteria	JTB255 marine benthic group	AR01	0.135
AR01_ASV_5	Cloacimonetes	MSBL8	AR01	0.121
AR01_ASV_6	Firmicutes	Unclassified genus in Syntrophomonadaceae	AR01	0.114
AR01_ASV_7	Proteobacteria	*Methylophaga*	AR01	0.113
AR01_ASV_8	Chloroflexi	ADurb.Bin120	AR01	0.110
AR01_ASV_9	Proteobacteria	*Methylomicrobium*	AR01	0.108
AR01_ASV_10	Proteobacteria	*Methylomicrobium*	AR01	0.105
COM01_ASV_1	Chloroflexi	Unclassified genus in Ardenticatenales	COM01	0.253
COM01_ASV_2	Euryarchaeota	Unclassified genus in Methanoperedenaceae	COM01	0.229
COM01_ASV_3	Gemmatimonadetes	BD2-11 terrestrial group	COM01	0.185
COM01_ASV_4	Euryarchaeota	*Haloparvum*	COM01	0.183
COM01_ASV_5	Euryarchaeota	ANME-2a-2b	COM01	0.172
COM01_ASV_6	Unclassified Bacteria	Unclassified Bacteria	COM01	0.163

Name of ASV	Phylum	Genus	Site	Proportion (%)
COM01_ASV_7	Euryarchaeota	ANME-2a-2b	COM01	0.162
COM01_ASV_8	Proteobacteria	MBNT15	COM01	0.156
COM01_ASV_9	Latescibacteria	Unclassified genus in Latescibacteraceae	COM01	0.141
COM01_ASV_10	Euryarchaeota	Unclassified genus in Methanoperedenaceae	COM01	0.138
PA01_ASV_1	Euryarchaeota	*Halanaeroarchaeum*	PA01	0.733
PA01_ASV_2	Euryarchaeota	*Halanaeroarchaeum*	PA01	0.699
PA01_ASV_3	Proteobacteria	Unclassified genus in Proteobacteria	PA01	0.698
PA01_ASV_4	Euryarchaeota	*Halanaeroarchaeum*	PA01	0.681
PA01_ASV_5	Proteobacteria	Unclassified genus in Proteobacteria	PA01	0.676
PA01_ASV_6	Euryarchaeota	*Halanaeroarchaeum*	PA01	0.389
PA01_ASV_7	Euryarchaeota	*Halanaeroarchaeum*	PA01	0.370
PA01_ASV_8	Unclassified Bacteria	Unclassified Bacteria	PA01	0.367
PA01_ASV_9	Proteobacteria	*Cupriavidus*	PA01	0.362
PA01_ASV_10	Euryarchaeota	*Halanaeroarchaeum*	PA01	0.312
PA02_ASV_1	Proteobacteria	Unclassified genus in Proteobacteria	PA02	1.020
PA02_ASV_2	Proteobacteria	Unclassified genus in Proteobacteria	PA02	0.891
PA02_ASV_3	Proteobacteria	Unclassified genus in Proteobacteria	PA02	0.756
PA02_ASV_4	Proteobacteria	Unclassified genus in Proteobacteria	PA02	0.748
PA02_ASV_5	Euryarchaeota	*Halodesulfurarchaeum*	PA02	0.543
PA02_ASV_6	Euryarchaeota	*Halodesulfurarchaeum*	PA02	0.541
PA02_ASV_7	Proteobacteria	Unclassified genus in Proteobacteria	PA02	0.518
PA02_ASV_8	Euryarchaeota	*Halodesulfurarchaeum*	PA02	0.471
PA02_ASV_9	Proteobacteria	*Methylohalobius*	PA02	0.411
PA02_ASV_10	Euryarchaeota	*Halodesulfurarchaeum*	PA02	0.357
AK03_ASV_1	Chloroflexi	ADurb.Bin120	AK03	0.745
AK03_ASV_2	Chloroflexi	ADurb.Bin120	AK03	0.611
AK03_ASV_3	Proteobacteria	Unclassified genus in Desulfuromonadaceae	AK03	0.582
AK03_ASV_4	Cloacimonetes	Candidatus Cloacimonas	AK03	0.556
AK03_ASV_5	Zixibacteria	Unclassified genus in Zixibacteria	AK03	0.512
AK03_ASV_6	Zixibacteria	Unclassified genus in Zixibacteria	AK03	0.480

Name of ASV	Phylum	Genus	Site	Proportion (%)
AK03_ASV_7	Chloroflexi	ADurb.Bin120	AK03	0.478
AK03_ASV_8	Cloacimonetes	Unclassified genus in Cloacimonadaceae	AK03	0.469
AK03_ASV_9	Proteobacteria	*Desulfuromusa*	AK03	0.462
AK03_ASV_10	Chloroflexi	ADurb.Bin120	AK03	0.437
GJ01_ASV_1	Cyanobacteria	Unclassified genus in Chloroplast	GJ01	1.314
GJ01_ASV_2	Cyanobacteria	Unclassified genus in Chloroplast	GJ01	1.143
GJ01_ASV_3	Cyanobacteria	Unclassified genus in Chloroplast	GJ01	1.142
GJ01_ASV_4	Cyanobacteria	Unclassified genus in Chloroplast	GJ01	0.843
GJ01_ASV_5	Cyanobacteria	Unclassified genus in Chloroplast	GJ01	0.838
GJ01_ASV_6	Cyanobacteria	Unclassified genus in Chloroplast	GJ01	0.834
GJ01_ASV_7	Cyanobacteria	Unclassified genus in Chloroplast	GJ01	0.828
GJ01_ASV_8	Cyanobacteria	Unclassified genus in Chloroplast	GJ01	0.826
GJ01_ASV_9	Cyanobacteria	Unclassified genus in Chloroplast	GJ01	0.809
GJ01_ASV_10	Cyanobacteria	Unclassified genus in Chloroplast	GJ01	0.760
QK01_ASV_1	Cyanobacteria	Arthrospira PCC-7345	QK01	1.030
QK01_ASV_2	Fusobacteria	*Hypnocyclicus*	QK01	0.809
QK01_ASV_3	Cyanobacteria	Geitlerinema PCC-7105	QK01	0.780
QK01_ASV_4	Fusobacteria	*Hypnocyclicus*	QK01	0.674
QK01_ASV_5	Cyanobacteria	Geitlerinema PCC-7105	QK01	0.636
QK01_ASV_6	Bacteroidetes	ML635J-40 aquatic group	QK01	0.516
QK01_ASV_7	Cyanobacteria	Geitlerinema PCC-7105	QK01	0.496
QK01_ASV_8	Bacteroidetes	ML635J-40 aquatic group	QK01	0.482
QK01_ASV_9	Bacteroidetes	ML635J-40 aquatic group	QK01	0.384
QK01_ASV_10	Planctomycetes	Unclassified genus in Planctomycetales	QK01	0.344
GG01_ASV_1	Euryarchaeota	*Halorubrum*	GG01	1.320
GG01_ASV_2	Euryarchaeota	*Halorubrum*	GG01	1.314
GG01_ASV_3	Euryarchaeota	*Halorubrum*	GG01	1.004
GG01_ASV_4	Euryarchaeota	*Halonotius*	GG01	0.989
GG01_ASV_5	Euryarchaeota	*Halonotius*	GG01	0.934
GG01_ASV_6	Euryarchaeota	*Halonotius*	GG01	0.908
GG01_ASV_7	Bacteroidetes	*Salinibacter*	GG01	0.764

Name of ASV	Phylum	Genus	Site	Proportion (%)
GG01_ASV_8	Euryarchaeota	*Halorubrum*	GG01	0.680
GG01_ASV_9	Cyanobacteria	Unclassified genus in Nodosilineaceae	GG01	0.624
GG01_ASV_10	Euryarchaeota	*Halorubrum*	GG01	0.548
TA01_ASV_1	Proteobacteria	*Desulfobacca*	TA01	2.763
TA01_ASV_2	Proteobacteria	*Desulfobacca*	TA01	2.759
TA01_ASV_3	Proteobacteria	Unclassified genus in Deltaproteobacteria	TA01	2.587
TA01_ASV_4	Proteobacteria	*Desulfobacca*	TA01	2.448
TA01_ASV_5	Proteobacteria	Unclassified genus in Deltaproteobacteria	TA01	2.354
TA01_ASV_6	Proteobacteria	*Desulfobacca*	TA01	2.188
TA01_ASV_7	Proteobacteria	Unclassified genus in Deltaproteobacteria	TA01	1.977
TA01_ASV_8	Proteobacteria	*Desulfobacca*	TA01	1.929
TA01_ASV_9	Proteobacteria	*Desulfobacca*	TA01	1.889
TA01_ASV_10	Proteobacteria	Unclassified genus in Deltaproteobacteria	TA01	1.877
SM22_ASV_1	Bacteroidetes	ML635J-40 aquatic group	SM22	1.127
SM22_ASV_2	Bacteroidetes	ML635J-40 aquatic group	SM22	1.103
SM22_ASV_3	Bacteroidetes	ML635J-40 aquatic group	SM22	1.049
SM22_ASV_4	Proteobacteria	*Desulfuromusa*	SM22	0.974
SM22_ASV_5	Bacteroidetes	ML635J-40 aquatic group	SM22	0.973
SM22_ASV_6	Bacteroidetes	ML635J-40 aquatic group	SM22	0.940
SM22_ASV_7	Bacteroidetes	ML635J-40 aquatic group	SM22	0.936
SM22_ASV_8	Bacteroidetes	ML635J-40 aquatic group	SM22	0.924
SM22_ASV_9	Bacteroidetes	ML635J-40 aquatic group	SM22	0.915
SM22_ASV_10	Bacteroidetes	ML635J-40 aquatic group	SM22	0.908
DSZ01_ASV_1	Bacteroidetes	*Phaeodactylibacter*	DSZ01	1.161
DSZ01_ASV_2	Bacteroidetes	*Phaeodactylibacter*	DSZ01	0.774
DSZ01_ASV_3	Bacteroidetes	*Phaeodactylibacter*	DSZ01	0.754
DSZ01_ASV_4	Bacteroidetes	*Phaeodactylibacter*	DSZ01	0.732
DSZ01_ASV_5	Proteobacteria	*Desulfatiglans*	DSZ01	0.702
DSZ01_ASV_6	Chloroflexi	Unclassified genus in Anaerolineaceae	DSZ01	0.683
DSZ01_ASV_7	Proteobacteria	Unclassified genus in Gammaproteobacteria	DSZ01	0.613
DSZ01_ASV_8	Bacteroidetes	*Algoriphagus*	DSZ01	0.586

Name of ASV	Phylum	Genus	Site	Proportion (%)
DSZ01_ASV_9	Proteobacteria	Sva1033	DSZ01	0.559
DSZ01_ASV_10	Bacteroidetes	*Phaeodactylibacter*	DSZ01	0.556
SH01_ASV_1	Nitrospirae	Uncultured genus in Thermodesulfovibrionia	SH01	1.032
SH01_ASV_2	Euryarchaeota	ANME-1b	SH01	0.742
SH01_ASV_3	Proteobacteria	Unclassified genus in Methylomonaceae	SH01	0.538
SH01_ASV_4	Chloroflexi	Unclassified genus in Ardenticatenales	SH01	0.493
SH01_ASV_5	Euryarchaeota	ANME-1b	SH01	0.481
SH01_ASV_6	Proteobacteria	*Desulfatiglans*	SH01	0.451
SH01_ASV_7	Proteobacteria	Unclassified genus in Methylomonaceae	SH01	0.451
SH01_ASV_8	Proteobacteria	*Desulfatiglans*	SH01	0.379
SH01_ASV_9	Chloroflexi	Uncultured genus in Ardenticatenales	SH01	0.305
SH01_ASV_10	Nitrospirae	*Phaeodactylibacter*	SH01	0.287
SI02_ASV_1	Firmicutes	*Ammoniphilus*	SI02	6.927
SI02_ASV_2	Firmicutes	*Ammoniphilus*	SI02	6.023
SI02_ASV_3	Firmicutes	*Ammoniphilus*	SI02	5.440
SI02_ASV_4	Firmicutes	*Ammoniphilus*	SI02	5.026
SI02_ASV_5	Firmicutes	*Ammoniphilus*	SI02	4.938
SI02_ASV_6	Firmicutes	*Ammoniphilus*	SI02	2.858
SI02_ASV_7	Firmicutes	*Ammoniphilus*	SI02	2.729
SI02_ASV_8	Firmicutes	*Ammoniphilus*	SI02	2.686
SI02_ASV_9	Firmicutes	*Ammoniphilus*	SI02	0.709
SI02_ASV_10	Firmicutes	*Ammoniphilus*	SI02	0.518
SYNH02_ASV_1	Proteobacteria	Unclassified genus in Pseudomonadaceae	SYNH02	2.243
SYNH02_ASV_2	Proteobacteria	Unclassified genus in Pseudomonadaceae	SYNH02	1.810
SYNH02_ASV_3	Proteobacteria	Unclassified genus in Pseudomonadaceae	SYNH02	1.517
SYNH02_ASV_4	Proteobacteria	Unclassified genus in Pseudomonadaceae	SYNH02	1.422
SYNH02_ASV_5	Proteobacteria	Unclassified genus in Pseudomonadaceae	SYNH02	1.293
SYNH02_ASV_6	Proteobacteria	Unclassified genus in Pseudomonadaceae	SYNH02	1.273
SYNH02_ASV_7	Proteobacteria	Unclassified genus in Pseudomonadaceae	SYNH02	1.142
SYNH02_ASV_8	Proteobacteria	Unclassified genus in Pseudomonadaceae	SYNH02	1.122
SYNH02_ASV_9	Proteobacteria	Unclassified genus in Pseudomonadaceae	SYNH02	1.041

Name of ASV	Phylum	Genus	Site	Proportion (%)
SYNH02_ASV_10	Proteobacteria	Unclassified genus in Pseudomonadaceae	SYNH02	0.993

Table S3. Multiple linear regression model[1] for Shannon index versus significant geochemical parameters.

	Estimate	Std. Error	t-value	*P*-value	Signif. code[2]
(Intercept)	4.02048	0.32306	12.445	< 0.0001	***
Methane	149.79125	76.02220	1.970	0.05106	.
TN	8.66657	3.13095	2.768	0.00652	**
TIC	0.43655	0.08123	5.374	< 0.0001	***

[1]The calculation was performed after the removal of collinear variables and with the application of Akaike information criterion (AIC). Variables were added to the model to generate the highest to lowest best fit from simple linear regression. R^2 values for multiple regression and adjusted regression were 0.2067 and 0.1872, respectively. F-statistic: 10.6 on 3 and 122 degrees of freedom, p-value: 3.047E-06.

[2]Significance codes: '***': 0-0.001; '**': 0.001-0.01; '.': 0.05-0.1.

Table S4. Simple linear regression of Shannon index versus individual geochemical parameters.

	Slope	Std. error	t-value	*P*-value	R^2
Sulfate	0.0004	0.0016	0.277	0.782	0.001
Chloride	0.0001	0.0001	1.808	0.0731	0.018
Methane	94.391	83.070	1.172	0.243	0.002
TN	2.733	3.267	0.837	0.404	0.002
TS	0.0985	0.1908	0.517	0.606	0.006
TIC	0.3577	0.0793	4.511	< 0.0001	0.134
TOC	0.3584	0.1862	1.925	0.0566	0.021

Table S5. Mantel test using Spearman's correlation (permutations = 999) for the Bray-Curtis dissimilarities between all communities, and geochemical parameters or geographic distance (km).

	Rho (ρ)	*p*	Signif. code[1]
km	0.322	< 0.001	***
env	0.178	< 0.001	***
Chloride	0.454	< 0.001	***
Sulfate	0.258	< 0.001	***
Methane	0.068	0.026	**
TIC	0.255	< 0.001	***
TOC	-0.081	0.986	-
TN	-0.001	0.481	-
TS	0.143	< 0.001	***

[1]Signifance codes: '***': 0-0.001; '-': 0.1-1.

Table S6. Permutational multivariate analysis of variance of beta diversity.

	Df	SumsOfSqs	MeanSqs	F.Model	R^2	*P*-value	Signif. code[1]
Chloride	1	1.504	1.50431	6.1690	0.02586	< 0.001	***
Sulfate	1	1.719	1.71883	7.0488	0.02954	< 0.001	***
Methane	1	0.509	0.50856	2.0856	0.00874	< 0.001	***
TIC	1	2.117	2.11708	8.6819	0.03639	< 0.001	***
TOC	1	1.377	1.37699	5.6469	0.02367	< 0.001	***
TN	1	1.828	1.82757	7.4947	0.03141	< 0.001	***
TS	1	1.177	1.17735	4	0.02024	< 0.001	***

[1]Significance codes: '***': 0-0.001.

[2]Analysis was performed on the basis of continuous variables only and Bray-Curtis dissimilarities.

Table S7. Coefficient of variation and the Pearson's correlation coefficient (r) for chloride and sulfate concentrations.

Sample name	Chloride	Sulfate	r	Signif. code[1]
AR01C1	8.97%	117%	0.15	-
COM01C1	8.37%	122%	0.05	-
PA01C1	3.61%	NA	NA	NA
PA02C1	13.24%	NA	NA	NA
AK03C1	7.55%	NA	NA	NA
GJ01C1	5.61%	NA	NA	NA
QK01C1	63.13%	NA	NA	NA
GG01C1	14.84%	17%	-0.35	-
TA01C1	2.15%	NA	NA	NA
SM22C1	7.08%	NA	NA	NA
DSZ01C1	25.93%	186%	0.46	-
SH01C1	15.58%	103%	0.91	***
SI02C1	3.45%	29%	0.24	.
LGH03C4	6.56%	NA	NA	NA
SYNH02C4	15.02%	126%	0.88	***
SYNH02C11	28.11%	138%	0.96	***

[1]Significance codes: '***': 0-0.001; '.': 0.05-0.1; '-': 0.1-1.
[2]NA means sulfate concentration below the level of detection.

Supplementary Figures

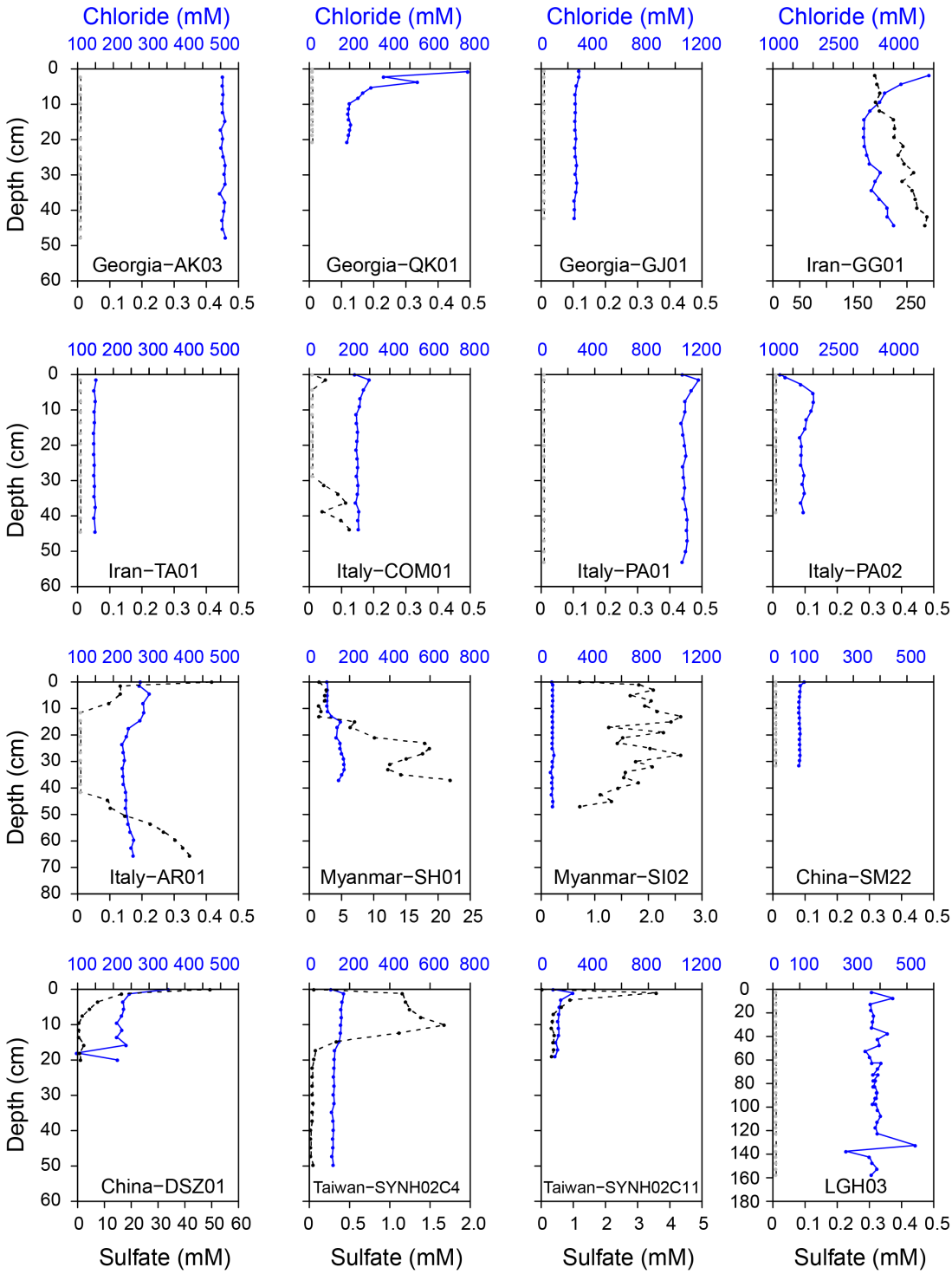

Figure S1: Chloride (in blue) and sulfate (in black) concentration profiles. Sulfate concentrations lower than the limit of detection (0.01 mM) are shown in gray dash. Data for SYNH02C4, SYNH02C11, and LGH03 were adopted from Tu et al. (2017) and Lin et al. (2018).

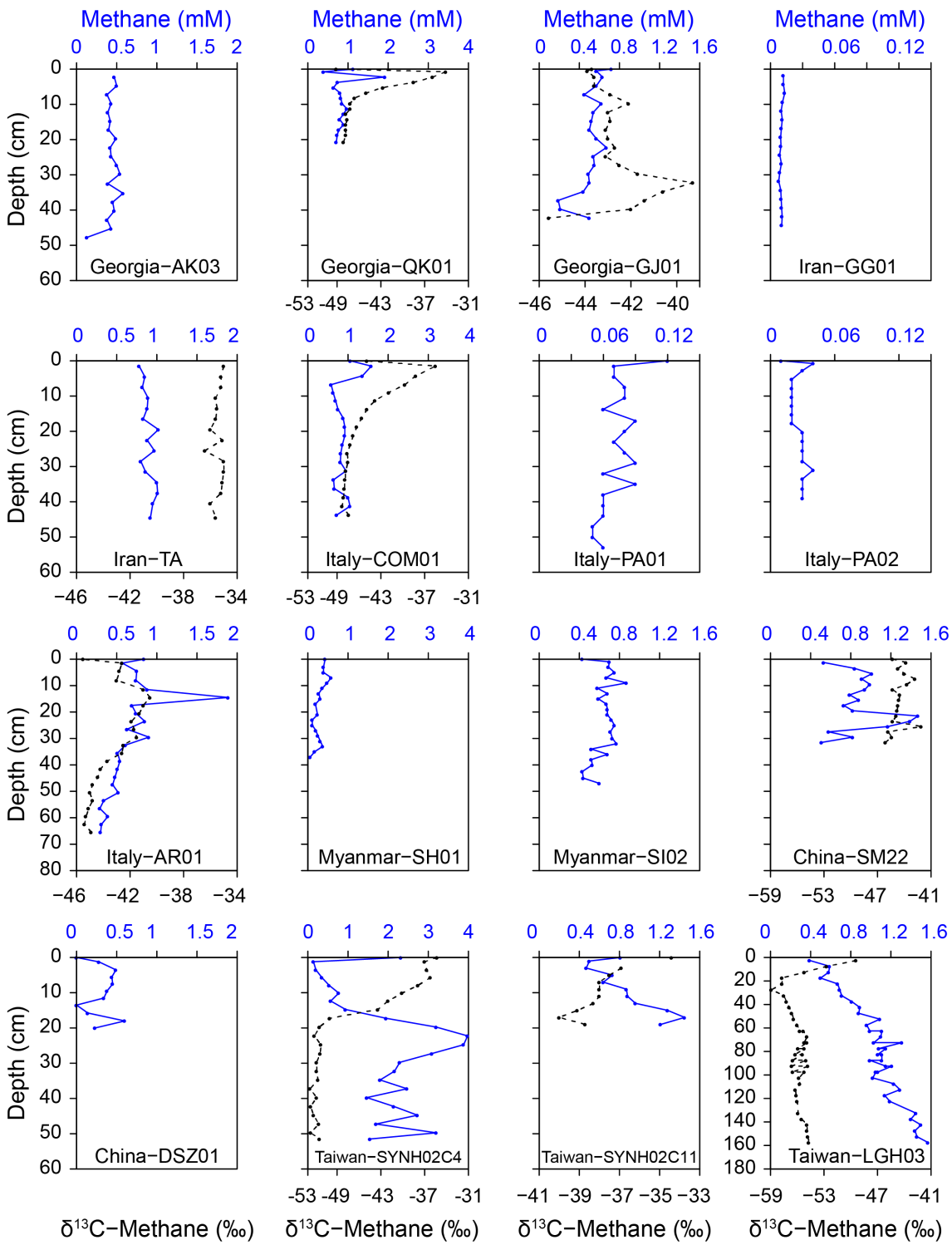

Figure S2: Profiles of methane concentrations (in blue) and $\delta^{13}C$ values of methane (in black). Samples collected from Georgia are not sufficient for porosity measurement. Therefore, their concentrations were calculated assuming that the weight proportion of pore water is 0.5. Data for SYNH02C4, SYNH02C11, and LGH03 were adopted from Tu et al. (2017) and Lin et al. (2018).

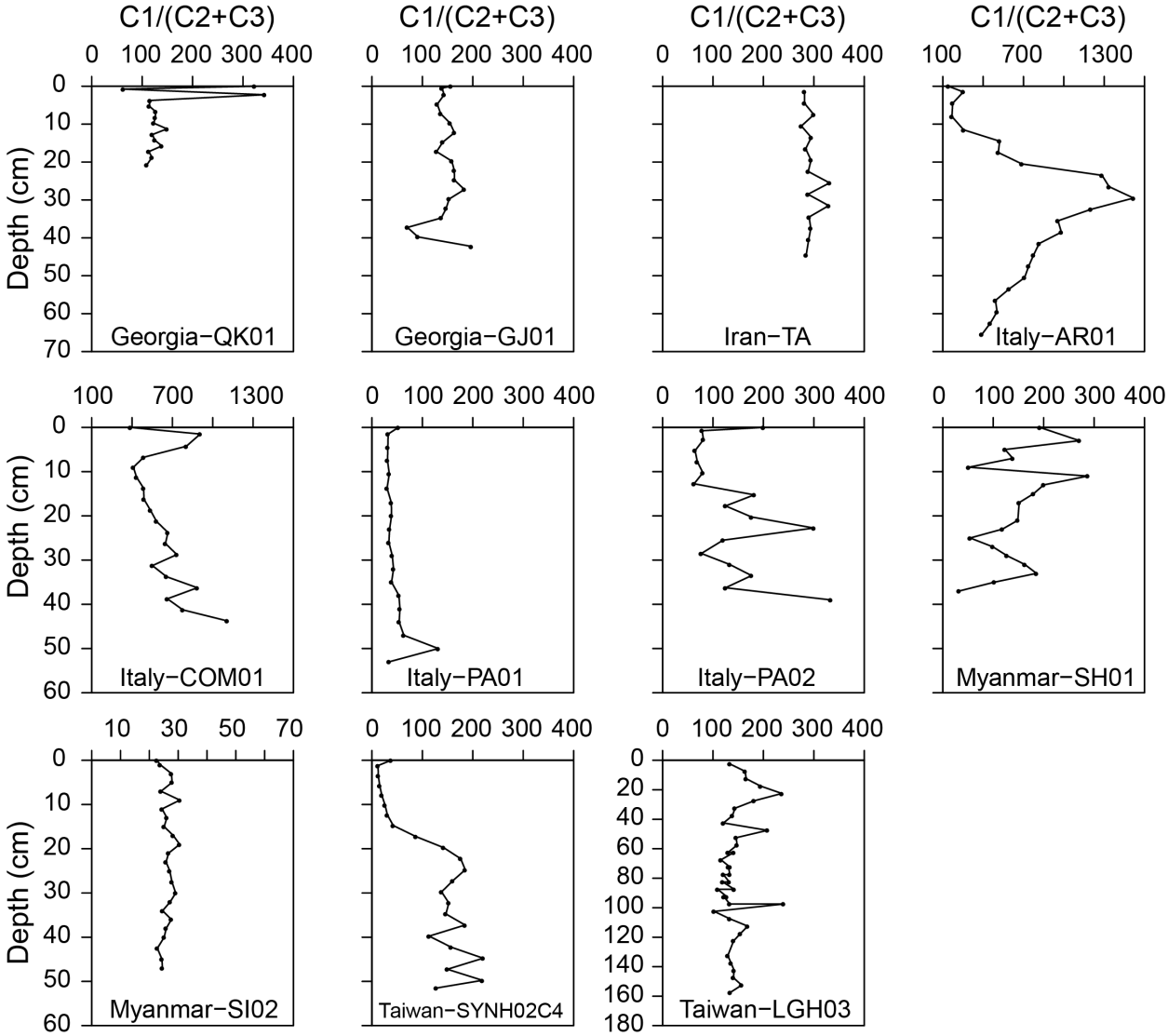

Figure S3: Plots of molar ratios of methane over the sum of ethane and propane (C1 / C2 + C3) versus depth. Data for SYNH02C4, SYNH02C11, and LGH03 were adopted from Tu et al. (2017) and Lin et al. (2018).

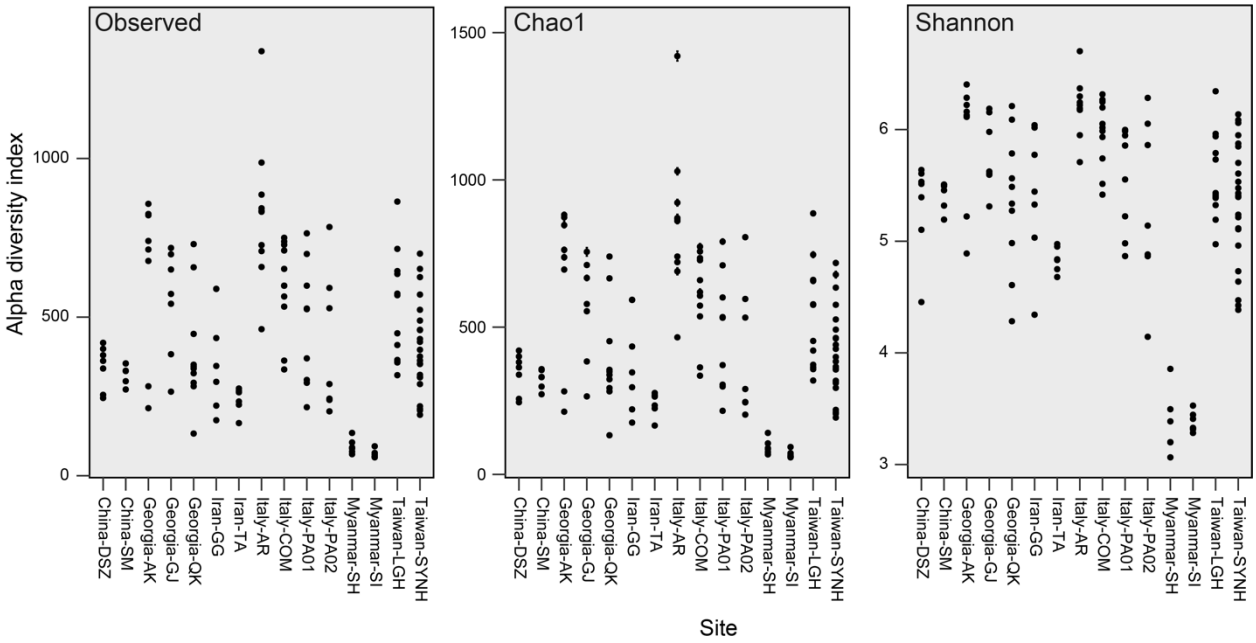

Figure S4: Alpha diversity indices calculated on the basis of rarefied dataset (n=9,413). No significant relationship was found between richness and sampling depth (Spearman's $\rho = 0.17$, $P > 0.01$ for the three indices).

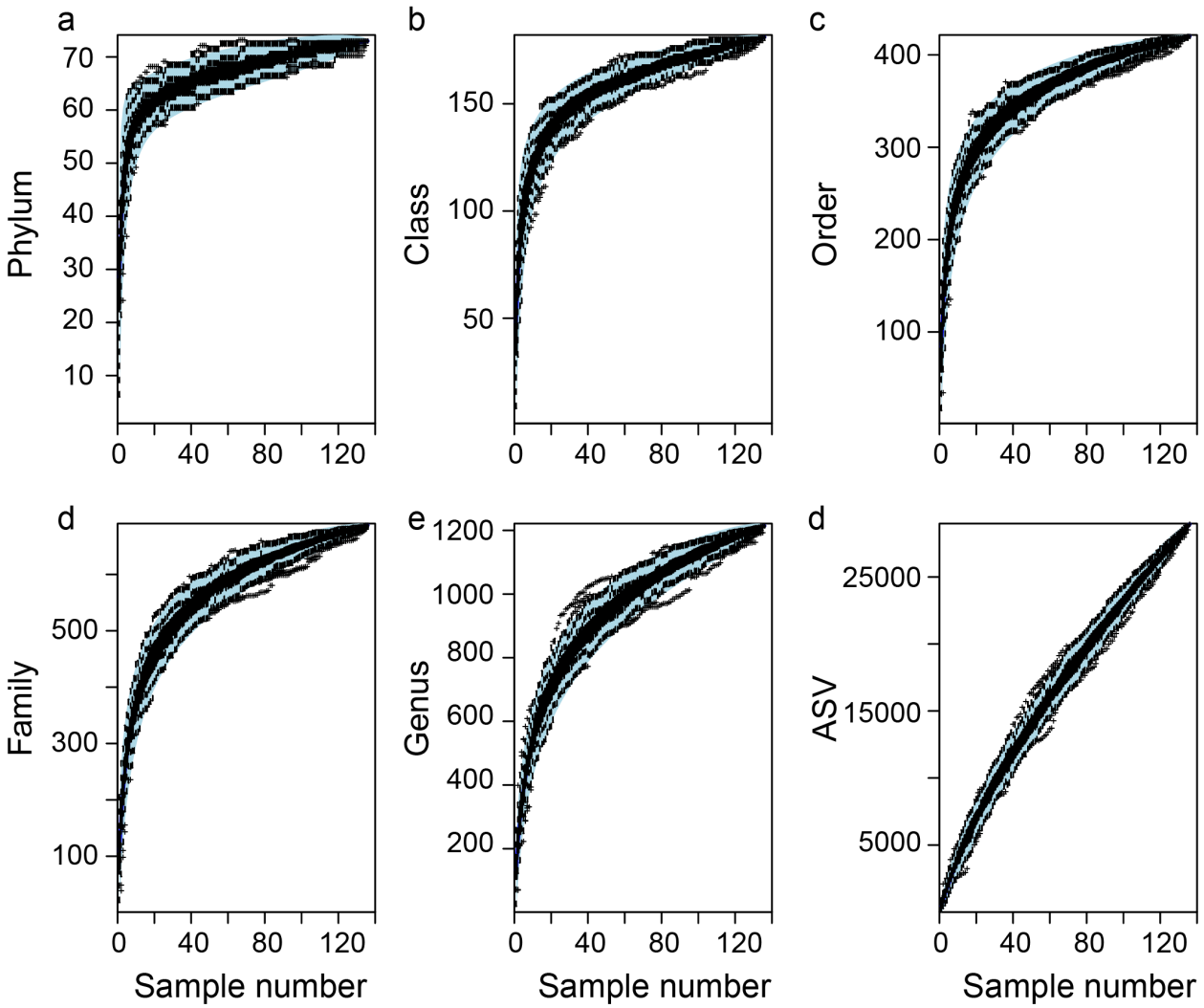

Figure S5: Accumulation curves for different taxonomic units: (a) phylum, (b) class, (c) order, (d) family, (e) genus, and (f) ASV. Boxplots show a summary of 100 permutations calculated with random subsampling. Absolute singletons were incorporated for comparison. Blue area depicts the 95% confidence interval.

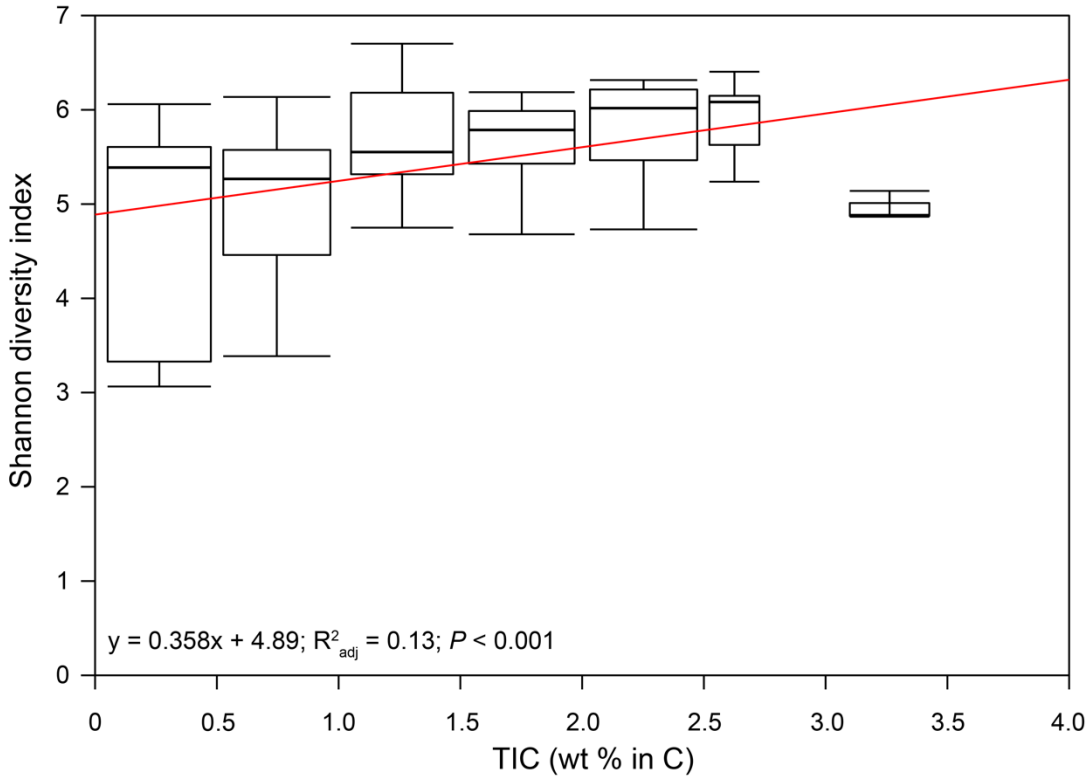

Figure S6: Plot of Shannon diversity versus TIC. Linear regression is shown in red (n=126). Box demonstrates the interquartile range that includes the first (25%), median (50%), and third (75%) quartiles. Lower and upper whiskers are the first and third quartiles minus and plus 1.5 times interquartile range, respectively.

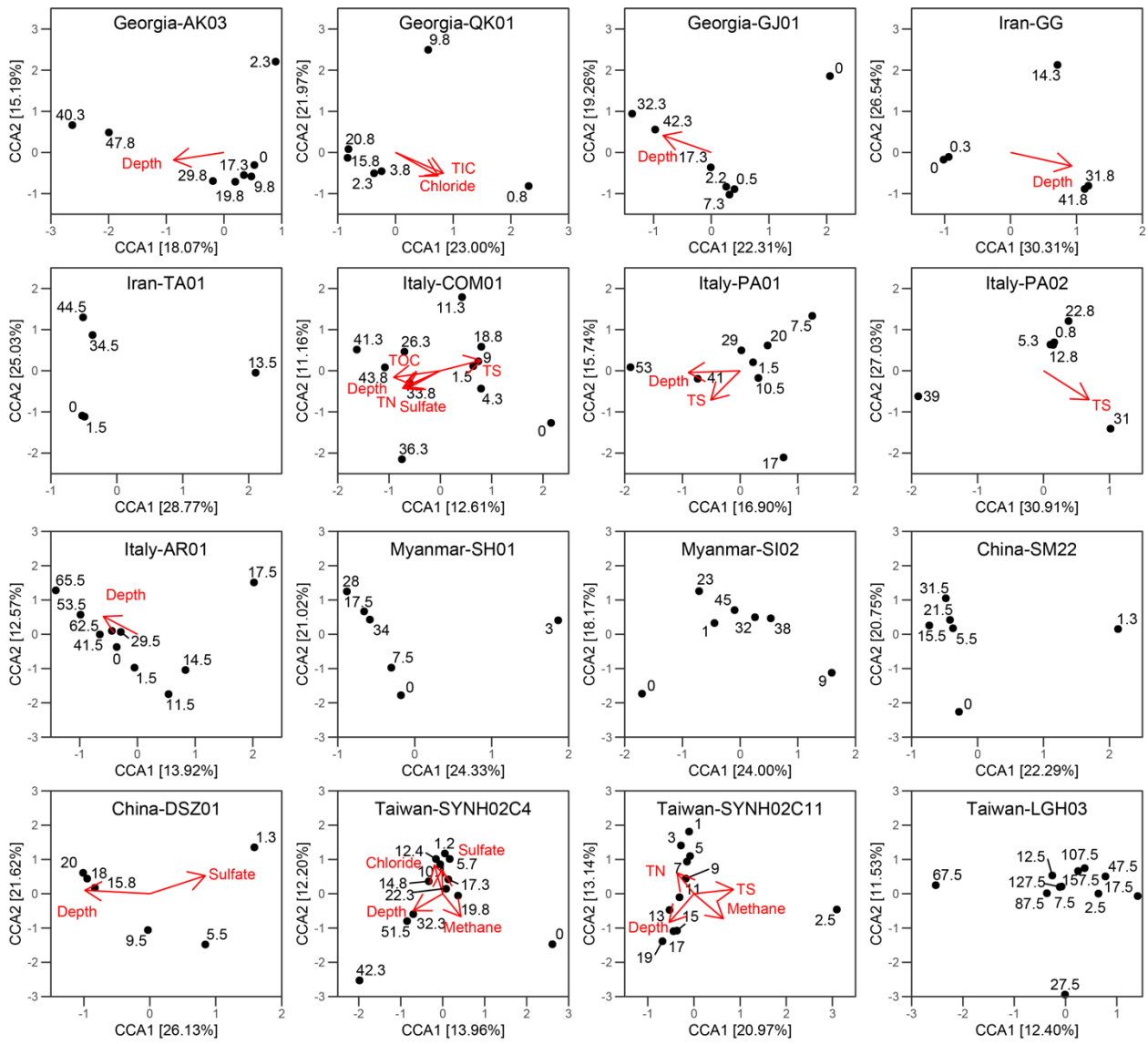

Figure S7: Constrained correspondence analysis of community relatedness quantified by the Chi-squared distance with the overlay of ordination for significant environmental parameters. Numbers next to each data point indicate the sampling depth (in centimeter).

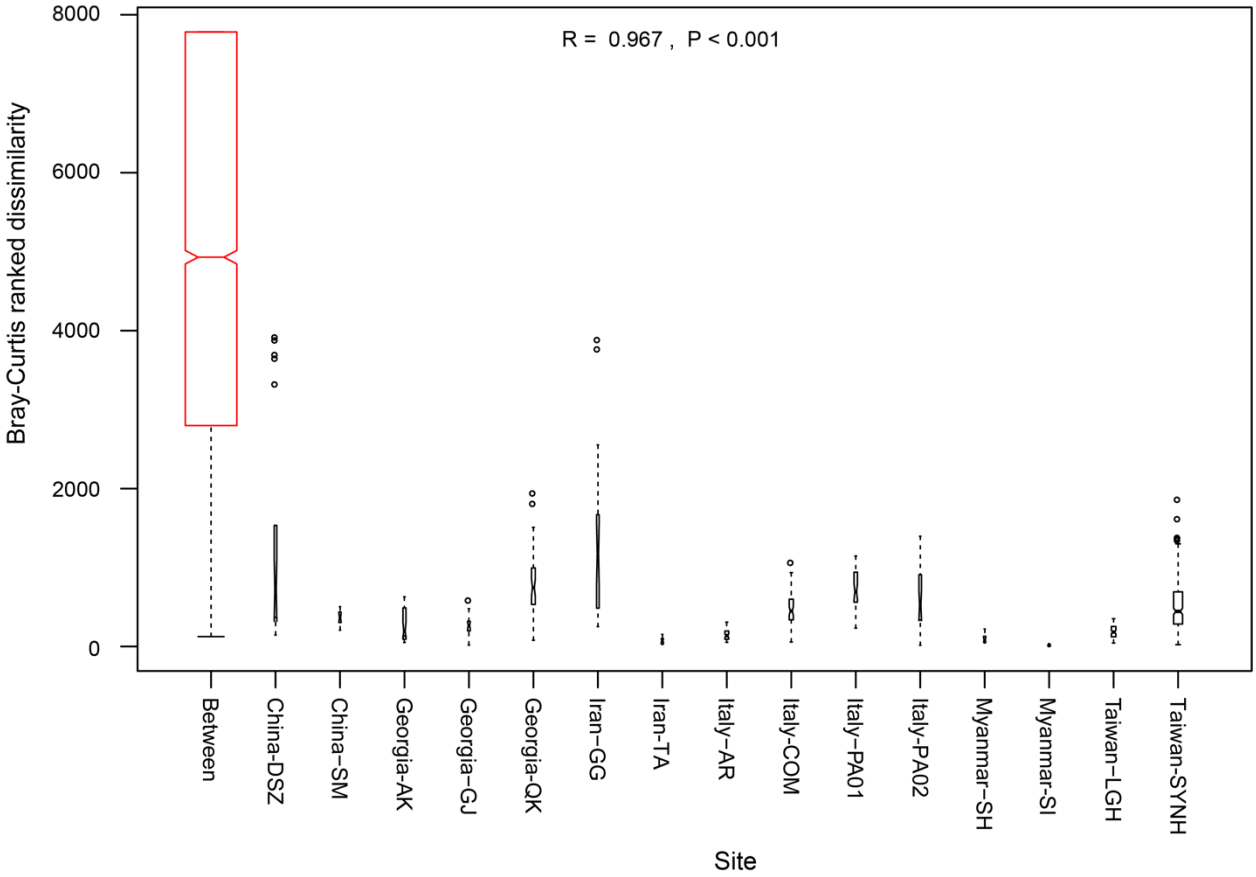

Figure S8: Analysis of similarities (ANOSIM: |R|) for community dissimilarity between all sites (in red) and within individual sites (in black). Lower and upper whiskers are first and third quartiles minus and plus 1.5 times interquartile range, respectively.

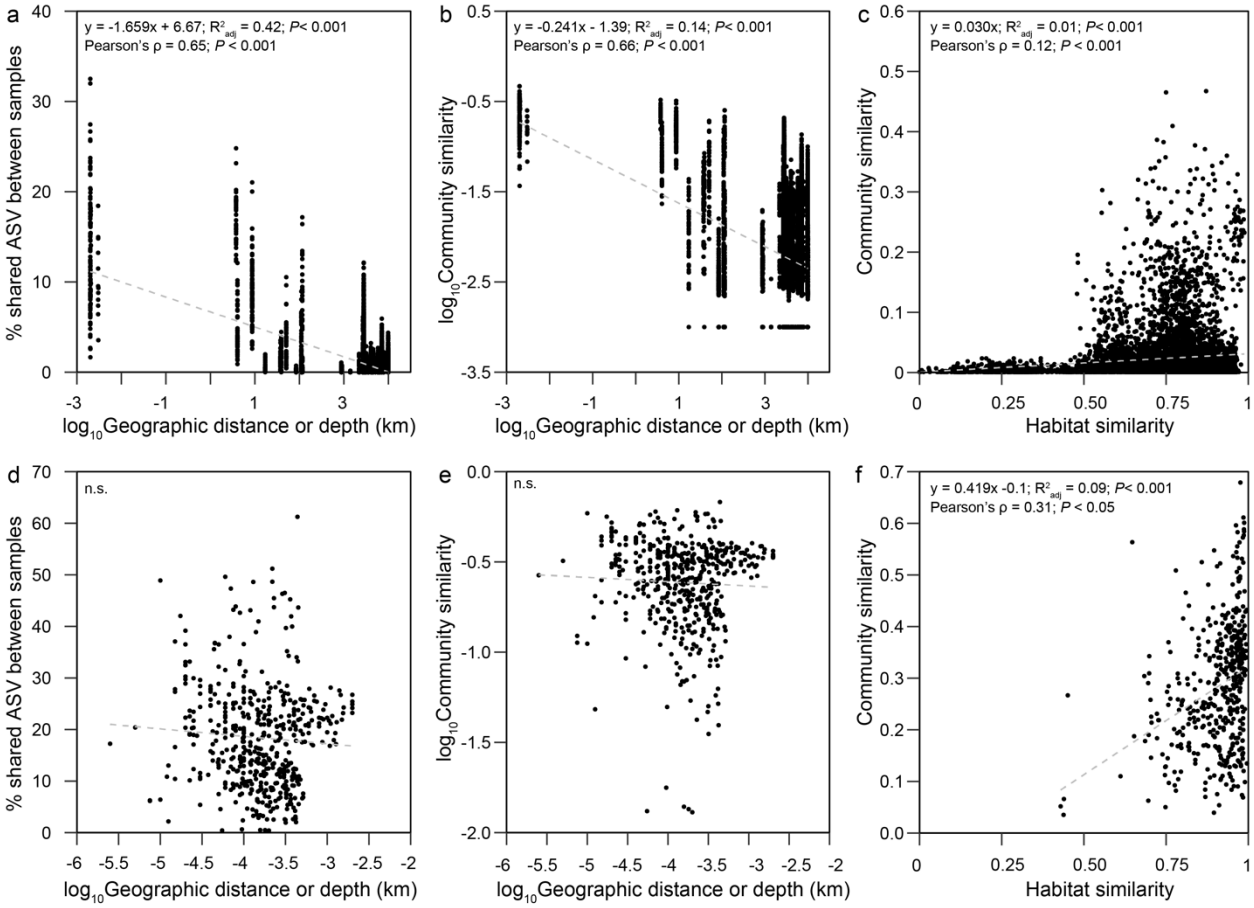

Figure S9: Geographic patterns and community similarity versus habitat similarity for communities across cores (a)–(c), and within cores (d)–(f).

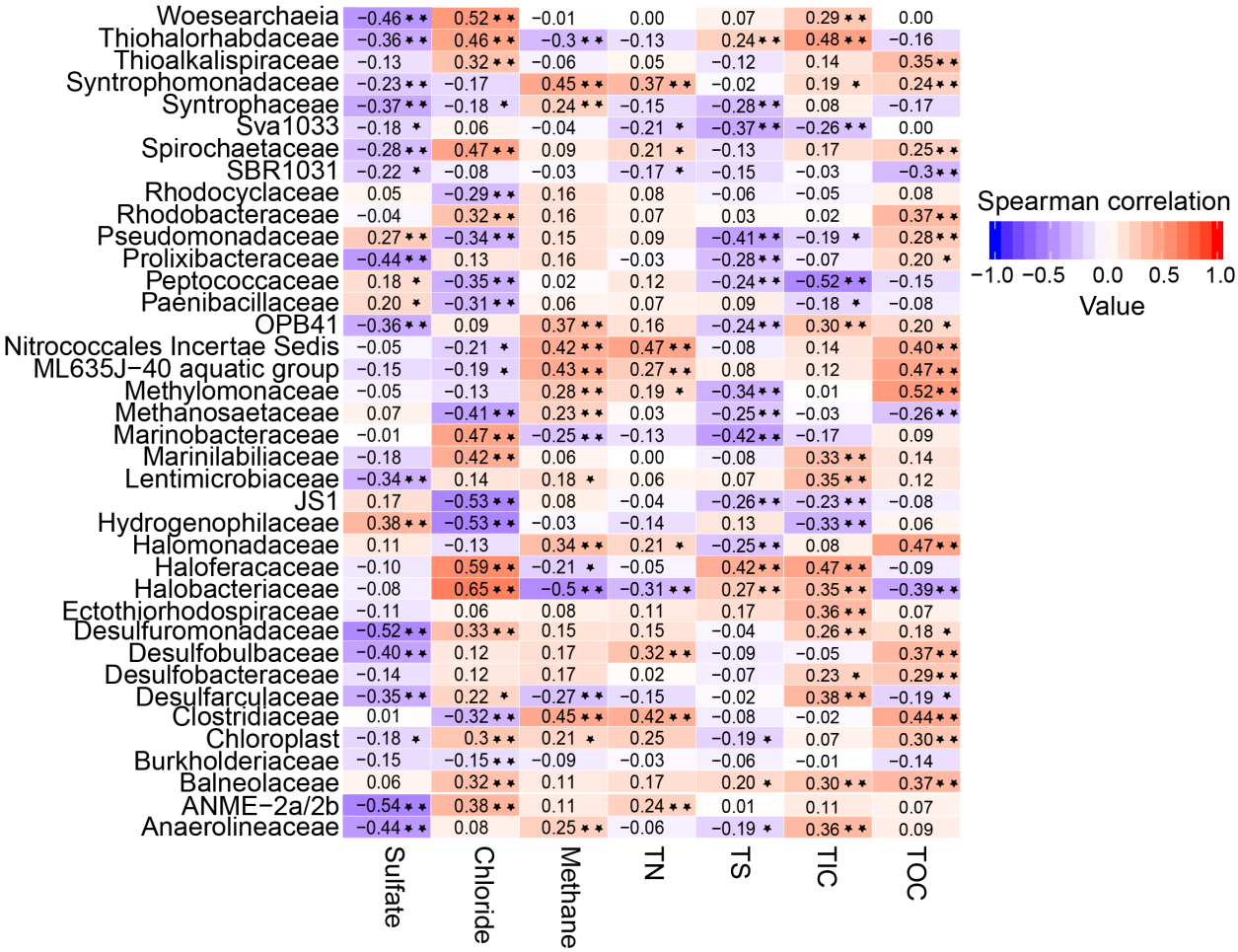

Figure S10: Color coded spearman correlation coefficients for concentrations of geochemical parameters and abundances of 38 major families. Major families are selected on the basis of the top 50 most abundant families. * and ** denote *P* values less than 0.01 and 0.05, respectively.

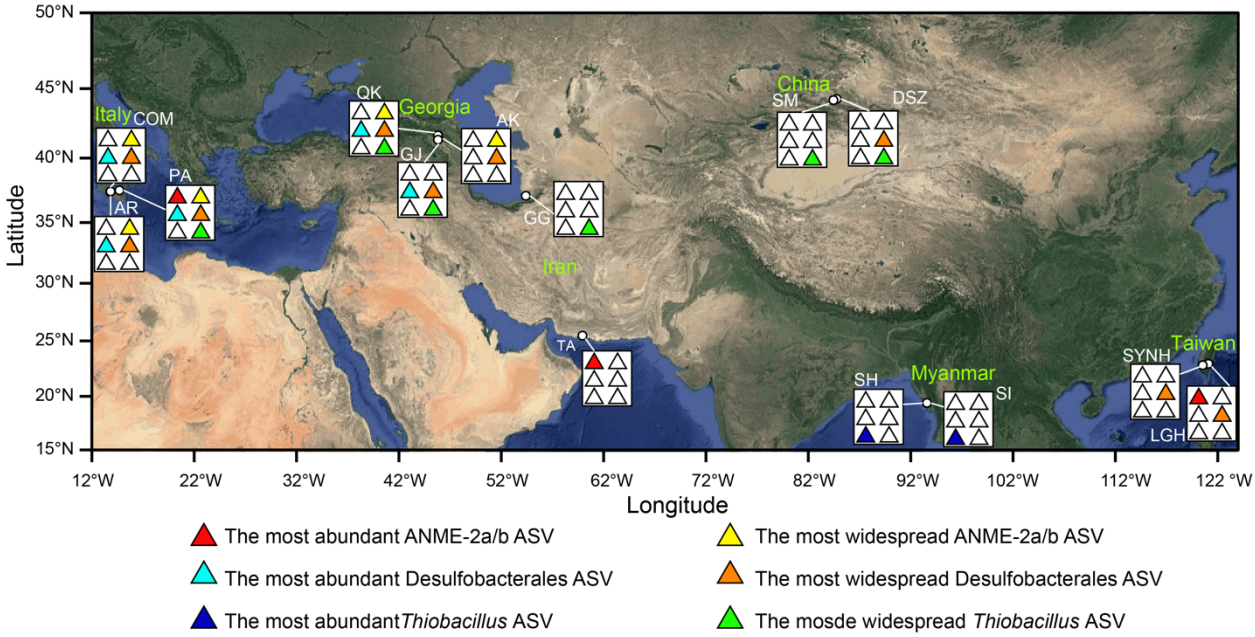

Figure S11: Occurrences of ASVs affiliated with key taxa likely involved in methane and sulfur cycling. Each sub-panel consists of six color codes indicating the presence or absence of six key ASVs likely involved in methane and sulfur cycling. These target ASVs include (1) the most abundant ANME-2a (in red), Desulfobacterales (in blue-green), and *Thiobacullus* (in blue) ASVs, and the most widespread ANME-2a (in yellow), Desulfobacterales (in orange), and Thiobacillus (in light green) ASVs. The basal map is from Google Maps © Google Maps 2021