The ISME Journal (2011) 5, 866–878

www.nature.com/ismej

ORIGINAL ARTICLE

Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage

James T Hollibaugh[1], Scott Gifford[1], Shalabh Sharma, Nasreen Bano and Mary Ann Moran[1]
Department of Marine Sciences, University of Georgia, Athens, GA, USA

Quantitative PCR (qPCR) analysis revealed elevated relative abundance (1.8% of prokaryotes) of marine group 1 Crenarchaeota (MG1C) in two samples of southeastern US coastal bacterioplankton, collected in August 2008, compared with samples collected from the same site at different times (mean 0.026%). We analyzed the MG1C sequences in metatranscriptomes from these samples to gain an insight into the metabolism of MG1C population growing in the environment, and for comparison with ammonia-oxidizing bacteria (AOB) in the same samples. Assemblies revealed low diversity within sequences assigned to most individual MG1C open reading frames (ORFs) and high homology with '*Candidatus* Nitrosopumilus maritimus' strain SCM1 genome sequences. Reads assigned to ORFs for ammonia uptake and oxidation accounted for 37% of all MG1C transcripts. We did not recover any reads for Nmar_1354–Nmar_1357, proposed to encode components of an alternative, nitroxyl-based ammonia oxidation pathway; however, reads from Nmar_1259 and Nmar_1667, annotated as encoding a multicopper oxidase with homology to *nirK*, were abundant. Reads assigned to two homologous ORFs (Nmar_1201 and Nmar_1547), annotated as hypothetical proteins were also abundant, suggesting that their unknown function is important to MG1C. Superoxide dismutase and peroxiredoxin-like transcripts were more abundant in the MG1C transcript pool than in the complete metatranscriptome, suggesting an enhanced response to oxidative stress by the MG1C population. qPCR indicated low AOB abundance (0.0010% of prokaryotes), and we found no transcripts related to ammonia oxidation and only one RuBisCO transcript among the transcripts assigned to AOB, suggesting they were not responding to the same environmental cues as the MG1C population.
The ISME Journal (2011) **5,** 866–878; doi:10.1038/ismej.2010.172; published online 18 November 2010
Subject Category: Integrated genomics and post-genomics approaches in microbial ecology
Keywords: Crenarchaeota; metatranscriptome; *amoA*; ammonia oxidation; *Nitrosopumilus*; Thaumarchaeota

Introduction

Marine Group 1 Crenarchaeota (MG1C) are abundant and widespread in mesopelagic, open ocean environments (Karner *et al.*, 2001; Fuhrman and Hagström, 2008). They have proved difficult to culture, so our knowledge of their metabolism is based primarily on culture-independent methods (Ouverney and Fuhrman, 2000; Teira *et al.*, 2004; Kirchman *et al.*, 2007). Metagenomic data (Venter *et al.*, 2004; Treusch *et al.*, 2005) suggested that MG1C might have a role in ammonia oxidation and more recent research (Wuchter *et al.*, 2006; Hallam *et al.*, 2006a, b; Beman *et al.*, 2008; Santoro *et al.*, 2010) and the successful isolation of a representative MG1C (Konneke *et al.*, 2005) have confirmed this. As a consequence of these findings, the paradigm that members of the β- and γ-Proteobacteria are responsible for most of the ammonia oxidation in the ocean has come into question (reviewed in Francis *et al.*, 2007; Prosser and Nicol, 2008).

Studies of the distributions of planktonic ammonia oxidizing organisms have shown that ammonia-oxidizing Crenarchaeota (ammonia-oxidizing archaea, AOA) tend to be numerically dominant in the open ocean (Wuchter *et al.*, 2006; Mincer *et al.*, 2007; Agogue *et al.*, 2008; Beman *et al.*, 2008; de Corte *et al.*, 2008; Kalanetra *et al.*, 2009; Santoro *et al.*, 2010) and fjords (Urakawa *et al.*, 2010; Zaikova *et al.*, 2010). Most studies of AOA populations in estuaries (Francis *et al.*, 2005; Beman and Francis, 2006; Caffrey *et al.*, 2007; Mosier and Francis, 2008; Santoro *et al.*, 2008; Magalhaes *et al.*, 2009; Bernhard *et al.*, 2010) have focused on sediments. From these studies, it appears that the

Correspondence: JT Hollibaugh or MA Moran, Department of Marine Sciences, University of Georgia, Athens, GA 30602-3636, USA.
E-mail: mmoran@uga.edu or aquadoc@uga.edu
[1]These authors contributed equally to this work.
Received 31 May 2010; revised 6 September 2010; accepted 21 September 2010; published online 18 November 2010

relative abundance of ammonia-oxidizing bacteria (AOB) increases in estuaries relative to coastal waters or the open ocean. The environmental factors responsible for the success of AOB vs AOA in estuarine and coastal waters are not known, but the shift correlates with salinity in some systems (Caffrey *et al.*, 2007; Mosier and Francis, 2008; Santoro *et al.*, 2008; Magalhaes *et al.*, 2009; Bernhard *et al.*, 2010). However, the success of one group over the other is not likely to be directly based on salinity, as AOA can be dominant in the oligohaline reaches of some estuaries (Mosier and Francis, 2008) and in soils (Prosser and Nicol, 2008).

Estuaries are distinct from mesopelagic open ocean environments in a number of important ways: salinity variation; trace metal availability; concentrations and types of organic carbon and other reduced substrates; and other factors known to influence microbes. As part of a program investigating the dynamics of microbial populations in estuarine waters and their response to fluctuating environmental variables (SIMO, http://simo.marsci.uga.edu/), samples of DNA and RNA from the plankton assemblage have been collected regularly at a station in Georgia coastal waters. Quantitative estimates of *amoA* gene abundance indicated elevated abundance of AOA in samples collected in August 2008. We analyzed the metatranscriptome of two samples collected at this time and studied the distribution of transcripts among MG1C ORFs to gain insight into the metabolism of a MG1C population growing in the environment. An additional goal was to understand the factors that regulate competition within the guild of ammonia oxidizing microorganisms. The data also allowed us to examine proposed pathways for ammonia oxidation in AOA.

Materials and methods

Sample collection

Near-surface water samples were collected quarterly from a floating dock at Marsh Landing on the Duplin River, Sapelo Island, Georgia (31° 25′ 4.08 N, 81° 17′ 43.26 W; Supplementary Figure 1), ~6 km from the mouth of Doboy Sound, as described in Gifford *et al.* (2010). Briefly, samples were collected twice per day at approximately noon and midnight, <1 h before high tide, over a 2-day period during each sampling campaign. Samples used for RNA extraction were collected in rapid succession in the middle of the first night of each sampling campaign. A sample (5.75 l) of surface (~0.5 m) sea water was pumped directly from the river through 3 μm pore size filters (Capsule Pleated 3 μm Versapor Membrane; Pall Life Sciences, Ann Arbor, MI, USA) then through 0.22 μm pore size filters (Supor polyethersulfone; Pall Life Sciences, Ann Arbor, MI, USA) using a peristaltic pump (Supplementary Figure 1). The 0.22 μm filter was placed in a Whirl-Pak (Nasco, Fort Atkinson, WI, USA) plastic bag and immediately flash-frozen in liquid nitrogen. Total time from the start of filtration to freezing was ~10 min. We began filtering the second sample (FN57) immediately (~5 min delay) after the filter from the first sample (FN56) was placed in liquid nitrogen. We collected samples for DNA extraction concurrently by filling 20-l carboys with surface water while the RNA samples filtered. Once the second RNA sample was frozen, we filtered 12 l of the DNA sample through 3- and 0.22-μm filters as above, and the 0.22 μm filters were flash frozen.

mRNA isolation

mRNA was isolated from the samples as described previously (Poretsky *et al.*, 2006, 2009). Before beginning the extraction, 25 ng of a 994 nt RNA standard (derived from the pGEM cloning vector) was added to the sample in lysis buffer to serve as an internal standard (Gifford *et al.*, 2010). Total RNA was extracted from the filters using an RNAEasy kit (Qiagen, Valencia, CA, USA) and any residual DNA was removed by treating the sample twice with a Turbo DNA-Free Kit (Applied Biosystems, Austin, TX, USA).

The purified RNA preparations (containing 14 and 32 μg of total RNA from samples FN56 and FN57, respectively; 2–5 μg of this RNA was taken through the rRNA removal steps) were treated in two ways to remove ribosomal RNA. Epicenter's mRNA-Only kit (Epicenter, Madison, WI, USA) was used first to decrease rRNA contamination enzymatically. The samples were then treated with MICROB*Express* and MICROB*Enrich* kits (both from Applied Biosystems) that couple rRNA oligonucleotide hybridization probes with magnetic separation to enrich for mRNA. Initial and final RNA extracts were analyzed on an Experion automated electrophoresis system (Bio-Rad, Hercules, CA, USA) to verify successful removal of most of the rRNA. RNA remaining in the samples was amplified linearly using the MessageAmp II-Bacteria kit (Applied Biosystems). The amplified RNA was then converted to cDNA using the Universal Ribo-Clone cDNA synthesis system (Promega, Madison, WI, USA) with random hexamer primers. Leftover reactants and nucleotides from cDNA synthesis were removed from the sample using the QIAquick PCR purification kit (Qiagen).

Sequencing and annotation

cDNA was sequenced in four GS-FLX (454 Life Sciences, Branford, CT, USA) runs. One half of each PicoTiter plate was loaded with cDNA from one replicate sample, resulting in each sample being sequenced to the equivalent of two full runs divided over four plates (Supplementary Figure 2; described in detail in Gifford *et al.*, 2010). Over 2 million sequence reads were produced. Ribosomal RNA sequences in these reads were identified by a BLASTn (Zhang *et al.*, 2000) search against the small and large subunit SILVA database (http://www.arb-silva.de) with a bit score cutoff of 50.

Sequences identified as rRNA (~50% of the total) were excluded from further processing.

The remaining non-rRNA sequences were queried against the NCBI's RefSeq database using BLASTx (Altschul *et al.*, 1997) with a bit score cutoff of 40. The top hit that exceeded this bit score was taken as the open reading frame (ORF) assignment for that sequence. Approximately 50% of the non-rRNA sequences were assigned to annotated ORFs by this procedure. Replicate reads (defined according to Gomez-Alvarez *et al.*, 2009) accounted for 24.4% of this total, but our sequencing protocol, which used technical replicates on different plates, allowed us to identify replicates arising from methodological artifacts as discussed in Gomez-Alvarez *et al.* (2009) vs biologically valid replicates (see Gifford *et al.*, 2010). Our analysis indicates that most of the replicates are not artifacts, so we retained them in the data set. As discussed in Gifford *et al.* (2010), for the purposes of the analyses that follow, we assume that the population of reads returned from the sequencing effort is an unbiased sampling of the transcripts present in the populations of Bacteria and Archaea *in situ*.

Quantitative PCR and sequencing of 16S rRNA and amoA *amplicons*

The abundance of MG1C and AOB *amoA* genes and of Crenarchaeota and bacteria 16S rRNA genes was determined by quantitative, real-time PCR (qPCR) as described previously (Caffrey *et al.*, 2007). Primers are given in Supplementary Table 1. The abundance of individual genes (copies per ng of DNA extracted from the sample) and epifluorescence counts of total prokaryote abundance were used to estimate the number of MG1C and AOB cells in the sample for the purposes of calculating the number of transcripts per cell. Crenarchaeota 16S rRNA and *amoA* genes were cloned and sequenced (primers listed in Supplementary Table 1) for phylogenetic analysis as described previously (Kalanetra *et al.*, 2009).

Analysis of MG1C ribotypes

We compared (using BLASTn) the rRNA reads removed from one of the pyrosequencing libraries (FN56) to the '*Candidatus* Nitrosopumilus maritimus' strain SCM1 (Konneke *et al.*, 2005) 16S rRNA gene sequence (Nmar_R0029) to identify MG1C 16S rRNA sequences in our data set. We then queried the top 250 hits against the NCBI nr/nt database to obtain information from the annotations of the top hits on the distribution by habitat of ribotypes (ecotypes) related to the MG1C in our samples. We also assembled these reads using Nmar_R0029 as a scaffold to obtain a consensus sequence that was compared with sequences obtained by cloning and then sequencing PCR amplicons of 16S rRNA genes from the DNA sample.

Assemblies

The Geneious (Biomatters Ltd, Auckland, New Zealand) software package version 4.8 was used to assemble reads into contigs, for sequence manipulations (for example, alignments) and for phylogenetic analyses. All assemblies were constructed using unedited cDNA sequences. Unless otherwise noted, the appropriate genomic reference sequence was used as a scaffold and assemblies required $\geqslant 25$ bp of overlap and $\geqslant 75\%$ identity between sequences in the overlapping portions. The gap/extend penalty was set at 18, mismatch score at -9 and match score at 5. With these assembly parameters, most of the reads assigned to MG1C ORFs assembled into one contig per ORF with good coverage over the entire length of the gene.

Consensus sequences for amoA *genes*

The sequences were derived from reads assembled against the '*Ca.* N. maritimus' strain SCM1 *amoA* gene (Nmar_1500) sequence as a scaffold as described above. The consensus sequence for the majority genotype was determined by requiring $>75\%$ agreement at each position. Reads representing this majority consensus sequence were removed from the data set manually, then remaining reads representing less abundant, minority sequences were assembled as before and the consensus sequence was again recorded. Although inspection suggested additional diversity in the reduced data set, we did not attempt to recover additional consensus sequences, as coverage was too low for reliable assembly and analysis.

Environmental data

The Georgia Coastal Ecosystems LTER program collects data on a variety of environmental variables from the area surrounding our sampling site. These data and their accompanying metadata are available on the Georgia Coastal Ecosystems-LTER website http://gce-lter.marsci.uga.edu/. The closest Georgia Coastal Ecosystems-LTER water quality monitoring station, GCE6, is located in Doboy Sound, ~4.5 km from our sampling site (Supplementary Figure 1).

Results and discussion

qPCR analysis

Analysis of the abundance of bacteria and MG1C 16S rRNA genes and of AOA and AOB *amoA* genes by qPCR indicated elevated abundance of MG1C and of ammonia oxidizers, especially AOA, in water samples collected on 6–7 August 2008 (Figure 1). MG1C *amoA* abundance was 35- to 781-fold greater in August than on other sampling dates, whereas MG1C 16S rRNA abundance was 43- to 1658-fold greater (Figures 1a and b). MG1C relative abundance in the prokaryotic community averaged 1.8% (range 1.1–2.6%) for the August samples vs an average of 0.026% (range

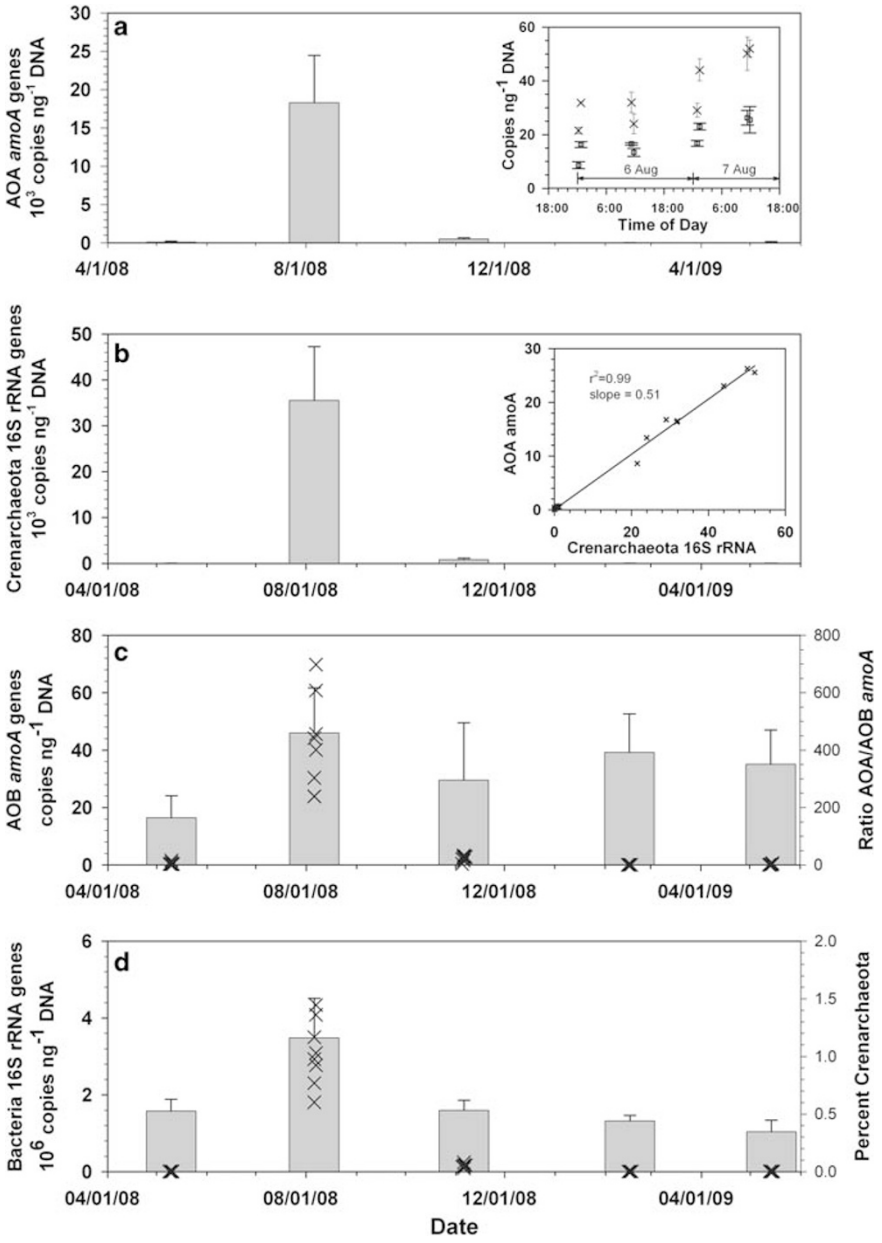

Figure 1 Time series of quantitative, real-time PCR (qPCR) estimates of the abundance of *amoA* and 16S rRNA genes at the sampling site. Means (wide bars) and s.d.'s (vertical lines) of eight samples collected over 2-day periods are shown. In some cases, the bars are smaller than the abcissa. (**a**) Archaeal *amoA* genes. Inset shows the time series of changes in amoA (□) and Crenarchaeota 16S rRNA (X) gene abundance on 6–7 August. Vertical bars are s.d.'s of triplicate qPCR determinations for each sample. (**b**) Marine Group 1 Crenarchaeota 16S rRNA gene abundance. Inset shows Archaeal *amoA* versus Crenarchaeota 16S rRNA gene abundance for each sample (regression line slope = 0.51, $r^2 = 0.99$). (**c**) Bacterial *amoA* gene abundance. Crosses show the ratio of Archaeal *amoA* to Bacterial *amoA* for each sample. (**d**) Bacteria 16S rRNA gene abundance. Crosses show the relative abundance of Crenarchaeota as a percentage of the prokaryotes (Bacteria + Crenarchaeota) in each sample assuming a gene dosage of 1 16S rRNA gene per genome for Marine Group 1 Crenarchaeota (from genomes annotated in DOE's IMG database) and 1.8 16S rRNA genes per genome as an average for marine bacteria (Biers *et al*., 2009).

0.0002–0.15%) on other dates (Figure 1d). Both MG1C *amoA* and Crenarchaeota 16S rRNA abundance increased during the 2-day sampling campaign in August 2008 (Figure 1a, inset).

There was no correlation between the abundance of *amoA* genes from MG1C and AOB in these samples (linear regression, $r^2 = 0.14$, $P > 0.1$). AOB *amoA* gene abundance was only 1.2- to 2.8-fold greater in August than on other sampling dates (Figure 1c), comparable to the increase in Bacteria 16S rRNA gene abundance (Figure 1d). AOB relative abundance (calculated assuming 2.5 *amoA* genes/AOB genome, an average from Norton *et al*., 2002) was 0.0024% (range 0.0018–0.0045%) in August vs 0.0044% (range 0.0009–0.013%) on other dates and AOB *amoA* gene abundance did not increase during the 2-day sampling campaign (not shown). The ratio of MG1C to AOB *amoA* abundance (398:1) in August

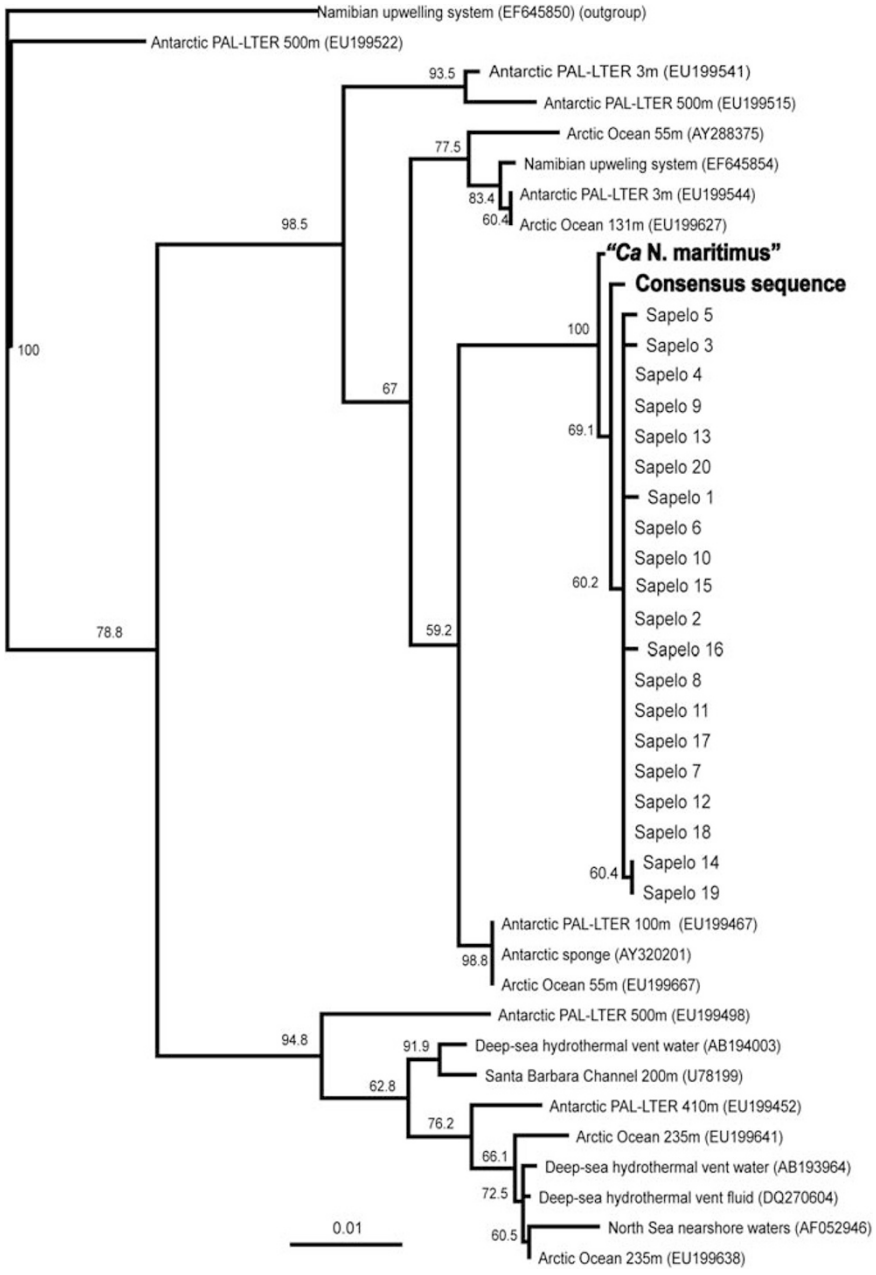

Figure 2 Phylogenetic analysis of the marine group 1 Crenarchaeota 16S rRNA sequences. The consensus sequence (bold) was obtained by assembling MG1C 16S rRNA reads contaminating the metatranscriptome. Sequences labeled 'Sapelo' are from cloned PCR amplicons produced with DNA from the same sample. The 16S rRNA gene sequence for '*Candidatus* Nitrosopumilus maritimus' strain SCM1 (Nmar_R0029, DQ085097) is shown in Bold. GenBank accession numbers for reference sequences are given in parentheses. This is a neighbor-joining tree based on 876 nt. Bootstrap analysis was used to estimate the reliability of phylogenetic reconstructions and support is shown if >50% (100 iterations).

was more than 20-fold greater than on other dates (Figure 1c), suggesting selective growth of MG1C over AOB at the time of sampling. With the exception of the August samples, the ratios of AOA to AOB abundance are similar to our previous observations in sediment samples collected from nearby sites (Caffrey *et al.*, 2007).

MG1C 16S rRNA ecotypes

MG1C 16S rRNA reads retrieved from our libraries were most similar (250 sequences, all >96.9% identity with 235>99% and 162 reads=100% identity, significance values $1*e^{-159}$ to $1*e^{-113}$) to environmental sequences taken from coastal waters, coral symbionts or sediments (13 different studies) or to the '*Ca.* N. maritimus' SCM1 16S rRNA gene. These reads assembled into one contig (not shown). We compared the consensus sequence from the contig to nearly full-length sequences obtained by cloning and sequencing PCR amplicons from the original sample (Figure 2). We detected two sequence variants by inspection of the assembly, but only one of these, corresponding to the consensus

Table 1 Distribution of cDNA reads among functional categories of the annotation pipeline and among marine group 1 Crenarchaeota and ammonia-oxidizing bacteria taxonomic groupings

Category	*Total*[a]
Total number of pyrosequencing reads	2 181 899
Total RefSeq hits	560 389
Reads assigned to Marine group 1 Crenarchaeota	17 386
Reads assigned to '*Candidatus* Nitrosopumilus maritimus' strain SCM1 ORFs	16 914
Number of different '*Ca.* N. maritimus' ORFs	786
Number of '*Ca.* N. maritimus' ORFs hit >50 times	32
Reads assigned to *Cenarchaeum symbiosum* A ORFs	472
Number of different *C. symbiosum* ORFs	84
Number of *C. symbiosum* ORFs hit >50 times	2
Reads assigned to ammonia oxidizing bacteria (AOB)	2651
Number of different AOB ORFs	1253
Reads assigned to *Nitrosococcus*	439
Reads assigned to *Nitrosomonas*	382
Reads assigned to *Nitrosospira*	272
Number of AOB ORFs hit >50 times	3

Abbreviations: *Ca.* N. maritimus, *Candidatus* Nitrosopumilus maritimus strain; ORF, open reading frame.
[a]Sum of reads from both libraries.

and with >99% identity to the '*Ca.* N. maritimus' strain SCM1 16S rRNA gene, was captured in the clone library.

Metatranscriptome properties

We retrieved ~2 million cDNA pyrosequencing reads from the two samples (Gifford *et al.*, 2010). Analysis of this data set (Table 1) revealed that 17 386 sequences (median length 236 bp, range 47–360 bp) could be assigned to coding regions in two MG1C genomes, '*Ca.* N. maritimus' strain SCM1 and *Cenarchaeum symbiosum* (Hallam *et al.*, 2006a, b). For simplicity, we will refer to this subset of reads as the 'MG1C metatranscriptome.' MG1C thus accounted for 3.1% of the reads identified as transcripts, which is comparable to their contribution to the population of prokaryotes (1.8%, Figure 1d). In contrast, only 46 reads were assigned to ORFs from Euryarchaeota. The remaining reads were assigned primarily to ORFs from Bacteria or viruses.

Of the 17 386 reads that were assigned to MG1C ORFs, 16 914 were assigned to *'Ca.* N. maritimus' strain SCM1, whereas 472 were assigned to *C. symbiosum* (Table 1). These reads were assigned to 786 different *'Ca.* N. maritimus' strain SCM1 ORFs (Figure 2), representing 44% of the 1797 coding regions annotated in this genome (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) and to 82 ORFs from the *C. symbiosum* genome (4% of 2017 annotated coding regions). As there are only two MG1C genomes currently represented in the RefSeq database and they share a great deal of homology (Walker *et al.*, 2010), reads assigned to one of them as the top hit were usually assigned (with similar significance values) to the other as the second hit. We noted gaps in the recruitment of reads against the '*Ca.* N. maritimus' strain SCM1 genome (Figure 2) that appear too long to be random consequences of low coverage (for example, Nmar_124 to Nmar_154, Nmar_1147 to Nmar_1173, Nmar_1323 to Nmar_1357), suggesting possible sites of indels.

We estimate that each MG1C cell contained 168 transcripts (averaged for the two libraries, Supplementary Table 2). This is similar to the value Gifford *et al.* (2010) obtained for the entire prokaryote metatranscriptome from these samples (190 transcripts per cell). The differences in the number of transcripts per MG1C cell between libraries (321 vs 79 for FN56 vs FN57) may reflect differences in the physiological state of MG1C community between samples. Ratios of the abundance of transcripts from specific ORFs between the two libraries were more variable than expected by random sampling error (Supplementary Figure 3), suggesting differences in the physiological state of the bacterioplankton in the two samples (see Gifford *et al.*, 2010, for a detailed analysis). Our estimate of the average number of transcripts in AOB cells was 49 241 transcripts per cell (Supplementary Table 2). This value is unreasonably high (see discussion in Gifford *et al.*, 2010) and suggests either that non-AOB reads were incorrectly assigned to AOB ORFs or that our qPCR estimates of AOB abundance are too low, or both. The qPCR estimates of AOB *amoA* gene abundance on which this calculation is based are typical of what we (Caffrey *et al.*, 2007) and others have reported for coastal waters. Comparison of the distributions of bit scores for hits to ORFs from these two populations (Supplementary Figure 4) suggests that read assignments are less reliable for AOB than for MG1C.

Most ORFs were represented by singletons or only a few reads (Figure 3, Supplementary Figure 5, Supplementary Tables 3 and 4); however, 34 MG1C ORFs were represented by 50 or more reads (Table 2), together accounting for 13 686 reads (78.8% of the reads in the MG1C metatranscriptome). The best-represented ORF (Nmar_1547, a hypothetical protein) accounted for 22% of the MG1C metatranscriptome.

Nmar_1547 homologs

Overall, 31% of the reads in the MG1C metatranscriptome were assigned to a group of six homologous ORFs: Nmar_1547, Nmar_1201, CENSYa_0159, CENSYa_0161, CENSYa_2159 and CENSYa_2160; with most assigned to Nmar_1547 and Nmar_1201 (Table 2). The sequences of these ORFs are very similar (bit scores >1140, *E*-value = 0) and the apparent duplication is noteworthy in genomes that appear to have undergone reduction as an adaptive strategy (Hallam *et al.*, 2006a; Walker *et al.*, 2010) and that contain many important genes (for example, *amoABC*) as single copies. Hallam *et al.* (2006a) first

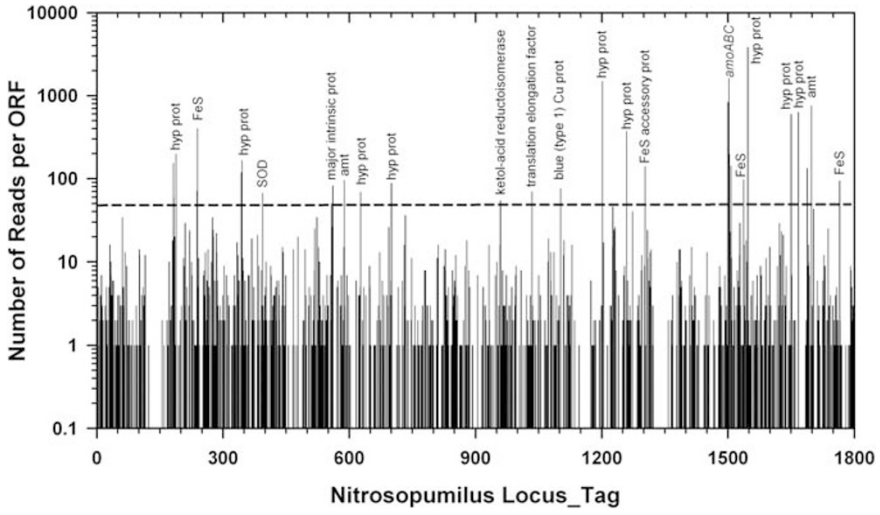

Figure 3 Distribution of pyrosequencing reads among *Nitrosopumilus* ORFs. The horizontal line is positioned at 50 hits per ORF and indicates the cutoff used to define highly expressed ORFs. Text over the longest bars identifies the annotation for that ORF: *amoABC*, ammonia monooxygenase subunits; amt, ammonium transporter; FeS, 4Fe–4S ferredoxin iron-sulfur binding domain-containing protein; FeS accessory protein, iron–sulfur cluster assembly accessory protein; hyp prot, hypothetical protein; SOD, superoxide dismutase; see Table 2 and Supplementary Tables 3 and 4 for a complete list.

Table 2 Crenarchaeota ORFs represented by 50 or more reads in metatranscriptomes retrieved from Georgia coastal waters

Locus	*Total count*	*Annotation*	*Bit scores*		
			Median	*Low*	*High*
Nmar_1547	3812	Hypothetical protein	141	40	176
Nmar_1502	1623	amoC	149	40	191
Nmar_1201	1492	Hypothetical protein	110	40	178
Nmar_1500	836	amoA	149	41	183
Nmar_1698	757	Ammonium transporter	98	40	149
Nmar_1667	633	Hypothetical protein (NirK?)	107	41	169
Nmar_1650	598	Hypothetical protein	144	40	169
Nmar_1501	588	Hypothetical protein	114	41	153
Nmar_0239	404	4Fe-4S Ferredoxin iron–sulfur binding domain-containing protein	154	40	175
Nmar_1259	373	Hypothetical protein (NirK?)	129	40	167
Nmar_0188	198	Hypothetical protein	161	40	185
Nmar_1503	198	amoB	127	40	176
Nmar_0345	166	Hypothetical protein	99	45	139
CENSYa_1453	166	Ammonia permease	111	42	135
Nmar_0182	154	Hypothetical protein	124	41	166
Nmar_1507	142	Hypothetical protein	86	42	164
Nmar_1303	140	Iron–sulfur cluster assembly accessory protein	143	48	157
Nmar_1688	133	H^+transporting two-sector ATPase C subunit	68	42	89
Nmar_0343	119	Hypothetical protein	120	42	164
Nmar_0344	117	Hypothetical protein	134	41	161
Nmar_1537	97	4Fe-4S Ferredoxin iron-sulfur binding domain-containing protein	154	50	186
Nmar_0588	95	Ammonium transporter	109	42	171
Nmar_1765	94	4Fe-4S Ferredoxin iron–sulfur binding domain-containing protein	165	42	188
Nmar_0700	88	Hypothetical protein	142	42	171
Nmar_0561	82	Major intrinsic protein	112	54	174
Nmar_1102	76	Blue (type1) copper domain-containing protein	110	40	164
Nmar_0238	71	4Fe-4S Ferredoxin iron–sulfur binding domain-containing protein	141	40	173
Nmar_1034	70	Elongation factor 1-alpha	145	44	181
Nmar_0627	69	Hypothetical protein	116	42	161
Nmar_0394	67	Superoxide dismutase	159	43	197
CENSYa_161	66	Hypothetical protein	67	44	115
Nmar_0183	59	Cytochrome c oxidase subunit II	131	40	193
Nmar_0959	54	Ketol acid reductoisomerase	145	52	181
Nmar_0558	50	Hypothetical protein	93	45	166

'Locus' refers to the Locus_Tag identifier assigned to the gene, 'Total Count' is the number of reads for which that locus was identified as the top hit by BLASTx. 'Annotation' gives the identity of the gene product. The median and range of the bit scores for all hits to a particular gene are also given.

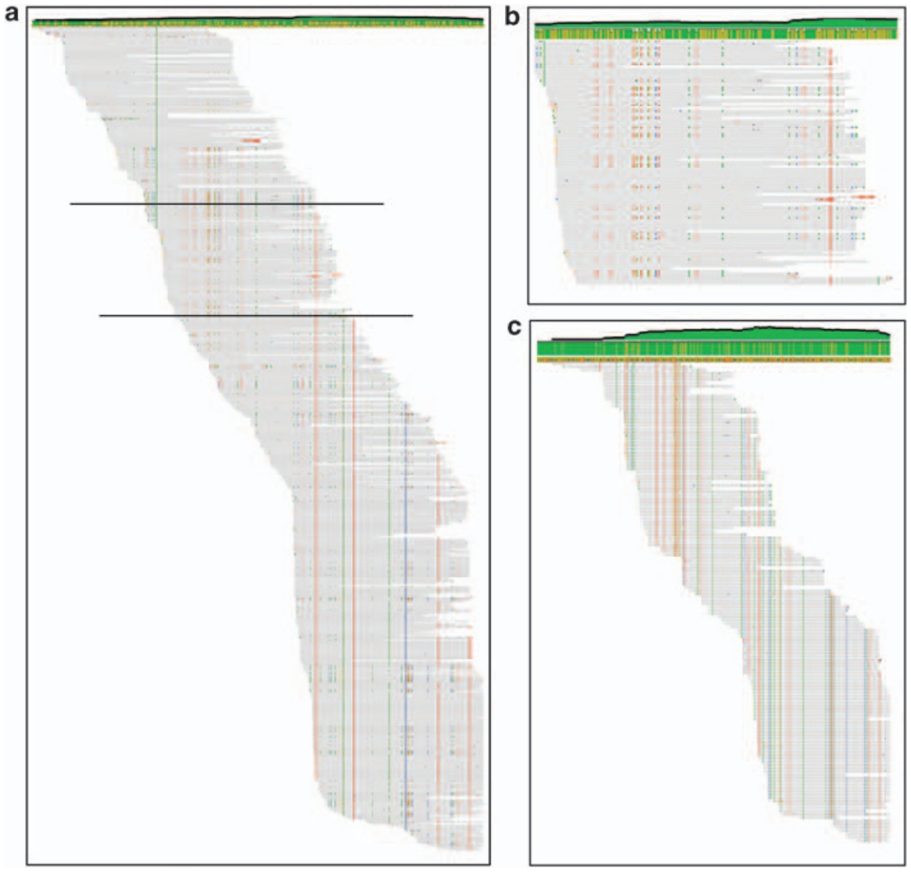

Figure 4 Assembly of 836 of 836 reads assigned to the '*Ca.* N. maritimus' strain SCM1 *amoA* ORF (Nmar_1500) against the Nmar_1500 sequence as a scaffold. (**a**) Coverage curve and assembly. (**b**). Close-up of a portion of the assembly (region between horizontal lines in **a**) showing primary and secondary sequence variants. (**c**) Assembly of 147 minority sequence variants. Objects from top to bottom: coverage curve (green shape, 0–558); identity at each position (green and olive bar, 0–100%), reference sequence (not shown in panel **b**), aligned reads. Highlighted positions in the aligned reads indicate disagreements with the reference sequence (code: red—A, blue—C, orange—G, green—T).

noted them in their analysis of the *C. symbiosum* metagenome and similar sequences (represented by ABZ07689) were reported to be abundant in cDNA libraries from 4000 m at station ALOHA by Shi *et al.* (2009). Searches (BLASTn) of the 'GOS All ORFS' data set in the CAMERA database (http://camera.calit2.net/) identified 25 sequences, all from coastal samples, with significant homology (*E*-values < -53) to Nmar_1547.

Nmar_1547 and Nmar_1201 are large ORFs, >5000 nt in length. In contrast to most other MG1C ORFs (see Figure 4 for example), reads assigned to Nmar_1547 and Nmar_1201 did not assemble into one contig against the respective genome sequence as a scaffold (Supplementary Figure 6). Although the assemblies otherwise have very high coverage (up to 229-fold and up to 1113-fold for Nmar_1201 and Nmar_1547, respectively), the regions in which the assemblies break have increased sequence variability (Supplementary Figure 6), resulting in insufficient homology to the RefSeq sequence and to other reads to support assembly. These 'ORFs' may not be protein-coding regions, though analysis of codon usage in a homologous sequence (ABZ07689, BLASTp score $= 604$, *E*-value $= 0$) by Shi *et al.* (2009) suggest that they are. Annotation and BLASTx searches of the GenBank non-redundant protein database suggest functions distantly related to adhesins or hemaglutinin and hemolysin proteins and annotation indicates that they have leader sequences and transmembrane domains. These features suggest that they may be cell-surface proteins involved in some interaction with other cells, detrital particles, high-molecular weight DOM and so on. By analogy with genes involved in ammonia processing (see next section below), their abundance in the transcript pool suggests that their unknown function is important to MG1C physiology or ecology.

Ammonia uptake and oxidation

Transcripts from ORFs related to ammonia uptake and oxidation were among the most abundant in the MG1C transcript pool (Table 2, Supplementary Tables 3 and 4). A total of 6455 reads (37% of the reads in the MG1C metatranscriptome) were assigned to ORFs identified by Walker *et al.* (2010) as being related to the ammonia oxidation pathway.

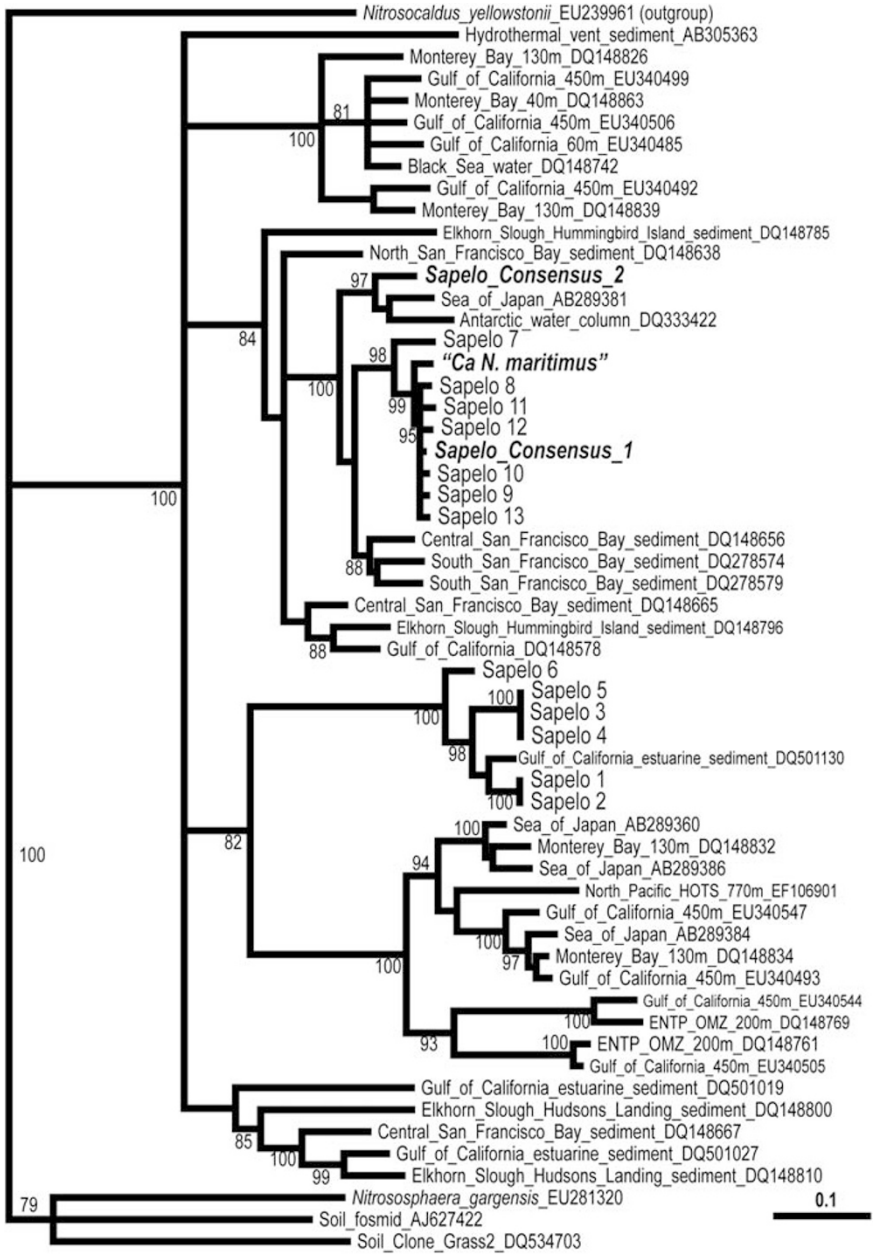

Figure 5 Phylogenetic analysis of marine group 1 Crenarchaeota *amoA* sequences in metatranscriptome and DNA samples. Consensus sequences (bold italics, 1 = dominant, 2 = minority) were obtained by assembling 16S *amoA* reads in the metatranscriptome. Sequences labeled 'Sapelo' are from cloned PCR amplicons produced with DNA from the same sample. Sequences were aligned with reference sequences using Clustal W. Minimum evolutionary distances were calculated using the Kimura two-parameter model. The '*Candidatus* Nitrosopumilus maritimus' strain SCM1 sequence (Nmar_1500, ABX13396). GenBank accession numbers of reference sequences are given in parentheses. This is a neighbor-joining tree based on 595 nt. Bootstrap analysis was used to estimate the reliability of phylogenetic reconstructions and support is shown if >50% (100 iterations).

This includes 2657 reads assigned to ammonia monooxygenase subunits: *amoA* (Nmar_1500, 836); *amoB* (Nmar_1503, 198); and *amoC* (Nmar_1502, 1623), giving relative abundances in the transcript pool of 4.2:1:8.2 for *amoA*:*amoB*:*amoC*, which differs from the stoichiometry of the subunits in native ammonia monooxygenase (1:1:1). Inspection of the assemblies of these reads revealed two dominant genotypes in the population of *amoA* reads (Figure 4) and at least two in the *amoB* and *amoC* (data not shown) populations. Phylogenetic analysis of the consensus sequence for the dominant *amoA* genotype (Figure 5) placed it in a clade containing the '*Ca.* N. maritimus' strain SCM1 gene and a variety of shallow water column and sediment environmental sequences. It is also >99.5% identical to MG1C *amoA* gene sequences retrieved previously from Georgia coastal waters (Hollibaugh, unpublished data) and to sequences obtained from the DNA samples collected in this study (Figure 5).

The minor consensus sequence grouped separately from the dominant consensus sequence and was not recovered in the (small) *amoA* clone library we sequenced. Half of the *amoA* gene sequences retrieved from the DNA sample were most similar to an environmental sequence from an eutrophic Mexican estuary (Beman and Francis, 2006). Reads corresponding to this clade were not found in the metatranscriptome, suggesting the presence of an inactive sub-population of MG1C in these samples.

Reads assigned to ORFs annotated as ammonia transporters and permeases were also abundant with a total of 1017 reads assigned to Nmar_1698 (757), Nmar_0588 (94) and CENSYa_1453 (166). The relatively high abundance of ammonia permeases in the MG1C metatranscriptome (6% of the transcripts) seems at odds with the model of ammonia oxidation as a cell-surface process proposed in Walker *et al.* (2010). Nmar_1698 and CENSYa_1453 are very similar to each other (BLASTx bit score 705, $E=0$; BLASTn bit score 803, $E=0$, 71% identity) and reads assigned to them assembled into one contig against an Nmar_1698 scaffold (not shown). Nmar_0588 is divergent with no similarity to other MG1C genes and a best BLASTx hit to an ammonium transporter from the slime mold *Polysphondylium pallidum* PN500 (bit score of 365, $E=5*e^{-99}$). Our data thus indicate transcription by the MG1C population of two different ammonia transporters, possibly with different kinetic properties. Inspection of the Nmar_1698 and Nmar_0588 assemblies indicates additional diversity in the ammonia transporter genes transcribed by the MG1C population, with at least three variants of Nmar_1698 and possibly two variants of Nmar_0588.

Walker *et al.* (2010) propose two alternative pathways for ammonia oxidation in '*Ca.* N. maritimus' strain SMC1. One of the proposed pathways proceeds by hydroxylamine, but depends on a Cu-based alternative to the AOB heme-based hydroxylamine oxidoreductase. The second alternative mechanism proposes ammonia oxidation by ammonia monooxygenase that produces a reactive nitroxyl intermediate instead of hydroxylamine. Both pathways transfer electrons to the quinone pool through a quinone reductase. Nmar_1226, proposed by Walker *et al.* (2010) to serve this function as an analog of the AOB quinone reductase, was well represented in the MG1C metatranscriptome (46 reads). However, we did not detect any transcripts from the genes (Nmar_1354 to Nmar_1357) proposed by Walker *et al.* (2010) to encode proteins involved in the nitroxyl-based alternative ammonia oxidation pathway. Instead, a total of 1006 reads was assigned to two other ORFs, Nmar_1259 and Nmar_1667 (Supplementary Table 4), that were similar to sequences retrieved from the Sargasso Sea (EAH96098 and EAI84410; bit scores >498, $E=0$). These sequences have been identified as crenarchaeote homologs of Cu-containing nitrite reductases (*nirK*) by Treusch *et al.* (2005) and Bartossek *et al.* (2010). Their function *in vivo* is in question because, at least under the aerobic growth conditions they reported, '*Ca.* N. maritimus' strain SCM1 stoichiometrically converts ammonia to nitrite in culture (Konneke *et al.*, 2005; Martens-Habbena *et al.*, 2009). An experiment performed by Bartossek *et al.* (2010) to test the relationship between transcription of these genes and the expected *nirK* activity (nitrous oxide production) failed to support a nitrite reductase function and Bartossek *et al.* (2010) speculated that the proteins encoded by these genes might exhibit 'other or additional activities besides nitrite reduction.' Walker *et al.* (2010) included them in the list of genes devoted to energy production from ammonia oxidation without specifying their function. These homologies and their elevated abundance in the transcript pool suggest that they have an important role in the primary ammonia oxidation pathway *in situ*.

Other metabolic functions

A total of 67 reads were assigned to an MG1C ORF annotated as superoxide dismutase (SOD, Nmar_0394, Table 2), which accounted for 26% of all hits to SOD in the complete (MG1C plus Bacteria) metatranscriptome. The proportion of reads assigned to SOD in the MG1C metatranscriptome (67 of 17 386 reads) is significantly ($X^2=314$, $P=0$) greater than the proportion (270 of 543 016) of SOD reads in the rest of the metatranscriptome. Assembly of the reads assigned to Nmar_0394 revealed a population with low diversity (only 2 of 67 reads differ from the consensus and 98.5% overall average pairwise identity) that differs slightly from the '*Ca.* N. maritimus' strain SCM1 gene (93% identity at the nucleotide level). SOD catalyzes the decomposition of superoxide radicals to yield hydrogen peroxide, which is broken down by catalase in many Bacteria. The '*Ca.* N. maritimus' strain SMC1 genome does not contain an ORF annotated as catalase; however, it contains four ORFs (Nmar_0275, Nmar_0560, Nmar_1438 and Nmar_1496) annotated as thiol-specific antioxidants (peroxiredoxins) that may serve the same function (Imlay, 2008). These ORFs were represented by a total of 45 hits in the MG1C metatranscriptome, with 34 of these hits assigned to one ORF, Nmar_0275. The complete metatranscriptome (Bacteria plus MG1C) contained 889 hits to ORFs annotated as 'catalase,' 'peroxiredoxin' or 'thiol-specific antioxidant.' Thus, a statistically significantly ($X^2=9.121$, $P=0.0025$) greater portion of reads in the MG1C metatranscriptome was assigned to ORFs with functions related to catalase than in the Bacteria metatranscriptome. Finally, 14 MG1C reads were assigned to the DNA repair gene *radA* (Nmar_1386), which was not different from the proportion of *recA* in the Bacteria metatranscriptome ($X^2=0.004$, $P=0.95$).

The overrepresentation of MG1C SOD and hydrogen peroxidase-related transcripts suggests that

MG1C may be subjected to greater exposure to superoxide or that they are more sensitive to it than the Bacteria in these samples. Increased exposure may be a consequence of reactions unique to their metabolism. In contrast, similar levels of transcripts for DNA repair enzymes (*radA* and *recA*) suggest that the two populations (MG1C and Bacteria) are responding similarly to agents that cause DNA damage, such as UV radiation.

MG1C are reported to fix carbon through the 3-hydroxypropionate/4-hydroxybutyrate pathway (Konneke *et al.*, 2005; Hallam *et al.*, 2006a, b; Berg *et al.*, 2007; Kockelkorn and Fuchs, 2009; Walker *et al.*, 2010). The MG1C metatranscriptome contained 146 reads assigned to ORFs from this pathway (Supplementary Tables 3 and 4). A total of 59 reads were assigned to MG1C TCA cycle genes (Supplementary Tables 3 and 4). There is no evidence in the metatranscriptome that heterotrophy had a significant role in the nutrition of this MG1C population. MG1C reads were not assigned to COGs for transporters of organic compounds (Supplementary Tables 3 and 4), in contrast to the high proportion of reads in the Bacteria metatranscriptome that were assigned to transporters ($\sim$20% of the Bacteria reads assigned to the top 50 COGs, Gifford unpublished data) or found previously in a Bacteria-dominated metatranscriptome retrieved from a near-by site (Poretsky *et al.*, 2010). Also, the ratio of MG1C *amoA* genes to MG1C 16S rRNA genes in these samples was 0.51 (Figure 1b). Although lower than values reported for ammonia-oxidizing enrichments (Wuchter *et al.*, 2006) or cultures (Konneke *et al.*, 2005) (1:1 to 2.8:1) or the gene dosage in MG1C genomes (1:1) (Hallam *et al.*, 2006a; Walker *et al.*, 2010), this ratio is much higher than ratios used to infer heterotrophy in other populations (Agogue *et al.*, 2008; de Corte *et al.*, 2008; Kalanetra *et al.*, 2009).

AOB transcripts

The metatranscriptome also contained 2651 reads (0.5% of reads assigned to ORFs) that were assigned to AOB ORFs (Table 1). Overall, 49% of the AOB reads were assigned to *Nitrosococcus* ORFs, with the remainder assigned to *Nitrosomonas* (27%) and *Nitrosospira* (24%) ORFs. None of the reads attributed to AOB were assigned to ORFs known to be involved in ammonia uptake or oxidation and only one AOB read (from *Nitrosomonas*) was assigned to RubisCO, the enzyme responsible for carbon fixation in AOB (data not shown). On the basis of our qPCR data, the abundance of AOB *amoA* transcripts in the metatranscriptome may have been below the limit of detection. The relative abundance of AOA vs AOB *amoA* genes in the sample as determined by qPCR averages 398:1 (Figure 1). Given the number of MG1C *amoA* transcripts in the metatranscriptome (836), assuming the relative abundance of *amoA* transcripts in the AOB transcript pool was comparable to that seen in the MG1C, we would expect to recover only 2.1 (836/398) AOB *amoA* transcripts from the complete metatranscriptome. In contrast, we would expect to encounter 131 AOB *amoA* reads if all of the transcripts assigned to AOB were actually from AOB and if the relative abundance of *amoA* transcripts in the AOB transcript pool was comparable to that seen in the MG1C metatranscriptome (4.9%). As discussed above, these calculations and the number of transcripts per cell implied by our data (Supplementary Table 2) suggests that the majority of the reads attributed to AOB were misassigned, likely because phylogenetically related but non-AOB sequences were binned to AOB genomes.

Implications for competition between AOA and AOB

Combined with the much greater abundance of MG1C 16S rRNA and *amoA* genes in the August sample relative to other sampling dates (Figure 1), the increasing abundance of Crenarchaeota 16S rRNA and AOA *amoA* genes over the 2 days we sampled (Figure 1), and the distribution of reads in the MG1C metatranscriptome (Supplementary Tables 3 and 4), our data suggest that the MG1C population was actively growing—blooming—when sampled. Assuming that the MG1C population developed locally rather than being advected into the study area, we examined environmental data collected by the Georgia Coastal Ecosystem LTER (http://gce-lter.marsci.uga.edu/) for the weeks preceding this sampling for potential explanations for the elevated MG1C population. There are no obvious perturbations in the records for weather (wind, rainfall, runoff, tides) or environmental variables (temperature, nutrients, chlorophyll and so on) that might indicate a re-suspension event, a pulse of nutrient-rich water from runoff or upwelling, a phytoplankton bloom and so on (data not shown).

Ammonia concentrations are variable in Georgia coastal waters (Supplementary Figure 7) and a late summer increase in ammonium concentration is a regular feature of this coastal environment (Verity, 2002). The abundance of Bacteria 16S rRNA genes was also greater in the August sample than on other sampling dates (Figure 1), consistent with an overall increase in heterotrophic metabolism, presumably leading to elevated ammonium regeneration at this time of the year. Although ammonium concentrations were elevated during the August 2008 sampling campaign (Supplementary Figure 7), elevated ammonium concentrations at other times of the year did not correspond to elevated MG1C abundance (compare Figure 1 with Supplementary Figure 7), suggesting that ammonium is not the sole driving variable. AOB abundance did not increase during the August sampling series and average August abundance was only slightly elevated compared with other sampling dates. Previous work (Caffrey

et al., 2007) documented a correlation between AOA (but not AOB) *amoA* gene abundance and potential nitrification rates in sediment samples from this site, suggesting that AOB are typically not very active at this site, even though they are present. One potential explanation for the difference in the response of AOA vs AOB is that the threshold ammonia concentration needed to stimulate growth of AOB may be higher than for AOA. Differences in ammonia uptake kinetics between MG1C and AOB (Martens-Habbena *et al.*, 2009) suggest that MG1C are better competitors at low ammonia concentrations (Martens-Habbena *et al.*, 2009). Alternatively, MG1C and AOB may be differentially limited or inhibited by environmental factors other than ammonia availability.

Acknowledgements

We thank R Newton for assistance with sample collection. L Tomsho and S Schuster provided 454 sequencing expertise. S Obrebski assisted with statistical analyses. The manuscript was greatly improved by the helpful comments of reviewers of earlier versions. We thank them for their efforts. This project was funded by grants from the Gordon and Betty Moore Foundation and the National Science Foundation (MCB0702125, OCE0620959 and OCE0352216).

References

Agogue H, Brink M, Dinasquet J, Herndl GJ. (2008). Major gradients in putatively nitrifying and non-nitrifying Archaea in the deep North Atlantic. *Nature* **456**: 788–791.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W *et al.* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Bartossek R, Nicol GW, Lanzen A, Klenk H-P, Schleper C. (2010). Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. *Environ Microbiol* **12**: 1075–1088.

Beman JM, Francis CA. (2006). Diversity of ammonia-oxidizing Archaea and Bacteria in the sediments of a hypernutrified subtropical estuary: Bahia del Tobari, Mexico. *Appl Environ Microbiol* **72**: 7767–7777.

Beman JM, Popp BN, Francis CA. (2008). Molecular and biogeochemical evidence for ammonia oxidation by marine Crenarchaeota in the Gulf of California. *ISME J* **2**: 429–441.

Berg IA, Kockelkorn D, Buckel W, Fuchs G. (2007). A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* **318**: 1782–1786.

Bernhard AE, Landry ZC, Blevins A, de la Torre JR, Giblin AE, Stahl DA. (2010). Abundance of ammonia-oxidizing Archaea and Bacteria along an estuarine salinity gradient in relation to potential nitrification rates. *Appl Environ Microbiol* **76**: 1285–1289.

Biers EJ, Sun S, Howard EC. (2009). Prokaryotic genomes and diversity in the surface ocean: interrogating the Global Ocean Sampling metagenome. *Appl Environ Microbiol* **75**: 2221–2229.

Caffrey JM, Bano N, Kalanetra K, Hollibaugh JT. (2007). Environmental factors controlling ammonia-oxidation by ammonia-oxidizing Bacteria and Archaea in Southeastern estuaries. *ISME J* **1**: 660–662.

de Corte D, Yokokawa T, Varela MM, Agogue H, Herndl GJ. (2008). Spatial distribution of Bacteria and Archaea and *amoA* gene copy numbers throughout the water column of the Eastern Mediterranean Sea. *ISME J* **3**: 147–158.

Francis CA, Beman JM, Kuypers MMM. (2007). New processes and players in the nitrogen cycle: the microbial ecology of anaerobic and archaeal ammonia oxidation. *ISME J* **1**: 19–27.

Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB. (2005). Ubiquity and diversity of ammonia-oxidizing Archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci USA* **102**: 14683–14688.

Fuhrman J, Hagström Å. (2008). Bacterial and archaeal community structure and its patterns. In: Kirchman DL (ed). *Microbial Ecology of the Oceans* Second edn. John Wiley & Sons, Inc.: Hoboken, NJ, pp 45–90.

Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA. (2010). Analysis of a marine microbial metatranscriptome as a function of sequencing depth. *ISME J*; doi:10.1038/ismej.2010.141.

Gomez-Alvarez V, Teal TK, Schmidt TM. (2009). Systematic artifacts in metagenomes from complex microbial communities. *ISME J* **3**: 1–4.

Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y-i, Sugahara J *et al.* (2006a). Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci USA* **103**: 18296–18301.

Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, Richardson PM *et al.* (2006b). Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol* **4**: e95.

Imlay JA. (2008). Cellular defenses against superoxide and hydrogen peroxide. *Annu Rev Biochem* **77**: 755–776.

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

Karner MB, DeLong EF, Karl DM. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**: 507–510.

Kirchman DL, Elifantz HD, Ana I, Malmstrom RR, Cottrell MT. (2007). Standing stocks and activity of Archaea and Bacteria in the western Arctic Ocean. *Limnol Oceanogr* **52**: 495–507.

Kockelkorn D, Fuchs G. (2009). Malonic semialdehyde reductase, succinic semialdehyde reductase, and succinyl-coenzyme A reductase from *Metallosphaera sedula*: enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in sulfolobales. *J Bacteriol* **191**: 6352–6362.

Konneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543–546.

Magalhaes CM, Machado A, Bordalo AA. (2009). Temporal variability in the abundance of ammonia- oxidizing

bacteria vs archaea in sandy sediments of the Douro River estuary, Portugal. *Aquat Microb Ecol* **56**: 13–23.
Martens-Habbena W, Berube PM, Urakawa H, de la Torre JR, Stahl DA. (2009). Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* **461**: 976–979.
Mincer TJ, Church MJ, Taylor LT, Preston C, Karl DM, DeLong EF. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ Microbiol* **9**: 1162–1175.
Mosier AC, Francis CA. (2008). Relative abundance and diversity of ammonia-oxidizing archaea and bacteria in the San Francisco Bay estuary. *Environ Microbiol* **10**: 3002–3016.
Norton J, Alzerreca J, Suwa Y, Klotz M. (2002). Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. *Arch Microbiol* **177**: 139–149.
Ouverney CC, Fuhrman JA. (2000). Marine planktonic archaea take up amino acids. *Appl Environ Microbiol* **66**: 4829–4833.
Poretsky RS, Bano N, Buchan A, Hollibaugh JT, Moran MA. (2006). Environmental transcriptomics: a method to access expressed genes in complex microbial communities. *Molecular Microbial Ecology Manual*, 3rd Edition, Section 8, pp 1892–1904.
Poretsky RS, Gifford S, Rinta-Kanto J, Vila-Costa M, Moran MA. (2009). Analyzing gene expression from marine microbial communities using environmental transcriptomics. *J Vis Exp*; doi: 10.3791/10860.
Poretsky RS, Sun S, Mou X, Moran MA. (2010). Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environ Microbiol* **12**: 616–627.
Prosser JI, Nicol GW. (2008). Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. *Environ Microbiol* **10**: 2931–2941.
Santoro AE, Casciotti KL, Francis CA. (2010). Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environ Microbiol* **12**: 1989–2006.
Santoro AE, Francis CA, de Sieyes NR, Boehm AB. (2008). Shifts in the relative abundance of ammonia-oxidizing bacteria and archaea across physicochemical gradients in a subterranean estuary. *Environ Microbiol* **10**: 1068–1079.
Shi Y, Tyson GW, DeLong EF. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459**: 266–269.
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.
Treusch AH, Leininger S, Kletzin A, Schuster SC, Klenk H-P, Schleper C. (2005). Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ Microbiol* **7**: 1985–1995.
Urakawa H, Martens-Habbena W, Stahl DA. (2010). High abundance of ammonia-oxidizing Archaea in coastal waters, determined using a modified DNA extraction method. *Appl Environ Microbiol* **76**: 2129–2135.
Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
Verity PG. (2002). A decade of change in the Skidaway River estuary. I. Hydrography and nutrients. *Estuaries* **25**: 944–960.
Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ *et al.* (2010). Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci USA* **107**: 8818–8823.
Wuchter C, Abbas B, Coolen MJL, Herfort L, van Bleijswijk J, Timmers P *et al.* (2006). Archaeal nitrification in the ocean. *Proc Natl Acad Sci USA* **103**: 12317–12322.
Zaikova E, Walsh DA, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ. (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* **12**: 172–191.
Zhang Z, Schwartz S, Wagner L, Miller W. (2000). A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7**: 203–214.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)