The ISME Journal (2014) 8, 685–698

www.nature.com/ismej

ORIGINAL ARTICLE

Seasonal variation in the metratranscriptomes of a Thaumarchaeota population from SE USA coastal waters

James T Hollibaugh[1], Scott M Gifford[1,3], Mary Ann Moran[1], Meredith J Ross[1,4], Shalabh Sharma[1] and Bradley B Tolar[2]

[1]*Department of Marine Sciences, University of Georgia, Athens, GA, USA and* [2]*Department of Microbiology, University of Georgia, Athens, GA, USA*

We used a combination of metatranscriptomic analyses and quantitative PCR (qPCR) to study seasonal changes in Thaumarchaeota populations from a salt marsh-dominated estuary. Surface waters (0.5 m depth) were sampled quarterly at Marsh Landing, Sapelo Island, GA, USA over a 3-year period. We found a mid-summer peak in Thaumarchaeota abundance measured by qPCR of either 16S rRNA or *amoA* genes in each of the 3 years. Thaumarchaeota were 100–1000-fold more abundant during the peak than at other times of the year, whereas the abundance of ammonia- and nitrite-oxidizing Bacteria varied <10-fold over the same period. Analysis of the microdiversity of several highly transcribed genes in 20 metatranscriptomes from a 1-year subset of these samples showed that the transcriptionally active population consisted of 2 or 3 dominant phylotypes that differed between successive summers. This shift appeared to have begun during the preceding winter and spring. Transcripts from the same genes dominated the Thaumarchaeota mRNA pool throughout the year, with genes encoding proteins believed to be involved in nitrogen uptake and oxidation, and two hypothetical proteins being the most abundant transcripts in all libraries. Analysis of individual genes over the seasonal cycle suggested that transcription was tied more closely to variation in growth rates than to seasonal changes in environmental conditions. Day–night differences in the relative abundance of transcripts for ribosomal proteins suggested diurnal variation in Thaumarchaeota growth.

The ISME Journal (2014) **8,** 685–698; doi:10.1038/ismej.2013.171; published online 17 October 2013
Subject Category: Microbial ecology and functional diversity of natural habitats
Keywords: ammonia oxidation; nitrification; estuary; Thaumarchaeota; nitrite; diurnal

Introduction

16S rRNA sequences characteristic of Archaea were first reported in low temperature marine environments in 1992 (DeLong, 1992; Fuhrman *et al.*, 1992), and a clade related to the Crenarchaeota was designated as the 'marine group 1 Archaea' by DeLong (1992). Based on genomic evidence, it has been proposed that these marine group 1 Archaea and related organisms should be assigned to a new phylum, the Thaumarchaeota, within the kingdom Archaea (Brochier-Armanet *et al.*, 2008; Kelly *et al.*, 2010; Spang *et al.*, 2010). Data collected in the 5–10 years following their discovery revealed that Thaumarchaeota are widely distributed and can be abundant in marine, fresh water (including hot springs) and terrestrial environments (reviewed by Francis *et al.*, 2005; Fuhrman and Hagström 2008; Erguder *et al.*, 2009; Nicol *et al.*, 2011; Ward, 2011; Biller *et al.*, 2012). Thaumarchaeota have proven difficult to culture, and our understanding of the details of their metabolism has been based primarily on inferences from culture-independent methods (for example, Ouverney and Fuhrman, 2000; Teira *et al.*, 2004; Venter *et al.*, 2004; Leininger *et al.*, 2006) and remained obscure until recently. Evidence from a variety of sources (reviewed in Francis *et al.*, 2007; Prosser and Nicol, 2008; Schleper and Nicol, 2010; Ward, 2011) strongly suggested that Thaumarchaeota were chemoautotrophs, depending on ammonia oxidation to supply the energy needed for carbon fixation; though the possibility that they might have heterotrophic or mixotrophic capabilities was also suggested (Ouverney and Fuhrman, 2000; Teira *et al.*, 2004; Kirchman *et al.*, 2007; Agogué *et al.*, 2008; Kalanetra *et al.*, 2009). Success in obtaining a pure culture of the Thaumarchaeote '*Candidatus* Nitrosopumilus maritimus' strain SCM1 has led

Correspondence: JT Hollibaugh or MA Moran, Department of Marine Sciences, University of Georgia, 226 Marine Sciences Building, Athens, GA 30602-3636, USA.
E-mail: aquadoc@uga.edu or mmoran@uga.edu
[3]Present address: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
[4]Present address: Biology Program, Vanderbilt University, Nashville, TN 37203, USA.
Received 1 May 2013; revised 22 July 2013; accepted 1 September 2013; published online 17 October 2013

to an understanding of the basic features of their ecophysiology (Könneke *et al.*, 2005; Martens-Habbena *et al.*, 2009; Walker *et al.*, 2010), including their ability to grow autotrophically by ammonia oxidation; however, the factors controlling their distribution and seasonal abundance cycles in the environment are not fully understood (Pitcher *et al.*, 2011; Biller *et al.*, 2012).

Analysis of a metatranscriptome from coastal waters of the South Atlantic Bight collected in August 2008 indicated the presence of a significant population of Thaumarchaeota in the water column in mid-summer (Gifford *et al.*, 2011; Hollibaugh *et al.*, 2011). Analysis of additional samples from 2008 by quantitative PCR (qPCR) revealed that Thaumarchaeota were ~3 orders of magnitude more abundant in August than at other times of the year (Hollibaugh *et al.*, 2011), suggesting that we had either sampled a random 'bloom' event, or that elevated Thaumarchaeota abundance in mid-summer is a normal part of the seasonal succession of prokaryotic communities in these waters.

In this paper, we extend our analysis of Thaumarchaeota abundance over 3 seasonal cycles to test the hypothesis that the mid-summer bloom is a recurrent feature of these populations. We also extend our analysis of metatranscriptomes to cover 1 full year with the goal of gaining insights into the environmental factors regulating the seasonal cycle Thaumarchaeota populations. We compare the dynamics of Thaumarchaeota populations to those of ammonia- (AOB) and nitrite-oxidizing Bacteria (NOB) to better understand the coupling between the processes of ammonia- and nitrite oxidation at this site, and present environmental data to provide a context for our analysis of the metatranscriptomes.

Materials and methods

Sample collection

Near-surface water samples were collected quarterly (February, May, August and November) from a floating dock at Marsh Landing, Sapelo Island, GA, USA (31° 25' 4.08N, 81° 17' 43.26W, Supplementary Figure 1), ~6 km from the mouth of Doboy Sound, as described in Gifford *et al.* (2011). Briefly, during each campaign, samples were collected twice per day <1 h before high tide (approximately noon and midnight), over a 2-day period (Supplementary Table 1). Duplicate samples for RNA extraction were collected in rapid succession by filtering ~6 l of surface (~0.5 m depth) sea water through 3-μm pore size filters (Capsule Pleated Versapor Membrane; Pall Life Sciences, Ann Arbor, MI, USA) then through 0.22-μm pore size filters (Supor polyethersulfone; Pall Life Sciences) using a peristaltic pump. The 0.22 μm filter was placed in a Whirl-Pak plastic bag (Nasco, Fort Atkinson, WI, USA) and immediately frozen in liquid nitrogen. Total time from the start of filtration to freezing was ~10 min. The second sample was filtered immediately (~5 min delay) after the filter from the first sample was placed in liquid nitrogen. We collected samples for DNA and nutrient analyses concurrently by filling 20 l carboys with surface water from the same location while the RNA samples were filtering. Once the second RNA sample was frozen, we filtered 12 l of the DNA sample through 3 and 0.22 μm filters as above, then the 0.22 μm filters were frozen in liquid nitrogen. Nutrient samples were filtered through GF/F glass fiber filters, frozen immediately and stored frozen at −80 °C until analysis.

Environmental data

Ammonium (NH_4^+), nitrite, nitrate plus nitrite (NO_x^-), dissolved organic nitrogen, total dissolved nitrogen, phosphate (P), dissolved organic phosphate, total dissolved P and dissolved organic carbon concentrations were measured using standard methods as specified in GCE-LTER metadata (accessible through http://gce-lter.marsci.uga.edu). The Sapelo Island National Estuarine Research Reserve (SINERR) maintains automated meteorological and water quality monitoring stations (Stations ML and LD, respectively) located ~15 and ~50 m, respectively, from the sample collection site. These data and the associated metadata are available through the National Estuarine Research Reserve System's Centralized Data Management Office (http://cdmo.baruch.sc.edu/). The water quality sensor package (a YSI 6600 data sonde) samples at 15 min intervals and is located at a fixed depth ~0.5 m above the bottom (~2 m depth at mean low water). The sensors are cleaned and calibrated on a bi-weekly schedule. SINERR also collects chlorophyll *a* (Chl*a*) and nutrient samples at this site at approximately bi-weekly intervals. Chl*a* concentrations were determined fluorometrically and nutrients are measured by flow injection analysis by SINERR personnel. Continuous monitoring and Chl*a* data recorded closest to the time we sampled, and meteorological data averaged over 48 h preceding the time we sampled, were used in the statistical analyses presented below. We used our own nutrient data in the statistical analysis, but the more highly resolved SINERR phosphate data are shown to illustrate seasonality.

qPCR

DNA was extracted from filters using PowerSoil (MoBio, Carlsbad, CA, USA) kits as per manufacturer's instructions and eluted into 4 ml of buffer. The abundance of ammonia monooxygenase subunit A (*amoA*) genes from ammonia-oxidizing Archaea (AOA) and AOB (primers targeted β-Proteobacteria) and of Thaumarchaeota and Bacteria 16S rRNA genes (*rrs*) was determined by real-time qPCR as

described previously (Hollibaugh *et al.*, 2011). The abundance of *Nitrospina rrs* genes was determined as described in Mincer *et al.* (2007). The *Nitrospina* primers do not capture all NOB; however, *Nitrospina* accounted for about 80% of the NOB population detected by tag sequencing *rrs* genes in samples from the study area (MIRADA project, http://vamps.mbl.edu/portals/mirada/mirada.php), *Nitrospira* and *Nitrococcus* were also present in this data set. The relative abundance (percent of prokaryotes) of Thaumarchaeota, AOB and *Nitrospina* was calculated from gene abundance (copies per ng of DNA extracted from the sample) and *rrs* or *amoA* gene dosage (copies per genome) as described in Kalanetra *et al.* (2009), and assuming two copies of the *rrs* gene per *Nitrospina* genome (Mincer *et al.*, 2007). All primers are given in Supplementary Table 2.

mRNA isolation

Metatranscriptomes were prepared for a subset of these samples collected quarterly from August 2008 to August 2009 (Supplementary Table 1). mRNA was isolated from the samples as described previously (Poretsky *et al.*, 2006, 2009; Gifford *et al.*, 2011). Before beginning the extraction, 25 ng of a 994-nt RNA standard (derived from the pGEM cloning vector) was added to the sample to serve as an internal standard (Gifford *et al.*, 2011). 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 were treated in two ways to remove ribosomal RNA. Epicentre's mRNA-Only kit (Epicentre, Madison, WI, USA) was used first to decrease rRNA contamination enzymatically. The samples were then treated with MICROBExpress and MICROBEnrich kits (both from Applied Biosystems) that couple rRNA oligonucleotide hybridization probes with magnetic separation to enrich for mRNA. RNA samples 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 converted to cDNA using the Universal RiboClone cDNA synthesis system (Promega, Madison, WI, USA) with random hexamer primers. Left over reactants and nucleotides from cDNA synthesis were removed from the sample using QIAquick PCR purification kits (Qiagen).

Sequencing and annotation

cDNA was sequenced using Roche 454 GS-FLX (454 Life Sciences, Branford, CT, USA) and Illumina (San Diego, CA, USA) GA-II single-end or paired-end technologies as indicated in Supplementary Table 1. We performed tests to confirm that metatranscriptomes and assemblies of genes from samples taken at the same time but sequenced on different platforms were not biased by sequencing method (other than differences in depth of coverage, summary given in Supplementary Table 3). Ribosomal RNA sequences in these libraries 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 were excluded from further processing. Sequences remaining after removing rRNA and internal standards have been deposited in the Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA; https://portal.camera.calit2.net) database under accession numbers CAM_PROJ_Sapelo2008, CAM_P_0000917 and CAM-P-0001108.

The remaining non-rRNA sequences were queried against a local copy of NCBI's RefSeq database (version 47, released May 7, 2011) 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 gene assignment for that sequence. We did not remove replicate reads (defined according to Gomez-Alvarez *et al.*, 2009) when determining hit counts in GS/FLX-pyrosequenced libraries, as our analysis (Gifford *et al.*, 2011; Hollibaugh *et al.*, 2011) indicates that most of the replicates in metatranscriptomes from these libraries are not artifacts. Thaumarchaeota reads were extracted from each metatranscriptome by filtering the data sets using taxonomic information in the hit descriptions. At the time the BLASTx search was initiated, the RefSeq database contained three Thaumarchaeota genomes: '*Ca. N. maritimus*' strain SCM1 (Nmar), '*Ca. Nitrosoarchaeum limnia*' strain SFB1 (Nlim) and Cenarchaeum symbiosum A, (CENSYa); thus, the Thaumarchaeota reads we recovered were assigned to genes from these genomes.

Seasonal variation in transcript relative abundance

We examined seasonal variation in Thaumarchaeota metabolism by comparing the relative abundance of transcripts from specific genes between seasons. The data set consists of four metatranscriptome libraries from each season (Supplementary Table 1: August 2008—Summer 1—Su1; November 2008—Fall—F; February 2009—Winter—W; May 2009—Spring—Sp; and August 2009—Summer 2—Su2) with technical replicates for two of these (FN101A/B,W-D1 and FN146A/B, Su2-N1). We calculated the fractional contribution of reads from a given gene to all of the Thaumarchaeota reads retrieved in a given library to normalize for seasonal and methodological (for example, 454 vs Illumina sequencing, recovery efficiency) differences in the depth of coverage. Hit counts were not normalized for gene length, as seasonal comparisons were made on a gene-by-gene basis. Instances of no hits to a given gene in a given

library are assigned a value of 0, regardless of the depth of coverage of that library.

Seasonal changes in the relative abundance of transcripts from a specific gene were then tested for statistical significance by a two-step approach. First, a Randomized Sum of Squares test (RSS; Simon 1992; implemented in R; http://www.r-project.org/) was used to test the null hypothesis that there was no difference in the relative abundance of transcripts of a given gene among seasons. This is a nonparametric, bootstrap test with no underlying assumption about distributions that treats each gene independently of others in the data set. If we rejected the null hypothesis of no difference among seasons ($P<0.05$), we then used *t*-tests of the relative abundances of transcripts between all pairs of seasons, treating each library from that season as a replicate ($n=4$, technical replicate libraries were averaged for the calculation), to determine which of the pairwise combinations were significantly different at $P<0.05$.

In order to improve the power of the analysis and in recognition of the fact that short reads of different parts of a given gene sequence might be binned to different taxa by BLASTx, we performed the same analysis using the Clusters of Orthologous Groups (COG) assignments from the IMG database (Integrated Microbial Genomes; http://img.jgi.doe.gov/cgi-bin/w/main.cgi) to combine transcripts from orthologous genes. IMG annotation did not include COG assignments for 901 of the 3075 genes in our data set. This 'unassigned COG' group included genes with some of the most abundant transcripts, with the result that most (67.8%) of the hits in the data set could not be assigned to COGs. These included some genes with likely physiological significance, such as the gene encoding ammonia monooxygenase subunit A (*amoA*). We examined the temporal variation in the relative abundance of hits to some of these genes, aggregated across taxa, by analysis of variance (ANOVA).

Genes and COGs for which seasonal differences in transcription were identified by the bootstrap test (at $P<0.05$) were plotted against '*Ca. N. maritimus*' strain SCM1 metabolic pathways using the Kyoto Encyclopedia of Genes and Genomes (KEGG) Mapper. We used '*Ca. N. maritimus*' strain SCM1 as the reference genome because 69% of the metatranscriptomic reads in our data set mapped best to it (Supplementary Table 1) and Sanger-sequenced, cloned amplicons of 16S rRNA and *amoA* genes were closely related to '*Ca. N. maritimus*' strain SCM1 reference sequences (Hollibaugh *et al.*, 2011). When necessary to group genes for statistical tests, assemblies and so on, hits assigned to genes mapping to C. symbiosum or '*Ca. N. limnia*' strain SFB1 were reassigned to the ortholog from '*Ca. N. maritimus*' strain SCM1 based on reciprocal BLASTx homologies with a cutoff of 1.0E–5.

Assemblies

The Geneious Pro software package version 6.1.4 (Biomatters Ltd., Auckland, NZ; Drummond *et al.*, 2012) was used to map reads against the reference sequences, for *de novo* assembly of reads into contigs and to host phylogenetic analyses. Assemblies were constructed using unedited cDNA sequences and combined transcripts from orthologous genes, regardless of taxonomic assignment. Unless otherwise noted, the appropriate '*Ca. N. maritimus*' strain SCM1 genomic sequence was used as a scaffold when reads were mapped against a reference sequence.

Microdiversity

We analyzed seasonal changes in the microdiversity of the transcriptionally active population as follows, using transcripts from a subset of genes that were well-represented in the metatranscriptomes. First, the reads assigned to a specific gene (for example *amoA*, which is represented by orthologs with locus tags Nmar_1500, Nlim_1890 or CENSYa_0402 in our data set) were mapped against the appropriate '*Ca. N. maritimus*' strain SCM1 reference sequence (here Nmar_1500) using the following assembly parameters: ungapped, at least 25 bp overlap, at least 75% identity in the overlapping region, word length 15 bp at 100% identity, maximum of 15% mismatches/read, maximum of four ambiguities/read. Under these conditions, most of the sequences assigned to the subject gene by BLASTx mapped to the reference gene, though this was not always the case (Supplementary Table 4). We then used the 'Find Variations/single-nucleotide polymorphisms (SNPs)' module in Geneious Pro 6.1.4 to identify all SNPs relative to the reference sequence, with the constraints that the substitution frequency at a given SNP position had to be at least 2.5% of the reads and coverage had to be >10 reads. We ran this analysis on the whole data set for that gene to generate a master list of SNPs, then ran the same analysis for subsets of transcripts for the same gene, sorted by season.

Phylogenetic analysis

We amplified (primers are given in Supplementary Table 2) and cloned *amoA* amplicons from eight different samples from Su1 and eight samples from Su2. Six cloned amplicons from each sample were selected randomly for sequencing (a total of 96 cloned sequences). Sequences were aligned with the 'Geneious Alignment' routine using iterations of a global alignment with free-end gaps with increasing stringency (the cost matrix increased from '51% similarity (5.0/−3.0)' to 'transition/transversion 5.0/1.0/−4.0'; with the gap opening penalty fixed at 15 and the gap extension penalty fixed at 4 for all iterations). Sequences were trimmed as needed after each iteration and realigned, then a neighbor-joining tree using the HKY model for genetic distances and no outgroup was constructed using the 'Geneious Tree Builder.' This tree was used to select

representative sequences for each of the *amoA* clades (>99.9% identity) in our samples.

We used *de novo* assembly under stringent rules (overlap = 35 bp, overlap identity = 98%, word length = 75 bp, word identity = 100% and ⩽15% mismatches/read, ambiguities = 1) of all reads binned to a specific gene (for example, *amoA*) in samples from Su1 or Su2 to obtain contigs. Consensus sequences from contigs that covered the full length of the reference sequence and for which individual reads were distributed unimodally over the contig were used for phylogenetic comparisons. The success of this approach varied depending on the degree of conservation within the target gene: it worked well for genes encoding ammonia monooxygenase subunits, but failed due to apparent hyperdiversity with (for example) Nmar_1667 and its homolog.

Consensus sequences were aligned with reference sequences and, for *amoA*, with representative cloned amplicons as described above. Final phylogenetic trees were constructed using MR BAYES (Huelsenbeck and Ronquist, 2001) implemented in Geneious. The number of reads assigned to a given contig as a percentage of all reads assigned to that gene from that set of samples (Su1 or Su2) was used to assess the relative contribution of that genotype to the transcriptionally active population. This was compared with the percentage of cloned *amoA* amplicons from the same set of samples that fell into each clade.

Statistical analyses

ANOVAs and individual *t*-tests were run using online tools (http://www.physics.csbsju.edu/stats/anova.html and http://graphpad.com/quickcalcs/ttest1.cfm; respectively). Model II ordinary least-squares pairwise regressions were calculated following Legendre and Legendre (1998) using software available from the R-Project web site (http://cran.r-project.org/web/packages/lmodel2/index.html). Coefficients of determination and confidence limits of Model II regressions were calculated from 999 bootstrap permutations. We used a combination of principal components analysis (PCA) and non-metric multi-dimensional scaling (MDS) to reduce the number of environmental and biological variables, respectively. Analyses were performed in R (http://www.r-project.org/) using the prcomp (stats package; PCA) and metaMDS (vegan package; MDS) commands. PCA analysis was run on three subsets of samples to reduce the number of samples eliminated due to missing [NH_4^+] or pH data. Variables included in the core data set are temperature; salinity; dissolved oxygen; turbidity; wind speed; precipitation; and the concentrations of Chl*a*, NO_x^-, phosphate, dissolved organic phosphate, dissolved organic nitrogen and dissolved organic carbon. In all cases, NO_x^- was log-transformed to achieve a normal distribution. To examine connections between the distribution of Thaumarchaeota genes and environmental conditions, we extracted loading values for each sample from PCA and MDS axes and used them in pairwise linear regressions using the lm command (stats package) in R.

Results

Seasonal variation in Thaumarchaeota abundance

The relative abundance (qPCR) of Thaumarchaeota 16S rRNA (*rrs*) and AOA ammonia monooxygenase subunit A (a*moA*) genes (Figure 1a) was elevated in summer (August) relative to other seasons of each year. The magnitude of summer relative abundance differed from year to year (ANOVA, $P<0.05$) with the highest values of both Thaumarchaeota *rrs* and AOA *amoA* gene abundance found in 2010 when Thaumarchaeota accounted for ~1.5% of prokaryotes. In contrast, the relative abundances of *Nitrospina rrs* and AOB *amoA* did not fluctuate seasonally (Figure 1b) and were ~3 orders of magnitude lower than AOA relative abundance in summer, though comparable at other times of the year.

An analysis of two metatranscriptomes from this series (libraries FN56 and FN57, August 2008) was presented in Hollibaugh *et al.* (2011). We processed 20 additional metatranscriptomes representing replicate, quarterly collections (Supplementary Table 1). The combined sequencing effort yielded 2.85×10^7 identifiable cDNA sequences (hits to RefSeq sequences with bit scores >40) that were ~100 (Illumina single-end) or ~220 bp (Illumina paired-end or 454 GS-FLX pyrosequencing) long. Hits to Thaumarchaeota genes accounted for 7.55×10^5 of these, or 2.6% of the total (averaged over all libraries), ranging from 183 to 400 714 (0.02–9.89%) reads per library (Supplementary Table 1). When corrected for the recovery of internal standards, the abundance of Thaumarchaeota transcripts in our samples averaged 1.3×10^{10} transcripts/l (range 9.6×10^7 to 6.5×10^{10}) and we captured, on average, 1 out of every 4.4×10^7 Thaumarchaeota transcripts in each sample (calculated from data in Supplementary Table S1). The time series of Thaumarchaeota transcript abundance (Figure 1c) revealed elevated abundance of Thaumarchaeota transcripts in the samples from August 2008 and August 2009 (average of 4.2% of RefSeq hits) versus at other times of the year (average of 0.14% of RefSeq hits). Transcripts most similar to genes encoded in the '*Ca. N. maritimus*' strain SCM1 genome dominated the Thaumarchaeota transcripts in each library regardless of season (mean of 69%, range of 51–86%), with transcripts most similar to '*Ca. N. limnia*' strain SFB1 genes, accounting for most of the remaining Thaumarchaeota hits (Supplementary Table 1).

In contrast to AOA, and consistent with qPCR data, the relative abundance of AOB and NOB (*Nitrobacter*, *Nitrococcus* and *Nitrospira*) transcripts was low (0.4% and 0.15%, respectively) and did not show strong seasonal dynamics

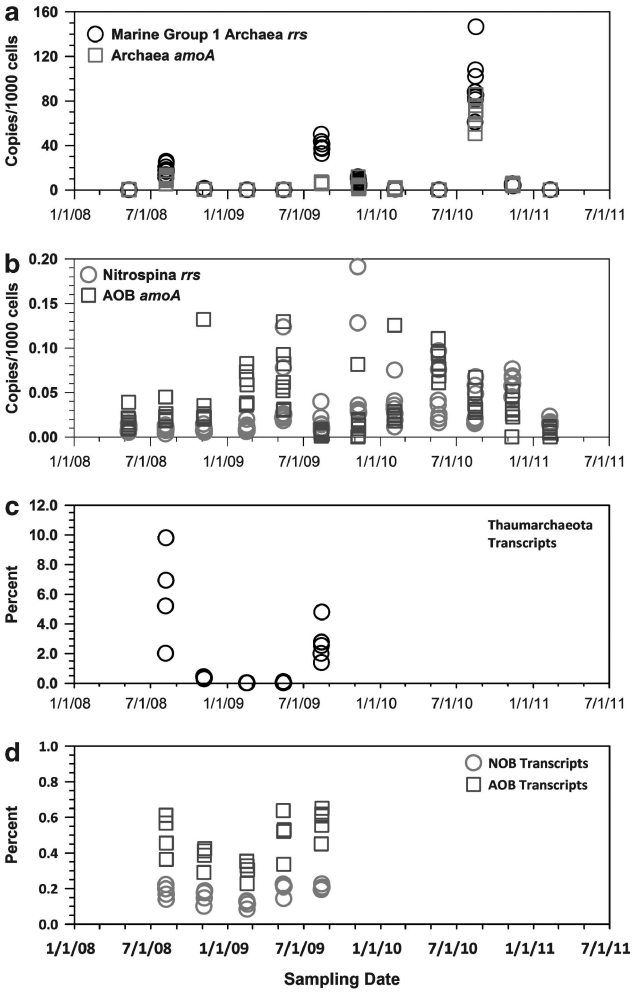

Figure 1 Time series of quarterly sampling at Marsh Landing, Sapelo Island, GA, USA. Data are from 4–8 replicate samples of surface water collected over a 48 h period and thus reflect variability associated with local patchiness. (**a**) qPCR measurements of the relative abundance of Thaumarchaeota *rrs* (○) and Archaeal *amoA* genes (□); (**b**) qPCR measurements of the abundance of *Nitrospina rrs* (○) and Bacterial *amoA* (□) genes; (**c**) contribution of Thaumarchaeota transcripts to metatranscriptomes (○); (**d**) contribution of transcripts from nitrite- (NOB; ○) or ammonia (AOB; □)-oxidizing Bacteria to metatranscriptomes. Relative abundances in **a** and **b** are calculated as in Kalanetra *et al.* (2009), hit counts in **c** and **d** are normalized as percentages of RefSeq hits (Bacteria plus Archaea, all prokaryotes) in each metatranscriptome.

(Figure 1d). The relative abundance of AOB transcripts is lower during the winter than at other times of the year, (ANOVA, $P<0.05$); however, the relative abundance of NOB transcripts did not change seasonally.

Correlation of Thaumarchaeota abundance with environmental variables

The August peak in AOA relative abundance appears to be accompanied by a biogeochemical signal in that nitrite concentrations increase relative to the background during the AOA bloom (Figure 2;

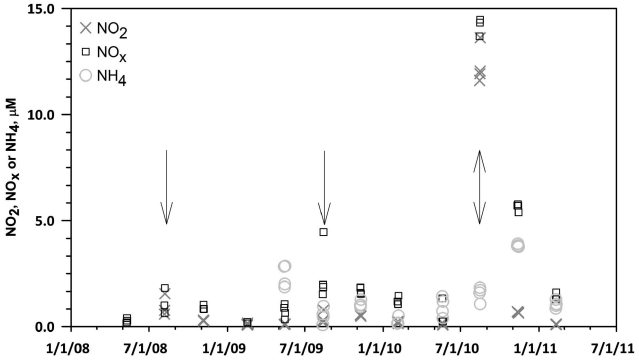

Figure 2 Time series of quarterly sampling of nitrite (NO_2; X), nitrite plus nitrate (NO_x; □), and ammonium (NH_4; ○) at Marsh Landing, Sapelo Island, GA, USA. Data are from 4–8 replicate samples of surface water collected over a 48 h period and thus reflect variability associated with local patchiness. Arrows indicate August sampling periods.

ANOVA, $P<0.05$). This was most obvious in August 2010 when AOA relative abundances were the highest observed in the data set, and NO_x and NO_2 concentrations measured in replicate samples taken over 48 h averaged 14.7 and 12.3 µM, respectively. NH_4 data were not available for the first four sampling periods; however, concentrations were significantly lower during Summer than in other seasons for the rest of the time series (Figure 2; ANOVA, $P<0.05$).

Of the variables measured by the SINERR monitoring program, temperature (Figure 3a), oxygen concentration (Figure 3b), phosphate concentration (Figure 3c) and pH (Figure 3d) display consistent seasonal cycles that are in phase with the August peak in AOA relative abundance. These patterns are consistent with the strong mid-summer, net heterotrophy reported for these waters by Wang and Cai (2004) and Wang *et al.* (2005). We performed PCA of the distribution of samples in parameter space defined by these and additional environmental variables. Samples were grouped both by season and by year (Figure 4, which shows the subset including pH). Samples were segregated by season along PC1 (41.1% of variance explained), whereas samples from individual years were separated along PC2 (15.3% of variance explained). MDS analysis based on the relative abundance of AOA *amoA*, Archaeal *rrs*, Bacterial *amoA*, Nitrospina *rrs* and Bacteria *rrs* separated samples along MDS axis 1 (Supplementary Figure 2), which was the only statistically significant axis, with August samples (corresponding to high AOA *amoA* and Thaumarchaeota *rrs* abundance) clearly differentiated from samples collected at other times of the year (Supplementary Figure 2). Linear regressions of PC1, PC2 and PC3 from each of the three PCA runs against MDS Axis 1 were highly significant (Supplementary Table 5), with many of the same variables showing significant loadings in all three analyses. In summary, most of the environmental factors covary, no single factor clearly explained the

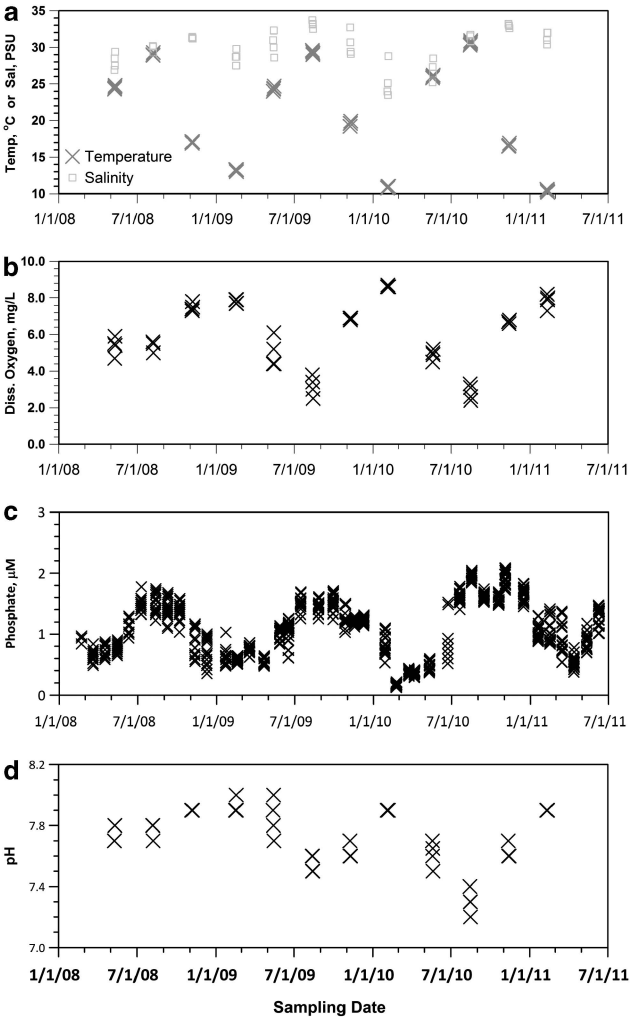

Figure 3 Time series of quarterly sampling at Marsh Landing, Sapelo Island, GA, USA. Data are from a continuous monitoring station adjacent to the sampling site. (**a**) Water temperature (X) and salinity (□); (**b**) dissolved oxygen concentration (X); (**c**) phosphate; (**d**) pH.

seasonal distribution and there was only a weak correlation with NH_4^+ concentration.

Characteristics of metatranscriptomes

The Thaumarchaeota transcripts in our data set were assigned to 3075 different genes, with the relative abundance of transcripts assigned to 420 genes varying significantly between seasons (bootstrap analysis, $P<0.05$ for tests of individual genes, Supplementary Table 6). These genes were found in several KEGG pathways, notably nucleotide synthesis and carbon fixation (not shown). The transcripts in our data set were assigned to 893 COGs (Supplementary Table 7) with the relative abundance of transcripts assigned to 159 COGs varying significantly among sampling periods ($P<0.05$) in the bootstrap analysis (Supplementary Table 7). Most of the seasonally varying COGs were elements of general metabolic pathways, rather than

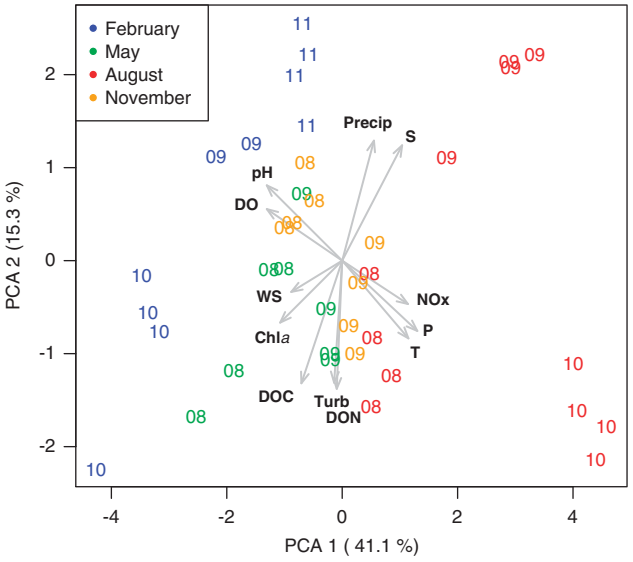

Figure 4 Plot of samples in parameter space defined by PC1 (41.4% of variance explained) and PC2 (15.3% of variance explained) from a PCA based on the environmental variables: water temperature (T); salinity (S); turbidity (Turb); wind speed (WS); precipitation (Precip); pH; and the concentrations of chlorophyll *a* (Chl*a*), dissolved oxygen (DO), nitrate plus nitrite (NO_x), phosphate (P), dissolved organic nitrogen (DON), and dissolved organic carbon (DOC). Samples are indicated by numbers in colors showing the year and month they were collected. Vector arrows indicate the contribution of that variable to each of the two axes.

components of subsystems with specific functions (for example, aromatic compound degradation). The distribution of transcripts among KEGG pathways showed little consistent seasonal variation (data not shown), even though the relative abundance of transcripts assigned to individual COGs did vary seasonally. For example, transcripts assigned to COG0004 (ammonia permeases/transporters) accounted for 5.8% of all Thaumarchaeota transcripts retrieved and were found in all samples (Supplementary Table 7), but were most abundant in the Summer 1 samples, where transcripts assigned to Nmar_1698 and Nlim_1421 contributed disproportionately to hits assigned to this COG. Transcripts assigned to COGs containing ribosomal proteins or DNA excision and repair had greater relative abundance in Summer 2 (Supplementary Table 7). Transcripts assigned to COG0106, the histidine biosynthesis module in KEGG, were also present in greater relative abundance in Summer 2 than during other seasons, whereas hits to COG3794, annotated as plastocyanin, were significantly lower in winter samples than in samples from either summer.

Thaumarchaeota ammonia monooxygenase subunits have not been assigned to COGs, and thus were not included in the analysis described above. Consequently, we investigated seasonal variation in transcription of these genes by comparing the relative abundance of their transcripts in the samples using ANOVA (Table 1). We extended

Table 1 Seasonal variation in the relative abundance of genes involved in ammonia uptake and oxidation

	Mean relative abundance (% of Thaumarchaeota hits)						*ANOVA: seasonal difference? (21 df)*
	All libraries	*Summer 1*	*Fall*	*Winter*	*Spring*	*Summer 2*	
Ortholog							
amoA[a]	3.50	**4.29**	1.71	1.05	1.62	1.81	**F = 17.87, P < 0.0001**
amoB[b]	0.95	1.10	**1.24**	0.12	0.35	0.51	**F = 6.281, P = 0.0027**
amoC[c]	7.18	8.72	6.62	3.94	5.11	5.91	F = 1.733, P = 0.19
nirK-like proteins[d]	9.48	**7.38**	3.74	1.68	4.08	5.52	**F = 3.113, P = 0.043**
amo-associated hypothetical[e]	1.03	**1.99**	0.48	0.50	0.58	1.19	**F = 4.098, P = 0.017**
Transporters/permeases[f]	5.71	**5.71**	2.56	2.94	2.42	4.16	**F = 8.699, P = 0.0005**
Ratios of ortholog abundance							
amoB/amoA	0.270	0.249	**0.755**	0.133	0.163	0.280	**F = 5.614, P = 0.0046**
amoC/amoA	2.049	1.980	3.888	3.733	2.940	3.341	F = 1.4, P = 0.28
nirK /amoA	2.706	1.798	2.684	1.633	2.355	3.090	F = 0.9561, P = 0.46
Nmar_1501/amoA	0.293	0.451	0.300	0.500	0.351	0.654	F = 1.219, P = 0.34
Transporters/amoA	1.631	1.388	1.752	**2.733**	1.497	2.289	**F = 3.022. P = 0.047**

Abbreviation: ANOVA, analysis of variance.
Bold signifies $P<0.05$.
[a]Nmar_1500, Nlim_1890, CENSYa_0402.
[b]Nmar_1503, Nlim_1893, CENSYa_0394.
[c]Nmar_1502, Nlim_1892, CENSYa_0399.
[d]Nmar_1259, Nmar_1667, Nlim_1007.
[e]Nmar_1501, Nlim_1891, CENSYa_0401.
[f]Nmar_0588, Nmar_1698, Nlim_1564, Nlim_1421, CENSYa_0526.

the analysis to include putative *nirK* homologs (COG2132, multicopper oxidases) and ammonia permeases/transporters (COG0004, ammonia permease) for comparison with the results of RSS tests of the seasonal variation of single genes (Supplementary Table 6). The relative abundances of transcripts for orthologs of *amoA*, the *nirK*-like proteins, a hypothetical protein associated with the *amoABC* genes (Nmar_1501, Nlim_1891 and CENSYa_0401) and ammonia permeases/transporters were greater among the Thaumarchaeota sequences for samples collected during Summer 1 than on other dates, while *amoB* transcripts were slightly more abundant in Fall samples (ANOVA, $P<0.05$, Table 1). The relative abundance of transcripts for *amoC* orthologs (Nmar_1502, Nlim_1892 and CENSYa0399) did not change significantly over the course of the year. These results are consistent with those of the RSS tests of the individual genes (Supplementary Table 6).

Ammonia monooxygenase subunits are believed to be present in a stoichiometry of 1:1:1 in the active enzyme (Walker *et al.*, 2010). In contrast, we observed a stoichiometry of 1:0.3:2.5 for *amoABC* transcripts, whereas the ratio of transcripts for the *amo*-associated hypothetical proteins homologous to Nmar_1501 to *amoA* was 0.3. The ratios of transcripts for *amoC*/*amoA* and the Nmar_1501 homologs/*amoA* did not vary significantly among seasons. However, the ratio of *amoB*/*amoA* transcripts was significantly higher in Fall than in other seasons. *amoA* and *amoB* are encoded on different strands of the DNA molecule in *Ca. N. maritimus* (Walker *et al.*, 2010) which, combined with our results, suggests that their transcription might be controlled independently. The ratio of transcripts for ammonia transporters to *amoA* transcripts was 1.6 and higher in Winter than other seasons (ANOVA, $P<0.05$, Table 1). The ratio of transcripts for the *nirK*-like proteins to *amoA* was 2.7 and did not change seasonally. The ratio of *nirK*/*amoA* transcript abundance we found was much lower than the 10–100-fold ratio reported by Lund *et al.* (2012) based on qPCR assays of Monterey Bay Thaumarchaeota populations. This difference may reflect hypervariability of these genes relative to *amoA* (reflected in Supplementary Table 4), and thus poor recovery and quantification by our BLASTx-based approach relative to qPCR, which depends on conservation of relatively short regions of the gene.

Diurnal variation in transcription

We also examined the seasonal and diurnal variation of transcription of Thaumarchaeota genes annotated as encoding ribosomal proteins as possible indicators of variations in protein synthesis rates, following Gifford *et al.* (2013). Sequences identified as ribosomal protein genes accounted for 0.8% of the Thaumarchaeota reads in all samples and were assigned to 117 different genes (Supplementary Tables 6 and 7). The number of ribosomal protein transcripts recovered from Winter and Spring samples was low, ranging from 0–12 hits per sample, compromising our ability to detect seasonal variations, should they exist (ANOVA, $P=0.39$). However, the contribution of ribosomal

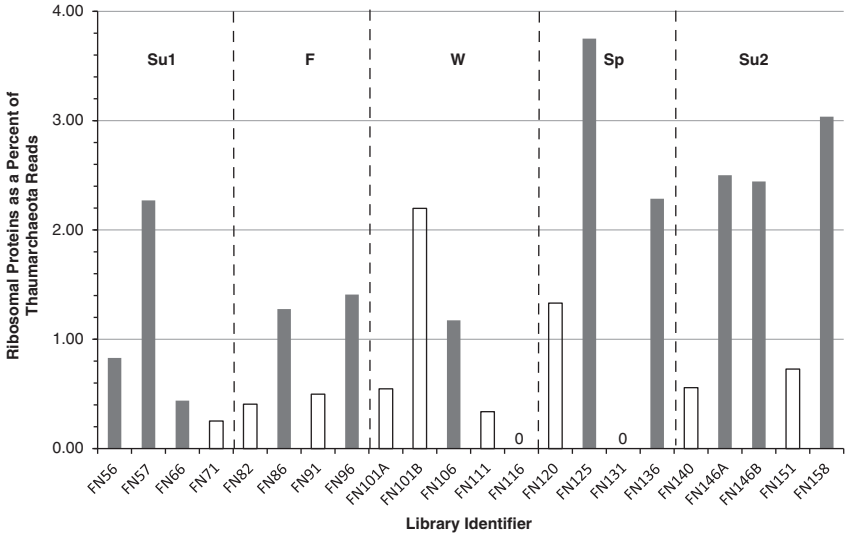

Figure 5 Relative abundance of transcripts encoding Thaumarchaeota ribosomal proteins in metatranscriptomes from samples collected at Marsh Landing, Sapelo Island, GA, USA from 2008 to 2009. Filled bars indicate relative abundance in samples collected at night, open bars are obtained from samples collected during the day (exact times are given in Supplementary Table 1). No Thaumarchaeota ribosomal protein transcripts were retrieved from libraries FN116 (N) and FN131 (D).

protein transcripts to the Thaumarchaeota metatranscriptome was significantly greater (2.7-fold, *t*-test $P = 0.012$) in samples collected at night versus day (Figure 5). For comparison, we performed the same analysis for the orthologs involved in ammonia oxidation listed in Table 1 and for each of the 30 most abundant individual genes in our metatranscriptomes (Supplementary Table 6), which cover a wide range of unrelated metabolic processes. We found no significant difference between day versus night samples in the relative abundance of transcripts assigned to any of these genes (data not shown). As increases in the relative abundance of transcripts for one gene or group of genes necessitates commensurate decreases in relative abundance elsewhere, we interpret this to indicate that transcription of genes for ribosomal proteins increases significantly at night, and that the requisite decrease in the relative abundance of other transcripts is spread relatively evenly over the rest of the transcriptome.

Seasonal changes in microdiversity of the transcriptionally active population

We next used assemblies of transcripts from selected genes to examine seasonal variation in the microdiversity of the Thaumarchaeota population (Supplementary Tables 4 and 8). This analysis revealed seasonal and interannual variation in the composition of the transcriptionally active Thaumarchaeota population at the study site. Low coverage (<10 reads per gene) limited our ability to analyze substitution frequencies to only summer samples for some genes. However, for those genes where the full-seasonal comparison could be made, SNP sites identified in Fall, Winter and Spring samples were consistent with those found during both Summers (Supplementary Table 8). We found that nucleotides were always different from the reference sequence at some SNP sites (~100% substitution rate), indicating consistent genetic differences between the thaumarchaeotes comprising this population and the reference organism, as shown previously by cloning and sequencing *amoA* and *rrs* genes (Hollibaugh *et al.*, 2011). For example, substitutions in the form of a transition at SNP position 241 and a transversion at SNP position 472 relative to the Nmar_1500 reference sequence occurred in 99% and 100% of all reads, respectively (Supplementary Table 8). In most cases, the substitutions were synonymous, generally resulting from differences at the third base of codons (Supplementary Table 8). The occurrence of nonsynonymous substitutions varied between the genes we examined, and may in some cases reflect misassignments of reads by BLASTx to closely related genes. For example, the highest frequency of nonsynonymous substitutions detected in our data set was with transcripts assigned to Nmar_0239 (4Fe-4S ferredoxin iron–sulfur-binding domain protein). The Nmar genome contains six similar genes with this gene product annotation.

Although the positions of SNP sites within a given gene were consistent across seasons, the relative abundance of transcripts with substitutions changed over the period of observation for most SNP sites, indicating the replacement of a dominant genotype present during Summer 1 with a different dominant genotype during Summer 2. This is most clearly shown with *amoC* transcripts (Figure 6) because coverage of that gene was deep enough to analyze patterns in winter and spring samples. A progressive shift in the frequency of reads containing

substitutions at SNP sites—reflected as a statistically significant deviation of regressions from the slope of 1.0 expected if substitution frequencies were the same in all seasons—suggests that the composition of the Thaumarchaeota population began to shift as early as the Winter and Spring samplings, culminating in a completely different pattern of substitution frequencies by Summer 2. Substitution frequency at most of the SNP sites with high (>80%) substitution frequencies during Summer 1 did not change between seasons and these were synonymous substitutions (Figure 6, Supplementary Table 8), as was the case for SNPs in *amoA* and *amoB* populations. These sites are not variable (and hence not 'SNPs') with regard to the local population of Thaumarchaeota.

This shift in substitution frequencies between Summer 1 and Summer 2 was observed in all of the genes we examined (Supplementary Table 8) with the exception of Nmar_0672, thioredoxin reductase, for which coverage was too low in Summer 2 to be reliable. Phylogenetic analysis of the consensus sequences of subpopulations of *amoA* and *amoC* transcripts (Figure 7) retrieved from Summer 1 and Summer 2 metatranscriptomes by *de novo* assembly of these reads showed that two genotypes accounted for most of the Thaumarchaeota reads in these samples, with the relative abundances switching between summers. A third genotype was retrieved from Summer 1 that was not detected in Summer 2, possibly as a consequence of differences in depths of coverage. The distribution of reads among clades in each season was not significantly different from the distribution of cloned amplicons among clades in each season (Figure 7, null hypothesis of no difference between estimates of relative abundance was tested by Model II ordinary least-squares regression: $y = 1.07 \times 8.93$; $n = 5$; $r^2 = 0.84$; $P = 0.013$; 95% confidence limits of slope = 0.23–1.9; 95% confidence limits of intercept = −48 to 30).

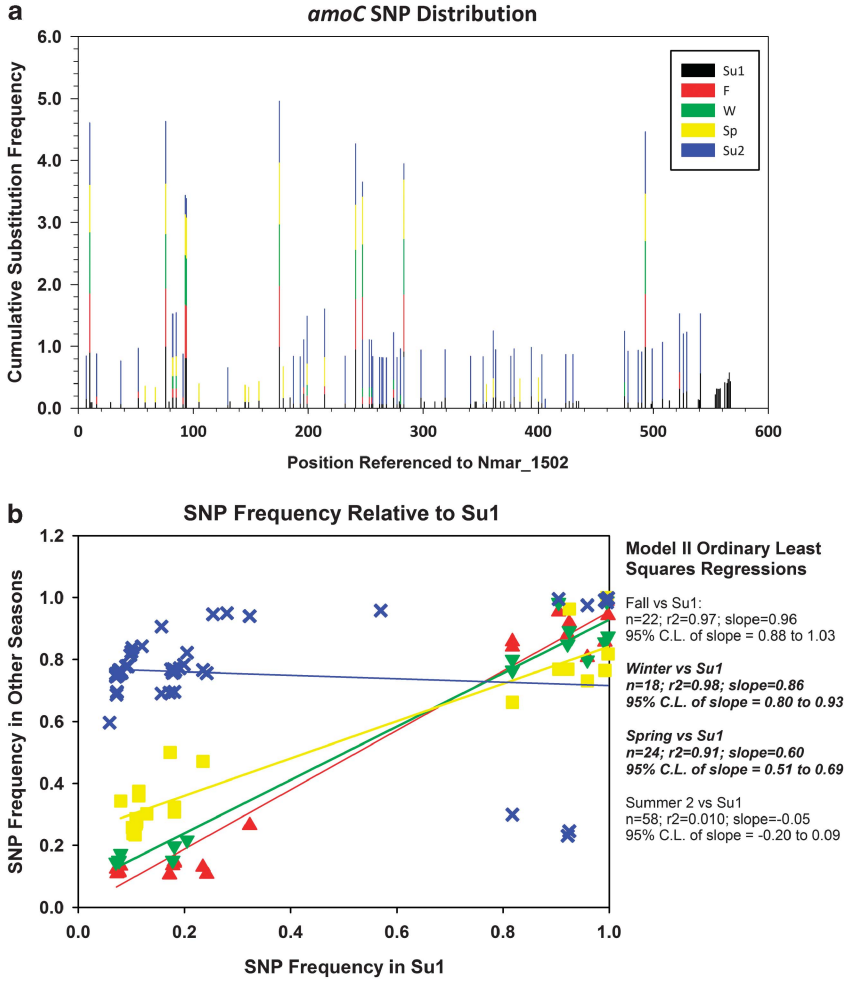

Figure 6 Comparison of SNP distributions in assemblies of sequences from *amoC* transcripts in libraries collected quarterly from August 2008 to August 2009. Panel **a** represents the frequency of substitution at SNP sites referenced to positions on the Nmar_1502 sequence, calculated by season (stacked bar for clarity so the ordinate scale gives the sum of substitution frequencies for all seasons, color key applies to both panels). Panel **b** compares the frequency of substitutions at SNP sites in Summer 1 with the frequency of substitutions at those sites in other seasons. Parameters of Model II ordinary least-squares regressions of the seasonal data are shown at the right of the graph.

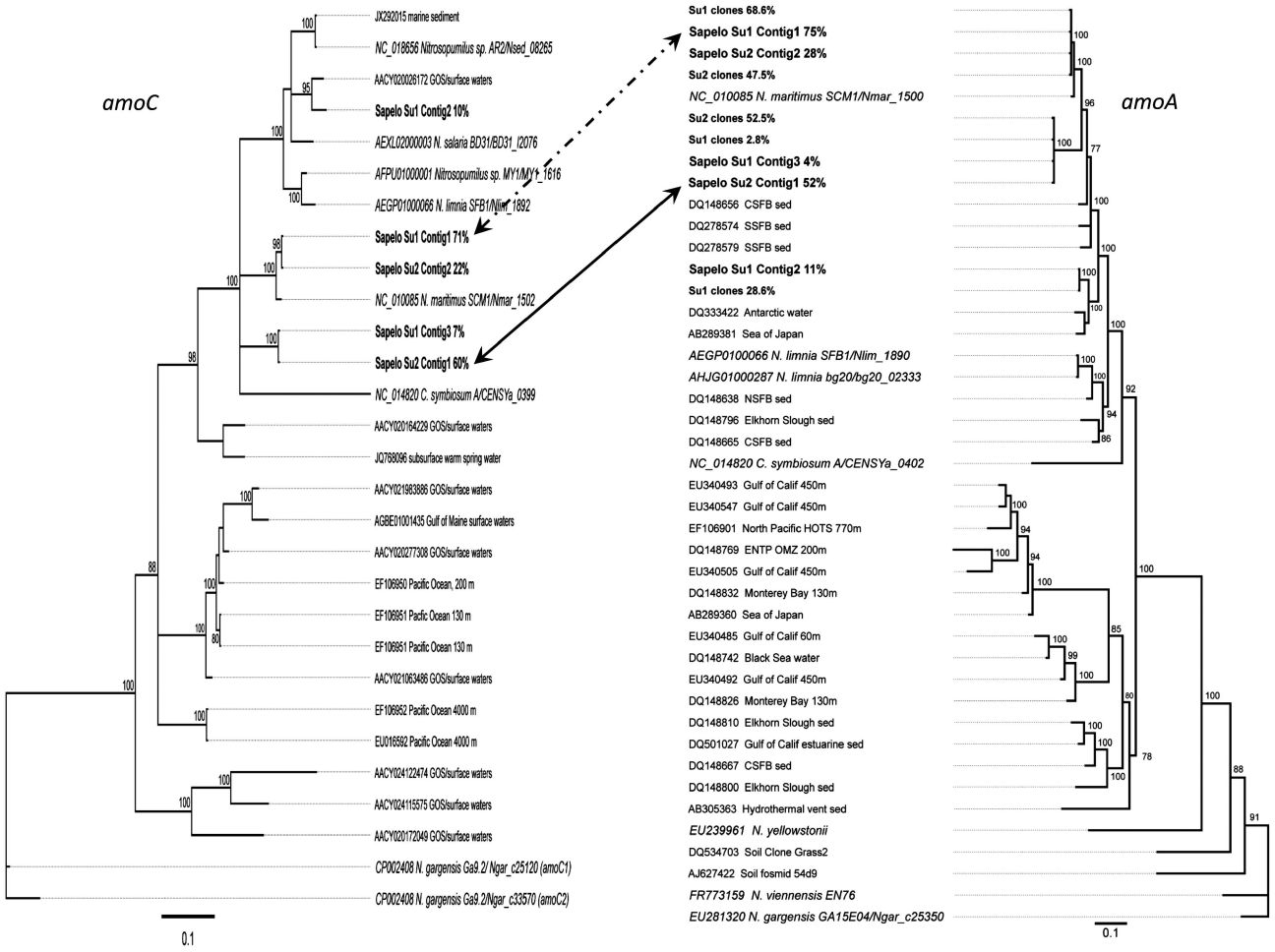

Figure 7 Comparison of the composition of Thaumarchaeota populations sampled in Summer 1 (Su1) versus Summer 2 (Su2). Consensus sequences were obtained from contigs retrieved by *de novo* assembly of transcripts assigned to Thaumarchaeota *amoA* and *amoC* genes. Percentages following the sample name indicate the portion of the transcript pool represented by that contig. Also shown are representative *amoA* gene sequences retrieved by cloning and sequencing PCR amplicons from Summer 1 (35 clones sequenced) and Summer 2 (40 clones sequenced) samples; the percentage of the library represented follows the sample identifier. These Bayesian consensus trees were constructed using MrBayes 3.2.1 (Huelsenbeck and Ronquist, 2001) from alignments of 590 bp (*amoA*) or 446 bp (*amoC*) using the HKY85 substitution model and the '*Ca. Nitrososphaera gargensis*' genes as the outgroups. Consensus support for branches is shown if >75%. Arrows connect dominant clades from Summer 1 (dashed) and Summer 2 (solid) assemblies.

Discussion

Both qPCR and metatranscriptomic data point to a recurrent, mid-summer peak in Thaumarchaeota abundance at the Sapelo Island coastal ocean study site. The magnitude of this peak varied from year to year, possibly due to either true interannual variation in the magnitude of the annual abundance peak or because we missed the abundance peak with the coarse temporal resolution (3 month intervals) sampling used here. Support for the latter explanation comes from a recent paper by Pitcher *et al.* (2011) who reported similar seasonal 'blooms' of Thaumarchaeota in the coastal North Sea that ramped up abruptly over a period of weeks to approximately the same standing stock each year. Elevated abundances persisted for a period of 2–3 months before returning to background levels. Although there was no obvious explanation for these population dynamics, they were consistent over four annual cycles. An interesting distinction between the seasonal dynamics of Thaumarchaeota populations in the North Sea and the Georgia coast is that maximum abundance of the North Sea populations occurs in mid-winter, whereas it occurs in mid-summer on the Georgia coast. The commonalities linking these two 'bloom conditions,' if there are any, are not obvious. Robidart *et al.* (2011) also report high temporal variability in Thaumarchaeota populations in Monterey Bay, CA, USA. Their time series is shorter, but more highly resolved, than the North Sea data presented in Pitcher *et al.* (2011) and the fluctuations they report appear to be tied to mesoscale or shorter oceanographic processes.

What causes Thaumarchaeota to 'bloom' in Georgia coastal waters? We first examined seasonal variation in environmental conditions at the study site but were not able to identify any single factor

that strongly correlated with, and thus might serve as the 'trigger' for the event, though combinations of factors generally capturing seasonal variation in net ecosystem metabolism at this site explained much of the variance. Next, we examined the metatranscriptomes for evidence of transcription patterns that might point to a more subtle trigger of the August blooms, for example, a signal indicating a change in a growth-limiting factor (ammonia availability, trace metal availability and so on) or environmental condition (salinity, oxygen, sulfide toxicity and so on). However, our analysis was limited by low relative abundance of Thaumarchaeota transcripts in winter and spring samples, and the problem of Type 1 errors arising from multiple tests (one for each gene) on the same data set (if corrected using Bonferroni or Benjamini–Hochstera procedures, none of the seasonal variation in the relative abundance of individual genes was statistically significant). We thus treat this analysis as a hypothesis generating exercise rather than as an effort to prove that one or another of the individual genes or COGs was up- or downregulated in our samples.

Of the 3075 genes tested, the null hypothesis (no difference in relative abundance of hits to that gene among sampling periods) was rejected for 420 of them. These significant hits mapped broadly over the metabolic pathways described for '*Ca. N. maritimus*' strain SCM1 in KEGG. A similar result was obtained when the data set was reduced by collapsing genes into COGs. Most of the significant differences involved Summer 2 samples (Supplementary Table 6), and an analysis of the distribution of SNPs among genes between populations indicated that the dominant strains present during Summer 1 were replaced by a closely related, but distinct, population in Summer 2. Although the metabolic properties of these two populations are similar, as indicated by the same genes being transcribed in both, increases in the relative abundance of transcripts for ribosomal proteins and biosynthetic pathways for carbon fixation and the synthesis of nucleic acids suggest that the Summer 2 population was growing faster than the Summer 1 population at the times we sampled them. This is consistent with the greater accumulation of nitrite during Summer 2 than Summer 1, assuming that loss rates are comparable, which seems reasonable, as water column NOB populations did not differ greatly (Figures 1b and d) and there is no evidence (for example, storms, wind events and salinity) to suggest that mixing or advection differed significantly between years. Transcription of genes identified as components of ABC transporters for phosphate, and phosphonate and iron was also greater in Summer 2 than in other samples. Phosphate concentrations are consistently high at this site during the summer (Figure 3c), suggesting that the increase in relative abundance of transcripts for phosphate transporters is not a response to nutrient limitation. Unfortunately, there are no data on seasonality of iron (or other trace metal) concentrations or bioavailability at this site. Relative abundance of transcripts from genes involved in DNA repair was also greater in Summer 2 samples than in samples from other seasons, suggesting that the Summer 2 population may have been subjected to greater DNA damage or possibly that cells have greater need of repair when they are growing faster.

We saw no large differences between bloom and non-bloom samples in the transcription of any of the genes we examined that could not be explained by low coverage in the non-bloom samples. We also found a consistent relationship between the contribution of specific *amoA* genotypes to Thaumarchaeota populations between clades and years, regardless of whether relative abundance of genotypes was assessed using genes or transcripts (Figure 7), suggesting little variation in transcription rates between subpopulations. This contrasts with our previous report (Hollibaugh *et al.*, 2011) of the possible occurrence of a clade of transcriptionally inactive organisms in the Summer 1 Thaumarchaeota population, based on recovery of a clade of *amoA* amplicons that were not represented by transcripts. We did not retrieve any amplicons from this clade when we re-sequenced these same samples, suggesting that this clade was reported in error. Our analysis thus suggests that, at least for genes that are transcribed highly enough for adequate coverage, variation in transcription may be more closely related to variation in growth rates of Thaumarchaeota than to seasonal changes in the influence of specific environmental factors. However, seasonal changes in environmental variables such as temperature or salinity would be expected to elicit a broad spectrum response that might be difficult to identify in analyses of single genes. Alternatively, it is possible that genes that responded to seasonally varying environmental factors, such as micronutrient availability, are not highly transcribed, and our ability to detect changes in their transcription levels was limited by low abundance of transcripts during winter and spring.

Finally, our analysis of day versus night transcription of genes for ribosomal proteins provides the first evidence for diurnal variation in the growth of field populations of Thaumarchaeota, and is counter to the pattern expected for microbial processes driven by coupling to photoautotrophy (greater substrate availability, faster growth and thus greater relative abundance of ribosomal proteins during daytime samples, for example, Gifford *et al.,* 2013). We did not find significant diurnal variation in the relative abundance of transcripts for genes involved in ammonia uptake and oxidation listed in Table 1, suggesting that the Thaumarchaeota populations are not responding to diurnal variation in ammonia availability. We tested other COGs, as well as the 30 most abundant individual genes in these metatranscriptomes, covering a wide range of metabolic processes, and found no significant difference in

their relative abundance in day versus night metatranscriptomes. The factors driving the diurnal variation in transcription of ribosomal proteins are unknown, but the response appears to be a general growth response that is not based on relative changes in the activity of any particular metabolic pathway in the Thaumarchaeota cells.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank R Newton for assistance with sample collection. L Tomsho and S Schuster provided 454 sequencing expertise. This project was funded by grants from the Gordon and Betty Moore Foundation and the National Science Foundation (MCB0702125, OCE0620959 and OCE0943278).

Author Contributions

SMG and MAM planned and orchestrated the metatranscriptome project. SS and SMG conducted the bioinformatic analyses; and SMG, BBT and JTH performed the statistical analyses. BBT and MJR performed the qPCR analyses. JTH, SMG, BBT and MAM interpreted the results; and JTH analyzed the data and wrote the paper.

Metatranscriptomic sequences have been deposited in the CAMERA Database (http://camera.calit2.net/) under the projects CAM_PROJ_Sapelo2008, CAM_P_0000917 and CAM_P_0001108. Sequences of the cloned amoA gene amplicons we obtained have been deposited in GenBank under accession numbers KF646597–KF6466668.

References

Agogué 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.

Biller SJ, Mosier AC, Wells GF, Francis CA. (2012). Global biodiversity of aquatic ammonia-oxidizing archaea is partitioned by habitat. *Front Microbiol* **3**: 252.

Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6**: 245–252.

DeLong EF. (1992). Archaea in coastal marine environments. *Proc Natl Acad Sci USA* **89**: 5685–5689.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C *et al.* (2012). *Geneious*. Biomatters Inc.: Auckland, NZ.

Erguder TH, Boon N, Wittebolle L, Marzorati M, Verstraete W. (2009). Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. *FEMS Microbiol Rev* **33**: 855–869.

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 editor. *Microbial Ecology of the Oceans*. John Wiley & Sons, Inc.: Hoboken, NJ, USA, pp 45–90.

Fuhrman JA, McCallum K, Davis AA. (1992). Novel major archaebacterial group from marine plankton. *Nature* **356**: 148–149.

Gifford SM, Sharma S, Booth M, Moran MA. (2013). Expression patterns reveal niche diversification in a marine microbial assemblage. *ISME J* **7**: 281–298.

Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA. (2011). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME J* **5**: 461–472.

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

Hollibaugh JT, Gifford S, Bano N, Sharma S, Moran MA. (2011). Metatranscriptomic analysis of ammonia-oxidizing organisms in an estuarine bacterioplankton assemblage. *ISME J* **5**: 866–878.

Huelsenbeck JP, Ronquist F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

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

Kelly S, Wickstead B, Gull K. (2010). Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc R Soc B Biol Sci* **278**: 1009–1018.

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.

Könneke 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.

Legendre P, Legendre L. (1998). *Numerical Ecology*, 2nd edn. Elsevier Science BV: Amsterdam, Netherlands.

Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW *et al.* (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**: 806–809.

Lund MB, Smith JM, Francis CA. (2012). Diversity, abundance and expression of nitrite reductase (*nirK*)-like genes in marine thaumarchaea. *ISME J* **6**: 1966–1977.

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.

Nicol GW, Leininger S, Schleper C. (2011). Distribution and activity of ammonia-oxidizing Archaea in natural environments. In: Ward BB, Arp DJ, Klotz MJ (eds) *Nitrification*. ASM Press: Washington, DC, USA, pp 157–178.

Ouverney CC, Fuhrman JA. (2000). Marine planktonic archaea take up amino acids. *Appl Environ Microbiol* **66**: 4829–4833.

Pitcher A, Wuchter C, Siedenberg K, Schouten S, Sinninghe-Damste JS. (2011). Crenarchaeol tracks winter blooms of ammonia-oxidizing Thaumarchaeota in the coastal North Sea. *Limnol Oceanogr* **56**: 2308–2318.

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 edn. Kowalchuk GA, de Bruijn FJ, Head IM, Akkermans ADL, van Elsas JD (eds) Springer Verlag: Dordrecht, The Netherlands, 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* **24**: doi:10.3791/1086.

Prosser JI, Nicol GW. (2008). Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. *Environ Microbiol* **10**: 2931–2941.

Robidart JC, Preston CM, Paerl RW, Turk KA, Mosier AC, Francis CA *et al.* (2011). Seasonal Synechococcus and Thaumarchaeal population dynamics examined with high resolution with remote in situ instrumentation. *ISME J* **6**: 513–523.

Schleper C, Nicol GW. (2010). Ammonia-oxidising Archaea—genomes, physiology and ecology. *Adv Microb Physiol* **57**: 1–41.

Simon JL. (1992). *Resampling: The New Statistics*. Resampling Stats: Arlington VA, USA.

Spang A, Hatzenpichler R, Brochier-Armanet C, Rattei T, Tischler P, Spieck E *et al.* (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends Microbiol* **18**: 331–340.

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.

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.

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* **107**: 8818–8823.

Wang AZ, Cai W-J. (2004). Carbon dioxide degassing and inorganic carbon export from a marsh-dominated estuary (the Duplin River): A marsh CO_2 pump. *Limnol Oceanogr* **49**: 341–354.

Wang AZ, Cai W-J, Wang Y, Ji H. (2005). The southeastern continental shelf of the United States as an atmospheric CO2 source and an exporter of inorganic carbon to the ocean. *Cont Shelf Res* **25**: 1917–1941.

Ward BB. (2011). Nitrification in the Ocean. In: Ward B, Arp DJ, Klotz MJ (eds) *Nitrification*. ASM Press: Washington, DC, USA, pp 325–345.

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 this paper on The ISME Journal website (http://www.nature.com/ismej)