Supporting Information Appendix

Piotr Starnawski, Thomas Bataillon, Thijs J. G. Ettema, Lara M. Jochum, Lars Schreiber, Xihan Chen, Mark A. Lever, Martin F. Polz, Bo B. Jørgensen, Andreas Schramm, and Kasper U. Kjeldsen (2017) Microbial Community Assembly and Evolution in Subseafloor Sediment.

Detailed Materials and Methods

Sampling. Sediment was sampled from 5 stations (M5, M24, M26, M27A and M29A) in Aarhus Bay by gravity coring (*SI Appendix* Table S1). Cores ranged from 5.5 to 10.8 m in length and were collected at water depths between 19 and 28 m (*SI Appendix* Table S1). The stations differed by the sediment depth of the sulfate-methane transition-zone (SMTZ: the depth at which sulfate becomes depleted and methane begins to accumulate). Subsamples for molecular biology analyses were collected with sterile, cut-off 5 mL syringes on ship or from intact cores in the laboratory less than 6 h upon their retrieval. The subsamples were immediately stored at -80 °C until processed for DNA extraction. Four to five subsamples from cores were used for molecular biology analyses, each subsample representing different geochemical zones of a given station (*SI Appendix* Table S1). Sediment subsamples for cell extraction were taken from intact core material stored at 4 ^{o}C for less than 24 h since coring. Sulfate and methane profiles and $^{35}S\text{-}SO_4^{2-}$ tracer-based sulfate reduction rates (1) were determined for all cores to map the geochemical zonation, these data are reported elsewhere (2).

DNA and cell extractions. DNA extractions for generating PCR amplicon sequence libraries were done according to a protocol, which included an initial washing step to remove extracellular DNA (3). For the metagenomic sequencing DNA was extracted from 5 g of sediment by the latter method, and in parallel by the PowerMax Soil DNA Isolation Kit (MoBio Laboratories) following the manufacturer´s instructions. Cells for single cell sorting were extracted from the sediment by density gradient centrifugation as previously described (4). Extracted cells were stored at -80 ^{o}C in 1x TE-buffer containing 5% (w/v) glycerol (final concentrations) until shipped on dry ice to the Bigelow Laboratory Single Cell Genomics Center (SCGC, https://scgc.bigelow.org).

16S rRNA gene sequence libraries. A ~283 bp- long fragment was PCR amplified from bacterial and archaeal 16S rRNA genes using the primers Univ519F/Univ802R (5, 6). This primer pair has excellent predicted coverage of archaeal and bacterial taxonomic diversity (7). PCR was performed in 25 µl-reactions using the KAPA HiFi PCR Kit (Kapa Biosystems) with 1-2 ng template DNA and 10 µM of each primer. Thermal cycling was as follows: initial denaturation at 95 °C for 5 min; 20 cycles of denaturing at 95 °C for 20 sec, annealing at 49 °C for 20 sec, and elongation at 72 °C for 25 sec; final elongation at 72 °C for 5 min. In a second round of PCR, each sample was supplied with unique Ion Xpress (Life Technologies) barcodes by using barcoded versions of the forward primer. PCR reaction conditions were modified as follows relative to the first round of PCR: 1 µl of the first round PCR product served as template in a total reaction volume of 50 µl, only 10 PCR cycles were run, and the annealing temperature was increased to 61 °C. PCR amplicons were purified using the Agencourt AMPure XP system (Beckman Coulter) with the recommended 1.8:1 bead-to-product ratio. For a few samples, additional E-Gel purification was performed (E-Gel iBase and E-Gel Safe Imager

Transilluminator system; Life Technologies) to remove short, unspecific PCR amplicons. Purified samples were quantified with the Qubit dsDNA BR Assay Kit and a Qubit 2.0 Fluorometer (Life Technologies). Amplicons were pooled in equimolar amounts and the size distribution and DNA concentration of the pooled amplicons were determined on a Bioanalyzer 2100 using the High Sensitivity DNA Analysis Kit (Agilent Technologies). Sequencing was done on an Ion Torrent PGM using 300 bp chemistry and two Ion 316 Chips (Life Technologies), with 15 samples per chip, according to the manufacturer's instructions.

The sequence data was initially analyzed within the Ion Torrent PGM software to remove reads of low quality and trim adaptor sequences. Subsequent sequence data analyses were performed using the usearch v7.0.959_i86osx32 (8, 9) and mothur v.1.36.1 (10) software. A high-quality (-fastq_maxee 1.0 and -fastq_trunclen 200) dataset for OTU clustering was prepared first by the fastq_filter command in uparse removing reads shorter than 200 bp. All reads were then called based on a perfect match with a barcode sequence using a custom shell script. Reads were clustered into operational taxonomic units (OTUs) with a 97% similarity cutoff. OTUs were assigned a taxonomy using mothur with the Silva SSU Ref NR release 123 database serving as a reference (11). All calculations and plots were done in the R statistical environment v3.2.0 (R Core Team, http://R-project.org) using RStudio v0.99.466 (RStudio Team, http://rstudio.com).

Mock community 16S rRNA gene sequence libraries. To test for contamination during our PCR and amplicon sequence library preparation procedures and to validate our OTU clustering and sequence analysis approach, we sequenced and analyzed a 3-species mock community in parallel with our environmental samples. The mock community consisted of 3 pGEM-T plasmids (Promega) mixed in equimolar amounts each to a final concentration of 10^6 plasmid copies μL^{-1}. Each plasmid carried a near full length 16S rRNA gene insert respectively of the archaeal species *Methanoccocus maripaludis* S2, an environmental alphaproteobacterium affiliated with the genus *Altererythrobacter* (for sequence, see *SI Appendix* Table S4) and an environmental Flavobacteriaceae bacterium (HM238120). The two environmental sequences were retrieved by PCR and cloning from a biofilter treating pig manure. A PCR product of this mock community was generated, sequenced and analyzed along with the PCR products from sediment samples using the procedures described in the section above. Our sequence analysis pipeline recovered 5 OTUs from the sequence library of the mock community. Four of the OTUs contained sequences 98-100% identical to the mock community input sequences. The *Altererythrobacter* affiliated input sequence gave rise to two OTUs. The 5th OTU was affiliated with *Escherichia coli* and thus represented a laboratory contaminant. *E. coli* affiliated sequence reads were not present in significant amounts in sequence libraries from sediment samples. The mock community sequence library consisted of 9,249 reads with the following distribution among detected OTUs: *M. maripaludis*-OTU: 3%; *Altererythrobacter*-OTUs 42% and 16%; Flavobacteriaceae-OTU: 38%; *E. coli*-OTU: 0.8%. The *M. maripaludis* input sequence was underrepresented in the sequence library likely as a result of PCR bias although the primers have perfect match to the sequence. To counteract this bias against Archaea, we treated bacterial and archaeal sequence reads separately in the analysis of our 16S rRNA gene sequence libraries and normalized their relative abundance by separate bacterial and archaeal qPCR assays. However the mock community sequence library showed that: (i) our amplicon sequence

libraries are minimally affected by laboratory contamination; (ii) the IonTorrent sequencing and our sequence analysis pipeline artificially inflates OTU richness only to a minor extent.

***dsrB* sequence libraries.** Approximately 350 bp long fragments of the *dsrB* gene, which encodes the beta subunit of the dissimilatory sulfite reductase and which represents a phylogenetic and functional marker gene for dissimilatory sulfate-reducing microorganisms (12), were PCR amplified from the same DNA extracts as used for generating the 16S rRNA gene PCR amplicon libraries. The PCR reactions were made with the primer variant mixtures dsrB-F1a-h and dsrB-4RSI1a-f (13), which targets most known *dsrB* sequence diversity (12). PCR reaction mixtures included 12.5 µL Hot Star Taq Master Mix (Qiagen), 1 µL of each primer variant mixture (containing 2 pmol μL^{-1} of each variant), 2 µL BSA (10 mg μL^{-1}) and 1 µL DNA template. Thermal cycling included 95 °C for 15 min; 40 cycles of 93 °C for 30 s, 56 °C for 30 s, 72 °C for 20 s and a final extension step of 72 ^{o}C for 10 min. PCR products were supplied with Ion Xpress barcodes (Life Technologies) and sequenced on an Ion Torrent PGM as described for the 16S rRNA gene sequence libraries.

Resultant *dsrB* sequence libraries were initially quality-filtered using mothur (10) removing sequences: (i) not carrying a perfect match to the barcode and one of the forward primer variants; (ii) sequences with ambiguous base calls or homopolymeric stretches exceeding 6 nt; and (iii) sequences shorter than 240 nt. The nucleotide sequences were corrected for homopolymer sequencing errors causing frameshifts in their inferred amino acid sequence using the FrameBot tool (14) of the Functional Gene Pipeline & Repository website (15). The inferred amino acid sequences were aligned against a custom-made reference set of aligned DsrAB sequences (12) using hmmalign of the HMMER 3.0 package (http://hmmer.org). The coding nucleotide sequences representing inferred amino acid sequences not aligning to the reference set were removed from further analysis. The *dsrB* nucleotide sequences were aligned using their respective aligned amino acid sequence as reference with a custom made Python script (http://python.org) kindly provided by Dr. Ian Marshall (Center for Geomicrobiology, Aarhus University, Denmark). The aligned nucleotide sequences were further denoised with the precluster algorithm implemented in mothur and then chimeric sequences were removed with the Chimera.uchime algorithm also implemented in mothur. The resultant final nucleotide sequence dataset were clustered into OTUs using the cluster_otus command in USEARCH v7.0.959_i86osx32 (8).

Quantitative PCR (qPCR). Total bacterial and archaeal 16S rRNA gene copies in the extracted DNA were quantified by SYBR-Green based qPCR. The primer pair Bac908F (5′-AACTCAAAKGAATTGACGGG-3′) (16) / Bac1075R (5′-CACGAGCTGACGACARCC-3′) (16), was used for the bacterial assay, and the primer pair Arch915F (5′-AATTGGCGGGGGAGCAC-3′) (17) / Arch1059R (5′-GCCATGCACCWCCTCT-3′) (18), for the archaeal assay. Reaction and thermal cycling conditions as well as qPCR standards were described previously (19). 16S rRNA gene copies of specific bacterial and archaeal taxonomic groups were quantified by SYBR-Green-based qPCR by designing primer pairs that specifically targeted a narrow range of predominant OTUs within these groups. The groups and respective primer sequences (5′-3′) were as follows: Atribacteria (F- GGTCTTAAAAGTCAGGTGTGA,

R- TCAGCGTCAGAGATAGACC); Marine Benthic Group-D archaea (F-GGTAATACCTGCAGCTCAG, R- CGTCAGACCCGTTCTAGC); Dehalococcoidea (F-GCGTAAAGAGGGCGTAGG, R- TCAGAAAYAGCCCAGGAGG). Thermal cycling, with an annealing temperature of 57 ^{o}C, and reaction conditions were similar to the total 16S rRNA gene quantification qPCR assays. Standards were prepared from pGEM-T (Promega) plasmids carrying a 16S rRNA gene insert of a representative member of the respective target groups by PCR amplifying the insert and flanking regions with a common M13 primer pair. The PCR products were purified (GenElute PCR Clean-up Kit, Sigma), and the concentrations of the purified products were determined by the Qubit dsDNA BR Assay Kit and a Qubit 2.0 Fluorometer (Life Technologies). The purified products were used for preparing 10-fold serial dilutions for standard curves. All standards were Sanger sequenced to ensure the integrity of the PCR priming sites using a commercial sequencing service (GATC Biotech). The standard curves were linear within a range of 10^2-10^8 target gene copies μL^{-1} template and PCR efficiencies as determined from the slope of standard curves ranged from 2.02-2.09. The specificity of the designed qPCR primer pairs was evaluated by cloning and sequencing end-point PCR products generated from DNA extracted from 25 cm sediment depth from station M5 in Aarhus Bay (*SI Appendix* Table S1). The insert of 31-36 clones from each clone library was Sanger sequenced on one strand (GATC Biotech). All sequenced clones represented the intended target group of a given assay.

Metagenome library construction and generation of single cell amplified genomes. For metagenomic sequence libraries DNA extracts were purified and concentrated with the DNA Clean and Concentrator kit (Zymo Research). Barcoded DNA sequence libraries were prepared using the Nextera DNA Library Preparation Kit (Illumina) according to the manufacturer´s protocol, with 50 ng of input DNA. Paired end sequencing was carried out on a HiSeq2500 instrument (Illumina) working in rapid mode using 2x 150 bp chemistry for libraries generated from 25 and 75 cm sediment depth and 2x 250 bp chemistry for libraries generated from 125 and 175 cm depth. We quantified sequencing error rates by analyzing the sequencing reads from the phiX174 phage genomic DNA added to all sequencing runs as an internal control as part of the standard Illumina sequencing protocol. These reads were mapped onto the reference phiX174 phage genome sequence (NCBI GI:764044028) using BBMap v34.x (Bushnell B. – http://sourceforge.net/projects/bbmap) and single nucleotide polymorphisms (SNPs) were identified with the SAMtools software v1.2 (20). Per site error rate were calculated by counting the number of SNPs per nucleotide position in the genome and dividing this with the total read coverage of that position. Mean sequencing error rates (±SD) ranged from 0.00022 (± 0.00015) to 0.00022 (±0.00010), which is much lower than the observed levels of genome nucleotide diversity (0.002-0.008; Fig. 3A). In addition, we observed a high proportion of SNP positions, which are conserved across sediment depth (*SI Appendix* Fig. S7*A*). These results prove that the influence of sequencing errors is negligible for the interpretation of our data.

Single-cell sorting, whole-genome amplification, and 16S rRNA gene PCR screening of single cells were performed at the SCGC using their established protocols as previously described (4). Single cells were sorted on two 384-well microplates and produced a total of 242 successful SAGs. Barcoded sequence library preparation from SAGs was done by the Nextera XT DNA

Library Preparation Kit (Illumina), with 1 ng of input DNA. The sequencing was performed on a MiSeq instrument with paired-end 2x 300 bp chemistry.

SAG and metagenome assembly and analysis. Metagenomic libraries were prepared and sequenced from DNA extracted by two independent methods from the same homogenized sediment samples. Reads from the two libraries from a given sample were pooled and analyzed. Sequence reads from metagenomes and SAGs were processed with the trimmomatic software v0.32 (21) for adapter trimming and quality filtering (flags used: Paired End [PE], ILLUMINACLIP:nextera_adaptors.fasta:2:30:10 for the adaptor trimming and SLIDINGWINDOW:5:30 for the quality filtering). SAGs were assembled using SPAdes v3.5.0 (22) with the --careful flag. Annotation was performed using the IMG annotation service (23). If 16S or 23S rRNA sequences were not present in a SAG assembly its taxonomy was inferred from its predicted set of conserved single copy genes identified with Phylosift v1.0.1 (24). Mapping of metagenomic reads on to SAG assemblies was done in BBMap v34.x with a lower identity cutoff of 0.95. Mapped reads in .sam format were converted to .bam format. SNPs were called using the SAMtools software v1.2 (20). All calculations were done in RStudio.

Estimation of genomic mutation rates. Per generation mutation rates (μ_g) were calculated according to $\pi = 2N_e \times \mu_g$ (25, 26) assuming that effective population size (N_e) equals the total observed population size. π values (SNPs base pair^{-1}) were calculated by mapping metagenomic reads onto predicted genes in SAG assemblies. Predicted genes included in the analyses were carefully selected by a stringent procedure. Only genes mapped by metagenomes from all four sediment depths and which had more than 5x metagenomic coverage across minimum 60 bp long gene fragments were considered in the analyses. Furthermore, only genes from SAGs with more than 10 genes that met the former criteria were included in the final reported results. The final dataset included 2,264 genes from 12 SAGs each contributing between 48 and 388 genes which were mapped with $4 \cdot 10^3$ to 10^6 metagenomic reads (*SI Appendix* Table S3). The π values were scaled to yield a total number of genome pairwise nucleotide differences (SNPs genome^{-1}) by multiplying them with the estimated size of the genomes represented by the individual SAGs. Size estimates were obtained by dividing the size of the SAG assemblies with their estimated respective completeness taking the average values for all SAGs of a given taxonomic lineage (*SI Appendix* Table S3). Finally, a total mutation rate per genome per generation (SNPs generation^{-1}) was calculated by dividing scaled π values with the estimated average number of microbial cell generations separating a given sediment depth from the surface sediment derived as described in the "Estimation of rates of microbial biomass turnover" section below. The total population size was calculated as the product of the relative abundance of metagenomic reads mapping to a given SAG and the total cell density as obtained from fluorescence microscopy counts of DNA stained cells (27).

P_n/P_s ratios. The ratio of nonsynonymous to synonymous number of SNPs (P_n/P_s) was calculated by mapping metagenomic sequence reads from 25 and 175 cm sediment depth onto genes predicted in the SAG assemblies. The genes included in the analyses were the same as those selected for estimating mutation rates. Given that genes had typically low counts of SNPs we standardized counts by adding 1 to all counts of SNPs in each gene. This reduces the

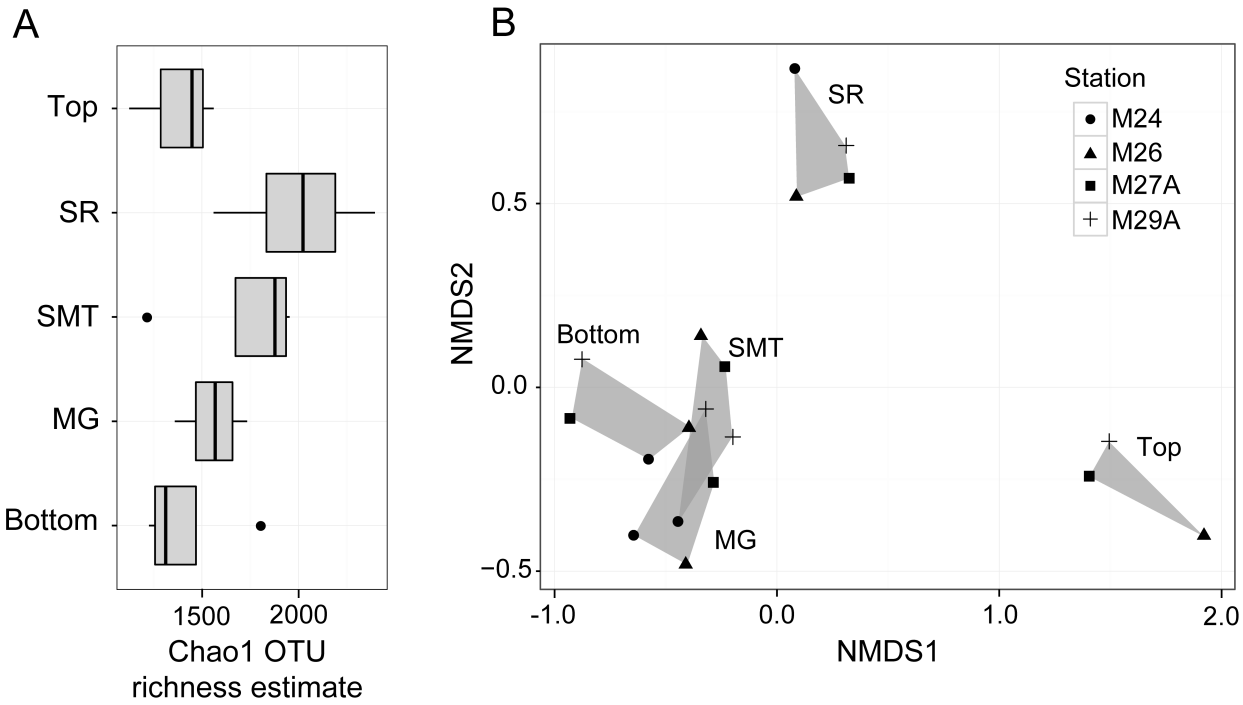
statistical noise of these ratios but for a given SAG-defined lineage still allows for proper comparisons of P_n/P_s values obtained from different sediment depths (see also reference [28] for further discussion). The P_n/P_s ratios were further standardized to account for the fact that, given the structure of the genetic code, a non-synonymous mutation event is approximately 3-fold more likely than a synonymous mutation event. Thus P_n/P_s was calculated by the following formula: $P_n/P_s = [(1+P_n)/3]/(1+P_s)$. To test the hypothesis of relaxation of purifying selection across depths, we counted the number of genes exhibiting a higher P_n/P_s ratio at 175 relative to 25 cm depth. A *P*-value for the hypothesis of "no change in purifying selection between depths" was obtained (binomial exact test). The predicted genes included in the analyses were grouped according to their inferred functional role by classifying them against the eggNOG database, release May 2015 (29). The sequence searches against the eggNOG database were performed with the hmmsearch program of the HMMER software package (30), recording only the best match based on the expected value.

Estimation of rates of microbial biomass turnover. The average rate by which cells turn over their biomass in subsurface sediments can be estimated from the rate of carbon oxidation in the sediment and the total cell densities (31). The rate of biomass turnover serves as a proxy for the generation times of cells assuming that the cells indeed are growing by cell division in the energy-depleted subseafloor as opposed to being in a non-growth state where cells merely repair and sustain their biomolecules (32). Depth profiles of organic carbon oxidation rates were calculated from a power law-based model relating sediment depth and sulfate reduction rates. The model is based on measurements of sulfate reduction rates by ^{35}S-tracer techniques at 12 different sampling Stations in Aarhus Bay (2). The rates of carbon oxidation were determined from the sulfate reduction rates assuming 2 moles of carbon oxidized per mole of sulfate reduced. Cell densities were derived from qPCR quantification of bacterial and archaeal 16S rRNA genes for stations M24, M26, M27A and M29A (33) assuming that bacterial and archaeal cells on average harbor 4.1 and 1.6 copies of the 16S rRNA gene, respectively (34). For station M5, cell densities were determined by epifluorescence microscopy cell counts (27). Average cell specific rates of carbon oxidation were converted to rates of cellular biomass turnover assuming a growth yield of 8% (2) and a cellular carbon content of 21.5 fg (35). The relationship between sediment depth and age was determined from ^{14}C dating of bivalve shells for station M5 (27). For station M24, M26, M27A and M29A the relationship was determined from average rates of sedimentation at the individual stations (2). The latter assumes that sedimentation was constant across time in Aarhus Bay, and is supported by the ^{14}C-based age model for station M5. By relating sediment age and the rate of biomass turnover we then calculated the cumulative number of cell generations across depths of the various sampling stations. Extensive sediment mixing by bioturbation extends down to around 7 cm below sea floor in Aarhus Bay (36). Therefore the residence time of a cell in the upper mixed layer of the sediment is unknown, and our calculation of cell generation times thus begins at 7 cm depth. For the same reason, the evolutionary analysis of the persister lineages does first start at 25 cm depth, well below the mixed layer.

SI Figures

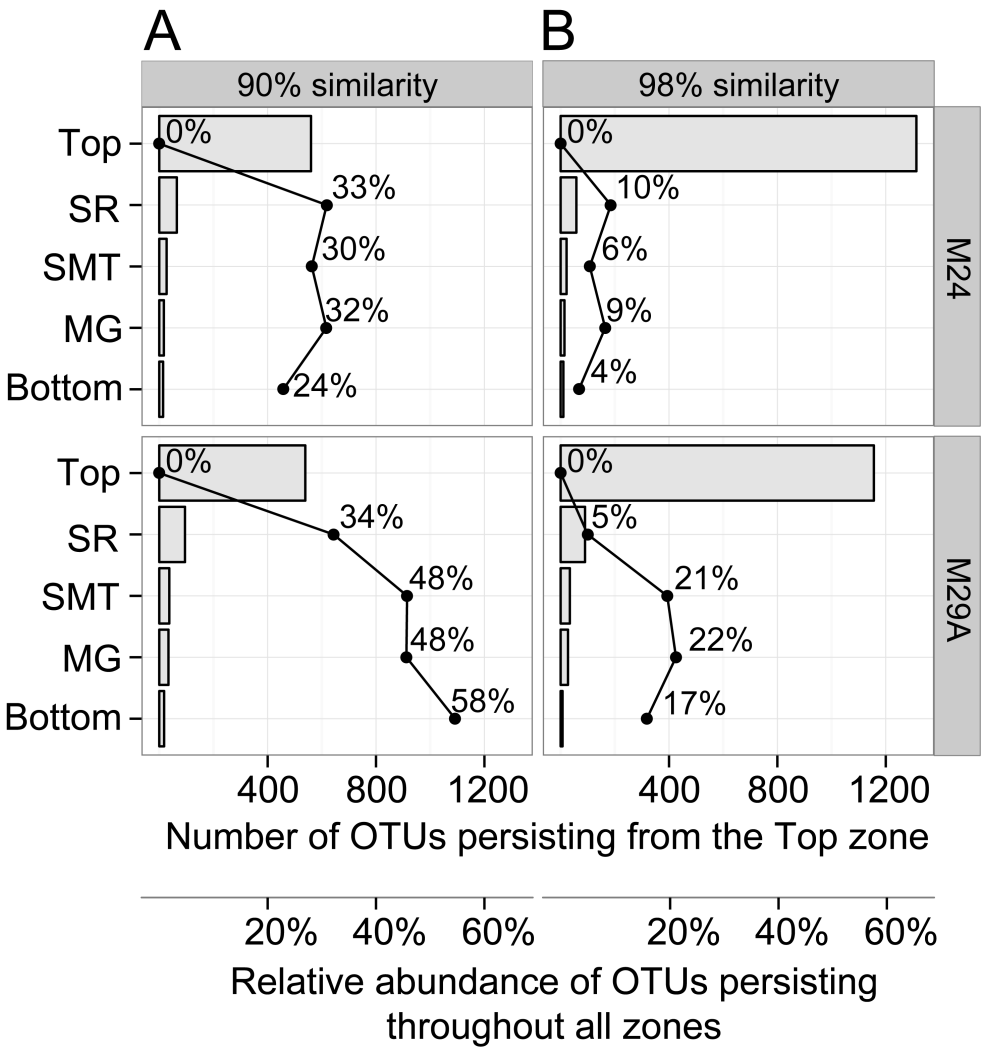
SI Figure S1.

Changes in species richness and composition across geochemical zones. (*A*) Chao1 estimates for 16S rRNA gene sequence OTU (97% cut-off) richness in different geochemical zones. In the Box-and-Whisker plots the middle line represents the median value for four different sampling stations and vertical lines minimum and maximum values. The box stretches from the lower to the upper quartile and dots show outliers (>1.5 times the corresponding quartile). The low Chao1 estimate for the Top zone is likely a sampling artifact caused by the uneven species distribution characteristic for this zone with a few highly predominant OTUs dominating the sequence libraries (*SI Appendix* Table S1). (*B*) Comparison of microbial community compositions from different geochemical zones and sampling stations by non-metric multidimensional scaling (NMDS) ordination of 16S rRNA gene sequence OTU frequencies. The NMDS ordination was made with a Wisconsin normalized Bray-Curtis dissimilatory matrix calculated from a subsampled data set (n=5,814 sequences per sample). Top = surface sediment; SR = upper sulfate-rich sediment; SMT = sulfate-methane transition zone; MG = methanogenic zone; Bottom = deep methanogenic zone.



SI Figure S2.

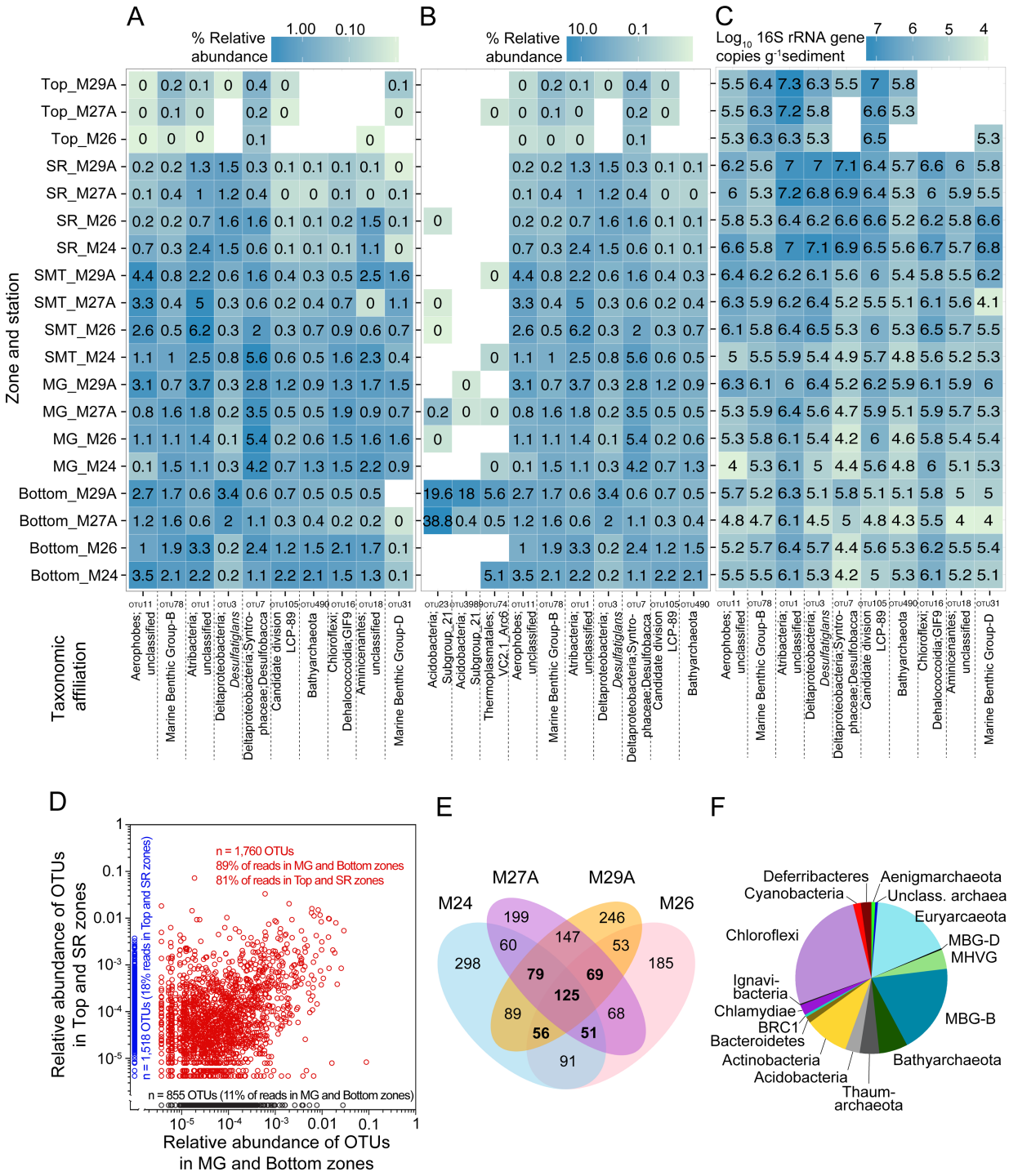
Persisting *dsrB* OTUs. Number of *dsrB* OTUs persisting across depth in sediment cores from two sampling stations (M24 and M29A). Bars show the number of OTUs shared with the Top zone (i.e. persisting OTUs). Filled circles connected by lines show the frequency of sequences assigned to persisting OTUs present in all zones of a given station. Sequences were clustered into OTUs based on either (*A*) a 90% sequence identity cut-off (corresponding to a 97% 16S rRNA gene sequence cut off (12) or (*B*) a 98% identity cutoff. Top = surface sediment; SR = upper sulfate-rich sediment; SMT = sulfate-methane transition zone; MG = methanogenic zone; Bottom = deep methanogenic zone.



SI Figure S3.

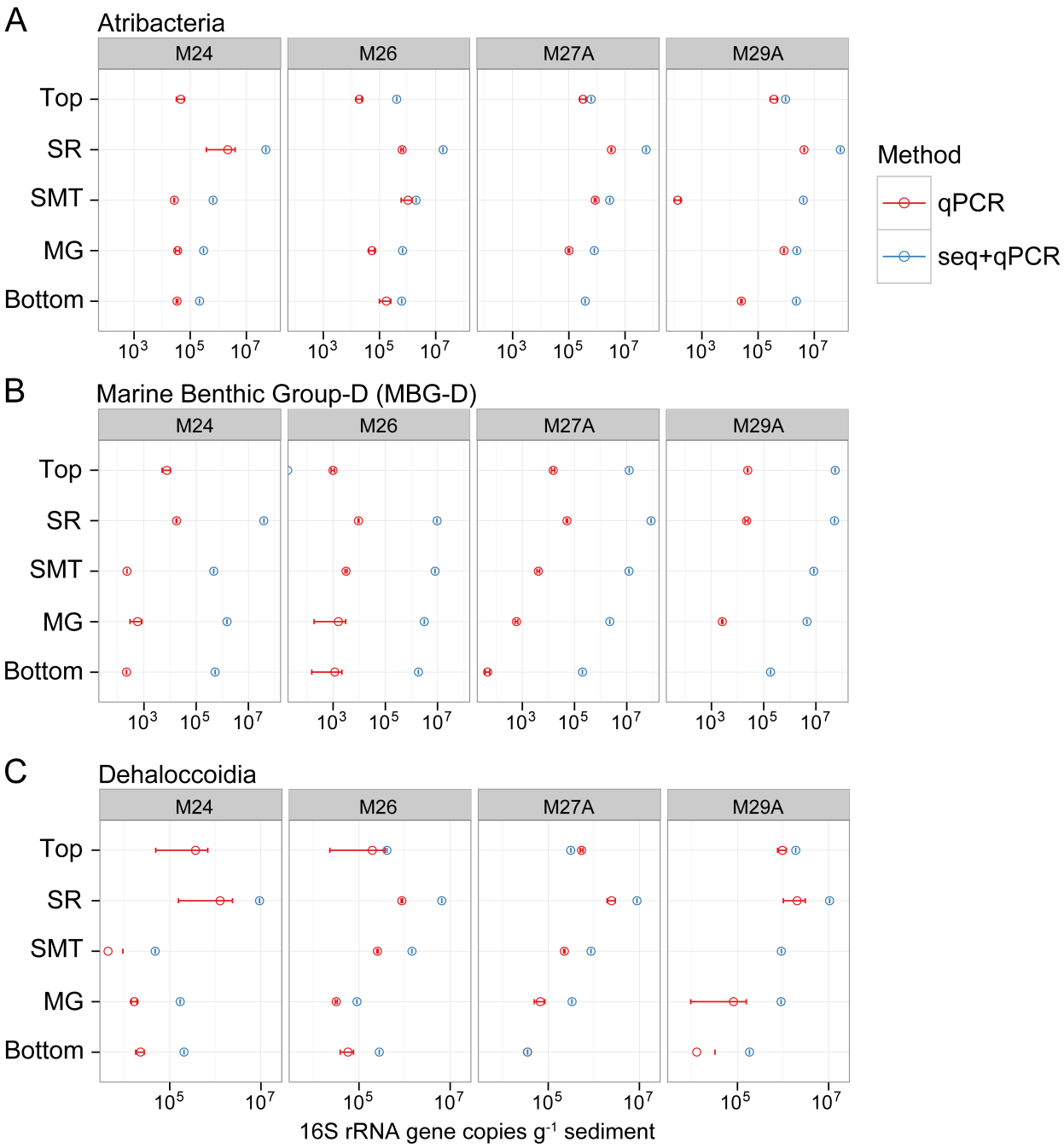
Depth distribution of persisting 16S rRNA gene sequence OTUs (A-C) and identity of OTUs uniquely present in subsurface part of sediments (D-F). Sequences were clustered into OTUs with a 97% cut-off and originate from PCR amplicon sequence libraries from different geochemical zones of sediment cores from four different sampling stations (M24, M26, M27A, M29A) in Aarhus Bay (SI Table S1). (*A*) The top ten most abundant OTUs persisting across all zones, relative abundances and (*B*) Top ten most abundant OTUs present in the deepest zone, relative abundances. In addition (*C*) absolute abundances of OTUs shown in *A* were estimated by multiplying the library-based relative abundance of individual OTUs with qPCR-derived total bacterial or archaeal 16S rRNA gene abundances. The validity of this approach was confirmed by independent qPCR quantification of selected groups (*SI Appendix* Fig. S4). White cells represent the absence of a given OTU at a given sample and relative/absolute abundance of "0" means that this value was below 0.1 but still detectable. (*D*) Relative abundance of 16S rRNA gene sequence OTUs in sequence libraries from the surficial sediment (Top and SR zones) and subsurface sediment (MG and Bottom zones). Shown values represent the average abundance of a given OTU in the Top and SR or the MG and Bottom zones across all 4 stations. Data points shown in blue color represent OTUs present in the surficial sediments but absent in the subsurface sediments. Conversely data points shown in black color represent OTUs only detected in the subsurface sediments. The axes show $\log_{10}$ scales, except for the origo, which show a value of 0 (i.e. the absence of OTUs). (*E*) Distribution of the 855 OTUs unique to the subsurface sediment across the zones of the 4 different stations. A set of 380 OTUs is present in the subsurface part of ≥3 of the four stations. These 380 OTUs make up 8-14% of the subsurface communities at the 4 individual stations; their taxonomic distribution is shown in panel (F), where shown values represent averages for the 4 stations. Top = surface sediment; SR = upper sulfate-rich sediment; SMT = sulfate-methane transition zone; MG = methanogenic zone; Bottom = deep methanogenic zone. Taxonomic classification according to Silva taxonomy (11).

SI Figure S3.


A
% Relative abundance
1.00
0.10
B
% Relative abundance
10.0
0.1
C
Log10 16S rRNA gene copies g-1sediment
7
6
5
4
Zone and station
Top_M29A
Top_M27A
Top_M26
SR_M29A
SR_M27A
SR_M26
SR_M24
SMT_M29A
SMT_M27A
SMT_M26
SMT_M24
MG_M29A
MG_M27A
MG_M26
MG_M24
Bottom_M29A
Bottom_M27A
Bottom_M26
Bottom_M24
Taxonomic affiliation
Aerophobes; unclassified
Marine Benthic Group-B
Atribacteria; unclassified
Deltaproteobacteria; Desulfatiglans
Deltaproteobacteria;Syntrophaceae;Desulfobacca
Candidate division LCP-89
Bathyarchaeota
Chloroflexi; Dehalococcoidia;GIF9
Aminicenantes; unclassified
Marine Benthic Group-D
Acidobacteria; Subgroup_21
Acidobacteria; Subgroup_21
Thermoplasmatales; VC2.1_Arc6
D
Relative abundance of OTUs in Top and SR zones
n = 1,760 OTUs
89% of reads in MG and Bottom zones
81% of reads in Top and SR zones
n = 1,518 OTUs (18% reads in Top and SR zones)
n = 855 OTUs (11% of reads in MG and Bottom zones)
Relative abundance of OTUs in MG and Bottom zones
E
M24
M27A
M29A
M26
199
246
60
147
53
298
79
69
185
125
89
68
56
51
91
F
Deferribacteres
Aenigmarchaeota
Cyanobacteria
Unclass. archaea
Euryarcaeota
Chloroflexi
MBG-D
MHVG
Ignavi-bacteria
Chlamydiae
MBG-B
BRC1
Bacteroidetes
Actinobacteria
Bathyarchaeota
Acidobacteria
Thaum-archaeota

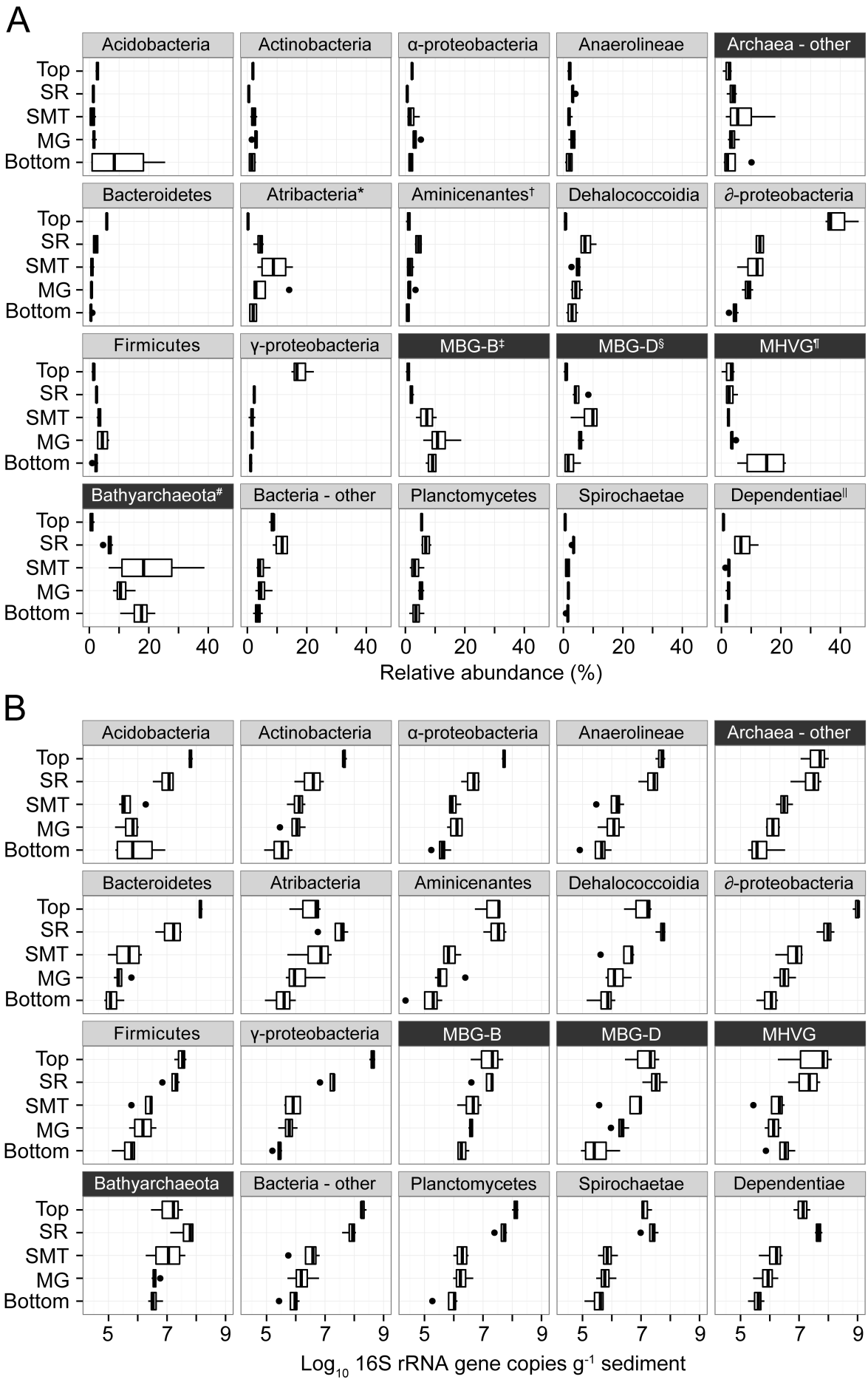
SI Figure S4.

Depth distribution of taxonomic groups in sediment cores from Aarhus Bay. Comparison of group-specific 16S rRNA gene qPCR-derived abundances (qPCR) and abundances estimated by multiplying the 16S rRNA PCR amplicon library-based relative abundance of individual groups with qPCR-derived total bacterial (Atribacteria [*A*] and Dehalococcoidia [*C*]) or archaeal (MBG-D [*B*]) 16S rRNA gene abundances (seq+qPCR). Top = surface sediment; SR = upper sulfate-rich sediment; SMT = sulfate-methane transition zone; MG = methanogenic zone; Bottom = deep methanogenic zone. Error bars show SD (n=3).

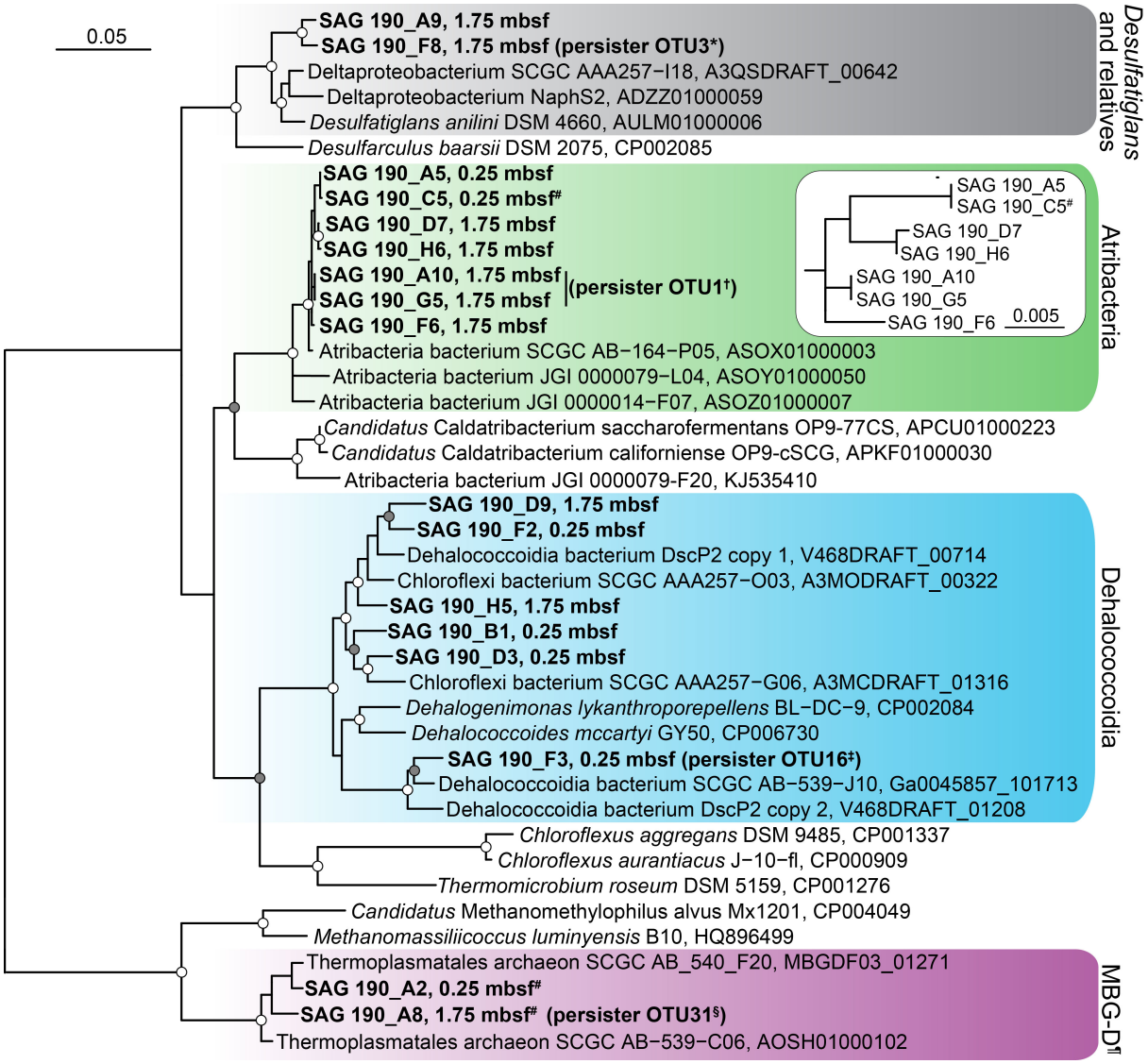


SI Figure S5.

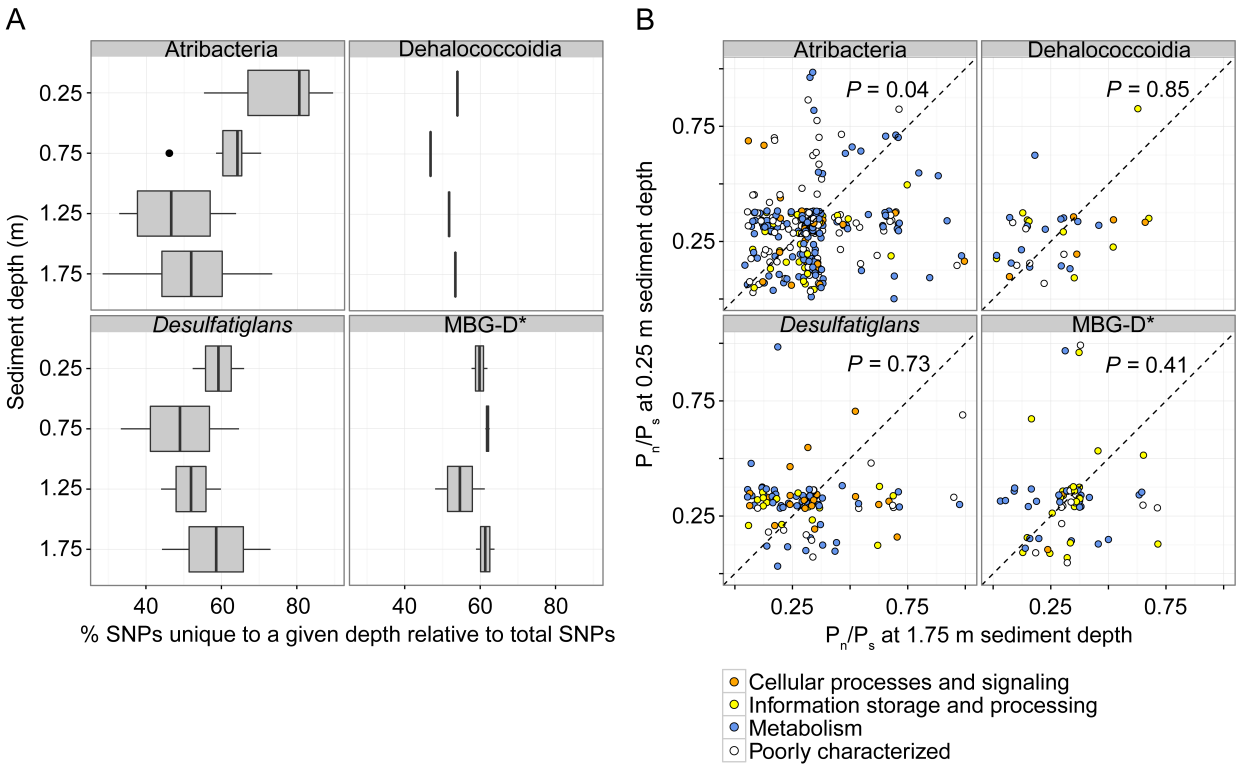
Depth distribution of major taxonomic groups across geochemical zones. (*A*) 16S rRNA PCR amplicon library-based relative abundances. (*B*) Absolute abundances estimated by multiplying the relative abundances shown in *A* by qPCR-derived total bacterial or archaeal 16S rRNA gene abundances. The Box-and-Whisker plot shows values for four different sampling stations as described for *SI Appendix* Fig. S1. The validity of this approach was confirmed by independent qPCR quantification of selected groups (*SI Appendix* Fig. S4). Top = surface sediment; SR = upper sulfate-rich sediment; SMT = sulfate-methane transition zone; MG = methanogenic zone; Bottom = deep methanogenic zone. Names of archaeal groups are show in white on a darkgrey background. *Formerly named: Candidate phylum JS1, †Formerly named: Candidate phylum OP8, ‡Marine Benthic Group-B, §Marine Benthic Group-D, ¶Marine Hydrothermal Vent Group, #Formerly named: Miscellaneous Crenarchaeotal Group, ||Formerly named: Candidate phylum TM6.


A
Acidobacteria
Actinobacteria
α-proteobacteria
Anaerolineae
Archaea - other
Bacteroidetes
Atribacteria*
Aminicenantes†
Dehalococcoidia
∂-proteobacteria
Firmicutes
γ-proteobacteria
MBG-B‡
MBG-D§
MHVG¶
Bathyarchaeota#
Bacteria - other
Planctomycetes
Spirochaetae
Dependentiae‖
Top
SR
SMT
MG
Bottom
0
20
40
Relative abundance (%)
B
Acidobacteria
Actinobacteria
α-proteobacteria
Anaerolineae
Archaea - other
Bacteroidetes
Atribacteria
Aminicenantes
Dehalococcoidia
∂-proteobacteria
Firmicutes
γ-proteobacteria
MBG-B
MBG-D
MHVG
Bathyarchaeota
Bacteria - other
Planctomycetes
Spirochaetae
Dependentiae
Top
SR
SMT
MG
Bottom
5
7
9
Log10 16S rRNA gene copies g-1 sediment

SI Figure S6. Phylogenetic affiliation of 16S rRNA gene sequences from single cell amplified genomes (SAGs, highlighted in bold). SAGs originate from 0.25 or 1.75 m sediment depth of Aarhus Bay station M5. The tree was inferred by FastDNA maximum likelihood analysis, as implemented in the ARB program package (37) considering 1,277 conserved alignment positions between *E. coli* 16S rRNA gene sequence positions 63-1,373. The bar shows 5% estimated sequence divergence. Bootstrap analysis was performed by Neighbor Joining analysis with 1000 resamplings, nodes receiving >50% and >80% support are indicated by grey and white circles, respectively. The insert shows a scale-up of the lineage formed by atribacterial SAGs. The phylogenetic affiliation of predominant persisting 16S rRNA gene sequence OTUs (*SI Appendix* Table S3) is indicated as follows: *SAG 190_F8 shares 97% sequence identity with OTU3. [†]SAG 190_A10 and G5 share 98% sequence identity with OTU1. [‡]SAG 190_F3 shares 96% sequence identity with OTU16. [§]SAG 190_A8 shares 99.4% identity with OTU 31. [¶]MBG-D = Marine Benthic Group-D. [#]Short sequence added to the tree without changing the overall tree topology with the Parsimony tool of the ARB program package The pairwise identities were based on a comparison between the 16S rRNA gene sequence of the SAGs and a representative sequence of a given OTU. Reference sequences are identified by GenBank accession numbers or locus tags.



SI Fig. S7. Unique genetic diversity within persisting populations and strength of purifying selection**.** (*A*) Proportion of single nucleotide polymorphism (SNPs) only observed at a single sediment depth. SNPs were recorded by mapping metagenome reads onto predicted genes from SAGs. Only genes mapped by the metagenomes of all 4 sediment depths were considered. SAGs from 4 different taxonomic lineages were included in the analyses. The Box-and-Whisker plot shows values for the different SAGs of a given lineages as described for *SI Appendix* Fig. S1. (*B*) Strength of purifying selection expressed as the ratio of non-synonymous to synonymous single nucleotide polymorphisms (P_n/P_s). P_n/P_s ratios were inferred for protein-coding genes from SAGs derived from 0.25 and 1.75 m sediment depth by mapping metagenomic reads onto the SAG-genes (see Materials and Methods). Only genes mapped by metagenomes at all sediment depths were considered in the analyses. Genes are colored according to their inferred functional role. SAGs from four different taxonomic lineages were included in the analyses. *P*-values indicate the likelihood that P_n/P_s ratios have a similar numerical distribution at the two sediment depths (binomial exact test). *Marine Benthic group-D.



SI Table S1. Sampling depths, geochemical zonation and summary of sequencing data

Station* (water depth)	Geochemical zone†	Sampling depth (mbsf)	Age (years bp)	Analysis‡	Sequence reads§	Number of OTUs observed¶	Evenness (16S rRNA¶)
M5 (28 m)	SR	0.25	250	ML & SCG	298,041,740	–	–
	Methane	0.75	750	ML	257,680,782	–	–
	Methane	1.25	1,250	ML	198,932,824	–	–
	Methane	1.75	1,750	ML & SCG	223,003,452	–	–
M24 (29 m)	Top	0.09	88	AL	163 rRNA/3,000 *dsrB*	84 rRNA/1,313 *dsrB*	–
	SR	0.44	438	AL	5,463 rRNA/3,000 *dsrB*	913 rRNA/755 *dsrB*	0.81
	SMT	4.15	3,282	AL	9,857 rRNA/3,000 *dsrB*	883 rRNA/631 *dsrB*	0.79
	Methane	5.05	3,785	AL	7,322 rRNA73,000 *dsrB*	980 rRNA/583 *dsrB*	0.83
	Bottom	5.60	4,091	AL	9,868 rRNA/3,000 *dsrB*	1,016 rRNA/859 *dsrB*	0.79
M26 (19 m)	Top	0.09	89	AL	11,330 rRNA	891 rRNA	0.72
	SR	0.44	441	AL	34,267 rRNA	1,727 rRNA	0.74
	SMT	3.64	3,002	AL	69,966 rRNA	1,705 rRNA	0.64
	Methane	4.14	3,280	AL	14,877 rRNA	1,270 rRNA	0.81
	Bottom	5.21	3,874	AL	20,212 rRNA	1,384 rRNA	0.78
M27A (19 m)	Top	0.12	122	AL	18,083 rRNA	1,232 rRNA	0.72
	SR	0.37	365	AL	26,416 rRNA	1,904 rRNA	0.78
	SMT	2.87	2,574	AL	37,951 rRNA	1,501 rRNA	0.67
	Methane	3.25	2,782	AL	34,255 rRNA	1,467 rRNA	0.74
	Bottom	7.12	4,935	AL	23,166 rRNA	1,069 rRNA	0.55
M29A (19 m)	Top	0.12	119	AL	7,935 rRNA/3,000 *dsrB*	947 rRNA/1,156 *dsrB*	0.74
	SR	0.36	360	AL	9,885 rRNA/3,000 *dsrB*	1,299 rRNA/883 *dsrB*	0.80
	SMT	2.45	2,341	AL	3,2196 rRNA/3,000 *dsrB*	1,346 rRNA/622 *dsrB*	0.73
	Methane	2.70	2,480	AL	23,604 rRNA/3,000 *dsrB*	1,281 rRNA/786 *dsrB*	0.75
	Bottom	7.20	4,980	AL	24,977 rRNA/3,000 *dsrB*	937 rRNA/658 *dsrB*	0.58

*Latitude-longitude. M5: 56.103333-10.457833. M24: 56.111883-10.41635. M26: 56.11155-10.41715. M27A: 56.111317-10.41805. M29A: 56.11075-10.419583.

†Top = surface sediment; SR = upper sulfate-rich sediment; SMT = sulfate-methane transition zone; MG = methanogenic zone; Bottom = deep methanogenic zone.

‡AL: PCR amplicon sequence library. ML: Metagenome libraries. SCG: Single cell genomes.

§Number of sequence reads after quality-filtering. Total number of forward + reverse sequence reads (Illumina NextSeq 150 bp chemistry) are shown for metagenomes. All *dsrB* sequence libraries were subsampled to a size of 3,000 reads.

¶PCR amplicon sequence libraries. rRNA: 16S rRNA gene. 16S rRNA and *dsrB* gene amplicon sequences were clustered into OTUs using a distance cutoff of 3.0 and 2.0%, respectively.

SI Table S2. Identity and function of predominant persisting 16S rRNA gene sequence OTUs

OTU	Domain	Taxonomic classification (alternative name)*	Inferred energy metabolism†	Reference
OTU1	Bacteria	Atribacteria (Candidate phylum JS1)	Organotrophic, fermentative	38, 39
OTU3	Bacteria	Deltaproteobacteria/Desulfarculaceae/*Desulfatiglans*	Organotrophic, dissimilatory sulfate reducer	40
OTU7	Bacteria	Deltaproteobacteria/Syntrophaceae/*Desulfobacca*	Organotrophic, dissimilatory sulfate reducer	41
OTU11	Bacteria	Aerophobetes (BHI80-139)/unclassified	Uncertain	42
OTU16	Bacteria	Chloroflexi/Dehalococcoidia/GIF9	Organotrophic, fermentative, sulfite-reducing	43, 44, 45, 46
OTU18	Bacteria	Aminicenantes (Candidate phylum OP8)	Organotrophic, fermentative	42
OTU78	Archaea	Marine Benthic Group-B (MBG-B)	Putative organotrophic, dissimilatory Fe- or Mn-reducer	47
OTU105	Bacteria	Candidate phylum LCP-89	Uncertain	–
OTU490	Archaea	Bathyarchaeota	Organotrophic, fermentative, acetogen, methanogen	4, 48, 49
OTU31	Archaea	Thermoplasmatales/Marine Benthic Group-D (MBG-D)	Organotrophic, fermentative	4

*Silva-based taxonomy (11).

†Based on the cited reference(s).

SI Table S3. Single cell amplified genome (SAG) assembly statistics and taxonomic identity

SAG name*	Depth (mbsf)	Sequence reads†	Assembly size (bp)	Contigs	G+C (%)	Complete-ness (%)‡	§ORFs/ mapped	Taxonomy¶	Metagenome reads mapped‖
190_A5	0.25	1,706,811	593,086	195	35.3	14	686/388	Atribacteria	23,050;39,969; 67,098;72,040
190_C5c	0.25	5,476,605	247,485	92	34.9	4	267/220	Atribacteria	10,204;16,892; 30,353;30,575
190_A10	1.75	4,190,296	413,447	167	35	15	488/76	Atribacteria	4,545;53,897; 378,210;673,177
190_D7	1.75	1,086,654	865,462	320	35.2	16	992/172	Atribacteria	8,043;72,225; 585,044;697,454
190_F6	1.75	2,592,013	1,146,637	440	35.2	50	1,352/256	Atribacteria	12,929;113,391; 976,491;1,078,552
190_G5	1.75	2,807,664	624,023	183	35.4	–	691/188	Atribacteria	8,046;79,878; 525,659;965,389
190_H6	1.75	1,450,546	621,098	245	35.7	11	767/124	Atribacteria	6,766;50,549; 373,985;463,871
190_B1	0.25	1,809,626	854,207	152	47.5	26	1,007	Dehalococcoidia	
190_D3	0.25	1,678,001	1,146,176	293	48.6	37	1,353	Dehalococcoidia	
191_B3c	0.25	4,458,811	1,556,758	338	48.5	51	1,835	Dehalococcoidia	
190_F3	0.25	1,698,338	353,271	64	47	31	399/152	Dehalococcoidia	13,275;36,206; 43,256;27,752
190_H5	1.75	1,461,517	848,699	287	47.1	34	1,061	Dehalococcoidia	
190_H1	0.25	2,405,842	793,811	308	47.4	40	1,055	Dehalococcoidia	
190_D9	1.75	1,703,977	1,068,743	227	49.7	36	1,276	Dehalococcoidia	
190_F2	0.25	1,613,790	712,387	123	48.3	28	791	Dehalococcoidia	
190_A9	1.75	1,754,307	897,160	334	42.2	43	1,104	∂-Proteobacteria /*Desulfatiglans*	
190_F8	1.75	1,708,599	487,358	181	45.9	14	618/380	∂-Proteobacteria /*Desulfatiglans*	17,225;20,380; 29,682;50,730
191_B12	1.75	1,779,247	1,199,698	327	48.4	19	1,430/48	∂-Proteobacteria /*Desulfatiglans*	24,412;21,973; 14,432;13,148
190_A2	0.25	1,592,081	2,173,979	711	36.6	56	2,347/80	Thermoplasmatales/ MBG-D#	9,581;9,263; 34,052;27,509
190_A8	1.75	1,593,526	1,064,389	290	35.5	35	1,220/180	Thermoplasmatales/ MBG-D#	25,851;28,132; 41,684;54,437

*SAGs were obtained from the uppermost part of the sulfate reduction zone (0.25 m depth) or the methanogenic zone (1.75 m depth) of sediment from station M5. SAGs highlighted in bold were used for estimating genetic diversity within persisting populations by mapping metagenomic reads onto the SAG assemblies and counting single nucleotide polymorphisms within predicted genes.
[†]Number of quality-filtered forward and reverse reads (Illumina MiSeq 300 bp chemistry) used for assembly.
[‡]Genome completeness as estimated from the presence of conserved single-copy genes by CheckM (50). –, not possible to calculate completeness.
[§]Predicted number of protein-coding genes (open reading frames, ORFs)/number of ORFs mapped by metagenomic reads for counting single nucleotide polymorphisms.
[¶]Silva-based taxonomy (11) as inferred from assembled 16S rRNA gene sequences.
[#]MBG-D, Marine Benthic Group-D.
[||]Number of metagenome sequence reads mapped to SAGs for mapping SNPs. The four numbers separated by semicolons represent mapped reads from metagenomes from 25, 27, 125 and 175 cm sediment depth, respectively.

SI Table S4. 16S rRNA gene sequence of the plasmid used for the mock community test and as bacterial qPCR standard

> clone Grp2_Bac1: *Altererythrobacter/Erythrobacter*

AGAGTTTGATCATGGCTCAGAACGAACGCTGGCGGCATGCCTAACACATG
CAAGTCGAACGAACCCTTCGGGGTGAGTGGCGCACGGGTGCGTAACGCGT
GGGAACCTGCCCTTAGGTTCGGAATAACAGTGAGAAATCGCTGCTAATAC
CGGATAATGTCTTCGGACCAAAGATTTATCGCCAAAGGATGGGCCCGCGT
TGGATTAGCTAGTTGGTGAGGTAAAGGCTCACCAAGGCGACGATCCATAG
CTGGTCTTAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGAC
TCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGAAAGCCTGA
TCCAGCAATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCTTT
TACTAGGGATGATAATGACAGTACCTAGAGAATAAGCTCCGGCTAACTCC
GTGCCAGCAGCCGCGGTAATACGGAGGGAGCTAGCGTTGTTCGGAAATAC
TGGGCGTAAAGCGCACGTAGGCGGCGTCGTAAGTCAGGGGTGAAATCCCA
GAGCTCAACTCTGGAACTGCCCTTGAAACTGCAATGCTAGAATATTGGAG
AGGTCAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAAG
AACACCAGTGGCGAAGGCGACTGACTGGACAATTATTGACGCTGAGGTGC
GAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTA
AACGATGATAACTAGCTGTTCGGGCTCATAGAGCTTGAGTGGCGCAGCTA
ACGCATTAAGTTATCCGCCTGGGGAGTACGGTCGCAAGATTAAAACTCAA
AGGAATTGACGGGGGCCTGCACAAGCGGTGGAGCATGTGGTTTAATTCGA
AGCAACGCGCAGAACCTTACCAGCCTTTGACATCCTTCGACGGTTTCTAG
AGATAGATTCCTTCCTTCGGGACGAAGTGACAGGTGCTGCATGGCTGTCG
TCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCC
TCATCCTTAGTTGCCATCATTCAGTTGGGCACTTTAAGGAAACTGCCGGT
GATAAGCCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATGGCCCTTACA
GGCTGGGCTACACACGTGCTACAATGGCATCTACAGTGAGTAGCGATCCC
GCGAGGGCTAGCTAATCTCCAAAAGATGTCTCAGTTCGGATTGTTCTCTG
CAACTCGAGAGCATGAAGGCGGAATCGCTAGTAATCGCGGATCAGCATGC
CGCGGTGAATACGTTCCCAGGCCTTGTACACACCGCCCGTCACGCCATGG
GAGTTGGTTTCACCCGAAGGTGGTGCGCTAACCGGTTTACCGGAGGCAGC
CAACCACGGTGGGATCAGCGACTGGGGTGAAGTCG

SI References

1. Røy H et al. (2014) Determination of dissimilatory sulfate reduction rates in marine sediment via radioactive 35-S tracer. *Limnol Oceanogr-Meth* 12(4):196-211.
2. Flury S et al. (2016) Controls on subsurface methane fluxes and shallow gas formation in Baltic Sea sediment (Aarhus Bay, Denmark). *Geochim Cosmochim Acta* 188:297-309.
3. Lever MA et al. (2015) A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. *Front Microbiol* 6:476.
4. Lloyd KG et al. (2013) Predominant archaea in marine sediments degrade detrital proteins. *Nature* 496(7444):215-218.
5. Wang Y, Qian PY (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS ONE* 4(10):e7401.
6. Claesson MJ et al. (2009) Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE* 4(8):pe6669.
7. Klindworth A et al. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41(1):e1.
8. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460-2461.
9. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 10(10):996-998.
10. Schloss PD et al. (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75(23):7537-7541.
11. Quast C et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41(D1):D590-596.
12. Müller AL, Kjeldsen KU, Rattei T, Pester M, Loy A (2015) Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *ISME J* 9(5):1152-1165.
13. Lever MA et al. (2013) Evidence for microbial carbon and sulfur cycling in deeply buried ridge flank basalt. *Science* 339(6125):1305-1308.
14. Wang Q et al. (2013) Ecological patterns of *nifH* genes in four terrestrial climatic zones. *MBio* 4(5): e00592-13.
15. Fish JA et al. (2013) FunGene: The functional gene pipeline and repository. *Front Microbiol* 4:291.
16. Ohkuma M, Kudo T (1998) Phylogenetic analysis of the symbiotic intestinal microflora of the termite *Cryptotermes domesticus*. *FEMS Microbiol Lett* 164(2):389-395.
17. Cadillo-Quiroz H et al. (2006) Vertical profiles of methanogenesis and methanogens in two contrasting acidic peatlands in central New York State, USA. *Environ Microbiol* 8(8):1428-1440.
18. Yu Y et al. (2005) Group-specific primer and probe sets to detect methanogenic

communities using quantitative real-time polymerase chain reaction. *Biotechnol Bioeng* 89(6):670-679.
19. Nielsen MB, Kjeldsen KU, Lever MA Ingvorsen K (2014) Survival of prokaryotes in a polluted waste dump during remediation by alkaline hydrolysis. Ecotoxicology 23(3):404-418.
20. Li H et al. (2009) The Sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
21. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.
22. Bankevich A et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455-477.
23. Markowitz VM et al. (2014) IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* 42(D1):D560-567.
24. Darling AE et al. (2014) PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2:e243.
25. Charlesworth B (2009) Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10(3):195-205.
26. Wright S (1942) Statistical genetics and evolution. *B Am Math Soc* 48(4):223-247.
27. Langerhuus AT et al. (2012) Endospore abundance and D:L-amino acid modelling of bacterial turnover in Holocene marine sediment (Aarhus Bay). *Geochim Cosmochim Acta* 99:87-99.
28. Stoletzki N, Eyre-Walker A (2011) Estimation of the neutrality index. *Mol Biol Evol* 28(1):63-70.
29. Huerta-Cepas J et al. (2016) eggNOG 45: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44(D1):D286-293.
30. Eddy SR (2008) A probabilistic model of local sequence alignment that simplifies statistical significance estimation. *PLoS Comput Biol* 4(5):e1000069.
31. Jørgensen BB, Marshall IP (2016) Slow Microbial Life in the Seabed. *Ann Rev Mar Sci* 8:311-332.
~~32.~~ Lever MA et al. (2015) Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations *FEMS Microbiol Rev* 39(5):688-728.
33. Chen X (2015) *Controls on microbial community zonation in coastal marine sediment (Aarhus Bay)* PhD dissertation, Department of Bioscience, Aarhus University, Aarhus, Denmark.
34. Stoddard SF et al. (2015) rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res* 43(D1):D593-598.
35. Braun S et al. (2016) Cellular content of biomolecules in sub-seafloor microbial communities. *Geochim Cosmochim Acta* 188:330–351.
36. Thamdrup B, Fossing H, Jørgensen BB (1994) Manganese, iron and sulfur cycling in a coastal marine sediment, Aarhus Bay, Denmark. *Geochim Cosmochim Acta* 58:5115-5129.
37. Ludwig W et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids*

Res 32(4):1363-1371.
38. Nobu MK et al. (2016) Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. *ISME J* 10(2):273-286.
39. Carr SA, Orcutt BN, Mandernack KW, Spear JR (2015) Abundant Atribacteria in deep marine sediment from the Adélie Basin, Antarctica. *Front Microbiol* 6:872.
40. Suzuki D, Li Z, Cui X, Zhang C, Katayama A (2014) Reclassification of *Desulfobacterium anilini* as *Desulfatiglans anilini* comb. nov. within *Desulfatiglans* gen. nov., and description of a 4-chlorophenol-degrading sulfate-reducing bacterium, *Desulfatiglans parachlorophenolica* sp. nov. *Int J Syst Evol Microbiol* 64(9):3081-3086.
41. Oude-Elferink SJ, Akkermans-van-Vliet WM, Bogte JJ, Stams AJ (1999) *Desulfobacca acetoxidans* gen. nov, sp. nov, a novel acetate-degrading sulfate reducer isolated from sulfidogenic granular sludge. *Int J Syst Bacteriol* 49(2):345-50.
42. Rinke C et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499(7459):431-437.
43. Wasmund K et al. (2014) Genome sequencing of a single cell of the widely distributed marine subsurface Dehalococcoidia, phylum Chloroflexi. *ISME J* 8(2):383-397.
44. Wasmund K et al. (2016) Single-cell genome and group-specific *dsrAB* sequencing implicate marine members of the class Dehalococcoidia (Phylum Chloroflexi) in sulfur cycling. *MBio* 7(3)pii:e00266-16
45. Kaster AK, Mayer-Blackwell K, Pasarelli B, Spormann AM (2014) Single cell genomic study of Dehalococcoidetes species from deep-sea sediments of the Peruvian Margin. *ISME J* 8(9):1831-1842.
46. Hug LA et al. (2013) Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* 1(1):22.
47. Jørgensen SL et al. (2012) Correlating microbial community profiles with geochemical data in highly stratified sediments from the Arctic Mid-Ocean Ridge. *Proc Natl Acad Sci USA* 109(42):E2846-2855.
48. Evans PN et al. (2015) Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* 350(6259):434-438.
49. He Y et al. (2016) Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat Microbiol* 1:16035.
50. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25(7):1043-1055.