PNAS Supporting information

Alonso-Sáez et al. '*Role for urea in nitrification by polar marine Archaea*'

Text S1

Quantitative PCR analysis (qPCR). qPCR was used to estimate the abundance of 16S rRNA, *amoA* and *UreC* genes from *Thaumarchaeota* in genomic DNA collected at different water masses in six stations from the Amundsen and Ross seas, and fifteen samples from Arctic surface or halocline waters collected from 15 January to 15 June 2008. For each sample, 4.5 to 7.5 L of seawater was filtered through a 0.22 µm Sterivex capsule filter (Millipore). Filters were amended with 1 ml lysis buffer (400 mM NaCl, 750 mM sucrose, 20 mM EDTA, 50 mM Tris-HCl, pH 7.6) and immediately frozen at -80ºC. Antarctic DNA was extracted using an enzyme/phenol-chloroform protocol as described elsewhere (1). Arctic DNA samples were extracted using a standard salt protocol (2). The primers used for amplifying *amoA* genes were Arch-amoA-for (5'-CTG AYT GGG CYT GGA CAT C-3') and Arch-amoA-rev (5'-TTC TTC TTT GTT GCC CAG TA-3') (3). The primers used for amplifying the clade of polar *Thaumarchaeota ureC* genes were designed in this study: Thaum-UreC forward (5′-ATGCAATYTGTAATGGAACWACWAC-3') and Thaum-UreC reverse (5' AGTTGTYCCCCAATCTTCATGTAATTTTA-3'). Thaumarchaeal 16S rRNA genes were amplified by the reverse primer MCGI-554r (5′-TGA CCA CTT GAG GTG CTG-3′) and a modified version of forward primer MCGI-391f (4), (i.e., MCGI-391mf: 5′-AAG GTT ART CCG AGT GRT TT-3′). We found that the widely used primer MCGI-391f severely underestimated the abundance of Antarctic thaumarchaeal 16S rRNA genes as it had a mismatch at the 3' end against a dominant thaumarchaeal phylotype in Antarctic waters (OTU I) (5). Therefore, in the modified primer, the last base (C) at the 3' end was deleted. For the Antarctic sampling, gene abundances in 1 µL of extract were measured in 3 to 6 replicates in a Bio-Rad system using IQ SYBR Green Supermix (Bio-Rad), except for the *ureC* gene, that was analyzed using the SYBR Green PCR Master Mix (Applied Biosystems) and an Applied Biosystems StepOnePlus instrument. Arch-amoA primers were added at 0.2 µM and MCGI and Thaum-UreC primers at 1 µM concentration. For the Arctic analysis, gene abundances in 5 µL of diluted extract were measured in a Chromo4 Real-Time Detector (Bio-Rad) with the Ssofast Evagreen Supermix (Bio-Rad) for Arch-amoA and MCGI primers and PerfeCTa SYBR Green

FastMix (Quanta Biosciences) for Thaum-UreC quantification. Arch-amoA and MCGI primers were added at 0.2 μM and Thaum-UreC primers were added at 0.3 μM concentration. Template DNA concentration varied between 1 and 10 ng per reaction. PCR conditions for MGI and Arch-amoA sets have been described earlier (3). For amplification with Thaum-UreC primers we used an annealing temperature of 53ºC in Arctic samples, and 59ºC in Antarctic samples. Standards for the Antarctic qPCR analysis were prepared from amplicons cloned from environmental samples. For cloning, products from 3 individual PCR reactions (20 μL) were pooled and purified using the QIAGEN PCR purification kit (Qiagen) and cloned using the TOPO-TA cloning kit (Invitrogen), following manufacturer instructions. After cloning, plasmids were extracted using the QIAprep Spin Miniprep kit (Qiagen), linearized with the PstI restriction enzyme (Fermentas) for 4 h at 37ºC and gel-purified (QIAquick Gel Extraction kit, Qiagen). For the Arctic qPCR standards preparation the StrataClone PCR cloning kit (Agilent Technologies) was used for cloning and amplicons were purified with a PCR purification kit (Feldan). DNA concentrations were measured using a Nanodrop spectrophotometer (Thermo Scientific). Standard curves were based on serial dilutions of six concentrations ranging from 10 to 10^5 copies μL^{-1}, and were run in triplicate for each qPCR run. Amplification efficiencies for Arch-amoA, MCGI and Thaum-UreC primers averaged 81%, 100%, and 98%, respectively. Gene abundances in the environment (copies mL^{-1}) were estimated from gene abundances in template DNA, the volume of seawater filtered, and the final volume of the DNA extract, assuming 100% extraction efficiency.

CARD-FISH. CARD-FISH was performed on Arctic and Antarctic water samples to quantify the abundance of *Bacteria* and *Thaumarchaeota* (6). Samples were fixed with formaldehyde (3.7% final conc.) overnight at 4ºC and filtered onto 0.2-μm polycarbonate filters (Millipore, GTTP), which were stored at -20ºC. Cells were permeabilized with lysozyme (37ºC, 1h) and achromopeptidase (37ºC, 30 min). Hybridizations were carried out overnight at 35°C following the addition of horseradish peroxidase labeled probes: Eub338-II-III (7-8), Cren554 (9) and Eury806 (10). Counter-staining of CARD-FISH preparations was done with 4'-6'-diamidino-2-phenylindole (DAPI, final conc. 1 μg mL^{-1}). At least 300 DAPI cells in a minimum of 10 fields were counted.

MAR-FISH. MAR-FISH was performed for Arctic and Antarctic samples in order to quantify the single-cell activity of *Bacteria* and *Thaumarchaeota* in the uptake of leucine and bicarbonate. Twenty mL samples were incubated with [^{3}H]-leucine (0.5 nM, Perkin Elmer, NET460A005, 140 Ci mmol^{-1}) or [^{14}C]-$NaHCO_3$ (3 microCi mL^{-1}, Amersham CFA3, 56 mCi mmol^{-1}) in ice-cold seawater in the dark. Incubation times were ca. 8h and 24h for leucine and bicarbonate incubations, respectively. After incubation, samples were fixed overnight with formaldehyde (1.8% final conc.) at 4ºC, and then filtered onto 0.2-µm polycarbonate filters (Millipore, GTTP). Samples killed with formaldehyde before the addition of the radioactive compounds served as controls. Filters were hybridized following the CARD-FISH protocol described above, and subsequently processed for microautoradiography (11). Optimal exposure times were determined for each compound and sampling. In the case of leucine uptake, optimal exposure time was 6 d for Antarctic samples and varied between 1 and 7 d for Arctic samples. Optimal exposure times for bicarbonate uptake were 22-23 d for both sampling locations. The slides were counted in a Nikon eclipse E600 epifluorescence microscope.

Urea uptake experiment. Urea uptake under light and dark conditions was examined on 5th of March 2008 at station 29D (latitude 71.0331, longitude -123.897), which was also selected for metagenomic sequencing (see below). Individual seawater samples (60 mL) were collected at a depth of 12 m for each dark and light (i.e., 30-33 microEinstein m^{-2} s^{-1}) incubation. Samples were spiked with [^{14}C]-Urea (GE Healthcare, CFA41, 61 mCi mmol^{-1}, 0.5 µM final conc.) and incubated in ice-cold seawater for ca. 6.5 h. Killed controls (20 mL) for each treatment were prepared by adding 3 mL of formaldehyde (3.7% final conc.) to the samples prior the addition of the isotope. At the end of the incubation, samples were immediately filtered through both 0.2 and 0.6 µm pore-size polycarbonate filters (25 mm, GE Water), with half of the sample volume filtered through each pore size. Filters were immediately rinsed with 0.2 µm filtered-seawater and placed in individual vials. Scintillation cocktail (Fisher Scientific, Scintiverse BD cocktail SX18-4) was added to each filter (8 mL), and the radioactivity was measured in a Packard Liquid Scintillation analyzer Tri-Carb 2900 TR onboard the ship.

Metagenomic sample collection and pyrosequencing. The sample for metagenomic analysis was collected on 10 March 2008 (station 29D) at a depth of 65 m. Microbial biomass was collected by filtering 11 L of seawater (prefiltered by a 52 µm mesh) through 0.2 µm polycarbonate filters (47mm diameter, Durapore) using a peristaltic pump. Filters were placed in sterile microcentrifuge tubes filled with 1.8 mL of lysis buffer (40 mM EDTA pH 8.0, 750 mM sucrose, 50 mM Tris-HCl pH 8.3) and kept at -80ºC. DNA was extracted and purified as previously described (1). Purified DNA was pyrosequenced using a 454 FLX system (454 Life Sciences, Branford, CT, USA) at the Royal Institute of Technology (KTH, Stockholm, Sweden). The total size of the metagenome was 197 Mbp, being the total number of reads 520 372, and the average length 369 bp.

Metagenome analysis. Reads were quality trimmed using Lucy (12) and loaded into SmashCommunity (13), removing duplicate reads in the process. Reads were assembled using the Celera Assembler (14) with SmashCommunity standard parameters. Genes were then predicted using MetaGeneMark (15) and the protein translations were mapped to eggNOG v2 using Blastp (Bit score cutoff of 60) (16). For phylogenetic analysis, Blastn (17) was used to map the trimmed reads to a custom reference genome database consisting of all complete microbial genomes in NCBI up to June 2010 augmented with partial marine genomes from the Moore Microbial Genomes project and marine fosmids, which could be phylogenetically assigned to the NCBI taxonomy. In order to correctly account for the genome size of different organisms when calculating their abundance, we assigned the genome size of the closest sequenced organism to each of the added fosmid or contig. For taxonomic assignment of the reads, we used the same set of cutoffs and parameters as Arumugam and colleagues (18), including a 65% nucleotide identity cutoff for assignments at the phylum level. The contigs that only had reads assigned to *Thaumarchaeota* using a 65% identity cutoff were assigned to this phylum (18). In addition, all contigs which had at least 50% of the reads assigned to *Thaumarchaeota* were manually checked for phylogenetic assignment. In order to link functional and phylogenetic annotation we used SmashCommunity to compute the contribution of different phyla to genes and orthologous groups (18). Gene and COG abundances (or copy number) were calculated by summing up the lengths of all reads which

were assigned to each gene or COG and then divided by the length of the gene, and therefore represent the abundance or number of copies of each full-length gene.

Comparative genomic analysis. The three available MGI thaumarchaeal genomes (*Nitrosopumilus maritimus* SCM1, *Candidatus "*Nitrosoarchaeum limnia SFB1" and *Candidatus "*Cenarchaeum symbiosum A") were used to build orthologous groups specific to *Thaumarchaeota* (referred to as thaumNOGs) using the eggNOG pipeline (19). Additionally, we functionally annotated the three genomes using eggNOG version 2 (16). A total of 1479 thaumNOGs (with at least two members) were identified. Thereafter, the genes identified in thaumarchaeal contigs and reads in the metagenome were assigned to the thaumNOGs using Blastp with a bit score cutoff of 60. All non-assigned thaumarchaeal genes were clustered using a 95% identity criterion into 868 non-redundant genes.

16S rRNA gene sequence phylogenetic analysis of metagenomic *Thaumarchaeota*. Thaumarchaeal 16S rDNA sequences with a minimum length of 200 bp were retrieved from the metagenome and imported into the ARB software package (20) loaded with the SILVA SSURef_106 non-redundant database. Polar thaumarchaeal sequences retrieved in previous studies (5,21) were also imported into ARB for the analysis. All imported sequences were manually checked in the alignment and added to the ARB reference tree using the Parsimony Quick Add Marked Tool, thereby maintaining the overall tree topology.

Nutrient analyses. Samples for nutrient analyses were collected in acid-cleaned tubes (stored with 10% HCl) and frozen or stored in the dark at 4°C prior to analysis within a few hours of collection. Urea was analyzed using the method of Mulveena & Savidge (22). Concentrations of nitrate, nitrite and silicate were determined using standard colorimetric methods (23) adapted for the AutoAnalyzer 3 (Bran + Luebbe). Determinations of ammonia were done on fresh samples using the fluorometric method (24). Reagents were added within minutes of sample collection.

References

1. Alonso-Sáez L, Galand PE, Casamayor EO, Pedrós-Alió C, Bertilsson S. (2010). High bicarbonate assimilation in the dark by Arctic bacteria. *ISME J* 4:1581–1590.

2. HardingT., Jungblut AD, Lovejoy C, Vincent WF. (2011) Microbes in High Arctic Snow and Implications for the Cold Biosphere. *Appl Environ Microbiol* 77: 3234-3243.

3. Wuchter C et al. (2006) Archaeal nitrification in the ocean. *Proc Natl Acad Sci USA* 103: 12317–12322.

4. Coolen MJL et al. (2007) Putative ammonia-oxidizing Crenarchaeota in suboxic waters of the Black Sea: a basin-wide ecological study using 16S ribosomal and functional genes and membrane lipids. *Environ Microbiol* 9: 1001-1016

5. Alonso-Sáez L, Andersson A, Heinrich F, Bertilsson S (2011) High archaeal diversity in Antarctic circumpolar deep waters. *Environ Microbiol Rep* 3: 689–697.

6. Pernthaler A, Pernthaler J, Amann R. (2004). Sensitive multi-color fluorescence in situ hybridization for the identification of environmental microorganisms. Molecular microbial ecology manual, 2nd ed. Kluwer Academic Publishers, Dordrecht, The Netherlands 711–726.

7. Amann RI et al. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* 56: 1919–1925.

8. Daims H, Brühl A, Amann R, Schleifer K, Wagner M. (1999). The domain-specific probe Eub338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* 22: 434–444.

9. Massana R, Murray AE, Preston CM, DeLong EF. (1997). Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl Environ Microbiol* 63: 50–56.

10. Teira E, Reinthaler T, Pernthaler A, Pernthaler J, Herndl GJ. (2004). Combining catalyzed reporter deposition-fluorescence in situ hybridization and microautoradiography to detect substrate utilization by bacteria and Archaea in the deep ocean. *Appl Environ Microbiol* 70: 4411–4414.

11. Alonso-Sáez L, Sánchez O, Gasol JM, Balagué V, Pedrós-Alió C. (2008). Winter-to-summer changes in the composition and single-cell activity of near-surface Arctic prokaryotes. *Environ Microbiol* 10: 2444–2454.

12. Chou HH, Holmes MH. (2001). DNA sequence quality trimming and vector removal. *Bioinformatics* 17: 1093–1104.

13. Arumugam M, Harrington ED, Foerstner KU, Raes J, Bork P. (2010). SmashCommunity: a metagenomic annotation and analysis tool. *Bioinformatics* 26: 2977–2978.

14. Miller JR et al. (2008). Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24: 2818–2824.

15. Zhu W, Lomsadze A, Borodovsky M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res* 38: E132.

16. Muller J et al. (2010) eggNOG v2.0: extending the evolutionary genealogy of genes

with enhanced non-supervised orthologous groups, species and functional annotations. *Nucleic Acids Res* 38: D190-D195

17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* 215: 403–410.

18. Arumugam M et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473: 174–180.

19. Powell et al. (2012) EggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* 40: D284-D289

20. Ludwig W et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* 32:1363–1371.

21. Kalanetra KM, Bano N, Hollibaugh JT (2009) Ammonia-oxidizing Archaea in the Arctic Ocean and Antarctic coastal waters. *Environ Microbiol* 11: 2434–2445.

22. Mulvenna PF, Savidge G (1992) A Modified Manual Method for the Determination of Urea in Seawater using Diacetylmonoxime Reagent. *Estuar Coast Shelf S* 34: 429-438.

23. Grasshoff K (1999) Methods of seawater analyses, 600 pp., Weinheim, New York.

24. Holmes RM, Aminot A, Kerouel R, Hooker BA, Peterson JB (1999) A simple and precise method for measuring ammonium in marine and freshwater ecosystems. *Can J Fishe Aquat Sci* 56: 1801-1808.

Table S1. Physico-chemical parameters at the stations sampled in the Southeast Beaufort Sea (Arctic Ocean) at the surface and variable depths along the halocline, and different water masses in the Amundsen and Ross seas (Antarctic Ocean). Q: quanta, nd: non detectable

Southeast Beaufort Sea (Arctic Ocean)

Station ID	Date	Longitude	Latitude	Depth (m)	Temperature (ºC)	Salinity	PAR (µmol Q m^{-2} s^{-1})	Nitrite (µM)	Nitrate (µM)	Silicate (µM)
17D	2008-01-17	-124.96	71.52	12	-1.648	30.811	0.0112	-	0.67	-
17D	2008-01-17	-124.96	71.52	60	-1.313	32.034	0.0018	-	7.46	-
19D	2008-02-04	-124.81	71.08	12	-1.696	30.930	0.0121	0.30	0.89	3.82
19D	2008-02-04	-124.81	71.08	160	-0.001	34.572	<0.0001	-	-	-
27D	2008-03-02	-123.46	70.76	9	-1.722	31.644	0.710	0.29	3.95	10.12
27D	2008-03-02	-123.46	70.76	71	-1.590	32.791	0.0001	0.13	13.69	27.01
D33	2008-03-27	-121.79	71.06	10	-1.721	31.662	0.499	0.24	3.33	8.24
D33	2008-03-27	-121.79	71.06	60	-1.318	32.416	0.173	0.14	11.87	26.34
D37	2008-04-10	-124.61	71.25	10	-1.699	31.918	3.29	0.15	4.44	9.84
D37	2008-04-10	-124.61	71.25	120	-0.971	33.827	0.0276	0.10	16.55	28.75
9016	2008-05-27	-129.04	74.35	10	-0.442	28.495	2.55	0.06	nd	2.84
FB-05	2008-06-15	-125.87	69.96	12	-0.786	31.550	0.402	0.20	2.82	9.32
FB-05	2008-06-15	-125.87	69.96	28	-1.425	32.357	0.472	0.25	8.43	18.69

Amundsen Sea (Antarctic Ocean)

Station ID	Date	Longitude	Latitude	Ice cover	Depth (m)	Temperature (ºC)	Salinity	Nitrite (µM)	Nitrate (µM)	Ammonia (µM)	Silicate (µM)
6	11/12/07	-105.15	-70.98	50%	9	-1.693	33.811	0.07	30.83	0.05	77.70
6	11/12/07	-105.15	-70.98	50%	50	-1.783	33.911	0.04	31.46	nd	77.90
6	11/12/07	-105.15	-70.98	50%	101	-1.775	33.960	0.02	32.08	nd	74.70
6	11/12/07	-105.15	-70.98	50%	251	0.611	34.489	nd	35.49	nd	95.80
6	11/12/07	-105.15	-70.98	50%	505	1.614	34.729	nd	34.49	nd	102.00
9	12/12/07	-106.00	-72.49	50%	10	-1.637	33.751	0.08	29.32	0.09	72.50
9	12/12/07	-106.00	-72.49	50%	51	-1.783	33.827	0.06	30.74	0.09	73.80
9	12/12/07	-106.00	-72.49	50%	253	-0.294	34.290	nd	36.39	nd	92.30
9	12/12/07	-106.00	-72.49	50%	455	1.395	34.719	nd	36.19	nd	105.00
11	13/12/07	-106.24	-73.53	70%	10	-1.718	33.818	0.09	28.41	0.18	77.10
11	13/12/07	-106.24	-73.53	70%	51	-1.714	33.962	0.07	29.43	0.28	79.60
11	13/12/07	-106.24	-73.53	70%	253	-1.273	34.095	nd	31.79	0.07	87.10
11	13/12/07	-106.24	-73.53	70%	507	1.262	34.684	nd	34.19	nd	106.00
16	18/12/07	-115.68	-73.94	0%	10	-1.503	33.942	0.07	24.93	0.20	88.70
16	18/12/07	-115.68	-73.94	0%	19	-1.504	33.942	0.07	25.03	0.26	88.60
16	18/12/07	-115.68	-73.94	0%	34	-1.501	33.942	0.06	24.94	0.20	87.80
16	18/12/07	-115.68	-73.94	0%	51	-1.501	33.942	0.07	24.93	0.22	87.70
16	18/12/07	-115.68	-73.94	0%	101	-1.557	33.950	0.07	26.34	0.46	89.30
16	18/12/07	-115.68	-73.94	0%	252	-1.750	34.017	0.08	30.42	0.22	88.40
16	18/12/07	-115.68	-73.94	0%	502	-1.218	34.184	nd	31.99	nd	94.00
16	18/12/07	-115.68	-73.94	0%	795	0.407	34.527	-	-	-	-

Amundsen Sea (Antarctic Ocean)

Station ID	Date	Longitude	Latitude	Ice cover	Depth (m)	Temperature (ºC)	Salinity	Nitrite (µM)	Nitrate (µM)	Ammonia (µM)	Silicate (µM)
21	21/12/07	-126.70	-71.72	100%	11	-1.830	34.042	0.06	31.14	nd	73.20
21	21/12/07	-126.70	-71.72	100%	21	-1.813	34.060	0.06	31.14	nd	73.10
21	21/12/07	-126.70	-71.72	100%	51	-1.811	34.066	0.05	31.05	nd	72.60
21	21/12/07	-126.70	-71.72	100%	102	-1.610	34.139	0.02	31.18	nd	73.40
21	21/12/07	-126.70	-71.72	100%	253	1.259	34.597	nd	33.39	nd	86.50
21	21/12/07	-126.70	-71.72	100%	507	1.603	34.725	nd	32.69	nd	96.40
21	21/12/07	-126.70	-71.72	100%	795	1.357	34.727	nd	33.09	nd	104.00
21	21/12/07	-126.70	-71.72	100%	1012	1.188	34.722	nd	32.89	nd	110.00

Ross Sea (Antarctic Ocean)

Station ID	Date	Longitude	Latitude	Ice cover	Depth (m)	Temperature (ºC)	Salinity	Nitrite (µM)	Nitrate (µM)	Ammonia (µM)	Silicate (µM)
23	27/12/07	-156.33	-76.23	80%	11	-1.468	33.830	0.11	28.89	0.16	81.10
23	27/12/07	-156.33	-76.23	80%	101	-1.825	34.021	0.02	30.38	nd	82.30
26	28/12/07	-163.58	-78.60	5%	10	-1.134	34.025	0.12	27.98	nd	82.80
26	28/12/07	-163.58	-78.60	5%	51	-1.383	34.036	0.10	28.70	0.13	82.60
26	28/12/07	-163.58	-78.60	5%	254	-1.806	34.204	0.10	31.30	nd	84.20
26	28/12/07	-163.58	-78.60	5%	405	-1.912	34.274	0.03	31.97	nd	85.10

Ross Sea (Antarctic Ocean)

Station ID	Date	Longitude	Latitude	Ice cover	Depth (m)	Temperature (ºC)	Salinity	Nitrite (µM)	Nitrate (µM)	Ammonia (µM)	Silicate (µM)
28	29/12/07	-176.77	-77.97	0%	10	-0.560	34.284	0.02	31.38	nd	81.40
28	29/12/07	-176.77	-77.97	0%	21	-0.852	34.238	0.15	25.25	0.14	75.10
28	29/12/07	-176.77	-77.97	0%	51	-1.399	34.528	0.08	29.12	0.11	79.20
28	29/12/07	-176.77	-77.97	0%	102	-1.676	34.723	0.07	31.03	nd	82.60
28	29/12/07	-176.77	-77.97	0%	255	-1.727	34.338	0.01	32.10	nd	85.20
28	29/12/07	-176.77	-77.97	0%	507	-1.944	34.767	nd	31.99	nd	81.10
29	30/12/07	-178.63	-77.85	0%	10	-0.206	34.178	0.18	24.43	0.13	74.90
29	30/12/07	-178.63	-77.85	0%	103	-1.389	34.245	0.11	29.19	0.44	80.30
31	02/01/08	170.50	-77.02	0%	21	-0.791	34.406	0.12	17.50	0.22	76.40
31	02/01/08	170.50	-77.02	0%	51	-0.721	34.421	0.03	31.30	nd	81.30
31	02/01/08	170.50	-77.02	0%	102	-1.466	34.538	0.12	24.68	1.55	78.70
31	02/01/08	170.50	-77.02	0%	253	-1.900	34.697	0.04	31.16	nd	79.60
31	02/01/08	170.50	-77.02	0%	507	-1.893	34.736	0.02	30.98	nd	78.50
31	02/01/08	170.36	-77.02	0%	806	-1.885	34.767	0.04	31.40	nd	82.30
32	02/01/08	168.33	-77.17	10%	10	-1.284	34.098	0.12	14.18	0.16	80.20
32	02/01/08	168.33	-77.17	10%	102	-1.792	34.502	0.13	30.07	0.45	82.60

Table S2. Description of the fifty most abundant orthologous groups (OGs) in the Arctic *Thaumarchaeota* metagenome, as analyzed with eggNOG, and the read counts retrieved for each OG. Cat: Categories. K: Transcription, H: Coenzyme transport and metabolism, C: Energy production and conversion, E: Amino acid transport and metabolism, F: Nucleotide transport and metabolism, G: Carbohydrate transport and metabolism, P: Inorganic ion transport and metabolism, O: Posttranslational modification, protein turnover, chaperones, N: Cell motility, I: Lipid transport and metabolism, L: Replication, recombination and repair, R: General function prediction only, J: Translation, ribosomal structure and biogenesis, Q: Secondary metabolites biosynthesis, transport and catabolism

OG	Counts	Cat	OG Description
COG1522	139	K	Transcriptional regulators
COG1060	101	H, R	Thiamine biosynthesis enzyme ThiH and related enzymes
COG0464	96	O	ATPases of the AAA+ class
COG0624	79	E	Acetylornithine deacetylase/Succinyl-diaminopimelate desuccinylase and related deacylases
COG1405	78	K	Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB
COG0046	74	F	Phosphoribosylformylglycinamidine (FGAM) synthase
COG0526	68	C, O	Thiol-disulfide isomerase and thioredoxins
COG0457	68	R	FOG: TPR repeat
COG0604	67	C, R	NADPH:quinone reductase and related Zn-dependent oxidoreductase
COG0115	66	E, H	Branched-chain amino acid aminotransferase/4-amino-4-deoxychorismate lyase
COG1331	66	O	Highly conserved protein containing a thioredoxin domain
COG0574	61	G	Phosphoenolpyruvate synthase/pyruvate phosphate dikinase
COG1032	61	C	Fe-S oxidoreductase
COG1063	60	E, R	Threonine dehydrogenase and related Zn-dependent dehydrogenases
COG0527	59	E	Aspartokinases
COG0667	59	C	Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)
COG1958	58	K	Small nuclear ribonucleoprotein (snRNP) homolog
COG2897	57	P	Rhodanese-related sulfurtransferase
COG1048	56	C	Aconitase A
COG0638	56	O	20S proteasome, alpha and beta subunits
COG5491	54	N	Conserved protein implicated in secretion
COG0365	54	I	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases
COG0644	52	C	Dehydrogenases (flavoproteins)
COG1651	52	O	Protein-disulfide isomerase
COG0417	51	L	DNA polymerase elongation subunit (family B)

COG0525	51	J	Valyl-tRNA synthetase
COG0484	51	O	DnaJ-class molecular chaperone with C-terminal Zn finger domain
COG0043	50	H	3-polyprenyl-4-hydroxybenzoate decarboxylase and related decarboxylases
COG1208	50	J,M	Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysaccharide biosynthesis/translation initiation factor 2B
COG1204	49	R	Superfamily II helicase
COG0004	49	P	Ammonia permease
COG0312	48	R	Predicted Zn-dependent proteases and their inactivated homologs
COG0119	48	E	Isopropylmalate/homocitrate/citramalate synthases
COG2132	48	Q	Putative multicopper oxidases
COG0495	47	J	Leucyl-tRNA synthetase
COG1064	47	R	Zn-dependent alcohol dehydrogenases
COG0137	47	E	Argininosuccinate synthase
COG1759	47	R	ATP-utilizing enzymes of ATP-grasp superfamily (probably carboligases)
COG0129	46	E, G	Dihydroxyacid dehydratase/phosphogluconate dehydratase
COG0065	46	E	3-isopropylmalate dehydratase large subunit
COG0162	46	J	Tyrosyl-tRNA synthetase
COG1053	45	C	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit
COG1042	45	C	Acyl-CoA synthetase (NDP forming)
COG0174	45	E	Glutamine synthetase
COG0142	44	H	Geranylgeranyl pyrophosphate synthase
COG0172	44	J	Seryl-tRNA synthetase
COG0719	43	O	ABC-type transport system involved in Fe-S cluster assembly, permease component
COG0652	43	O	Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family
COG0373	43	H	Glutamyl-tRNA reductase

Table S3. Arctic thaumarchaeal metagenome-unique Orthologous Groups (OGs, i.e., present in Arctic *Thaumarchaeota* and absent in the genomes of *Nitrosopumilus maritimus* SCM1, *Candidatus* "Nitrosoarchaeum limnia SFB1", and *Candidatus* "Cenarchaeum symbiosum A*"),* as analyzed with eggNOG v2. G: Carbohydrate transport and metabolism, E: Amino acid transport and metabolism, J: Translation, ribosomal structure and biogenesis, I: Lipid transport and metabolism, Q: Secondary metabolites biosynthesis, transport and catabolism, M: Cell wall/membrane/envelope biogenesis, U: Intracellular trafficking, secretion, and vesicular transport, V: Defense mechanisms, O: Posttranslational modification, protein turnover, chaperones, S: Function unknown, R: General prediction only.

OG	Category	OG description
COG0337	E	3-dehydroquinate synthetase
COG0403	E	Glycine cleavage system protein P (pyridoxal binding)
COG0752	J	Glycyl-tRNA synthetase, alpha subunit
COG0751	J	Glycyl-tRNA synthetase, beta subunit
COG2097	J	Ribosomal protein L31E
COG1472	G	Beta-glucosidase-related glycosidases
COG0304	IQ	3-oxoacyl-(acyl-carrier-protein) synthase
COG2885	M	Outer membrane protein and related peptidoglycan-associated (lipo)proteins
COG5523	S	Predicted integral membrane protein
COG1238	S	Predicted membrane protein
NOG149265	R	Macrophage receptor with collagenous structure
NOG145280		Collagen triple helix repeat protein
COG4796	U	Type II secretory pathway, component HofQ
COG3510	V	Cephalosporin hydroxylase
COG1132	V	ABtype multidrug transport system, ATPase and permease
COG1123	R	ATPase of various ABCtype transport systems
COG2214	O	DnaJ-class molecular chaperone
NOG72401	R	Calcium ion binding protein
COG1011	R	Predicted hydrolase (HAD superfamily)
NOG119150	R	Ubia prenyltransferase
COG2802	R	Uncharacterized protein, similar to the N-terminal doma
COG1540	R	Uncharacterized proteins, homologs of lactam utilizatio
NOG135128	S	Flavin mononucleotide (FMN)-binding protein
NOG25588	S	Tetratricopeptide
NOG136474	S	Annotation not available
NOG79303	S	Annotation not available
NOG87924	S	Annotation not available
COG2013	S	Uncharacterized conserved protein
COG1284	S	Uncharacterized conserved protein
COG1801	S	Uncharacterized conserved protein

Table S4. Copy number of genes encoding Marine Group I *Archaea* 16S rRNA, monooxygenase subunit A (*amoA*), and urease alpha subunit (*ureC*) in surface and halocline samples of the Arctic Southeast Beaufort Sea.

Water mass	Date	Depth (m)	MGI *Archaea* 16S rRNA (10^3 copies mL^{-1} ± SD)	*amoA* (10^3 copies mL^{-1} ± SD)	*ureC* (10^3 copies mL^{-1} ± SD)
Surface	2008-01-15	10	1.69 ± 0.07	4.80 ±1.65	0.25 ± 0.02
	2008-02-03	30	9.11 ±1.64	20.42 ± 0.56	1.44 ± 0.08
	2008-03-02	12	7.05 ± 0.65	20.48 ± 1.21	1.30 ± 0.19
	2008-03-26	10	9.18 ± 0.66	9.89 ± 6.88	1.13 ± 0.03
	2008-04-08	10	2.98 ± 1.62	10.17 ± 0.46	0.69 ± 0.06
	2008-05-28	12	0.81 ± 0.15	1.21 ± 0.19	0.13 ± 0.01
	2008-06-15	12	1.07 ± 0.09	3.81 ± 0.28	0.39 ± 0.05
Halocline	2008-01-15	55	3.83 ± 1.83	24.52 ± 0.97	3.80 ± 2.71
	2008-02-03	160	4.60 ± 0.44	17.69 ± 1.51	6.95 ± 0.70
	2008-03-02	70	5.33 ± 0.69	16.99 ± 2.21	3.32 ± 0.19
	2008-03-10	65	18.13 ± 2.48	49.42 ± 3.07	6.28 ± 0.62
	2008-03-26	60	9.21 ± 2.58	20.50 ± 1.97	7.12 ± 0.34
	2008-04-08	80	7.72 ± 0.53	16.07 ± 1.32	1.22 ± 0.19
	2008-05-28	55	3.68 ± 1.05	7.18 ± 7.43	5.47 ± 1.18
	2008-06-15	30	3.73 ± 0.58	11.85 ± 0.60	3.32 ± 0.46

Table S5. Copy numbers of genes encoding MGI *Archaea* 16S rRNA, ammonia monooxygenase subunitA (*amoA*) and urease alpha subunit (*ureC*) in different Antarctic water masses as analysed by quantitative PCR. AASW: Antarctic Surface Waters, THE: Thermocline, SW: Shelf Waters, CDW: Circumpolar Deep Waters, n.d.: non detectable

Sample Code	**MGI *Archaea* 16S rRNA genes** (10^3 copies mL^{-1} ± SD)	**MGI *Archaea* *amoA* genes** (10^3 copies mL^{-1} ± SD)	**MGI *Archaea* *ureC* genes** (10^3 copies mL^{-1} ± SD)
AASW			
A6-10	34.08 ± 0.58	28.93 ± 0.26	22.96 ± 2.43
A6-50	59.98 ± 2.49	38.97 ± 0.53	23.67 ± 1.18
A6-100	17.07 ± 0.07	18.98 ± 0.22	13.39 ± 0.17
A9-10	1.89 ± 0.08	11.11 ± 0.06	n.d.
A9-50	29.79 ± 1.14	16.01 ± 0.31	0.71 ± 0.21
A11-50	20.71± 0.35	21.59 ± 0.20	0.69 ± 0.05
R26-10	5.10 ± 0.17	4.02 ± 0.03	0.97 ± 0.12
R26-50	4.70 ± 0.11	5.29 ± 0.05	2.51 ± 0.27
R28-10	0.10 ± 0.003	0.06 ± 0.001	n.d.
THE			
R26-250	26.02 ± 0.51	29.46 ± 0.23	7.10 ± 2.03
R26-400	22.35 ± 0.56	29.99 ± 0.29	10.15 ± 4.00
R28-50	9.93 ± 0.27	8.93 ± 0.17	2.58 ± 0.20
R28-100	113.99 ± 2.29	138.49 ± 0.69	1.07 ± 0.03
R28-250	30.84 ± 0.94	36.81 ± 0.65	9.10 ± 1.47
SW			
R28-500	0.74 ± 0.02	0.99 ± 0.01	0.25 ± 0.07
R31-250	23.53 ± 1.24	25.75 ± 0.16	4.01 ± 0.37
R31-500	20.25 ± 0.51	25.98 ± 0.15	12.46 ± 0.68
CDW			
A6-250	10.06 ± 0.19	13.99 ± 0.19	21.47 ± 6.15
A6-500	16.60 ± 0.48	5.67 ± 0.08	8.93 ± 1.70
A9-250	2.39 ± 0.05	6.83 ± 0.22	21.95 ± 10.09
A11-250	26.38 ± 0.81	35.39 ± 0.47	18.40 ± 0.38
A11-500	18.07 ± 0.50	10.84 ± 0.12	14.08 ± 3.43

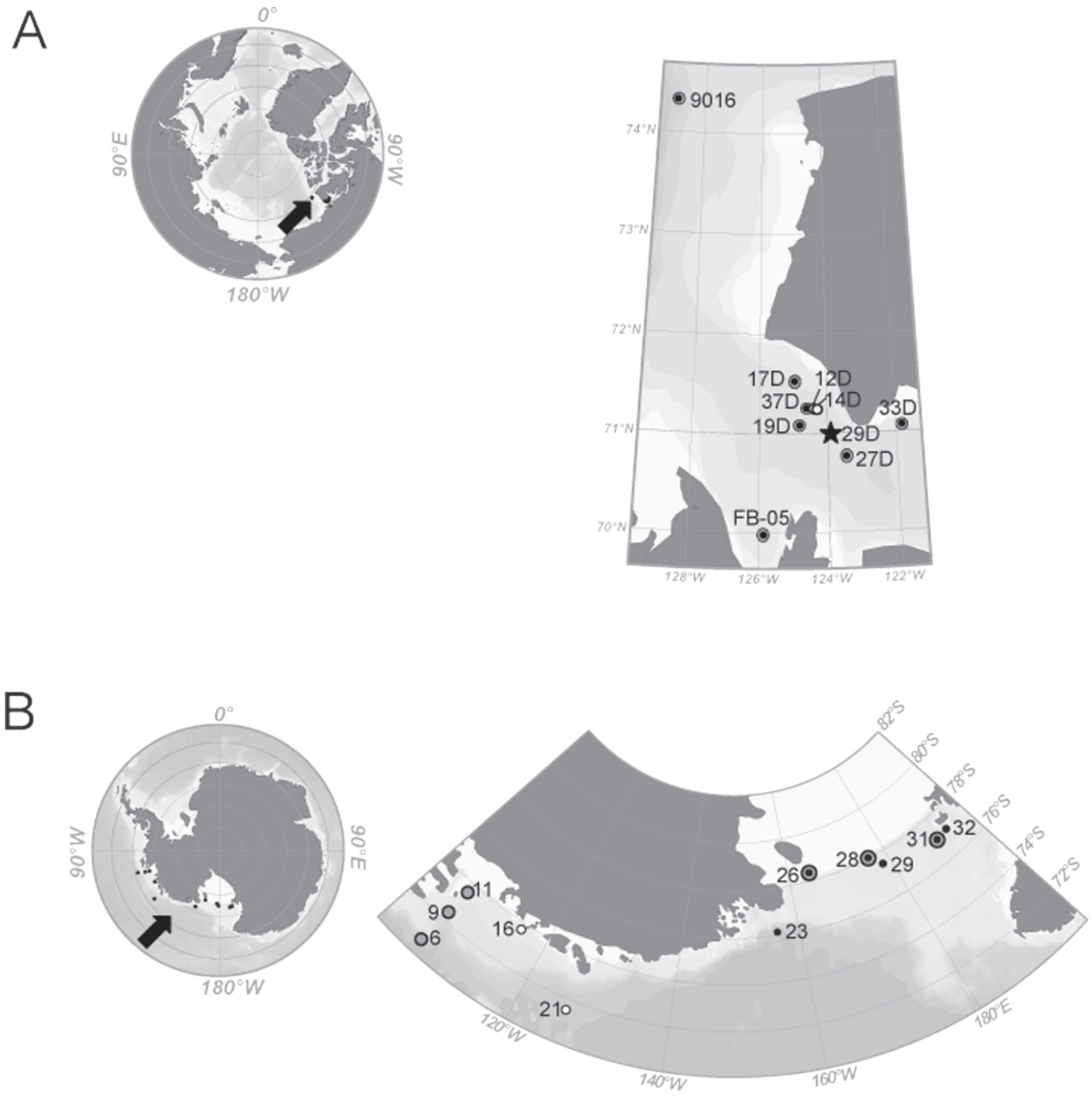

Figure S1. (A) Map showing the stations sampled during the Arctic CFL cruise. CARD-FISH, MAR-FISH and qPCR analyses were done in all stations, except for stations 12D and 14D (marked with an open circle) where only CARD-FISH was done and 29D (marked by a star symbol), where qPCR, the metagenomic analysis and the urea uptake experiment were carried out. Stations were sampled at different months , including December 2007 (Sta 12D), January (Sta 14D and 17D), February (Sta 19D), March (Sta 27D, 29D and 33D), April (37D), May (9016) and June (FB-05) 2008. (B) Map showing the stations sampled during the Antarctic OSO cruise during the austral summer 2007/2008. Black circles denote stations where both CARD-FISH and MAR-FISH analyses were carried out, open circles represent stations where only CARD-FISH analyses were performed, and grey circles indicate stations where qPCR analyses were carried out.

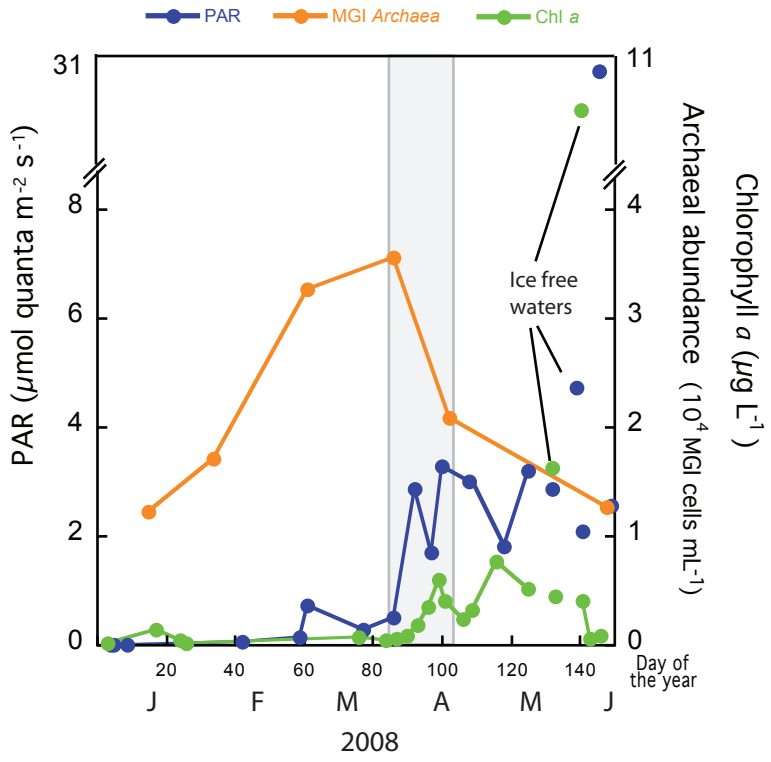

Figure S2. Dynamics of the abundance of Marine Group I (MGI) archaeal cells, chlorophyll *a* (Chl *a*) concentration and photosynthetically active radiation (PAR) in Arctic surface waters during the CFL cruise. The grey area represents the period of marked decrease in archaeal abundance.

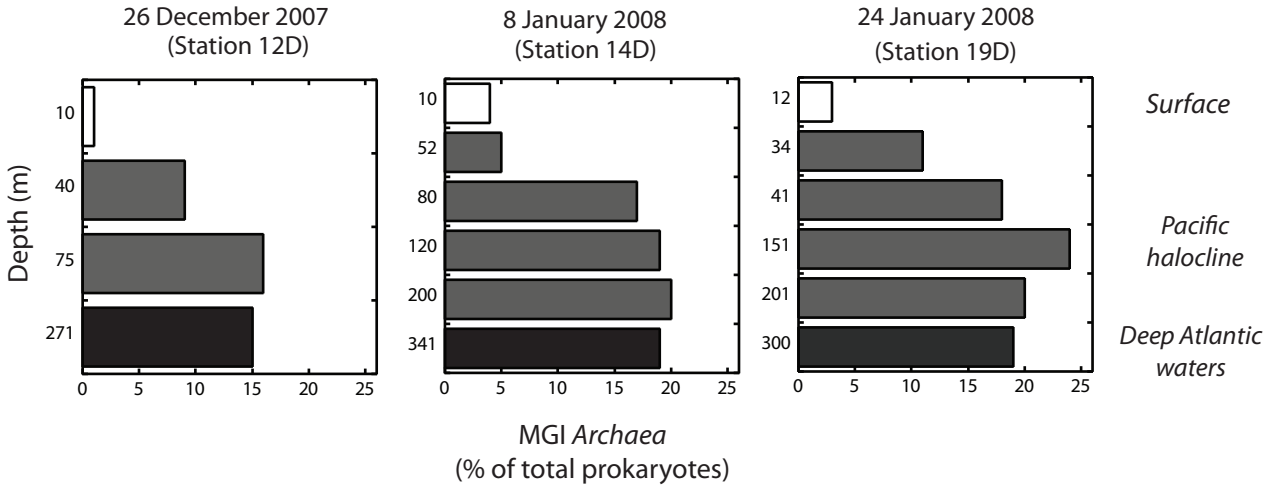

Figure S3. Depth profiles of the abundance of MGI *Archaea* (expressed as percentages of total prokaryotes) from 26 December 2007 to 24 January 2008 in the Southern Beaufort Sea (Arctic). White, grey and black bars represent samples collected in surface, halocline and deep Atlantic waters, respectively.

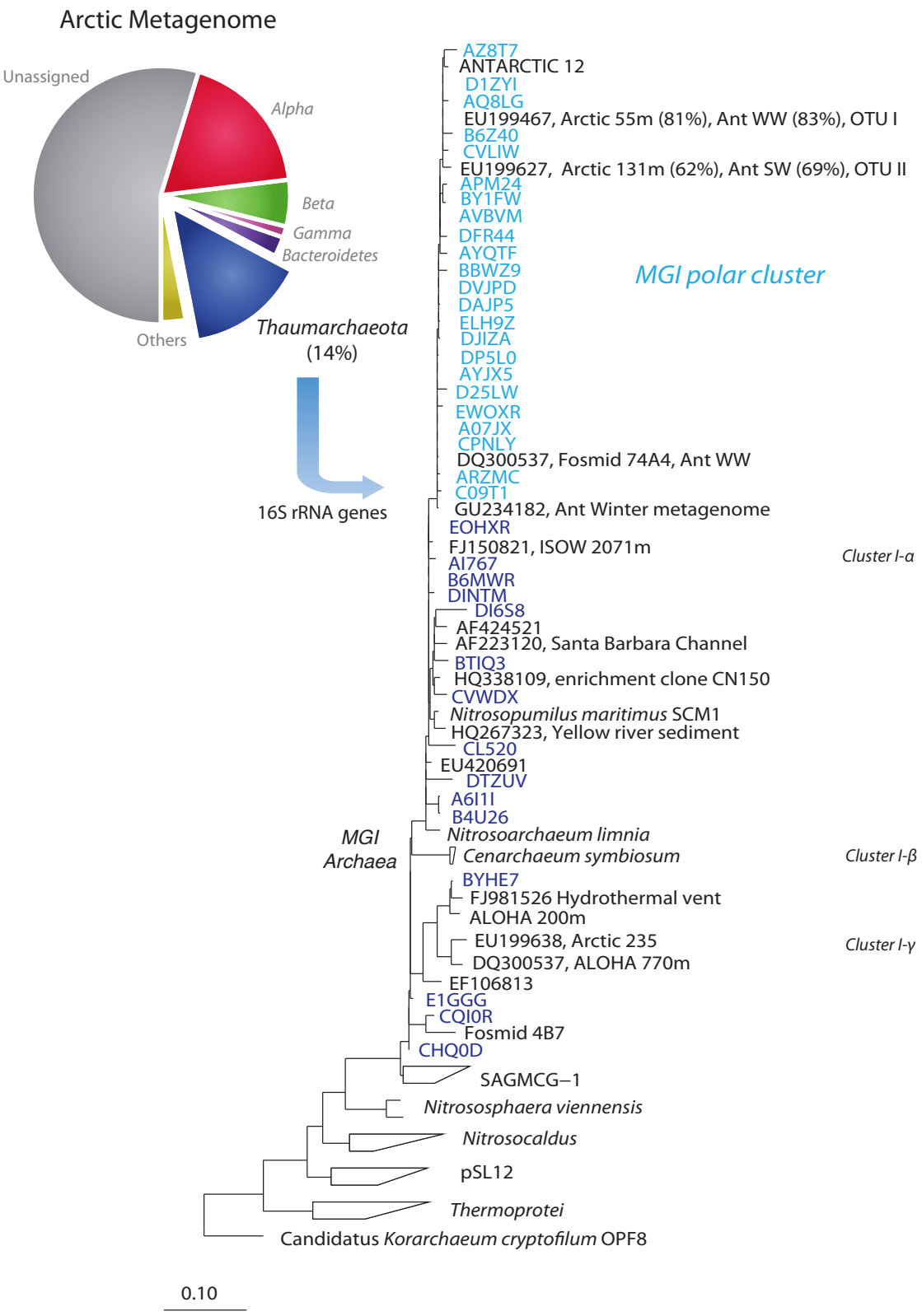

Figure S4. Contribution of different classes or phyla to the reads obtained in the metagenome (pie chart), and phylogenetic tree showing the 16S rRNA genes of thaumarchaeal reads retrieved from the metagenome (in blue). Light blue sequences represent a cluster of Arctic thaumarchaeal sequences found in the metagenome with >98% nucleotide similarity, which represent a closely related polar thaumarchaeal population. The fosmid 74A4 and sequences EU199467, EU199627, and GU234182 (Genbank accession numbers) previously retrieved in high abundance from polar waters also belonged to this cluster. The percent contribution of these phylotypes to total thaumarchaeal clones at the sites of study (Kalanetra et al. 2009) appear in parentheses. Ant: Antarctic, WW: Winter waters, SW: Shelf Waters, ISOW: Iceland-Scotland Overflow Water, Alpha: Alphaproteobacteria, Beta: Betaproteobacteria, Gamma: Gammaproteobacteria

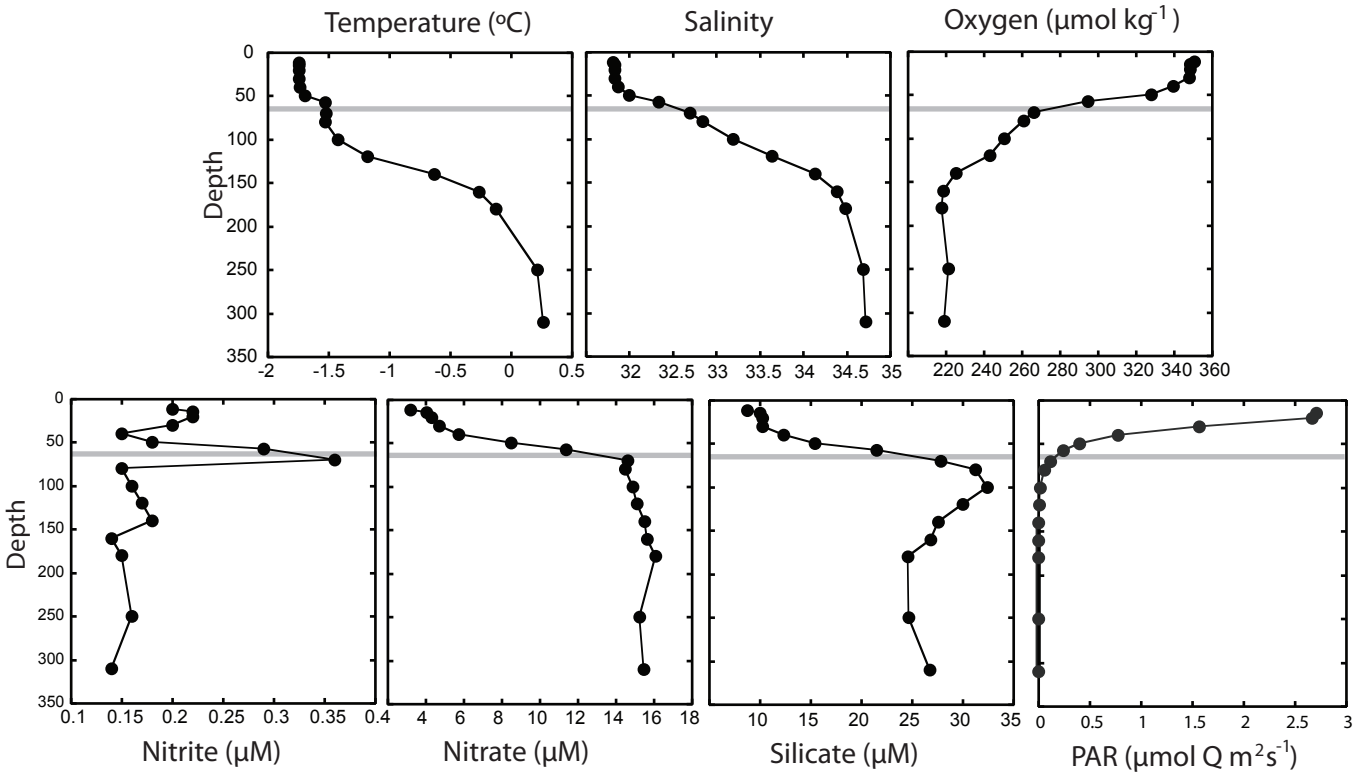

Figure S5. Depth profiles of temperature (ºC), salinity, oxygen (µmol kg^{-1}), nutrient concentrations (µM) and PAR radiation (µmol quanta $m^2 s^{-1}$) at the station where the metagenome was retrieved (Sta 29D). The grey line indicates the depth where the metagenome was isolated. Q: quanta.

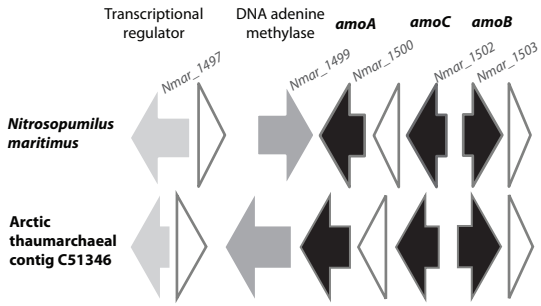

Figure S6. Comparison of the contig C51346 retrieved from the Arctic metagenome containing the amo operon and the corresponding genomic region of *Nitrosopumilus maritimus* SCM1. White arrows represent genes encoding hypothetical proteins

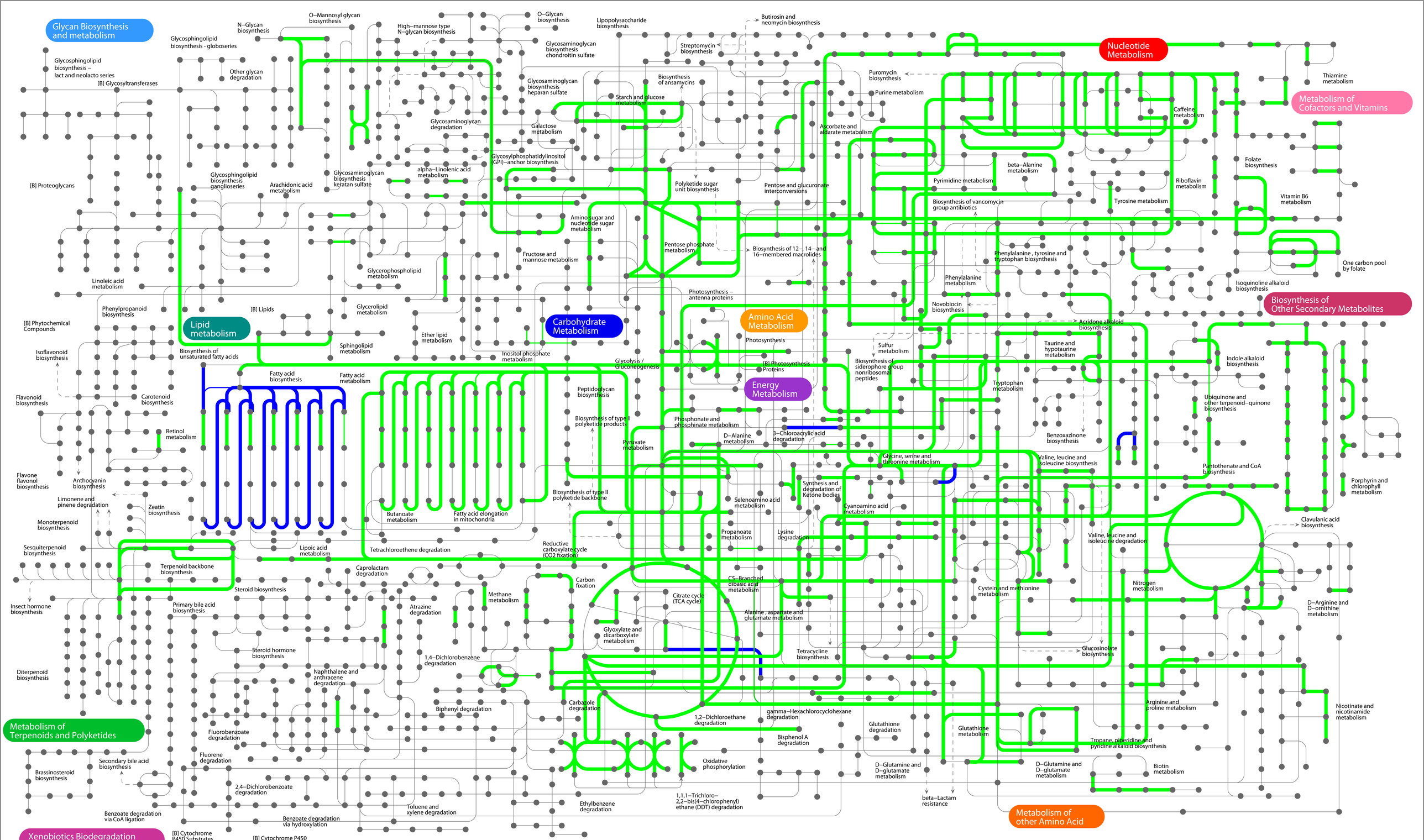

Figure S7. Metabolic map showing in green the common metabolic pathways present in the MGI *Thaumarchaeota* analyzed and in blue the pathways exclusive for the Arctic *Thaumarchaeota*.

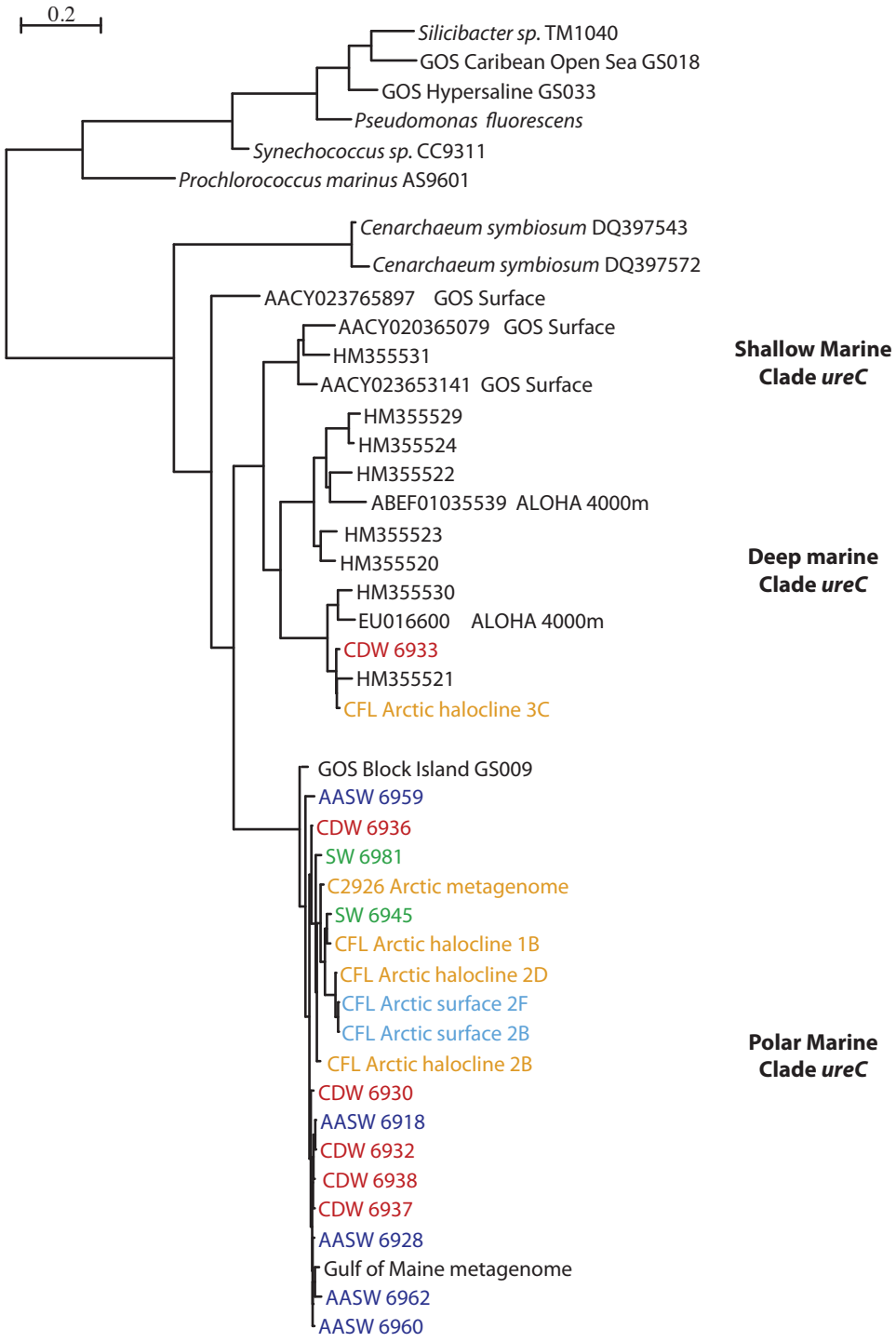

Figure S8. Maximum likelihood phylogenetic tree of *ureC* nucleotide sequences. Archaeal *ureC* sequences retrieved in Arctic surface (12 m) and halocline (65 m) waters from samples collected on the 10th March 2008 appear in light blue and orange, respectively. The *ureC* sequence found in a contig (C2926) from the Arctic metagenome analyzed in this study is also shown. Sequences retrieved in Antarctic suface waters (AASW), deep Shelf waters (SW) and Circumpolar Deep waters (CDW) appear in dark blue, green and red, respectively. Selected *ureC* sequences detected in Mediterranean waters (44), and metagenomes from the Gulf of Maine , Hawaii Ocean Time Series station ALOHA, and several sites of the Global Ocean Survey (GOS) are also shown in the tree. Shallow and Deep marine *ureC* clades have been named following the designation by Yakimov et al. (2011)(44).

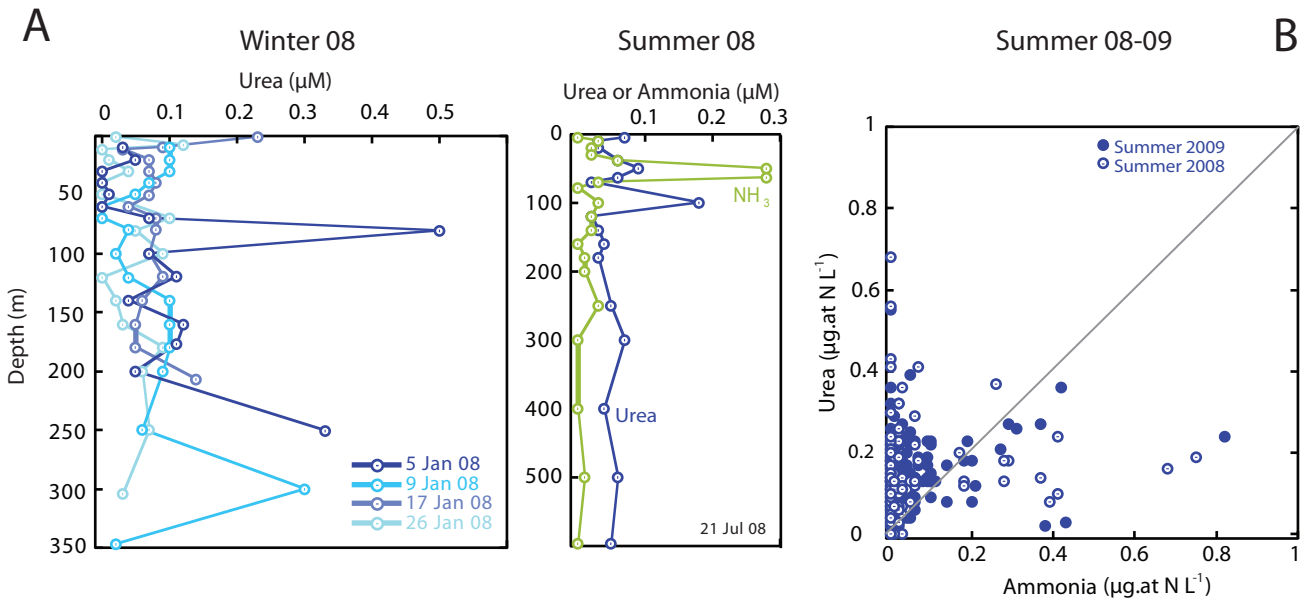

Figure S9. Concentration of urea and ammonia in the Arctic Southeast Beaufort Sea. A) Depth profiles of urea and ammonia concentration in samples collected in winter and summer during the CFL cruise. B) Concentration of ammonia versus urea in summer samples collected in the years 2008 and 2009 in the Southeast Beaufort Sea.