The ISME Journal (2012) 6, 2269–2279

www.nature.com/ismej

ORIGINAL ARTICLE

Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume

Brett J Baker[1], Ryan A Lesniewski[1,4] and Gregory J Dick[1,2,3]
[1]*Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA;*
[2]*Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA and*
[3]*Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA*

Ammonia-oxidizing Archaea (AOA) are among the most abundant microorganisms in the oceans and have crucial roles in biogeochemical cycling of nitrogen and carbon. To better understand AOA inhabiting the deep sea, we obtained community genomic and transcriptomic data from ammonium-rich hydrothermal plumes in the Guaymas Basin (GB) and from surrounding deep waters of the Gulf of California. Among the most abundant and active lineages in the sequence data were marine group I (MGI) Archaea related to the cultured autotrophic ammonia-oxidizer, *Nitrosopumilus maritimus*. Assembly of MGI genomic fragments yielded 2.9 Mb of sequence containing seven 16S rRNA genes (95.4–98.4% similar to *N. maritimus*), including two near-complete genomes and several lower-abundance variants. Equal copy numbers of MGI 16S rRNA genes and ammonia monooxygenase genes and transcription of ammonia oxidation genes indicates that all of these genotypes actively oxidize ammonia. *De novo* genomic assembly revealed the functional potential of MGI populations and enhanced interpretation of metatranscriptomic data. Physiological distinction from *N. maritimus* is evident in the transcription of novel genes, including genes for urea utilization, suggesting an alternative source of ammonia. We were also able to determine which genotypes are most active in the plume. Transcripts involved in nitrification were more prominent in the plume and were among the most abundant transcripts in the community. These unique data sets reveal populations of deep-sea AOA thriving in the ammonium-rich GB that are related to surface types, but with key genomic and physiological differences.
The ISME Journal (2012) **6,** 2269–2279; doi:10.1038/ismej.2012.64; published online 14 June 2012
Subject Category: geomicrobiology and microbial contributions to geochemical cycles
Keywords: Archaea; ammonia; nitrification; metatranscriptomics; metagenomics

Introduction

Marine group I (MGI) Archaea are a diverse group of Archaea that are ubiquitous in marine environments and are thought to have a significant role in global nitrification (DeLong, 1992; Fuhrman *et al.*, 1992; DeLong *et al.*, 1994; Francis *et al.*, 2005; Wuchter *et al.*, 2006; Kalanetra *et al.*, 2009). Originally classified as Crenarchaeota, recent phylogenetic analysis suggests that the MGI are part of the distinct and deeply branching phylum Thaumarchaeota (Brochier-Armanet *et al.*, 2008; Pester *et al.*, 2011). These Archaea are particularly abundant in the deep, dark ocean (Church *et al.*, 2010), where they account for up to 40% of microbial communities (Karner *et al.*, 2001). Despite their abundance and biogeochemical importance, fundamental questions remain regarding the physiology and metabolism of MGI. Several studies of MGI have provided evidence for both autotrophic ammonia oxidation (Konneke *et al.*, 2005; Ingalls *et al.*, 2006) and heterotrophy (Ouverney and Fuhrman, 2000; Agogué *et al.*, 2008; Mußmann *et al.*, 2011; Tourna *et al.*, 2011). Autotrophic ammonia oxidation has now been confirmed in a few cultured representatives (De La Torre *et al.*, 2008; Tourna *et al.*, 2011), including *Nitrosopumilus maritimus* (Konneke *et al.*, 2005). Physiological characterization of *N. maritimus* showed that it has a high affinity for ammonia, providing a mechanism of niche differentiation with ammonia-oxidizing bacteria (AOB) (Martens-Habbena *et al.*, 2009), which are active in soils and other environments with higher ammonium concentration (Verhamme *et al.*, 2011).

Correspondence: GJ Dick, Department of Earth and Environmental Sciences, University of Michigan, 2534 CC Little Building, 1100 N. University Avenue, Ann Arbor, MI 48109, USA.
E-mail: gdick@umich.edu
[4]Present address: Department of Marine Biology and Biological Oceanography, University of Southern California, 3616 Trousdale Parkway, Los Angeles, CA 90089, USA.
Received 16 December 2011; revised 1 May 2012; accepted 11 May 2012; published online 14 June 2012

Given the difficulty in culturing MGI, only two genomes have been fully sequenced. Both come from shallow waters, the sponge symbiont *Cenarchaeum symbiosum* (Hallam *et al.*, 2006) and the aquarium isolate *N. maritimus* (Walker *et al.*, 2010). Recently, draft genome sequence has been obtained from single cells and San Francisco Bay sediment enrichments of *Nitrosoarchaeum limnia*, recovered from an estuary in San Francisco bay (Blainey *et al.*, 2011) and from a soil isolate, *Nitrososphaera viennensis* (Tourna *et al.*, 2011). Characterization of these genomes suggests they use a modified 3-hydroxypropionate/4-hydroxybutyrate pathway for carbon fixation, and have a copper-dependent system for ammonia oxidation and electron-transfer that is distinct from AOB. Additionally, comparison of these genomes with marine metagenomic data sets revealed widespread conservation of gene content, highlighting the ubiquity of these oligophiles throughout the world. A recent genomic characterization of communities of MGI Archaea from surface waters in the Gulf of Maine revealed that *N. maritimus* has several genomic islands that are not present in marine populations (Tully *et al.*, 2012).

Although MGI are particularly abundant in the deep oceans (Karner *et al.*, 2001), these deep populations are not well studied compared with those from shallower depths. A recent PCR-based study found that deep waters (>1000 m depth) of the North Atlantic have lower ratios of MGI *amoA* to 16S rRNA gene copies than subsurface waters, suggesting that most deep-sea MGI are heterotrophic (Agogué *et al.*, 2008). However, metagenomic sequencing of North Pacific waters at 4000 m depth revealed an equal ratio of MGI *amoA* to 16S rRNA genes (Konstantidis *et al.*, 2009). Furthermore, it has recently been shown that expression of ammonia monoxygenase does not always signify autotrophy (Muβmann *et al.*, 2011). In this study, we utilize deep-sea hydrothermal vent plumes in the Guaymas Basin (GB) of the Gulf of California as natural laboratories in which to study ecological and physiological responses of deep-sea MGI to ammonium inputs. Sedimented hydrothermal systems such as the GB are enriched in ammonium due to interactions of hydrothermal fluids with organic-rich sediments as they ascend *en route* to the water column (Von Damm *et al.*, 1985). As a result, ammonium concentrations in the GB end-member fluids (10.3–15.6 mM) (Von Damm *et al.*, 1985) are considerably higher than unsedimented ridge discharge fluids (<0.01 mM; Lilley *et al.*, 1993). These hydrothermal inputs contribute to ammonium concentrations of up to 3 μM in GB deep waters (1800–2000 m depth; Lam, 2004). Gene-based surveys have shown that the MGI dominate the GB plume archaeal community (Dick and Tebo, 2010, Lesniewski *et al.*, 2012), and that MGI are more abundant in plumes than background seawater in the deep Indian Ocean and Okinawa Trough (Takai *et al.*, 2004).

Here we use community genomics and transcriptomics to survey the genomic diversity and activity of MGI populations in ammonium-enriched GB plumes compared with surrounding background waters. Community genomics and transcriptomics have proved to be valuable in understanding ecology of microbial communities (Hallam *et al.*, 2006; Frias-Lopez *et al.*, 2008; Shi *et al.*, 2009; Baker *et al.*, 2010). To date, metatranscriptomic studies have relied almost entirely on comparisons with public genomic databases (Frias-Lopez *et al.*, 2008; Shi *et al.*, 2009; Stewart *et al.*, 2011), isolate genomes (Hollibaugh *et al.*, 2011) and unassembled DNA sequence (Shi *et al.*, 2011). Instead, we utilized *de novo* genomic assembly of community DNA to evaluate the genomic diversity of MGI and provide a framework for recruitment of transcripts to closely relate the gene variants from plume and background waters. These analyses provide a unique glimpse into deep-sea MGI genomic diversity and suggest that a cluster of closely related Archaea dominate nitrification in the deep waters of the Gulf of California.

Materials and methods

Sample collection and processing

Samples were obtained by CTD Rosette from the GB and Carmen Basin on three cruises aboard the *R/V New Horizon* in 2004 and 2005. Once on deck, plume and background waters were immediately filtered by N_2 gas pressure onto 0.2-μm pore size, 142-mm diameter polycarbonate filters, and fixed and frozen in RNAlater as previously described (Dick *et al.*, 2009b; Dick and Tebo, 2010). Further details of sample processing, locations and environmental conditions are provided in Supplementary Table S1 and in Lesniewski *et al.* (2012). Plume-1 and Plume-2 were used for genomics, whereas Plume-3 and Plume-4 were transcriptomics samples from the plume. Two background samples were each used for both metagenomics and metatranscriptomics. As it is not possible to obtain true background samples from sub-sill depths of the GB, Background-1 was taken from just above the GB plume and Background-2 was from the next basin south of Guaymas, Carmen Basin (Lesniewski *et al.*, 2012).

RNA was isolated using a modification of the mirVana miRNA Isolation kit (Ambion, Grand Island, NY, USA) as described previously (Hollibaugh *et al.*, 2011; Stewart *et al.*, 2011). The RNA was then purified and concentrated using the RNeasy MinElute Cleanup kit (Qiagen, Valencia, CA, USA). cDNA synthesis was conducted as described previously (Hollibaugh *et al.*, 2011). Genomic DNA and cDNA libraries were prepared for sequencing using standard protocols (454 Life Sciences, Roche, Branford, CT, USA) and randomly shotgun sequenced by 454 Titanium

pyrosequencing. All of the cDNA reads presented here are available in the NCBI Sequence Read Archive under accession number SRA045655. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AJXC00000000. The version described in this paper is the first version, AJXC01000000. Gene annotations are available at IMG under the accessison 2263196145.

Genomic analyses

Genomic reads were assembled using MIRA 3 (http://chevreux.org/projects_mira.html), and resulting contigs were manually checked with consed (Gordon *et al.*, 1998) and annotated using the JGI IMG/MER system (Markowitz *et al.*, 2009). Initial binning of the assembled fragments was done using tetra-nucleotide frequencies signatures and ESOM mapping as detailed in Dick *et al.* (2009a). As this binning only obtained fragments larger than 2.5 kb, we also searched the entire assembly for additional fragments using reciprocal BLAST searches with the *N. maritimus* genome. GB fragments were then checked for synteny by manually comparing the gene order with that of *N. maritimus*. Fragments were chosen to be added to the bin manually based on synteny and sequence-match qualities. All phylogenetic trees were generated using maximum likelihood method within ARB software package (Ludwig *et al.*, 2004).

cDNA analyses

Transcript reads were mapped to predicted proteins using BLASTX (cutoff of bit score >45 and $>70\%$ similarity). Previous transcriptome studies have relied on publically available databases to recruit cDNA reads using bit score of >40–45 (Frias-Lopez *et al.*, 2008; Shi *et al.*, 2009; Gifford *et al.*, 2011; Hollibaugh *et al.*, 2011). We used a bit score cutoff of >40 to assign mRNA reads to the *de novo* assembled MGI genomes, resulting in recruitment of a total of 10 747 reads (Supplementary Figure S4). Of these, 4382 hits have less than 70% sequence similarity to the MGIC proteins. Comparison of these transcripts with Genbank revealed that most are not archaeal. In fact, only 13% had top hits to Archaea and just 10% of those matched MGIC (bit score >40). Furthermore, recruitment of reads originally identified as archaeal (bit score >40, % ID <70) to the entire GB genomic assembly revealed that the majority of them (72%) are more accurately assigned to other members of the community (Supplementary Figure S5). In contrast, 93% of the cDNA reads recruited at $>70\%$ sequence similarity had top hits to Archaea in NCBI, and 88% of those had top hits to MGI. Comparisons of recruitment patterns between the samples within the plume and the two backgrounds did not reveal significant differences in transcription profiles (data not shown); therefore, we pooled the two plume and two background samples for analyses.

To confirm the absence of AOB, we searched the transcript libraries using *amoA* from *Nitrosomonas marinus* and *Nitrosococcus oceani* with *e*-value cutoffs of 1×10^{-10}. Some genes like ammonia monooxygenases are highly conserved, for example there is 97–99% similarity (at the protein level) between *amoA* genes in the community. Therefore, to accurately differentiate expression among variants, we found it necessary to recruit reads at the DNA level. For normalized comparison analyses, the total number of mapped transcripts was divided by the length of the gene and the total number of transcripts from each sample for comparison between them. A total of 1 651 287 (696 718 from plume-1 and 954 569 from plume-2) and 1 117 284 (570 580 from background-1 and 546 704 from background-2) reads were recruited from the plume and background libraries, respectively. The circular diagram for comparative genomics and transcriptomics was generated using Circos (Krzywinski *et al.*, 2009). Suspected replicate reads due to artifacts of 454 pyrosequencing were manually removed from ammonia monooxygenase gene analyses and from all DNA read coverage-based analyses.

Results and discussion

Community genomics and transcriptomics reveals multiple populations of MGI Archaea in the deep Gulf of California

Plume samples yielded more RNA, as well as more DNA and cDNA reads, than the background samples (Supplementary Table S1). *De novo* genomic assembly and binning by tetranucleotide frequency and emergent self-organizing maps (Dick *et al.*, 2009a) revealed a well-defined MGI bin that contained 449 DNA sequence fragments with length greater than 2.5 kb, totaling 1.79 Mb of consensus sequence (Supplementary Figure S1). On the basis of BLAST searches of the whole community versus the *N. maritimus* genome, we identified 790 additional fragments belonging to MGI, bringing the total length of assembled fragments identified as MGI to 2.9 Mb. The average GC content of the bin was 31%, similar to that of *N. maritimus* (34%). The average fragment size was 2.3 kb, and there were 80 sequences (of 1239 total) longer than 5 kb.

Seven different MGI 16S rRNA genes were identified. Phylogenetic analyses placed them all in group I.1a (Supplementary Figure S2). 16S rRNA sequence similarity to *N. maritimus* ranged from 95.4–98.4% (pair-wise gene aligned), indicating that these GB populations likely represent distinct species of the *Nitrosopumilus* genus. To estimate the abundance and overall metabolic activity of each phylotype across samples, metagenomic and metatranscriptomic reads were mapped to the MGI 16S

rRNA genes (Supplementary Figure S3). MGI 16S rRNA genes consistently recruited more cDNA reads from plumes than background. The balance of DNA reads in background versus plume was more variable, and in some cases, much higher in background than plume (Supplementary Figure S3). In terms of the whole community, these MGI small subunit (SSU) rRNA genes were the most abundantly represented Archaea in both plume and background metatranscriptomic data sets (Lesniewski *et al.*, 2012).

The GB genomic assembly also contained seven different sequence types of MGI *amoA* genes. The average coverage of *amoA* genes was 6× (294 total genomic reads), comparable to the 4× coverage of MGI 16S rRNA genes (390 total genomic reads). Normalization of read numbers by gene length resulted in a roughly equivalent copy number of *amoA* and 16S genes (*amoA*:16S ratio of ~1.4), indicating that the majority of GB MGI cells have *amoA* and are thus capable of ammonia oxidation. These findings are consistent with those of Konstantinidis *et al.* (2009) from 4000 m depth at station ALOHA and distinct from those of Agogué *et al.* (2008), which found smaller *amoA*:16S ratios in deep Atlantic waters.

Comparison of the metagenome and metatranscriptome to N. maritimus

Comparison of the GB MGI metagenome to the genome of *N. maritimus* (Walker *et al.*, 2010) revealed both similarities and differences. 85% of the ORFs in the GB MGI bin (4875 of 5744) are homologous to proteins from *N. maritimus*. These homologs average 78% protein sequence similarity, and appear to stem primarily from four different genotypes, two of which are well-covered in the GB metagenome (Figure 1). The remaining 15% of putative proteins in the MGI bin do not have homology to proteins in *N. maritimus* (*e*-value cutoff 1×10^{-10}). Most of these GB-unique proteins (76%) could not be assigned any putative function. There were also predicted proteins from the *N. maritimus* genome that could not be identified in the GB genomic data (205 of the 1799), the majority of which were annotated as hypothetical proteins. Many of these *N. maritimus*-unique genes clustered in certain regions of the *N. maritimus* genome (Figure 1), which correspond to recently identified genomic islands that are also absent in MGI populations from surface waters of the Gulf of Maine (Tully *et al.*, 2012). This suggests that these regions are unique to the *N. maritimus* genome and that our metagenomic assembly contains near-complete genomes of the MGI populations.

The advent of transcriptomic sequencing of microbial communities is advancing knowledge of the transcriptional activity organisms in the environment (Frias-Lopez *et al.*, 2008; Stewart *et al.*, 2011). However, accurate assignment and phylogenetic placement of transcripts from natural populations of uncultivated microorganisms is hindered by a lack of coverage of genomes present in the environment. We applied a stringent threshold (>70% sequence identity and bit score >45) to recruit 8520 high similarity reads to the GB MGI metagenome (Figure 2). A total of 6363 of these transcripts came from the plume and 2157 from the background samples, with an average of 94% amino acid similarity. Using the same parameters, only 6849 transcripts were mapped to *N. maritimus* genes (Supplementary Figure S4), highlighting the value of genomes assembled directly from the same environment where metatranscriptomic data was collected. Given the considerable diversity of MGI in deep Gulf of California waters (Supplementary Figure S2) and the modest quantity of mRNA transcripts recovered, the metatranscriptomic data presented here likely represents only the most abundantly transcribed genes of MGI populations.

Enhancement of ammonia-oxidizing Archaea in plumes and dominance over AOB

Several recent studies have investigated how the balance of ammonia-oxidizing Archaea (AOA) and AOB varies as a function of ammonium concentration (Martens-Habbena *et al.*, 2009; Verhamme *et al.*, 2011). Hydrothermal inputs into the deep GB lead to ammonium concentrations of 0.2–3 μM in plumes (Lam, 2004), which spans the range proposed to delineate niches of AOA and AOB (Martens-Habbena *et al.*, 2009). We found that transcripts of the MGI genes encoding ammonia monooxygenase (*amoA*) and an ammonium transporter were among the most abundant protein-coding transcripts in the deep GB microbial community (total of 405 and 1713 transcripts, respectively) and were more abundant in plume samples compared with the background (Figure 3). In contrast, no bacterial ammonia monooxygenase genes were identified in any of the GB metagenomic or metatranscriptomic data sets (plume or background). This suggests that ammonia oxidation in the deep Gulf of California, including ammonium-enriched hydrothermal plumes, is dominated by AOA.

Species-resolved transcriptomics of ammonia oxidation genes

Detailed analysis of *amoA* transcripts revealed dynamic transcription patterns of particular AOA populations. The GB metagenome contains 27 contigs that have *amo* genes from at least seven different genotypes (Figure 4). These well-assembled *amo* loci represent the dominant AOA genotypes present in the genomic data. To assess the ammonia-oxidizing transcriptional activity of each of these genotypes in ammonia-rich and ammonia-poor settings, we compared transcript recruitment from plume and background samples with all ammonia monooxygenase genes (*amoA*, *amoB*,

Figure 1 Mapping of GB metagenomic fragments to the *Nitrsospumilus maritimus* genome. The outer-most ring is the complete genome of *N. maritimus* with open-reading frames (ORFs) colored based on clusters of orthologous genes (COG) categories. The black tiles inside are assembled genomic fragments from the Guaymas assembly that map by BLASTn to regions of the *N. maritimus* genome. The outer gray-shaded circle is a histogram showing percent sequence identity of top GB proteins that match *N. maritimus* proteins (scaled from 50% inside to 100% outside). The inner-most gray circle is the raw number of transcripts that map to those homologous proteins; only the range of 0–30 is shown ito highlight regions with little or no transcript recruitment. Genomic islands (>1 kb) missing in archaeal metagenomic data from Gulf of Maine surface waters (Tully *et al.*, 2012) are highlighted with light red wedges. Note that nearly all the gaps in the GB genomic data occur in these genomic islands.

amoC and the *amo*-associated hypothetical) from all genotypes. Transcription of *amoA* genes from three of the abundant GB genotypes (c1374, c45409 and c51705) is dramatically higher in plume compared with background (Figure 4). Interestingly, there are several low-abundance variants that are highly active in the plume (*amoA* and hypothetical from c51705, *amoC* from c113214 and hypothetical from c225589).

The four GB *amoA* variants that are most active in the plume (c1235, c1374, c45409 and c51705) fall within a tight phylogenetic group (Figure 5). Interestingly, the most abundant transcript type (c1374) is most closely related to a clone recovered from

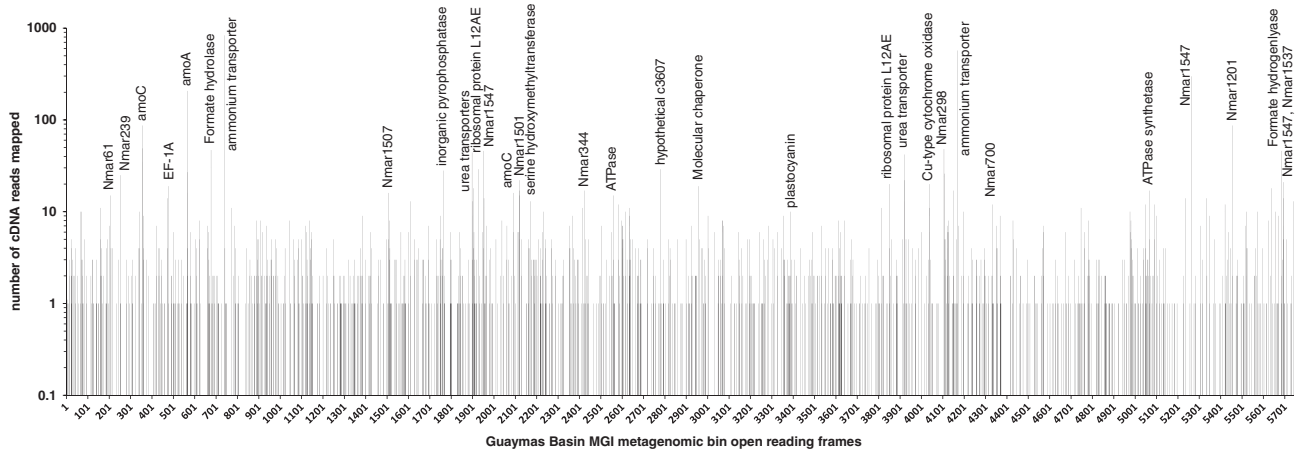

Figure 2 Abundance of raw (not normalized) transcripts mapped to genes in the GB MGI metagenomic bin (5744 total genes). Predicted hypothetical proteins that have matches to *N. maritimus* genes are labeled 'Nmar'.

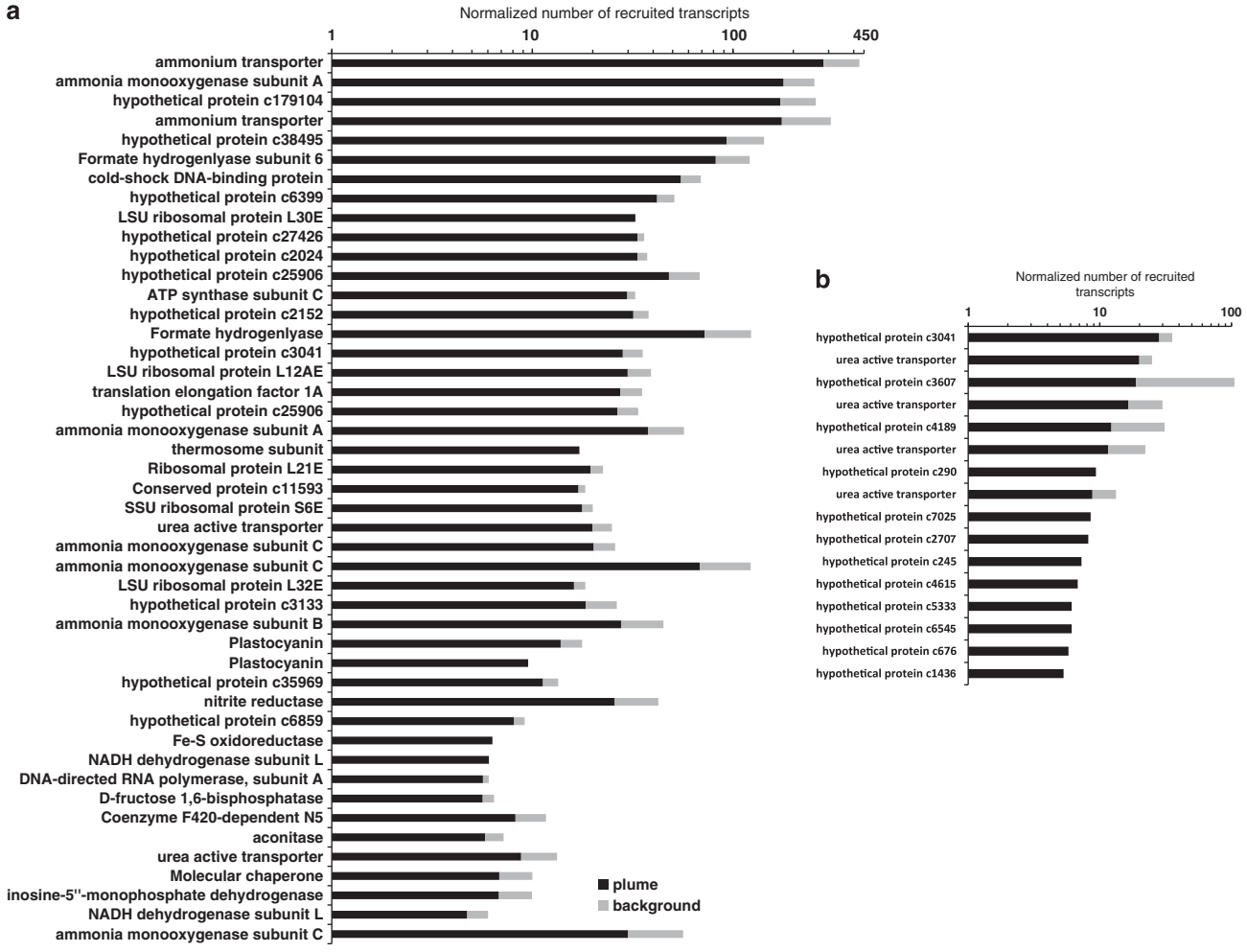

Figure 3 Stacked bar graph showing the number of transcripts recruited to MGI Archaea genes in the plume and background samples, and sorted by difference between the plume and background recruited, with the greatest being at the top. Numbers are normalized to length of the genes as well as the total number of transcripts per sample (raw number transcripts of recruited divided by gene length and library size, then multiplied by a million to it comparable to the raw number of reads). (**a**) shows transcripts that are most abundant in the plume. (**b**) shows transcripts of genes not present in *Nitrosompumilus maritimus* that are most up-regulated in the plume.

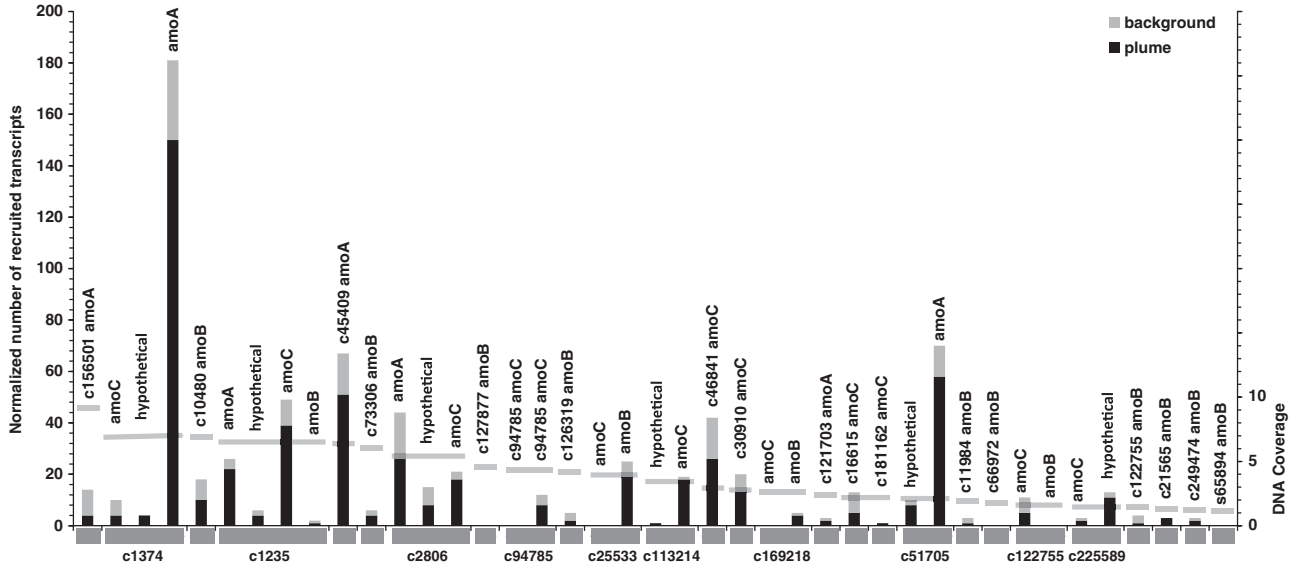

Figure 4 Transcript levels of sequence variants of ammonia monooxygenase genes in plume and background samples. Transcript numbers are normalized to gene length and library size. DNA fragments (contigs) with more than one gene are designated with gray bars on the *x* axis. Individual genes are labeled on top. Thin gray horizontal lines indicate contig coverage in the genomic libraries (see scale on the right).

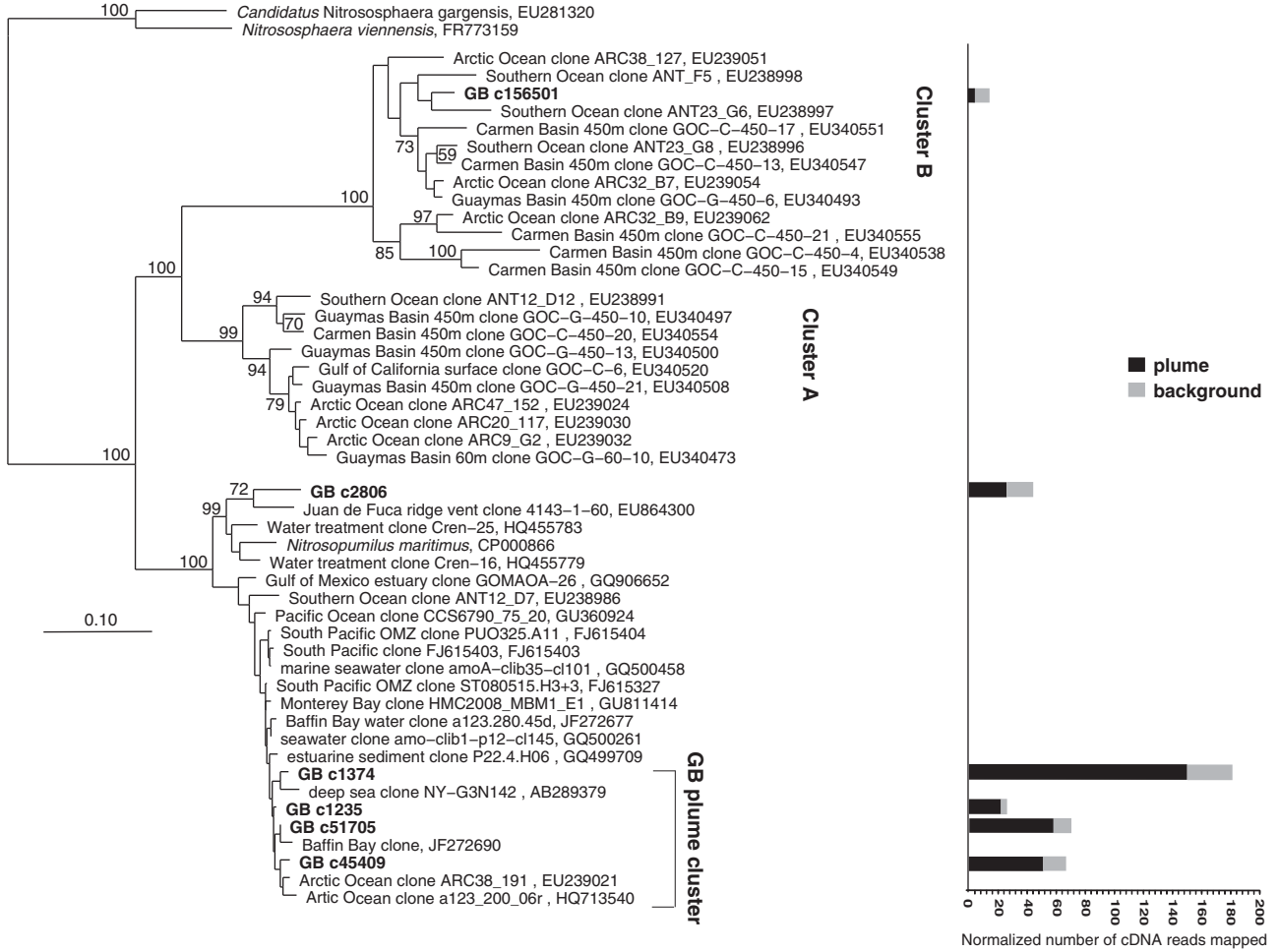

Figure 5 Phylogenetic tree of ammonia monooxygenase (*amoA*) genes and abundance of their transcripts in plume and background datasets. Recruitment numbers were normalized to gene lengths and the library sizes. Notice the Gulf of California sequences recovered from the surface waters (Beman *et al.*, 2008) are not related to the types found in this study.

deep waters (2956 m) of the Japan Sea (Nakagawa *et al*., 2007). Furthermore, these deep-sea genotypes are distinct from those that have been recovered from the upper 650 m of the water column at GB and Carmen Basins (Beman *et al*., 2008). These GB gene sequences are >97% similar to one another and 92.3 to 93.4% similar to *N. maritimus*. Thus, the genotypes that dominate *amoA* transcription in the deep GB likely represent strains of a novel species of *Nitrosopumilus*, a notion that is supported by sequence similarity and phylogeny of the dominant 16S rRNA genes (Supplementary Figure S2). Our data suggests that expression of *amoA* genes from this deep GB group is enhanced in ammonium-rich hydrothermal plumes of the GB. Several other *amoA* sequences in this phylogenetic cluster were recovered from a site in the Arctic Ocean that has high ammonium concentrations (Kalanetra *et al*., 2009). Taken together, this evidence reveals a cluster of MGI that thrives in geographically widespread ammonium-rich marine environments.

Genomic insights into the carbon metabolism of GB MGI

Given their abundance in the oceans and potential role in the carbon cycle, defining the carbon metabolism of MGI is an important yet unfinished task. Conflicting results leave open the question of whether individual MGI are capable of both heterotrophy and autotrophy or there are sub-groups that specialize in each. Whereas studies of cultures and surface waters indicate autotrophy, observations of a lower ratio of MGI *amoA* to 16S rRNA gene copies (Agogué *et al*., 2008) and decreasing MGI carbon fixation with depth (Varela *et al*., 2011) in the Atlantic suggest that deep-sea MGI are predominantly organoheterotrophic. Recent studies also show that the presence (and expression) of *amoA* genes does not necessarily indicate CO_2 fixation (Muβmann *et al*., 2011; Tourna *et al*., 2011). The GB metagenome contains genes homologous to *N. maritimus* genes encoding the 3-hydroxypropionate/4-hydroxybutyrate pathway for CO_2 fixation (Berg *et al*., 2007), including 4-hydroxybutyryl-CoA dehydratase, methylmalonyl-CoA epimerase and mutase, and acetyl-CoA carboxylase. Genes for the acetyl-CoA carboxylase recruited 25 transcripts (16 from plume and 9 from background), as did genes for methylmalonyl-CoA epimerase and mutase (19 plume and 6 background). Representation of these autotrophy genes in the metatranscriptomic data supports the idea that MGI fix CO_2 in the deep Gulf of California.

The GB MGI genomes contain 57 predicted ABC-type transporters for uptake of amino acids, which might be an important source of carbon, nitrogen and energy for marine heterotrophs, including MGI (Fuhrman, 1987; Suttle *et al*., 1991; Ouverney and Fuhrman, 2000). Although this indicates genomic potential for heterotrophy in the GB MGI, these transporters recruited few or no transcripts ($\leqslant 2$), suggesting that transcription of genes encoding MGI amino acid transporters was lower than those for carbon fixation.

Nitrogen and energy metabolism of GB MGI

Although MGI show high affinity for ammonium (Martens-Habbena *et al*., 2009), low ammonium concentration still presents a potential bottleneck for energy metabolism of MGI. The GB MGI show evidence of several strategies for ammonia acquisition. First, genes encoding ammonium transporters are the most abundant protein-coding transcripts in the MGI metatranscriptome (Figure 2). Such high transcription of MGI ammonium transporters is consistent with prior observations from surface waters (Hollibaugh *et al*., 2011; Stewart *et al*., 2011) and likely reflects the much higher concentration of ammonium than ammonia at seawater pH. The fact that ammonium transporters are the most highly expressed protein coding gene of deep GB MGI suggests that ammonium must be first transported into the cell for oxidation to occur. Regardless, this gene is clearly critical to the MGI's success in the community, and may account for their high N affinity (Martens-Habbena *et al*., 2009).

Second, the deep-sea GB MGI metagenome contains three operons of *ure* genes for urea utilization. One genomic fragment (c229) has *ureE*, *ureF*, *ureG* and *ureH* genes, and another (c464) has *ureB*, *ureG*, *ureE* and urease-associated metallopeptidase genes. Two additional fragments contain urea active transporters and one of these has a second urease-associated metallopeptidase gene. It has been recently shown that the soil AOA isolate, *N. viennensis*, is capable of growth on urea (Tourna *et al*., 2011). Both *C. symbiosum* and *N. viennensis* contain urease genes (Hallum *et al*., 2006; Tourna *et al*., 2011); however, *N. maritimus* lacks any recognizable genes for urea utilization. Thus, our results and other recent environmental studies (Konstantinidis *et al*., 2009; Yakimov *et al*., 2011; Tully *et al*., 2012) highlight an important difference in N acquisition between natural populations of MGI and *N. maritimus*.

All genes for the proposed AOA respiratory pathway (Walker *et al*., 2010) are present in the GB genomic data, except the plastocyanin-like subunit of complex III. Genes present include those encoding NADH dehydrogenase (NuoABCDHIJKMLN), ATP F_oF_1-type synthetase, complex III, multicopper oxidases and the terminal oxidase (complex IV). Many of the respiratory pathway genes have multiple variants (up to seven) in the GB, but in nearly every instance, one specific genotype recruited the majority of transcripts (see Supplementary Table S2 for complete list).

Nitrite, the product of ammonia oxidation, inhibits growth of AOA (Tourna *et al*., 2011). However, a recent study suggests that AOA reduce nitrite

through a pathway known as 'nitrifier-denitrification', resulting in globally significant production of nitrous oxide (N_2O), an important greenhouse gas (Santoro *et al.*, 2011). Although culture-based studies of MGI physiology have not demonstrated nitrite reduction, genes with homology to nitrite reductase (*nirK*) and several cupredoxin domain-containing multicopper oxidases thought to be involved in nitrite reduction were identified in the *N. maritimus* genome (Walker *et al.*, 2010). We identified single copies of *nirK*-like genes Nmar_1259 and Nmar_1667 in the GB genomic data. The Nmar_1259 *nirK* homolog (c632) is well-represented in both plumes (61 transcripts) and background (35 transcripts), whereas only a few transcripts of the Nmar_1667 homolog were detected. Nearly all the *nirK*-associated multicopper oxidases are also present in the GB MGI bin (Nmar_1354 was not found), but they are not expressed at significant levels. A potential source of electrons for nitrite reduction is formate (Ruiz-Herrera and DeMoss, 1969), which is likely present in the GB plume. Some of the most abundant MGI transcripts that are highly enriched in the plume come from two variants of formate dehydrogenase (c1456 and c85331) that are highly similar (100% and 97%) to this protein from *N. maritimus* (Figure 3). Taken together, the evolutionary conservation and abundant transcriptional activity of this formate hydrogenase suggests that it serves a critical role in the GB MGI. The overall magnitude and extensive enrichment of transcripts of formate dehydrogenase and nitrite reductase genes that we observe in the GB plume implies that AOA actively reduce nitrite in these deep waters.

Conclusions

It is becoming increasingly apparent that MGI are widespread and globally significant factors in the nitrogen and carbon cycles, yet the extent and implications of their influence are unclear because of questions surrounding their physiology and ecology. This is especially true for deep-sea MGI, which are numerically dominant, but not well studied. In this study, *de novo* assembly of community genomic sequence provided a framework for investigating the activity of naturally occurring populations of MGI in the Gulf of California. This approach proved to be especially useful for differentiating transcriptional activity among closely related genotypes. Additionally, it provided a catalog of genes not present in reference genomes, including those for urea utilization and many hypothetical genes.

Our findings show that the dominant Archaea in the deep Gulf of California are ammonia oxidizers. Archaeal genes for ammonia oxidation are among the most highly transcribed protein-coding genes in microbial communities inhabiting ammonium-enriched GB deep-sea hydrothermal plumes, suggesting vigorous MGI-mediated nitrification. This is surprising in light of the prevailing view that bacteria tend to dominate at higher ammonium concentrations. Instead, we found a dominant clade of deep-sea AOA that thrive under ammonium-rich conditions, perhaps indicating that the marine AOA niche has a broader range of ammonia concentration than previously recognized. This group is closely related to *N. maritimus*, sharing with it the ability to oxidize ammonia and fix carbon, but is also characterized by genomic novelty reflecting important physiological differences such as acquisition of nitrogen via urea. These insights highlight populations of MGI Archaea in the deep Gulf of California that are distinct from those in surface waters and deep Atlantic waters, and that respond to geochemical perturbation in the plume environment.

Acknowledgements

This project is funded in part by the Gordon and Betty Moore Foundation and the National Science Foundation (OCE 1029242). We also thank the University of Michigan Rackham Graduate School Faculty Research Fellowship Program for their support and Kathryn Iverson and Sunit Jain for assistance with metagenomic analyses. DNA and cDNA sequencing was conducted in the laboratory of Stephan Schuster at Penn State University, and we are grateful for the assistance of Lynn Tomsho. We also thank Dr Cody Sheik and Karthik Anantharaman for their helpful comments and revisions to the manuscript.

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**: 788–792.

Baker BJ, Comolli LR, Dick GJ, Hauser LJ, Hyatt D, Dill BD *et al.* (2010). Enigmatic, ultra-small uncultivated Archaea. *Proc Natl Acad Sci* **107**: 8806–8811.

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.

Blainey PC, Mosier AC, Potanina A, Francis CA, Quake SR. (2011). Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analyses. *PLoS One* **6**: 1–12.

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

Church MJ, Wai B, Karl DM, DeLong EF. (2010). Abundances of crenarchaeal amoA genes and transcripts in the Pacific Ocean. *Environ Microbiol* **12**: 679–688.

De La Torre J, Walker CB, Ingalls AE, Konneke M, Stahl DA. (2008). Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol* **10**: 810–818.

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

DeLong EF, Wu KY, Prézelin BB, Jovine RVM. (1994). High abundance of Archaea in Antarctic marine picoplankton. *Nature* **371**: 695–697.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC *et al.* (2009a). Community-wide analysis of microbial genome sequence signatures. *Genome Biology* **10**: R85.

Dick GJ, Clement BG, Webb SM, Fodrie FJ, Bargar JR, Tebo BM. (2009b). Enzymatic microbial Mn(II) oxidation and Mn biooxide production in the Guymas Basin deep-sea hydrothermal plume. *Geochim et Cosmo Acta* **73**: 6517–6530.

Dick GJ, Tebo BM. (2010). Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. *Environ Microbiol* **12**: 1334–1347.

Francis CA, Roberts KF, 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.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al.* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci* **105**: 3805–3810.

Fuhrman JA. (1987). Close coupling between release and uptake of dissolved free amino acids in seawater studied by an isotope dilution approach. *Mar Ecol* **37**: 45–52.

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

Gordon D, Abajian C, Green P. (1998). Consed: a graphical tool for sequence finishing. *Genome Res* **8**: 186–194.

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

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

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

Ingalls AE, Shah SR, Hansman RL, Aluwihare LI, Santos GM, Drufel ERM *et al.* (2006). Quantifying archaeal community autotrophy in the mesopelagic using natural radiocarbon. *Proc Natl Acad Sci* **103**: 6442–6447.

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.

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

Konstantinidis KT, Braff J, Karl DM, DeLong EF. (2009). Comparative metagenomic analysis of a microbial community residing at a depth of 4000 meters at station ALOHA in the North Pacific Subtropical Gyre. *Appl Environ Microbiol* **75**: 5345–5355.

Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D *et al.* (2009). Circos: an information aesthetic for comparative genomics. *Genomic Res* **19**: 1639–1645.

Lam P. (2004). *Microbial ammonia oxidation in deep-sea hydrothermal plumes.* PhD Thesis, University of Hawaii: Hawaii, USA.

Lesniewski R, Jain S, Anantharaman K, Schloss PD, Dick GJ. (2012). The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J;* e-pub ahead of print; doi:10.1038/ismej.2012.63.

Lilley MD, Betterfield DA, Olson EJ, Lupton JE, Macko SA, McDuff RE. (1993). Anomalous CH_4 and NH_4^+ concentrations at an unsedimented mid-ocean-ridge hydrothermal system. *Nature* **364**: 45–47.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yakhukumar *et al.* (2004). ARB: a software environment for sequence data. *Nucl Acids Res* **32**: 1363–1371.

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**: 976–981.

Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. (2009). IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25**: 2271–2278.

Muβmann M, Brito I, Pitcher A, Sinninghe Damste JS, Hatzenpichler R, Richter A *et al.* (2011). Thaumarchaeotes abundant in refinery nitrifying sludges express *amoA* but are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci* in press.

Nakagawa T, Mori K, Kato C, Takahashi R, Tokuyama T. (2007). Distribution of cold-adapted ammonia-oxidizing microorganisms in the deep-ocean of the northestern Japan Sea. *Microbes Environ* **4**: 365–372.

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

Pester M, Schleper C, Wagner M. (2011). The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol* **14**: 300–306.

Ruiz-Herrera J, DeMoss JA. (1969). Nitrate reductase complex of *Escherichia coli* K-12: participation of specific formate dehydrogenase and cytochrome b1 components in nitrate reduction. *J Bacteriol* **99**: 720–729.

Santoro AE, Buchwald C, McIlvin MR, Cascitti KL. (2011). Isotopic signature of N_2O produced by marine ammonia-oxidizing Archaea. *Science* **333**: 1282–1285.

Suttle CA, Chan AM, Fuhrman JA. (1991). Dissolved free amino acids in the Sargasso Sea: uptake and repiration rates, turnover times, and concentrations. *Mar Ecol Prog Ser* **70**: 189–199.

Shi YM, Tyson GW, DeLong EF. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459**: 266–269.

Shi Y, Tyson GW, Eppley JM, DeLong EF. (2011). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J* **5**: 999–1013.

Stewart FJ, Ulloa O, DeLong EF. (2011). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ Microbiol* **14**: 23–40.

Takai K, Oidam H, Suzuki Y, Hirayama H, Nakagawa S, Nunoura T *et al.* (2004). Spatial distribution of marine Crenarchaeota group I in the vicinity of deep-sea

hydrothermal systems. *Appl Environ Microbiol* **70**: 2404–2413.

Tourna M, Stieglmeier M, Spang A, Konneke M, Schintlmeister A, Urich T *et al.* (2011). *Nitrosophaera viennensis*, an ammonia oxidizing archaeon from soil. *Proc Natl Acad Sci USA* **108**: 8420–8425.

Tully BJ, Nelson WC, Heidelberg JF. (2012). Metagenomic analysis of a complex marine planktonic thaumarchaea community from the Gulf of Maine. *Environ Microbiol* **14**: 254–267.

Varela MM, van Aken HM, Sintes E, Reinthaler T, Herndl GJ. (2011). Contribution of Crenarchaeota and Bacteria to autotrophy in the North Atlantic interior. *Environ Microbiol* **13**: 1524–1533.

Verhamme DT, Prosser JI, Nicol GW. (2011). Ammonia concentration determines differential growth of ammonia-oxidising archaea and bacteria in soil microcosms. *ISME J* **5**: 1067–1071.

Von Damm KL, Edmond JM, Measures CI, Grant B. (1985). [Chemistry of submarine hydrothermal solutions at Guaymas Basin, Gulf of California]. *Geochim et Cosmochim Acta* **49**: 2221–2237.

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.

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* **103**: 12317–12322.

Yakimov MM, La Cono V, Smedile F, DeLuca TH, Juarez S, Ciordia S *et al.* (2011). Contribution of crenarchaeal autotrophic ammonia oxidizers to the dark primary production of Tyrrhanian deep waters (Central Mediterranean Sea). *ISME J* **5**: 945–961.

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