Planktonic Euryarchaeota are a significant source of archaeal tetraether lipids in the ocean.

Sara A. Lincoln, Brenner Wai, John M. Eppley, Matthew J. Church, Roger E. Summons, and Edward F. DeLong

Supplementary Information

Sampling

Samples were collected on the BioLINCs cruise:

http://hahana.soest.hawaii.edu/cmorebiolincs/biolincs.html

We collected suspended particulate matter (SPM) using in situ pumps equipped with three-tiered filter housings for size fractionation (WTS-LV08, McLane Laboratories). Seawater was filtered through 57 μm nylon mesh (Nitex, U.S.A) and two 142 mm diameter glass fiber filters: a 3 μm glass fiber filter (Pall) and a 0.3 μm glass fiber filter (Sterlitech). Between 648 and 2096 L were filtered during each deployment (2-4 h). Actual in situ pump deployment depths were measured using a submersible data logger (Vemco, Canada) attached to the pumps. Upon retrieval of pumps, filter segments designated for DNA analysis were excised and frozen in lysis buffer (40 mM EDTA, 50 mM Tris, 0.73 M sucrose) at -80 ºC until DNA extraction. Remaining filter material was wrapped in combusted aluminum foil and stored at -20 ºC until lipid extraction.

In one instance suspended particulate matter was sampled on ship for metagenomic analysis. Seawater was collected from a depth of 130 m at Station ALOHA using a conductivity-temperature-depth rosette water sampler equipped with 24 12-L sampling bottles. 20 L were filtered in series through a combusted 3μm glass fiber filter (47 mm diameter, Pall) and a 0.22 μm Sterivex filter unit (Millipore) using a peristaltic pump and platinum-fired silicone tubing (Masterflex, Cole Parmer). After filtering,

excess water was removed from the Sterivex filter unit, 1.8 ml lysis buffer were added, and the sample was frozen at -80 ºC until DNA extraction.

DNA extraction

DNA was extracted with a Quick-Gene 610 l system (Fujifilm, Japan) and DNA tissue kit L using a modified lysis protocol. Cryovials containing filter sections stored in lysis buffer were thawed on ice and additional lysis buffer was added to bring the volume to 1.6 ml. Cryovials were vortexed for 2 minutes. Lysozyme (Fisher) dissolved in lysis buffer was added to a final concentration of 5 mg/ml, and vials were incubated with rotation at 37 ºC for 45 min. 90 μl proteinase K EDT-01 and 90 μl tissue lysis buffer MDT-01 (Fujifilm, Japan) were added and vials were incubated with rotation at 55 ºC for 2 hours. Lysate was decanted from filter material and transferred to a falcon tube. 1.8 ml Lysis Buffer LDT-01 were added, and samples were incubated with rotation at 55 ºC for 15 min. Finally, 2.4 ml >99% ethanol were added, and samples were vortexed and loaded on the Quick-Gene instrument. DNA was eluted in 400 μl and quantified using the PicoGreen dsDNA assay (Invitrogen).

DNA was extracted from Sterivex filter units using the same protocol, except that lysis buffer and reagents were added directly to the filter cartridge.

qPCR

qPCR was performed in duplicate 25 µl reactions containing: 12.5 µl 2X SYBRGreen Master Mix (Applied Biosystems), 8 µl of nuclease-free water, 2 µl of environmental DNA, and 0.5 µM final concentration of both QPCR reactions were analyzed using an Applied Biosystems Real-Time PCR system 7300, following the thermal cycling reaction conditions described in QPCR standards consisted of serial 10-fold dilutions of a plasmid

containing a SSU rRNA gene amplified from Station ALOHA environmental DNA using the primers Ar20F (23) and 1390R (24). The resulting gene copies in the standard dilution series ranged from $3x10^{-1}$ to $3x10^{5}$ copies per reaction.

SSU rDNA amplification and sequencing

We amplified the V1-V3 variable region of the archaeal SSU rRNA gene using the forward primer 20F 5'-TCCGGTTGATCCYGCCRG-3' (23) and a barcoded reverse primer 519R, 5'-GGTDTTACCGCGGCKGCT-3' (69). 20F was linked to a 454 adaptor, and 519R was linked to both a 454 adaptor and a barcode tag consisting of 5-7 nucleotides. 400 ng of each barcoded amplicon set were pooled in two batches, each containing 16 unique barcodes. Each DNA sample was assigned a unique barcode designed by the Human Microbiome Project Working Group (http://www.hmpdacc.org/doc/HMP_MDG_454_16S_Protocol.pdf). PCR reactions were carried out in triplicate 20 μl volumes, with each containing 2 μl environmental DNA template, 1 μl forward primer (10 μM), 1 μl B-adapter (10 μM), 13.8 μl DNAse-free water, 0.2 μl Taq DNA polymerase (Accuprime Taq Hifi, Invitrogen), and 2 μl AccuPrime Buffer II (Invitrogen) The thermal cycling program was as follows: initial hold at 95 ºC; 20 repetitions of a cycle of 95 ºC for 20 s, 60 ºC for 30 s, and 72 ºC for 5 minutes. PCR products were run on 1% agarose gel to verify amplification and confirm amplicons were of appropriate size. Plasmid DNA containing a cloned SSU MG-I sequence and DNAse-free water were used as positive and negative controls, respectively. PCR products were pooled, purified using a QIAquick PCR purification kit (Qiagen) and quantified using the PicoGreen dsDNA assay (Invitrogen). Additional PCR reactions were performed for samples that initially yielded <400 ng amplicons.

Amplicon library preparation followed the Titanium Rapid Library Preparation protocol (Roche), except that adaptor-ligated libraries were not diluted with AMPure XP beads before size selection, and 1/4 of the recommended volume of amplification primers was used for emulsion PCR. Libraries were quantified using the Titanium Slingshot kit (Fluidigm), added to emulsion PCR reactions at a concentration of 0.1 molecules per bead, and sequenced in a half plate run on a GS FLX system (Roche).

rDNA amplicon data analysis

In the demultiplexing step, reads were filtered and any low quality or ambiguous reads were removed using default parameters (minimum quality score = 25, minimum/maximum length = 200/1000, no ambiguous bases allowed, and no mismatches allowed in the primer sequence). Primer and adapter sequences were removed in this step. The number of amplicon reads per sample ranged from 4 to 22,364 (Table S2). Low amplicon counts likely reflect low archaeal cell numbers at shallow depths, but may have also have stemmed from inaccuracies in template DNA quantification or differences in amplification efficiency at any depth. To reduce noise caused by comparison of samples containing significantly different numbers of amplicon sequences, we focused the study on samples from which ≥1000 archaeal sequences were amplified.

To examine the effect of this cutoff on community composition, we used BLASTn to assign the reads in each demultiplexed set to their closest relatives in the ARB-SILVA SSU reference database (top hit, e values <0.001, bit score >50, ARB-SILVA release 111). All archaeal reads identified below the domain level could be binned into one of three categories: Thaumarchaeota (primarily MG-I), MG-II Euryarchaeota, and "other Euryarchaeota" (Fig. S5-S6). The last category, including

reads identified as belonging to MG-III and Unidentified Hydrothermal Vent Euryarchaeon PVA, was generally detected at low abundance at depths <130 m, but reached a maximum of 26% of the archaeal population in the 23 m, 3-57 μm size fraction sample. When a threshold of ≥100 archaeal amplicons in both size fractions was adopted, however, this category was eliminated. We adopted a still more stringent criterion requiring ≥1000 amplicons in both size fractions at a given depth because it reduced sample size disparity while preserving the general patterns of MG-I and MG-II as seen in the ≥100 amplicon plots (Fig S5-S6).

Representative sequences for the 20 most abundant OTUs (each representing >500 amplicon sequences across the focus depths) were extracted for further phylogenetic analysis. The heatmap of the 10 most abundant OTUs (representing >2000 amplicon sequences) detected at the focus depths was modified using the statistical software R (www.R-project.org) so that OTU abundance was ranked within each depth and size class, rather than across the entire matrix.

Representative sequences were aligned with the SILVA Incremental Aligner (SINA, http://www.arb-silva.de/aligner/) using the archaea variability profile. The three nearest neighbors and selected archaeal SSU rRNA sequences (e.g. pSL12, Marine Group III) were added to the dataset for phylogenetic context. Aligned sequences were imported to a pared SSU database (SILVA release 106, non-redundant) in ARB (72). A maximum likelihood tree was constructed in using PhyML with the Jukes-Cantor 69 substitution model and a customized base frequency filter ignoring positions in which special characters and ambiguity codes appear most often. Bootstrap values >80 (percent of 100

replicates) are displayed. The tree was visualized and annotated using the Interactive Tree of Life (73).

Metagenomic Data and Analyses

DNA from the 130 m (0.22-3 μm) sample collected by CTD rosette was prepared and sequenced in a quarter plate run following the GS FLX Titanium protocol.

For Illumina, samples were prepared using a Nextera XT DNA kit and Trimmomatic (76) identified reads in which the original DNA sequences were shorter than the output read length. Remaining reads were joined using PandaSeq (77); unjoined read pairs were retained and tracked as single records. All reads were trimmed to remove low quality bases (< 5) from the ends. Reads containing fewer than 40 bases, or in which >90% of bases were a single nucleotide, were removed. Remaining reads were filtered using SortMeRNA (78) to identify rRNA sequence-containing reads using the following databases: rfam-5.8s, rfam-5s, silva-arc-16s, silva-arc-23s, silva-bac-16s, silva-bac-23s, silva-euk-18s, and silva-euk-28s. Taxonomic identities were determined by searching against Silva release 115 using lastal (79) with default parameters. ***454 Datasets:***

The reference database used to determine taxonomic affiliation of protein-coding reads was comprised of NCBI RefSeq release 54, protein sequences from the Moore Marine Microbial Genomes project, and several recently published marine microbial genomes. Matches with BLAST bit scores of at least 50 or LAST alignment scores of at least 100 were retained. Protein sequence matches with a score within 10% of the best match to each read were exported to MEGAN (81) and the putative taxonomic affiliation of each read was determined using the Least Common Ancestor algorithm.

Lipid extraction, hydrolysis and sample preparation

For the modified Bligh-Dyer protocol, sections of filters were placed in 250 ml Teflon bottles (Nalgene Nunc), submerged in a monophasic solution of 2:1:0.8 methanol:dichloromethane:phosphate-buffered saline (PBS, Sigma-Aldrich) and sonicated for 20 minutes. Solvent was decanted and this step repeated twice. The decanted solvents were combined in a separatory funnel and phase separation achieved with the addition of 1:1 dichloromethane:water. The organic phase was removed and the aqueous phase extracted with dichloromethane three times. Combined extracts were rinsed with dichloromethane-extracted Nanopure water three times.

Filters were then submerged in a monophasic solution of 10:5:4 methanol:dichloromethane:trichloroacetic acid (Sigma-Aldrich, 2.5% in Nanopure water) and sonicated for 20 minutes. Solvent was decanted and this step repeated twice. Phase separation was achieved as above, and the resultant organic phase was washed with Nanopure water three times. Total lipid extracts (TLEs) were combined with those from the previous step, evaporated under a gentle stream of N_2, and stored at -20°C until analysis.

Acid hydrolysis

1 ml of a mixture of 6 M HCl:methanol:dichloromethane (1:9:1, v/v) was added to dried TLE and vials were sealed and incubated at 70°C for 12 h. The hydrolysate was evaporated under N2 and prepared as core GDGT aliquots were.

Lipid analysis

Core GDGT analysis: TLE aliquots designated for core GDGT analysis were dissolved in 99:1 hexane:isopropanol, filtered (0.45 μm syringe filters, Millipore), evaporated, and

redissolved in 99:1 hexane:isopropanol containing 1 ng/μl of a synthetic C_{46} tetraether lipid internal standard.

LCMS: We analyzed core GDGTs in APCI+ mode using an Agilent 1260 Infinity series LC coupled to an Agilent 6130 mass spectrometer. 10 μl of extract were injected, typically corresponding to ~150 L of seawater. Separation was achieved on a Prevail Cyano column (150mm x 2.1mm, 3µm, Grace) maintained at 30°C, GDGTs were detected by selected monitoring of [M+H]+ ions m/z 1302.3, 1300.3, 1298.3, 1296.3, 1294.3, 1292.2 and 743.7. Compounds were eluted isocratically with 100% eluent A (hexane:isopropanol 99:1) for 5 minutes, followed sequentially by: 1) a linear gradient to 15% eluent B (hexane:isopropanol 9:1) over 15 minutes; 2) a linear gradient to 100% eluent B over 15 minutes; and 3) a 5 min isocratic hold at 100% eluent B, at a constant flow rate of 0.4 ml/min. The column was re-equilibrated with 100% eluent A for 10 min between analyses. APCI-MS conditions were: gas temperature 350°C, vaporizer temperature 380°C, drying gas flow 6 l/min., nebulizer pressure 30psi, capillary 2000V, corona 5µA. Full scans (m/z 400-2000) were run periodically to monitor background and to confirm compound identification (m/z 1200-1750).

IPL analysis:

TLE aliquots designated for IPL analysis were dissolved in 9:1 dichloromethane:methanol, filtered, dried and redissolved in the same solvent mixture. We used an Agilent 1200 series HPLC system coupled to an Agilent 6520 Accurate-Mass Quadrupole Time-of-Flight (Q-TOF; Agilent Technologies) mass spectrometer operated in positive mode to measure IPL GDGTs. Chromatographic separation was achieved using a UPLC BEH ESI) column (Waters; 2.1 x 150 mm, 1.7 µm particle size) following

the method of Wörmer et al. (51). methods for IPL ESI) methods for IPL Mass spectrometer source parameters were as follows: gas temperature 200°C, drying gas flow 6 L/min, nebulizer pressure 40 psi, capillary voltage 3000, and fragmentor voltage 175. The Q-TOF was set to a range of m/z 400-2000 in MS^1 and 100-2000 in MS^2. In MS^2 the scan rate was set to 2.99 and a maximum of 3 precursors were selected per cycle, with active exclusion after 5 spectra. Collision energy was set to 70 over the m/z range of IPL GDGTs (1400-2000).

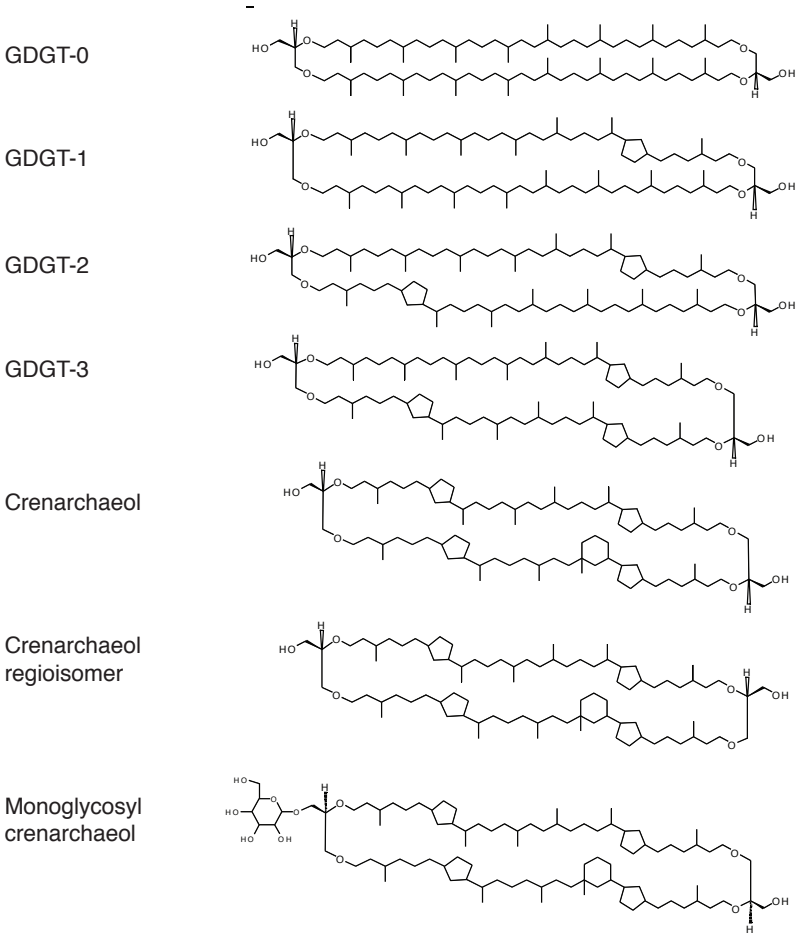
IPL GDGT identification by single quad LCMS

Variations in head group structure and composition are likely to cause the observed chromatographic separation of putative IPLs with identical core GDGTs, resulting in the multiple series of late eluting peaks observed in these two samples. To investigate head group composition, we used the core GDGT method (using atmospheric pressure chemical ionization (APCI)) to analyze several environmental extracts in which IPL GDGTs with either glycosyl or phosphatidyl head groups were previously detected using electrospray ionization (ESI) methods for IPL analysis. Only monoglycosyl GDGTs appeared in the analytical window, creating late-eluting peaks with m/z ratios identical to those seen in our two NPSG samples. Thus, the NPSG samples likely contain IPL GDGTs with several different hexose isomers as head groups. We cannot, however, exclude the possibility that they also contain more polar IPL GDGTs that are outside the window of detection of the core method.

The electrospray IPL method could not resolve the putative diversity of GDGT head groups detected using the extended core GDGT method. The extended core method may thus prove useful in further exploration of the identity of these polar moieties.

Cell density inference: Two assumptions enabled inference of MG-II cell densities (equation1): 1) The number of MG-I thaumarchaea/L (n_{MG-I}) is equal to qPCR-derived MG-I SSU rRNA copy numbers/L, and 2) The number of archaeal cells in a sample equals the sum of MG-I + MG-II, since the population is binary across all depths analyzed. By extension, the sum of the population fractions of MG-I (f_{MG-I}) and MG-II (f_{MG-I}) equals 1.

Fig. S1. Structures of glycerol dialkyl glycerol tetraether (GDGT) lipids discussed in text.



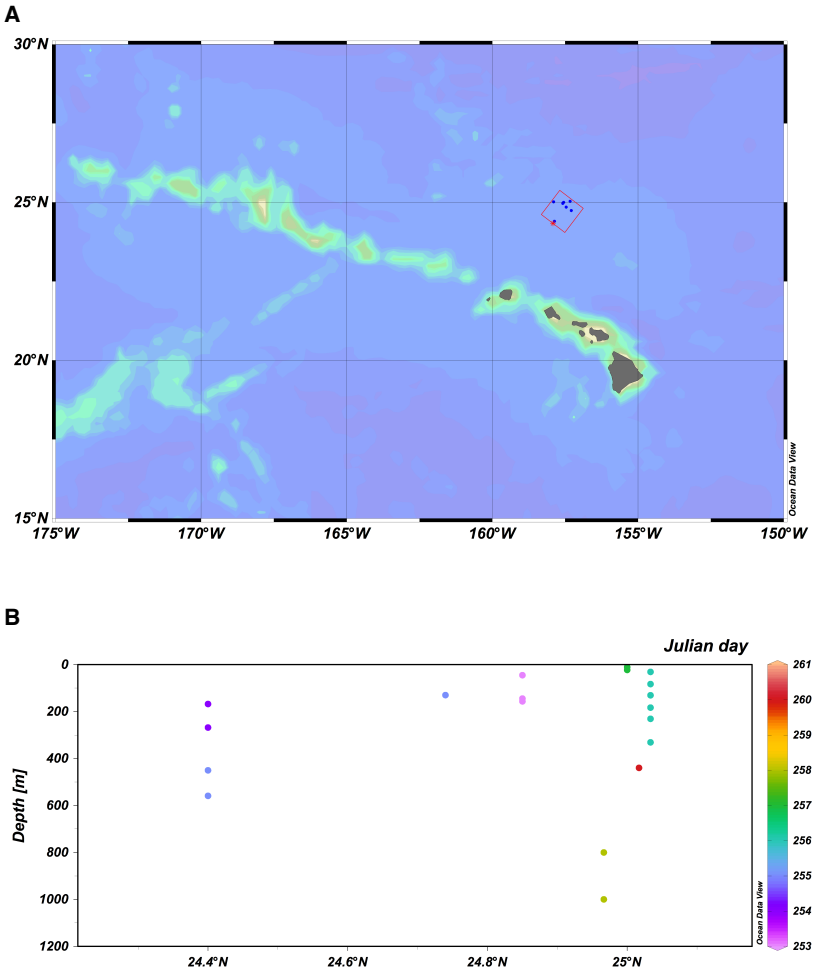


Fig. S2. Spatial and temporal distribution of BioLINCS in situ pump sampling locations: (A) map view, (B) section view. Colored points in (B) indicate Julian days during which samples were collected at the depth and latitude shown; color key at right. The Julian day span of 253-261 corresponds to 9/10/2011-9/17/2011.

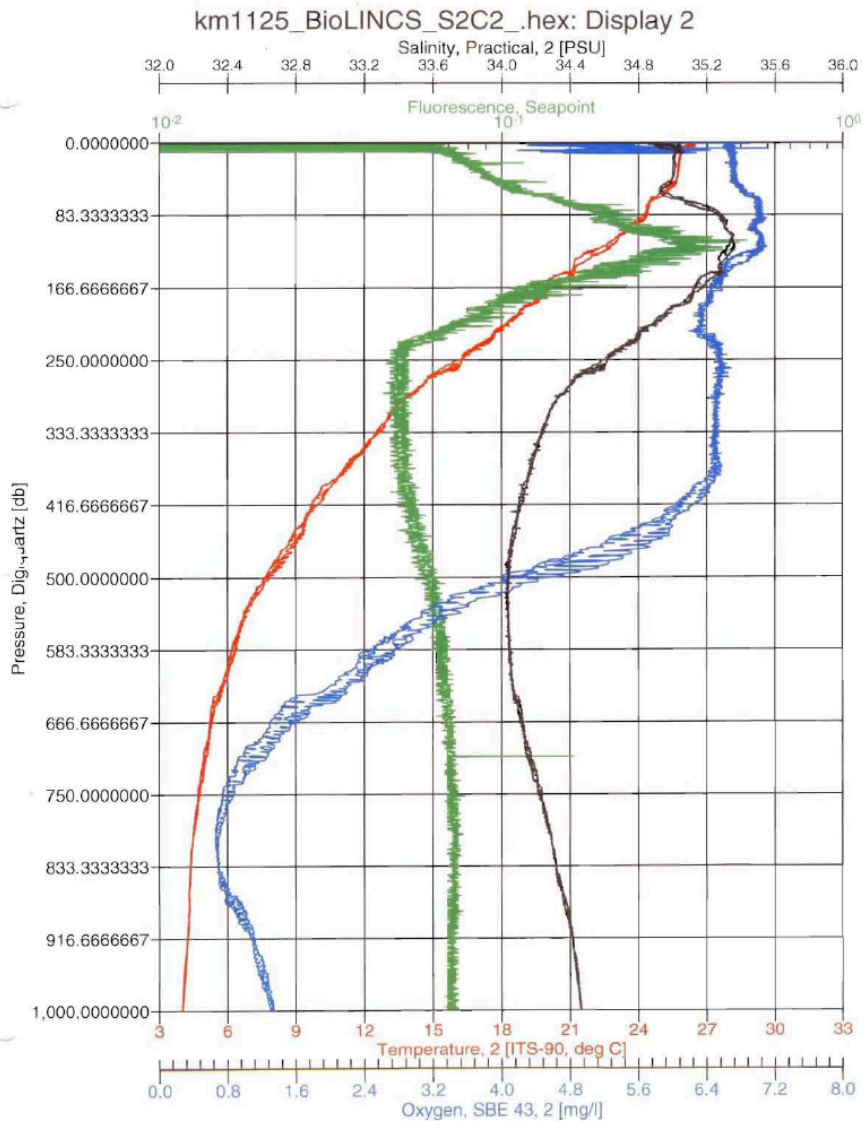


Fig. S3. CTD profile from the BioLINCS cruise, showing temperature (red), salinity (black), fluorescence (green) and oxygen concentration (blue) vs. pressure.

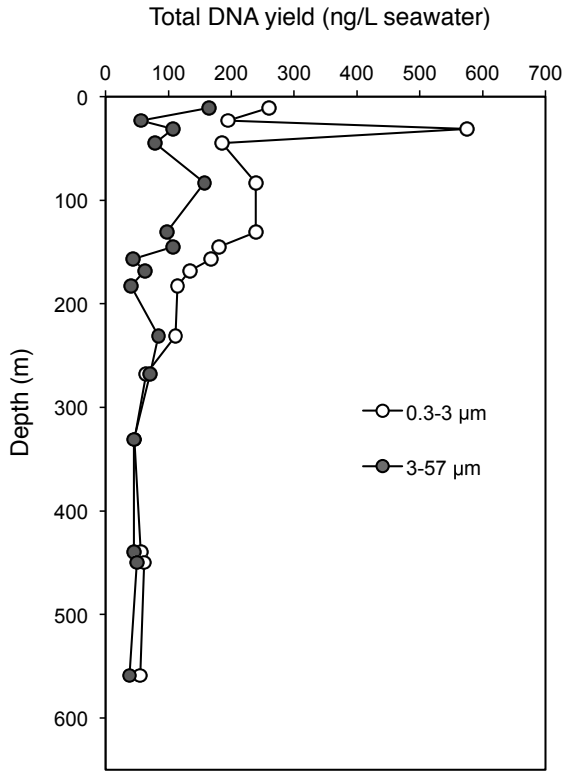


Fig. S4. Total DNA yield from glass fiber filters.

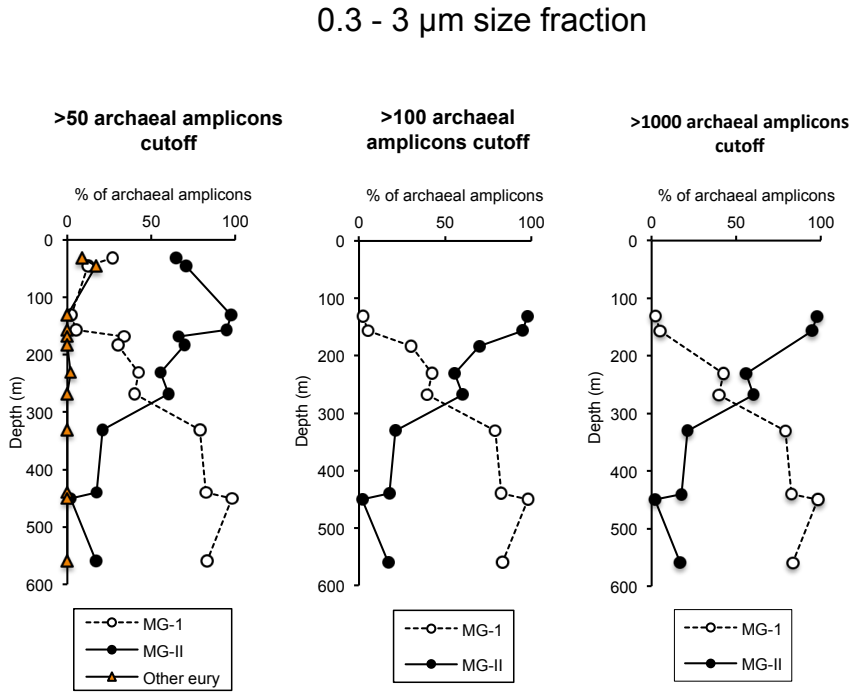


Fig. S5. Effect of different amplicon cutoff criteria on archaeal community composition profiles, 0.3-3 μm size fraction. The category "other euryarchaeota" includes sequences with top BLAST hits to MG-III Euryarchaeota and Unidentified Hydrothermal Vent Archaeon PVA.

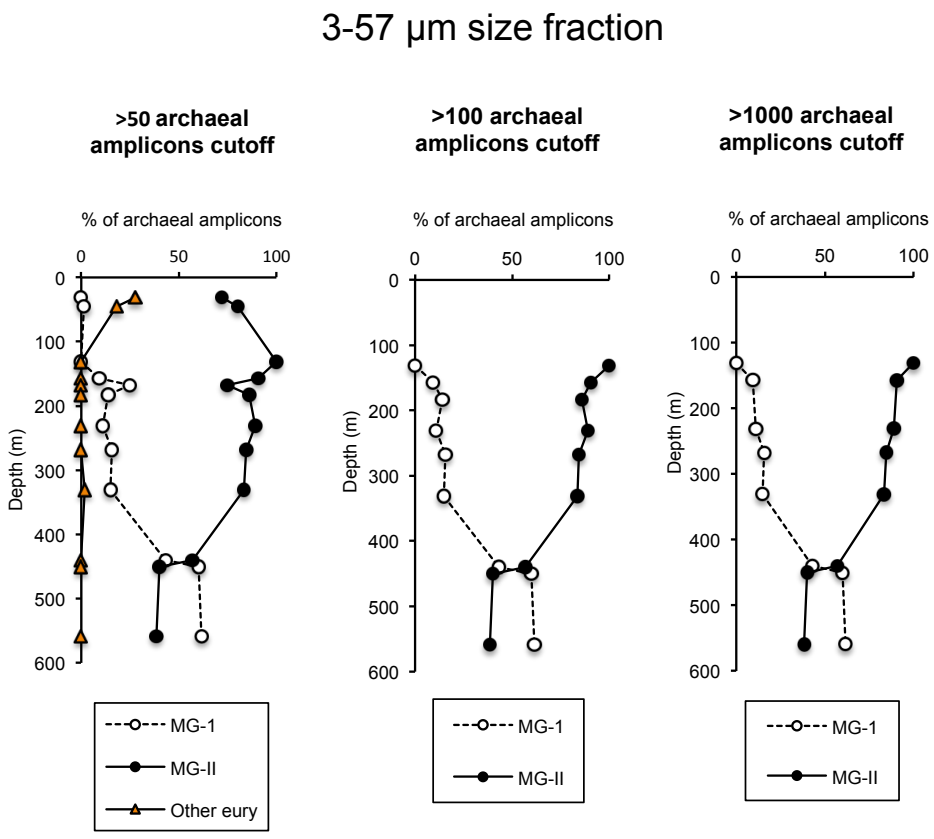


Fig. S6. Effect of different amplicon cutoff criteria on archaeal community composition profiles, 3-57 μm size fraction. As in the small size fraction, the category "other euryarchaea" includes sequences with top BLAST hits to MG-III euryarchaea and the Unidentified Hydrothermal Vent Archaeon PVA.

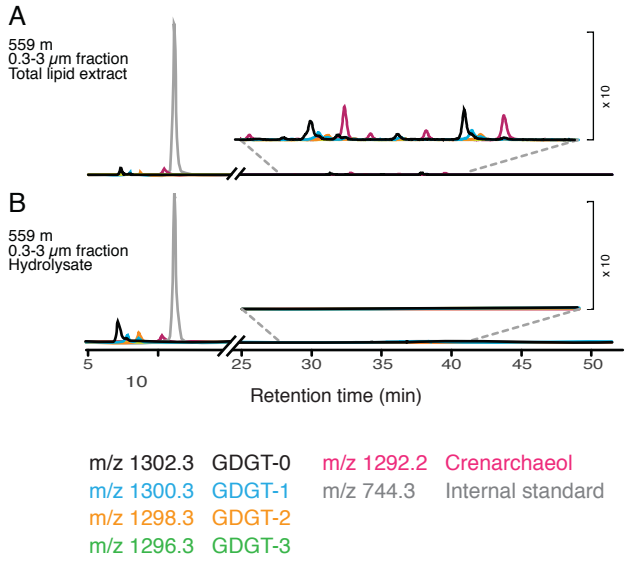


Fig. S7. HPLC-APCI-MS composite extracted ion chromatograms (EIC) of TLE from 559 m, 0.3-3 μm size fraction before (A) and after (B) acid hydrolysis. Colored traces are EICs of m/z values of individual core GDGTs and the hydrolysis-resistant internal standard. Late eluting peaks in A represent three putative series of core GDGTs released from IPL GDGTs by fragmentation in the APCI source. After acid hydrolysis, late-eluting peaks disappeared and the peak area of core GDGTs increased, supporting this interpretation. We analyzed environmental samples known to contain GDGTs with either predominantly mono- and diglycosyl GDTS or predominantly phosphatidyl GDGTs (not shown), and only monoglycosyl GDGTs eluted using this method. Thus, the three late-eluting series (A) may be core GDGTs released from monoglycosyl GDGTs containing different sugar moieties, or unknown IPL GDGTs of similar polarity. A contribution of IPL GDGTs outside of this method's window of detection to the core GDGTs released by acid hydrolysis is possible. The relative response factor of IPL and core GDGTs detected using this method has not been determined.

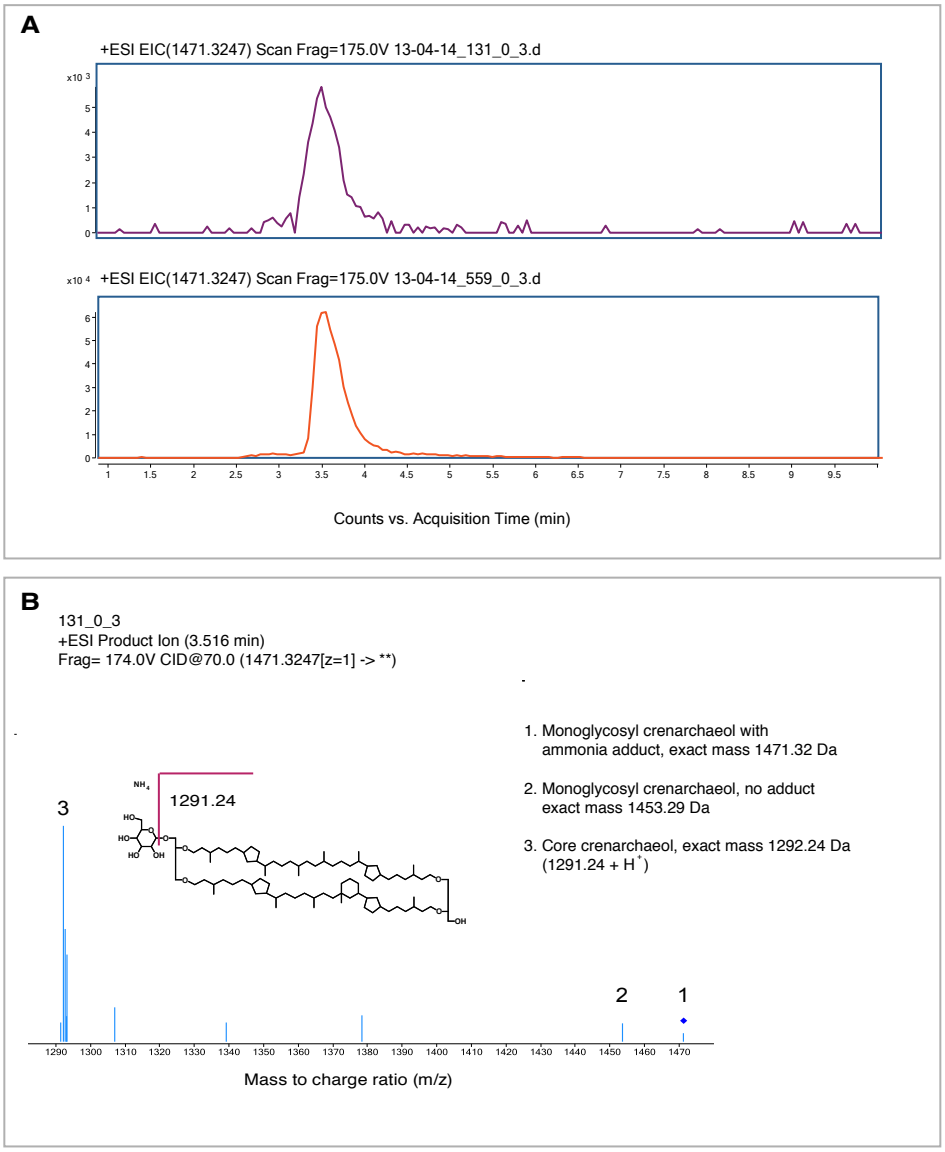


Fig. S8. HPLC-ESI-MS analyses of the 559 m sample, small size fraction. Extracted ion chromatograms (A) of m/z 1471.3247 corresponding to monoglycosyl crenarchaeol + ammonia adduct (formula mass 1471.3244), indicate that this compound is present in both samples; characteristic fragments are visible in the MS^2 spectrum of the 131 m sample (B).

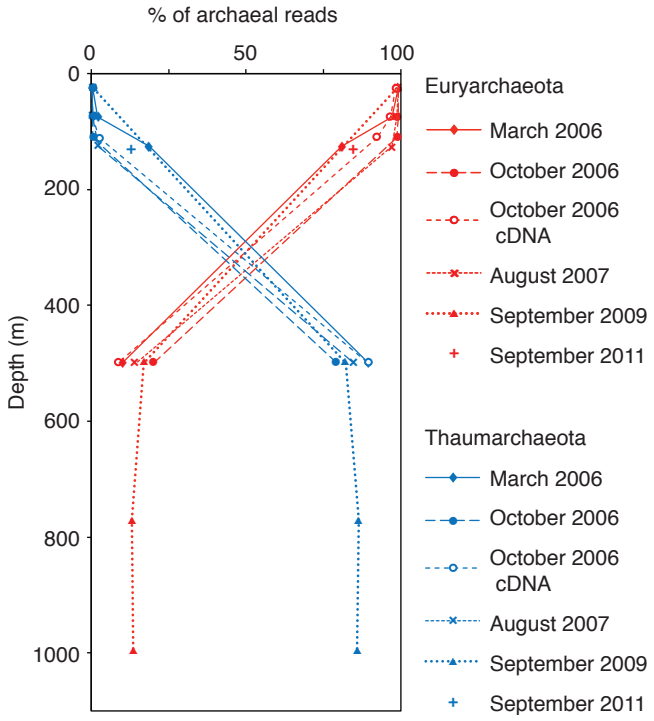


Fig. S9. Euryarchaeal and thaumarchaeal representation in metagenomic and metatranscriptomic datasets generated from SPM (0.22-2.6 μm) collected at Station ALOHA from March 2006 to September 2011. Putative protein-coding reads assigned at or below each phylum level are shown as a percent of total archaeal reads assigned.

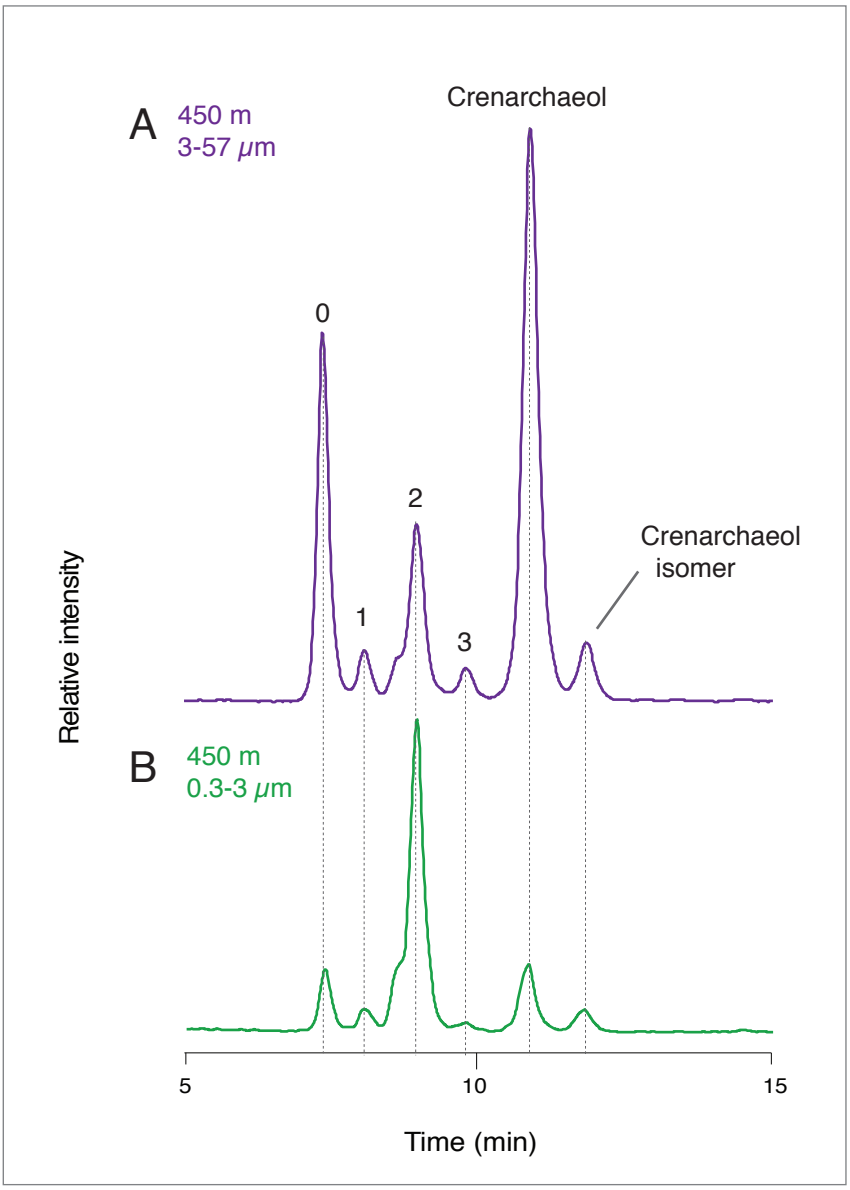


Fig. S10. Composite extracted ion chromatograms of core GDGTs in the small (bottom, green trace) and large size fractions (top, violet trace) of SPM collected during the same in situ pump deployment at 450 m. Numbered peaks correspond to structures in Fig. S1 (e.g. "0" denotes GDGT-0). MG-I Thaumarchaeota comprised 98% of the archaeal community in the small size fraction and 60% of the community in the large size fraction.

Table S1. Sequence statistics for metagenomic samples.

Depth (m)	Size fraction (µm)	Reads	Mean read length (bp)	Mean joined read length (bp)	Archaeal rRNA reads (SSU + LSU)
83	0.3 - 3	1359546	539.0	356.3	36
83	3 - 57	1203705	525.8	309.4	10
131	0.3 - 3	2851413	529.8	337.9	129
131	3 - 57	2404371	527.6	318.4	34
157	0.3 - 3	1868209	510.1	297.9	150
157	3 - 57	239555	528.3	342.9	4
231	0.3 - 3	3121798	538.6	346.3	587
231	3 - 57	2344582	527.9	296.1	95
331	0.3 - 3	4047249	507.3	308.2	847
331	3 - 57	3432366	503.4	266.9	93
450	0.3 - 3	1461112	463.0	264.3	225
450	3-57	1405905	535.5	305.8	

Table S2. Amplicon statistics. Boxes indicate samples that meet the threshold criterion of >1000 amplicons in both size fractions. Bold blue type denotes replicate samples amplified using different tags.

Depth (m)	Size fraction (µm)	Total reads (archaeal + eukaryotic)	Archaeal reads	Mean length of archaeal reads
11	0.3 - 3	82	51	392.55
11	3 - 57	33	26	393.76
23	0.3-3	1132	209	393.05
23	3-57	5	4	400.00
31	0.3 - 3	98	82	390.85
31	3 - 57	1172	613	395.91
45	0.3 - 3	132	82	394.55
45	3 - 57	2421	1768	398.41
83	0.3 - 3	615	106	397.51
83	3 - 57	33	32	389.61
130	0.22 – 1.6	22364	21805	399.69
131	0.3 - 3	7261	5514	396.77
131	**0.3 – 3**	**21642**	**15187**	**398.62**
131	3 - 57	16166	2477	381.57
145	0.3 - 3	25229	16662	394.97
145	3 - 57	57	41	395.28
157	0.3 - 3	4013	2226	392.58
157	3 - 57	3023	1605	379.85
168	0.3 - 3	760	83	389.52
168	3 - 57	1524	1381	396.01
183	0.3 - 3	426	408	397.38
183	3 - 57	324	172	396.02
231	0.3 - 3	16136	14075	399.03
231	3 - 57	11066	6449	396.08
268	0.3 - 3	3073	2990	398.02
268	3 - 57	7891	3260	391.15
331	0.3 - 3	16004	14080	397.15
331	3 - 57	1437	1115	397.87
440	0.3 - 3	18690	16839	392.33
440	3 - 57	7183	4050	395.80
450	0.3 - 3	19480	18818	399.56
450	**0.3 – 3**	**9598**	**9385**	**396.20**
450	3 - 57	17562	9038	391.50
559	0.3 - 3	19794	18379	393.04
559	3 - 57	15739	8437	393.77
800	0.22 - 1.6	5199	4194	399.07
1000	0.22 – 1.6	3712	2716	399.62

Table S3. Comparison of archaeal community composition as determined through analysis of SSU rDNA amplicons from SPM collected on Sterivex filters (0.22-2.6 μm size fraction) and glass fiber filters (0.3-3 μm) and metagenomic data. Two amplicon samples (131 and 450 m, 0.3–3 μm) were processed and sequenced in duplicate using different tags.

Depth (m)	Size fraction	Replicate	Analysis type	% Thaum.	% MG-II Eury.	% Other Eury	% Other archaea
130	0.22-2.6		SSU rRNA	5.1	94.8	0.0	0.0
130	0.22-2.6		metagenome	12.9	79.7	4.9	0.2
131	0.3-3	a	SSU rRNA	2.3	97.7	0.0	0.0
131	0.3-3	b	SSU rRNA	0.0	100.0	0.0	0.0
450	0.3-3	a	SSU rRNA	98.1	1.9	0.0	0.0
450	0.3-3	b	SSU rRNA	97.3	2.7	0.0	0.0