The putative functional ecology and distribution of archaeal communities in sponges, sediment and seawater in a coral reef environment

Ana R. M. Polónia, Daniel F. R. Cleary, Rossana Freitas, Nicole J. de Voogd, Newton C. M. Gomes

MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 409–423 doi: 10.1111/mec.13024

The putative functional ecology and distribution of archaeal communities in sponges, sediment and seawater in a coral reef environment

ANA R. M. POLÓNIA,* DANIEL F. R. CLEARY,* ROSSANA FREITAS,* NICOLE J. DE VOOGD† and NEWTON C. M. GOMES*

*Department of Biology, CESAM, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal, †Naturalis Biodiversity Center, PO Box 9517, 2300 RA Leiden, The Netherlands

Abstract

Archaea play crucial roles in a number of key ecological processes including nitrification and methanogenesis. Although several studies have been conducted on these organisms, the roles and dynamics of coral reef archaeal communities are still poorly understood, particularly in host and nonhost biotopes and in high (HMA) and low microbial abundance (LMA) sponges. Here, archaeal communities detected in six distinct biotopes, namely, sediment, seawater and four different sponge species *Stylissa carteri*, *Stylissa massa*, *Xestospongia testudinaria* and *Hyrtios erectus* from the Spermonde Archipelago, SW Sulawesi, Indonesia were investigated using 454-pyrosequencing of 16S rRNA genes (OTU cut-off 97%). Archaeal communities from sediment and sponges were dominated by Crenarchaeota, while the seawater community was dominated by Euryarchaeota. The biotope explained almost 75% of the variation in archaeal composition, with clear separation between microbial assemblages from sediment, *X. testudinaria* and *H. erectus* (HMA). In contrast, samples from seawater and both *Stylissa* species (LMA) showed considerable overlap in the ordination and, furthermore, shared most abundant OTUs with the exception of a single dominant OTU specifically enriched in both *Stylissa* species. Predicted functional gene content in archaeal assemblages also revealed significant differences among biotopes. Different ammonia assimilation strategies were exhibited by the archaeal communities: *X. testudinaria*, *H. erectus* and sediment archaeal communities were enriched for glutamate dehydrogenase with mixed specificity ($NAD(P)^+$) pathways, while archaeal planktonic communities were enriched for specific glutamate dehydrogenase ($NADP^+$) and glutamate synthase pathways. Archaeal communities in *Stylissa* had intermediate levels of enrichment. Our results indicate that archaeal communities in different biotopes have distinct ecophysiological roles.

Keywords: archaea, coral reef, glutamate, nitrogen, sponge metabolomics

Received 20 April 2014; revision received 23 November 2014; accepted 26 November 2014

Introduction

The Archaea domain consists of five phyla: Crenarchaeota, Thaumarchaeota, Euryarchaeota, Korarchaeota and Nanoarchaeota, and several new phyla have been proposed [Diapherotrites (pMC2A384), Parvarchaeota, Aenigmarchaeota (DSEG), Nanohaloarchaeota, Aigarchaeota; Rinke *et al*. 2013]. Although originally believed to be restricted to extreme environments, the domain is now known to be present in a wide range of ecosystems under varying environmental conditions (Hoppert *et al*. 2013). In tropical marine environments, mesophilic Crenarchaeota (Thaumarchaeota) and Euryarchaeota are the most frequently found phyla (Pires *et al*. 2012; Polónia *et al*. 2013; Yin *et al*. 2013). The recently described Thaumarchaeota phylum is the most ubiquitous (Offre *et al*. 2013) and can be abundant in aer-

Correspondence: Ana R. M. Polónia, Fax: (+351) 234 370004; E-mails: ritapolonia@ua.pt or ritapolonia@gmail.com

obic terrestrial and marine environments (soil, sediment, seawater, hot springs, hydrothermal vents, marine sponges; Reigstad *et al.* 2008; Wang *et al.* 2009; Tourna *et al.* 2011; Dang *et al.* 2013; Polónia *et al.* 2013), whereas Euryarchaeota are found predominantly in seawater and to a lesser extent in sediment (Wemheuer *et al.* 2012).

In addition to being abundant members of the vast marine microbial community, Archaea are also important players in processes including the geochemical cycling of carbon, nitrogen and sulphur (Offre *et al.* 2013). For oligotrophic coral reefs, this cycling activity, and particularly the nitrogen cycle, is of crucial importance to degrade organic matter and maintain high levels of primary production (Schöttner *et al.* 2011). The importance of nitrogen for organisms and ecosystems is critical; nitrogen is an essential component of proteins, nucleic acids and cell wall constituents and limits marine ecosystem primary productivity (Francis *et al.* 2007). Despite the increasing number of archaeal studies, the geochemical cycling of nitrogen is still less understood in Archaea than in Bacteria. Archaea are involved in nitrification, denitrification and nitrogen fixation (Offre *et al.* 2013). Denitrification, however, has only been detected in halophilic (e.g. *Haloferax denitrificans*) or extreme thermophilic (e.g. *Pyrobaculum aerophilum*) Archaea, very few of which are cultivable (Offre *et al.* 2013). The first observation of nitrification in Archaea was reported by Könneke *et al.* (2005). These authors reported the detection of a chemolithoautotrophic archaeon, which aerobically oxidized ammonia to nitrite (*Nitrosopumilus maritimus*). This finding was unexpected given that until then lithotrophic bacteria were thought to be virtually the only microbes involved in nitrification (Offre *et al.* 2013). Currently, Archaea are thought to play a critical role as ammonia oxidizers. Due to their low tolerance to high NH_3 concentrations, ammonia-oxidizing Archaea (AOA) are believed to be able to out-compete their bacterial counterparts in ocean waters (Radax *et al.* 2012).

Few studies have analysed archaeal composition or putative functions in coral reef biotopes and even less have studied the influence of environmental variables on those factors. In this study, we assessed the composition and function of Archaea in six biotopes including four host (*Stylissa carteri*, *Stylissa massa*, *Xestospongia testudinaria* and *Hyrtios erectus*) and two nonhost (seawater and sediment) biotopes. For the functional analysis, we chose to focus on a single nutrient cycle, the nitrogen cycle, one of the most perturbed nutrient cycles (Offre *et al.* 2013). We, furthermore, compared the composition of Archaea and enrichment of genes involved in the nitrogen metabolism in four biotopes (*S. massa*, *X. testudinaria*, seawater and sediment) from the Spermonde reef to a different coral reef system, the Kepulauan Seribu reef system. Both, the Spermonde and Kepulauan Seribu reefs, suffer from anthropogenic disturbance but to different degrees. The anthropogenic disturbances to which these coral reefs are exposed are related to the population densities of the proximate cities. The Spermonde reef system is situated adjacent to the city of Makassar, a city of more than 2 million inhabitants (Renema 2010) and the Kepulauan Seribu reef system is adjacent to the city of Jakarta, one of the most densely populated conglomerations on earth with more than 12 million inhabitants (Renema 2010). Much of the inshore reefs adjacent to the city of Jakarta have long disappeared or are in a highly degraded state (Cleary *et al.* 2014), whereas inshore reefs proximate to Makassar still have relatively high coral cover (A. R. M. Polónia, D. F. R. Cleary, N. J. de Voogd, W. Renema, B. W. Hoeksema, A. Martins & N. C. M. Gomes, unpublished).

Marine sponges are abundant and ecologically important components of coral reefs (Diaz & Rützler 2001) and have been shown to harbour exceptionally high microbial densities, which can make up from 35% to 40% of sponge biomass (Hentschel *et al.* 2002, 2012; Taylor *et al.* 2007). However, sponge species can differ substantially in the abundance of their microbial symbiont communities (Kamke *et al.* 2010). 'High microbial abundance' (HMA) sponges putatively rely more heavily on their elevated number of micro-organisms to acquire energy than 'low microbial abundance' sponges (LMA), which rely more heavily on their high pumping rates (Weisz *et al.* 2007). The roles of coral reef microbial communities inhabiting these distinct sponges, however, remain unclear, particularly in Archaea. Sponge prokaryotic diversity is, in most cases, dominated by bacterial species (Taylor *et al.* 2007; Fan *et al.* 2012). In some sponge species, however, Archaea are the dominant group. The microbial communities of *Axinella mexicana* and *Inflatella pellicula*, for example, are dominated by Archaea (Preston *et al.* 1996; Jackson *et al.* 2013).

Using 454-pyrosequencing of 16S rRNA genes (OTU cut-off 97%) and predictive functional analysis (PICRUSt), we aimed to assess to what extent: (i) Archaeal communities in HMA and LMA sponges differed compositionally from those in nonhost biotopes (seawater and sediment); (ii) biotopes hosted phylogenetically distinct lineages; (iii) nitrogen metabolic pathways differed among Archaea from different biotopes.

Materials and methods

Study site

All sampling took place in the Spermonde Archipelago, South Sulawesi, Indonesia. This Archipelago consists of 160 fringing, barrier and patch reefs (de Voogd *et al.*

2006; Fig. 1) situated adjacent to the city of Makassar. Its proximity to a city of more than 2 million inhabitants (Renema 2010) leaves these coral reefs exposed to anthropogenic disturbances including river discharge (sedimentation, agricultural runoff), oil spills, destructive fisheries, tourism and coral mining (de Voogd & Cleary 2007).

Sampling

Four sponge species, sediment and seawater were collected in different reef sites surveyed using SCUBA in August 2012. Different reef sites were used to better characterize the archaeal community of the Spermonde reef as a whole. The reef sites Lae Lae, Samalona, Kudingkareng Keke, Bone Baku and Langkai were sampled. At each site, one sample of each biotope (if present) was taken, namely, sediment, seawater, the LMA sponges *Stylissa carteri* and *S. massa* (order Halichondrida: family Dictyonellidae) and the HMA sponges *X. testudinaria* (order Haplosclerida: family Petrosiidae) and *H. erectus* (order Dictyoceratida: family Thorectidae). Only *S. carteri* was collected in Bone Baku, and this sponge was not found at Langkai and Lae Lae. Because of this, we took two *S. carteri* samples at Kudingkareng Keke. Samples of *S. massa*, *X. testudinaria*, sediment and seawater from the Kepulauan Seribu reef system were sampled in August 2011 (Polónia *et al.* 2013). The sediment samples were taken using the mini core method. Mini-cores, consisting of the top 5 cm of sediment, were collected using a plastic disposable syringe from which the end had been cut to facilitate sampling (Capone *et al.* 1992). Cores of all sponge species were sampled including segments of surface and interior to sample, as much as possible, the whole archaeal community (Pires *et al.* 2012; Polónia *et al.* 2013). The seawater samples were collected by filtering 1 L (Bowen *et al.* 2012) of seawater through a Millipore® White Isopore Membrane Filter (0.22 μm pore size). All samples were stored in 96% EtOH (Cleary *et al.* 2013) and kept at temperatures lower than 4 °C immediately after collection. Once in the laboratory, samples were stored at −20 °C until DNA extraction.

DNA extraction and pyrosequencing

We isolated PCR-ready genomic DNA from seawater, sediment and sponge samples using the FastDNA® SPIN Kit (MP Biomedicals) following manufacturer's instructions. This is an extraction method frequently used for this purpose (Cleary *et al.* 2013; Costa *et al.* 2013). Briefly, the whole membrane filter and 500 mg of sediment and sponge were transferred to Lysing Matrix E tubes containing a mixture of ceramic and silica particles. The microbial cell lysis was performed in the FastPrep® Instrument (Q Biogene) for 80 s at speed 6.0. The extracted DNA was eluted into DNase/pyrogen-free water to a final volume of 50 μL and stored at −20 °C until use. Prior to pyrosequencing, the amplicons of the archaeal 16S rRNA gene were obtained using Archaea-specific primers ARC344f-mod and Arch958R-mod (Pires *et al.* 2012). After a denaturation step at 94 °C

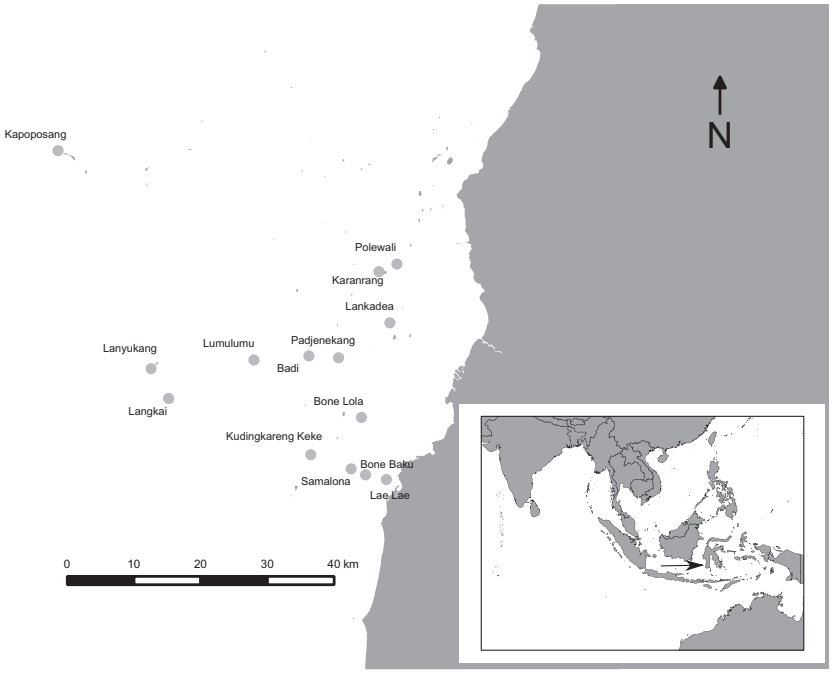

Fig. 1 Map of the study area (Spermonde Coral Reef System) showing the study sites.

during 5 min, 35 thermal cycles of 1 min at 94 °C, 1 min at 56 °C and 1 min at 72 °C were carried out followed by an extension step at 72 °C for 7 min (Pires *et al.* 2012). Using the amplicons of the archaeal 16S rRNA gene as template, the V3V4 region was amplified using barcoded fusion primers [524F-10-ext (5′- TGY CAGCCGCCGCGGTAA -3′) and Arch958R-mod (5′-C CGGCGTTGAVTCCAATT -3′); Pires *et al.* 2012] with the Roche-454 A and B Titanium sequencing adapters, an eight-base barcode sequence in adaptor B and specific sequences for the ribosomal region. The barcoded pyrosequencing libraries were analysed using the QIIME (Quantitative Insights Into Microbial Ecology; Caporaso *et al.* 2010) software package (http://www.qiime.org/; last checked 2014-01-20) on a computer running the BIOLINUX 7 operating system (http://nebc.nerc.ac.uk/; checked 2014-01-20). In QIIME, FASTA and QUAL files were used as input for the split_libraries.py script. OTUs were selected using UPARSE with USEARCH7 (Edgar 2013). Chimera checking was performed using the UCHIME algorithm, which is the fastest and most sensitive chimera checking algorithm currently available (Edgar *et al.* 2011). OTU clustering was performed using the -cluster_otus command (cut-off threshold at 97%). An additional chimera check was subsequently applied using the -uchime_ref command with the gold.fa database (http://drive5.com/uchime/gold.fa). In QIIME, representative sequences were selected using the pick_rep_set.py script using the 'most_abundant' method. Reference sequences of OTUs were assigned taxonomies using default arguments in the assign_taxonomy.py script in QIIME with the rdp method (Wang *et al.* 2007). In the assign_taxonomy.py script, we used a FASTA file containing reference sequences from the Greengenes 13_5 release and the rdp classifier method. We used a modified version of the taxonomy file supplied with the Greengenes 13_5 release to map sequences to the assigned taxonomy. Finally, we used the make_otu_table.py script in QIIME to generate a square matrix of OTUs × samples. This was subsequently used as input for further analyses using R (R Core Team 2013). In the most recent official Greengenes release (gg_13_5;http://greengenes.lbl.gov/cgi--bin/nph-index.cgi; checked 2014 04 01), the recently adopted phylum Thaumarchaeota is still considered a class of the Crenarchaeota phylum; in the results section of this study, we follow the Greengenes taxonomy for practical purposes.

BLAST, phylogenetic and predictive functional analysis

Briefly, sequence identifiers of closely related taxa of numerically dominant OTUs (≥200 sequences) were downloaded using the NCBI Basic Local Alignment Search Tool (BLAST) command line 'blastn' tool with the −db argument set to nt (Zhang *et al.* 2000). A phylogenetic tree including all dominant OTUs (≥200 sequences) was constructed using the MEGA6 program (http://www.megasoftware.net/; last checked 2014/09/11; Tamura *et al.* 2011) with the nearest-neighbor-interchange and generalised time-reversible model (Tavaré 1986) with Gamma distributed and invariant sites. In the results, we present a bootstrap consensus tree based on 100 replicates (Felsenstein 1985). Branches reproduced in <50% of the bootstrap replicates are collapsed. The bootstrap value is shown next to each branch when this exceeds 50%. This value represents the percentage of replicate trees in which the associated taxa clustered together.

To predict the metagenome of each sample, we used PICRUSt (Langille *et al.* 2013). PICRUSt is a bioinformatics tool that uses marker genes, in this case 16S rRNA, to predict metagenome gene functional content. A detailed description of these methods has been published previously (Cleary *et al.* 2013; Langille *et al.* 2013; Polónia *et al.* 2013) and can be found in the Appendix S1 (Supporting information). In this study, we used the KEGG database and focused on KEGG orthologs (KOs) in the nitrogen energy metabolism pathway. In addition to metagenomic data, we also used the -a option in the predict_metagenomes.py script to obtain weighted Nearest Sequenced Taxon Index (NSTI) scores for each sample. NSTI scores are a means of quality control, which provide a summary of the extent to which OTUs in a given sample are related to reference OTUs. NSTI scores represent the average branch length separating an OTU from a reference OTU. We used R to generate bargraphs showing the percentage of total genes for each sample. In addition to this, we used the metagenome_contributions.py script to assess the relative contribution of selected orders. The metagenome_contributions.py script partitions functional contributions to function, OTU and sample. Results of this analysis are presented using barplots for each biotope.

Comparison between the Spermonde and Kepulauan Seribu reefs

We compared samples of *S. massa*, *X. testudinaria*, sediment and seawater from Kepulauan Seribu reef and Spermonde reef systems with respect to higher taxon relative abundance, archaeal composition and KO enrichment in the nitrogen energy metabolism pathway. This enabled us to assess to what extent archaeal communities differ in the same biotopes over large geographical distances. Although, both the Spermonde and Kepulauan Seribu reefs suffer from anthropogenic disturbance the degree of disturbance is much more pronounced in Jakarta (Cleary *et al.* 2005, 2014; A. R. M. Polónia, D. F. R. Cleary, N.

J. de Voogd, W. Renema, B. W. Hoeksema, A. Martins & N. C. M. Gomes, unpublished). Our comparison will, therefore, give us some insight into the impact of a major conurbation on the archaeal community.

Statistical analysis

A square matrix containing the presence and abundance of all OTUs per sample was imported into R (R Core Team 2013) using the read.table() function. Sequences not classified as Archaea were removed prior to statistical analysis. The OTU abundance matrix was $\log_{10}(x + 1)$ transformed (to normalize the distribution of the data) and a distance matrix constructed using the Bray–Curtis index with the vegdist() function in the VEGAN package (Oksanen *et al.* 2009) in R. The Bray–Curtis index is one of the most frequently applied (dis) similarity indices used in ecology (Legendre & Gallagher 2001; Cleary 2003). Variation in archaeal composition among biotopes (sediment, seawater, *S. carteri*, *S. massa*, *X. testudinaria* and *H. erectus*) was assessed with principal coordinates analysis (PCO) using the cmdscale() function in R with the Bray–Curtis distance matrix as input. Variation among biotopes was tested for significance using the adonis() function in VEGAN. In the adonis analysis, the Bray–Curtis distance matrix of species composition was the response variable with biotope as independent variable. The number of permutations was set at 999; all other arguments used the default values set in the function. Weighted averages scores were computed for OTUs on the first two PCO axes using the wascores() function in the VEGAN package. We performed an additional compositional analysis comparing *S. massa*, *X. testudinaria*, seawater and sediment from the Spermonde and Kepulauan Seribu reefs.

Results

The sequencing effort yielded 95 019 sequences, which were assigned to 656 OTUs after quality control, OTU picking, removal of chimera and removal of OTUs not assigned to the domain Archaea. All archaeal OTUs were assigned to three phyla, Crenarchaeota (63 519 sequences), Euryarchaeota (31 379 sequences) and Parvarchaeota (18 sequences). In addition to this, OTUs were assigned to 14 classes and 17 orders. Of these, the classes Thaumarchaeota[1] (61 990 sequences) and Thermoplasmata (30 940 sequences), the orders Cenarchaeales (61 625 sequences) and E2 (30 930 sequences) were the most abundant.

[1]Recently recognized as an archaeal phylum.

Higher taxon abundance

There were marked differences in the abundance of higher archaeal taxa among biotopes (Fig. 2). The Euryarchaeota achieved their greatest abundance in seawater where they comprised more than 98% on average of all sequences; Euryarchaeota were also abundant in both *Stylissa* hosts. In all other biotopes, more than 60% of sequences belonged to the Crenarchaeota. In the Spermonde reef system (adjacent to Makassar), more than 98% of OTUs assigned to the Euryarchaeota phylum were assigned to the Thermoplasmata class. Likewise, more than 97% of OTUs assigned to the Crenarchaeota phylum were assigned to the Thaumarchaeota1class and Cenarchaeales order. There was also a marked difference in dominance between host and nonhost biotopes. In nonhost biotopes, the single most dominant OTU made up <28% of all sequences. In host biotopes, in contrast, the single most dominant OTU made up more than 60%, on average, of all sequences. The third most abundant class (Miscellaneous Crenarchaeotal Group; MCG) was virtually restricted to the sediment biotope, and the third most abundant order (Nitrososphaerales) was found in sediment, *X. testudinaria* and *H. erectus*.

Interestingly, the relative abundance of *Thermoplasmata* (*Euryarchaeota*) was much higher in samples from the Spermonde reef as opposed to the Kepulauan Seribu reef system. In addition to seawater, this included samples from the sponge species *S. massa* and surprisingly *X. testudinaria*. Thermoplasmata were found in all samples of *X. testudinaria* in Spermonde where they made up more than 8% of all sequences in the biotope as opposed to <0.3% for *X. testudinaria* in Kepulauan Seribu; this was despite identical sampling strategies by the same person (NjdV) in both locations. Despite this trend, there was also some marked intrabiotope variation, for example, specimens from the same host (*S. massa*) and same site (Kud) in the Spermonde (Makassar) contained a very high (>95%) vs. very low (<0.4%) abundance of Thermoplasmata (Fig. 2). Likewise, one seawater sample in Kepulauan Seribu (Jakarta) was dominated (>74%) by Thaumarchaeota in contrast to all other seawater samples from the Spermonde and Kepulauan Seribu reefs that overwhelmingly contained sequences assigned to Thermoplasmata (>75%).

OTU composition analysis

BLAST was used to find closely related organisms to the most abundant (≥200 sequences) OTUs (see Appendix S2, Supporting information for OTU tables). The most abundant OTU overall was OTU-1, assigned to the

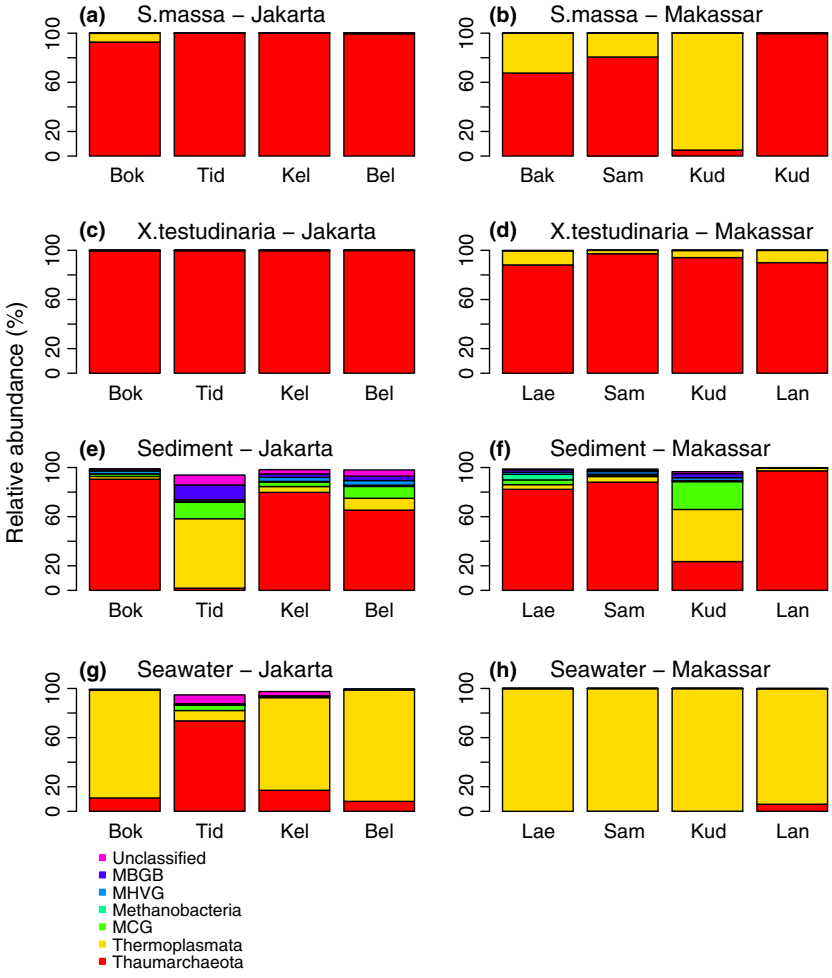

Fig. 2 Mean relative abundance of the most abundant archaeal classes (Marine Benthic Group B (MBGB); Marine Hydrothermal Vent Group (MHVG); Methanobacteria; Miscellaneous Crenarchaeota Group (MCG); *Thermoplasmata*; *Thaumarchaeota*; Unclassified) in the Kepulauan Seribu reef system [Jakarta—data published in Polónia *et al.* 2013; sites: Bokor (Bok), Tidung Kecil (Tid), Pulau Kelapa (Kel), Belanda (Bel)] and Spermonde reef [Makassar; sites: Lae Lae (Lae), Bone Baku (Bak), Samalona (Sam), Kudingkareng Keke (Kud), Langkai (Lan)]. *Stylissa massa* in (a) Jakarta and (b) Makassar, *Xestospongia testudinaria* in (c) Jakarta and (d) Makassar, sediment in (e) Jakarta and (f) Makassar and seawater in (g) Jakarta and (h) Makassar.

genus *Cenarchaeum* and found exclusively in *S. massa* and *Stylissa carteri* hosts and represented by 21 559 sequences. OTU-1 was closely related to an organism previously detected in *Axinella carteri* hosts in Israel (Y. Schechtman, E. T. Sieradzki & M. Ilan, unpublished; GI:404160733), and *Phakellia fusca* in the South China Sea (Han *et al.* 2012; GI:340764424).

Importance of biotope in structuring composition

There was a highly significant difference in archaeal composition among biotopes ($F_{5,18} = 10.73$, $P < 0.001$, $R^2 = 0.749$). Variation among biotopes thus explained almost 75% of the variation in archaeal composition. A PCO ordination (Fig. 3) of the first two axes shows four distinct clusters representing samples from the six biotopes. One cluster consisted of samples from seawater, and both *Stylissa* hosts with other clusters consisting of samples from sediment, *X. testudinaria* and *H. erectus*. The main axis of variation separated OTUs found predominantly in seawater, and both *Stylissa* hosts from OTUs found predominantly in sediment.

In a comparison of samples from the Spermonde and Kepulauan Seribu (Fig. 4), there was a highly significant difference in archaeal composition among locations ($F_{1,24} = 6.85$, $P < 0.001$, $R^2 = 0.088$), biotopes ($F_{3,24} = 12.98$, $P < 0.001$, $R^2 = 0.501$) and a significant interaction ($F_{3,24} = 2.64$, $P = 0.002$, $R^2 = 0.102$). Variation among location, biotopes and the interaction thus explained almost 70% of the variation in archaeal composition. The main axis separated samples of sediment from both areas from samples from other biotopes. The second axis separated samples of *S. massa* and *X. testudinaria* sampled in Kepulauan Seribu from seawater samples from both areas and *S. massa* and *X. testudinaria* sampled in Spermonde. Samples of *S. massa* and *X. testudinaria* and seawater (but not sediment) thus contained distinct archaeal communities in the Spermonde and Kepulauan Seribu reefs.

Phylogeny

In the phylogenetic tree (Fig. 5), there were two main clusters, (i) a cluster consisting of OTUs belonging to

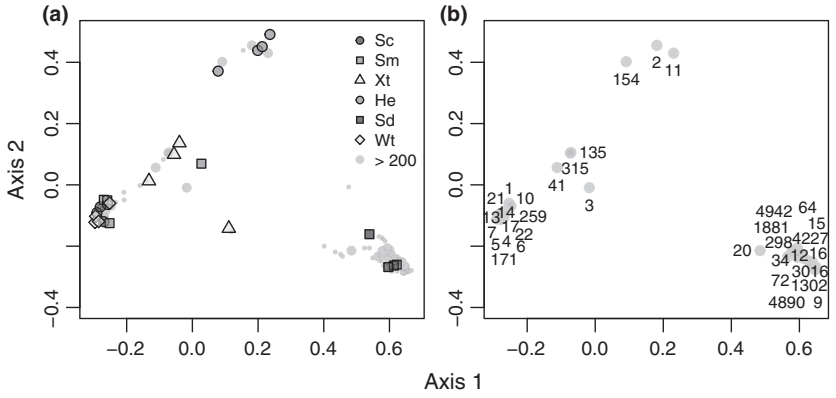

Fig. 3 Ordination showing the first two axes of the PCO analysis of Spermonde samples. (a) Symbols represent samples from *Stylissa carteri* (Sc), *Stylissa massa* (Sm), *Xestospongia testudinaria* (Xt), *Hyrtios erectus* (He), sediment (Sd) and seawater (Wt). Very small circles represent OTUs <200 sequence reads. (b) Numbers represent abundant (≥200 sequence reads) OTUs referred to in Appendix S2 (Supporting information). PCO, principal coordinates analysis.

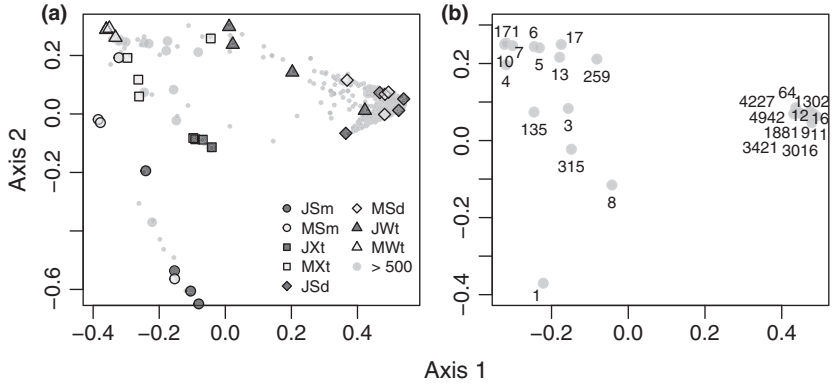

Fig. 4 Ordination showing the first two axes of the PCO analysis comparing the archaeal communities present in Spermonde (Makassar; M) and Kepulauan Seribu (Jakarta; J—data published in Polónia *et al*. 2013) reefs. (a) Symbols represent samples from *Stylissa massa* in Jakarta (JSm) and Makassar (MSm), *Xestospongia testudinaria* in Jakarta (JXt) and Makassar (MXt), sediment in Jakarta (JSd) and Makassar (MSd) and seawater in Jakarta (JWt) and Makassar (MWt). Very small circles represent OTUs <500 sequence reads. (b) Numbers represent abundant (≥500 sequence reads) OTUs. PCO, principal coordinates analysis.

the Crenarchaeota phylum and (ii) a cluster consisting of OTUs belonging to the Euryarchaeota phylum. Inside the main cluster of Euryarchaeota, the most abundant OTUs found in seawater and *Stylissa* hosts belonged to Marine Group II and formed a distinct cluster. Inside the Crenarchaeota main cluster, OTUs found in *X. testudinaria* and *H. erectus* formed two distinct and well-supported clusters that grouped with OTUs found in sediment and assigned to the genus *Nitrosopumilus* (sequence alignment file can be found in Appendix S3, Supporting information).

Predictive functional analysis

Nitrogen metabolism. Mean (and range) NSTI values for the biotopes sampled in Makassar were 0.079 (0.065–0.093) for *S. carteri*, 0.083 (0.057–0.142) for *S. massa*, 0.035 (0.025–0.051) for *X. testudinaria*, 0.145 (0.062–0.335) for *H. erectus*, 0.108 (0.051–0.227) for sediment and 0.125 (0.076–0.180) for seawater. Only 21 of the 54 KOs involved in the nitrogen metabolism pathways were detected. The most abundant of these are related to ammonia and are presented in Fig. 6. Almost all of these KOs are, however, shared with other pathways. For example, the majority of the detected KOs also participate in the amino acid metabolism. In addition to KOs related to ammonia, some KOs related to urea were also found (UreA, UreB, UreC); these were enriched in sponge biotopes (data not shown).

Particularly intriguing was the absence of genes for ammonia oxidation (amoA) in a data set with a high number of sequences assigned to known ammonia oxidizing Archaea (e.g. 3041 assigned *Nitrosopumilus* sequences). As PICRUSt predicts the functional potential of microbial communities from their phylogeny, the presence of *Nitrosopumilus* sequences in the analysed data set should have resulted in at least some counts in the KOs for ammonia monooxygenase A (K10944). The reason for this absence was due to a discrepancy with respect to the presence/absence of this KO in the genome database. In particular, the KEGG entry for *Nitro-*

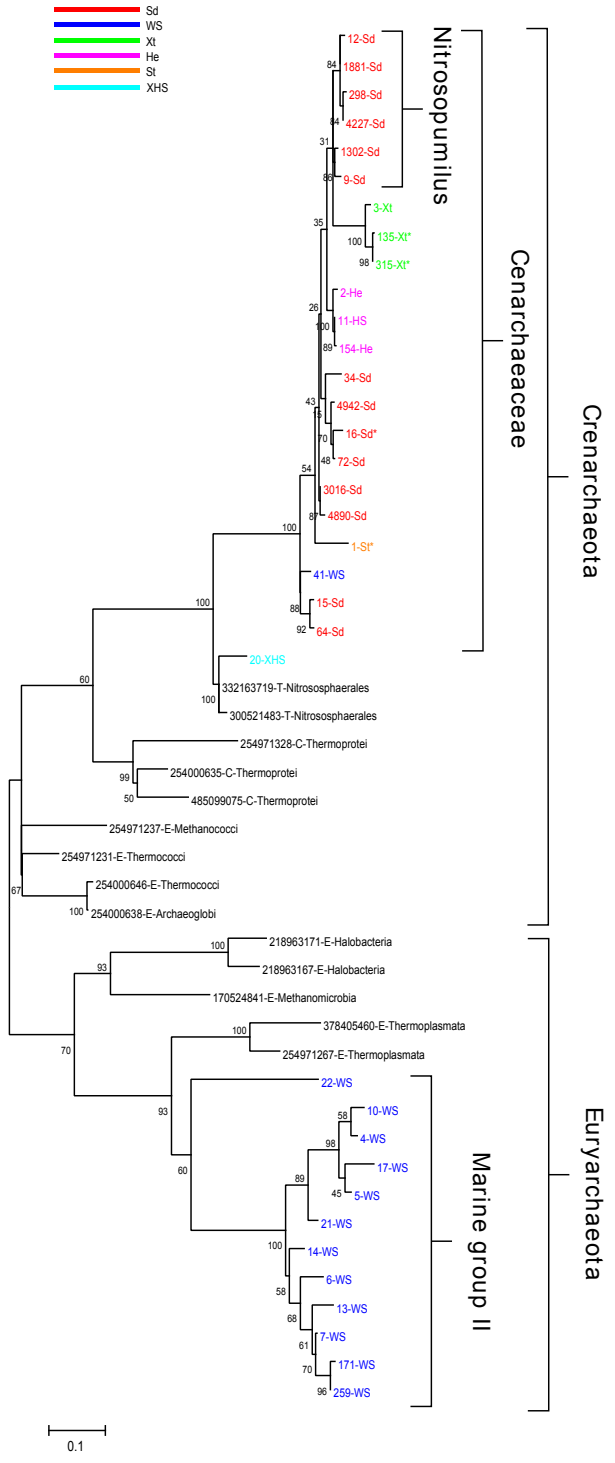

Fig. 5 Phylogenetic tree (constructed using the nearest-neighbor-interchange and generalised time-Reversible model) of the archaeal 16S rRNA gene sequences recovered from the studied biotopes (*Stylissa carteri*, *Stylissa massa*, *Xestospongia testudinaria* and *Hyrtios erectus*, sediment and seawater); bootstrap values lower than 50% were omitted. The number of each OTU is indicated as are GenBank GenInfo sequence identifiers of cultured archeal sequences. OTUs are assigned to the following clusters: Sd: mainly found in sediment biotope; WS: mainly found in sponges belonging to the genus *Stylissa* and seawater biotopes; Xt: mainly found in *X. testudinaria*; He: mainly found in *H. erectus*. St: OTU only found in *S. massa* and *S. carteri;* HS: OTU mainly found in *H. erectus* and sediment; XHS: OTU mainly found in *X. testudinaria*, *H. erectus* and sediment.*Represents an OTU only present in a single biotope or in a single sponge genus.

sopumilus showed K10944 (methane/ammonia monooxygenase subunit A), but in the cached IMG table, the same KO with the same genome accession (*Nitrosopumilus maritimus* SCM, NC_010085) was absent (J. Zaneveld, personal communication). This is something that will be addressed in the future, but given the relative abundance of known ammonia-oxidizing taxa (e.g. *Nitrosopumilus*), the gene count for K10944 would have been highest in sediment, intermediate in sponges and virtually absent in seawater.

The relative abundance of those KOs that were present revealed several differences among biotopes. Both *Stylissa* species, for example, were enriched with respect to glutamate dehydrogenase ($NADP^+$; K00262), glutamate synthase (NADPH/NADH; K00266) and carbamate kinase (K00926) compared with *X. testudinaria*. *Xestospongia testudinaria* had very low relative abundances of glutamate dehydrogenase ($NADP^+$) and glutamate synthase (NADPH/NADH) but the highest relative abundances of glutamate dehydrogenase ($NAD(P)^+$; K00261) and nitrate reductase beta subunit. Seawater was enriched with respect to glutamate dehydrogenase ($NADP^+$; K00262), carbamate kinase (K00926) and glutamate synthase (NADPH/NADH; K00266).

OTUs belonging to the Cenarchaeales order were the only ones contributing to the abundance of glutamate dehydrogenase ($NAD(P)^+$; K00261) and nitrate reductase beta subunit (K00371) enzymes (Appendix S4, Supporting information). In contrast, OTUs belonging to the E2 order were the only ones contributing to the abundance of glutamate dehydrogenase ($NADP^+$, K00262) and glutamate synthase (NADPH/NADH, K00266). OTUs belonging to the E2 order were also major contributors to the abundance of carbamate kinase (K00926). For glutamine synthetase (K01915), both orders, Cenarchaeales and E2, had similar contributions, with OTUs belonging to the Cenarchaeales order being the major contributors to the presence of this enzyme in sponge and sediment biotopes and OTUs belonging to the E2 order the major contributors to the abundance of this enzyme in seawater.

When comparing Spermonde and Kepulauan Seribu reefs (Fig. 7), samples of *S. massa*, *X. testudinaria* and

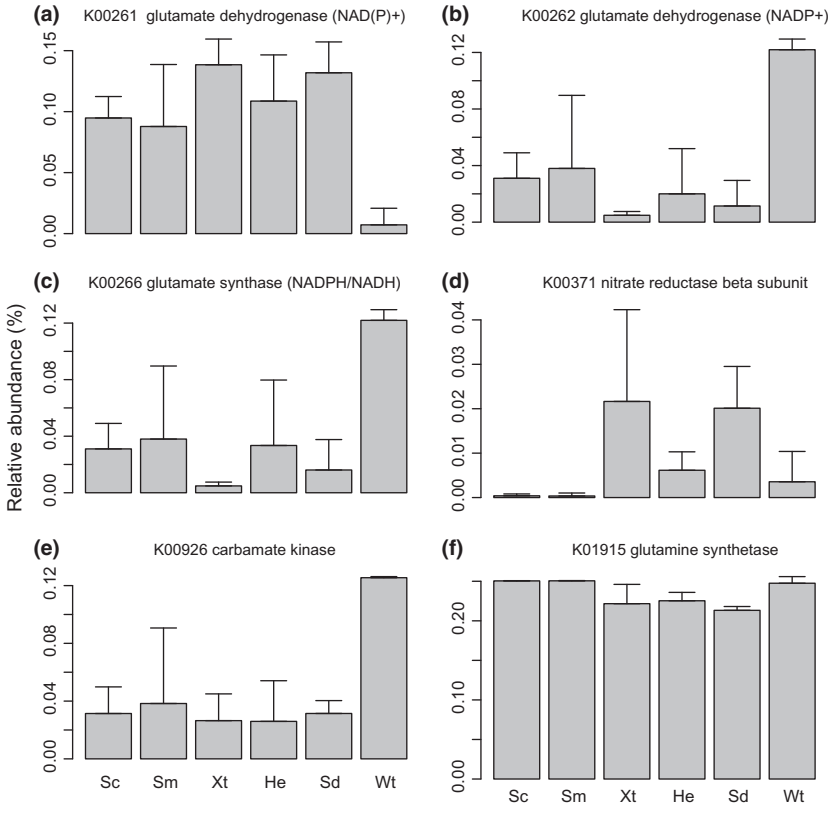

Fig. 6 Mean relative abundance of KEGG orthologs (KOs) involved in the nitrogen metabolism pathways for samples from *Stylissa carteri* (Sc), *Stylissa massa* (Sm), *Xestospongia testudinaria* (Xt), *Hyrtios erectus* (He), sediment (Sd) and seawater (Wt). Error bars represent a single standard deviation. The KOs shown include the following: (a) K00261 glutamate dehydrogenase ($NAD(P)^+$); (b) K00262 glutamate dehydrogenase ($NADP^+$); (c) K00266 glutamate synthase (NADPH/NADH); (d) K00371 nitrate reductase beta subunit; (e) K00926 carbamate kinase; (f) K01915 glutamine synthetase.

seawater were enriched in Kepulauan Seribu reefs for K00261 glutamate dehydrogenase ($NAD(P)^+$). In contrast, samples of *S. massa* and seawater in the Spermonde reefs were enriched for K00262 glutamate dehydrogenase ($NADP^+$) and K00266 glutamate synthase (NADPH/NADH).

Discussion

With the exception of the seawater community, which was dominated by Euryarchaeota, the archaeal community of sediment and sponges was dominated by Crenarchaeota. Sediment was the richest biotope (three phyla; 12 classes, 16 orders and 15 families; see Appendix S5, Supporting information for figures showing rarefied richness) in the coral reef environment. Moreover, all the unclassified OTUs at the phylum level (36) were found in sediment.

Crenarchaeota have been previously shown to be the dominant phylum in sponge biotopes (Holmes & Blanch 2007), as was the case in our study. However, the percentage of Crenarchaeota communities detected here was lower than in previous studies. In the Kepulauan Seribu (Polónia *et al.* 2013), 98% and 99.8% of the archaeal communities in *S. massa* and *X. testudinaria*, respectively, were assigned to the Crenarchaeota phylum. In the Spermonde reef (present study), only 63% of the archaeal community inhabiting *S. massa* and 92% of the archaeal community inhabiting *X. testudinaria* were assigned to Crenarchaeota. Polónia *et al.* (2013) also found that 29% of the seawater community in Kepulauan Seribu was assigned to Crenarchaeota. In the Spermonde reef system <2% of the seawater community was assigned to this phylum. This suggests that in seawater environments dominated by Euryarchaeota, the proportion of this phylum in sponge tissues tends to increase, indicating a clear influence of the environment on the sponge microbial communities. Importantly, this was not only the case for *S. massa*, an LMA sponge known to host seawater microbes, but was also true for the HMA *X. testudinaria*.

In the tropical surface waters of the Georgetown coast (Penang, Malaysia) and Kepulauan Seribu reef system (Java, Indonesia), 65–70% of the archaeal community was assigned to the Euryarchaeota phylum (Chan *et al.* 2013; Polónia *et al.* 2013). However, in the South Pacific Gyre, considered one of the cleanest oceanic regions of the world, due to its isolation from sources of pollution, the seawater community was almost entirely composed of Euryarchaeota (Yin *et al.* 2013). These authors suggested that the low ammonium concentration in the seawater of the South Pacific Gyre is the reason for the very low Crenarchaeota abundance. Crenarchaeota have been shown to be important players in many

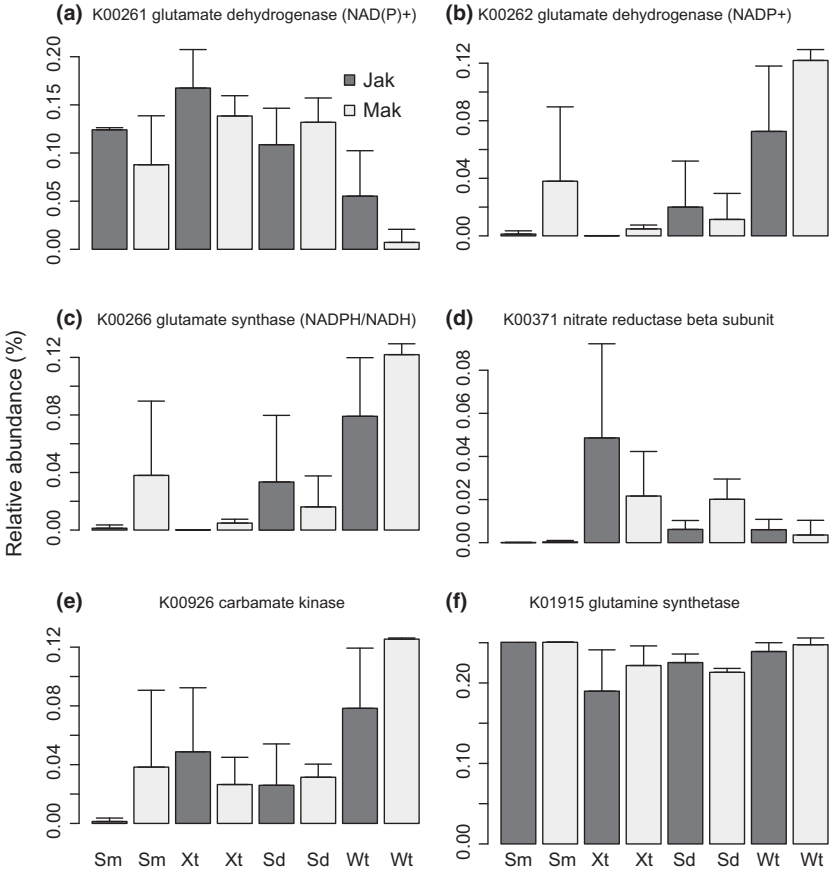

Fig. 7 Comparison of the mean relative abundance of KEGG orthologs (KOs) involved in the nitrogen metabolism pathways for samples from *Stylissa massa* (Sm), *Xestospongia testudinaria* (Xt), sediment (Sd) and seawater (Wt) in the Spermonde (Makassar—Mak) and Kepulauan Seribu (Jakarta—Jak) reefs. Error bars represent a single standard deviation. The KOs shown include the following: (a) K00261 glutamate dehydrogenase ($NAD(P)^+$); (b) K00262 glutamate dehydrogenase ($NADP^+$); (c) K00266 glutamate synthase (NADPH/NADH); (d) K00371 nitrate reductase beta subunit; (e) K00926 carbamate kinase; (f) K01915 glutamine synthetase.

geochemical cycles. All the cultivated members of the phylum Thaumarchaeota (Mesophilic Crenarchaeota), for example, obtain energy through ammonia oxidation (Offre *et al.* 2013) and thus play an important role in the nitrogen cycle. A reduced concentration of ammonia may thus limit Crenarchaeota abundance. The low abundance of seawater Crenarchaeota sequences in the Spermonde is an indication of lower pollution levels when compared, for example, to the Kepulauan Seribu reef system or Georgetown coast. Although rare in the Spermonde seawater samples, Crenarchaeota remained the dominant phylum in all sponge species and sediment. Sponges offer their symbionts a stable and nutrient rich environment, namely a constant supply of ammonia, a metabolic waste product excreted by sponges. This makes sponges suitable habitats for ammonia-oxidizing Archaea (AOA).

Xestospongia testudinaria and *Hyrtios erectus* shared a higher number of OTUs with sediment than with seawater. In contrast, *Stylissa carteri* and *S. massa* shared a higher number of OTUs with seawater when compared to sediment. This was reflected both in the phylogenetic tree and in the PCO where samples from both *Stylissa* sponges clustered together with seawater samples. This result may be related to the different morphologies and life strategies of each of the sponge species. The skeleton of the slow growing and long-lived *X. testudinaria* consists of a very dense network of silicious spicules (Desqueyroux-Faúndez & Valentine 2002). Sponges belonging to the genus *Stylissa*, in turn, are fast growers and have a loose skeleton of very large spicules (Van Soest *et al.* 2002), which results in higher amounts of water in their tissues. Indeed, when squeezed, a copious amount of water is expelled from the sponge oscules. *Hyrtios erectus* lives cryptically embedded in the sediment covered in coral sand and the build-up of its skeleton is composed by the incorporation of exogenous materials such as sediment grains. Additionally, sponges belonging to the genus *Xestospongia* have been considered HMA sponges in contrast to sponges belonging to the genus *Stylissa* (Moitinho-Silva *et al.* 2013). Previous studies have noted that LMA sponges tend to filter much larger volumes of water than HMA sponges, hosting communities with lower specificity and diversity and similar to that of seawater (Thacker & Freeman 2012; Moitinho-Silva *et al.* 2013). This is consistent with the high amount of water found in the tissues of sponges belonging to the genus *Stylissa* and may account for the greater percentage of shared symbionts in both *Stylissa* species. Our results, however, did not

allow us to determine whether the sponge OTUs shared with seawater and sediment belonged to the sponge microbiome (as result of environmental selection) or whether they were merely contaminants.

An ongoing debate in sponge microbial studies is the degree to which sponge microbes are transferred horizontally as opposed to vertically (Hentschel *et al.* 2002; Taylor *et al.* 2007). In the present study, the most abundant OTU (21 591 sequences) was found exclusively in both *Stylissa* species (OTU-1) and represented a phylogenetically distinct lineage. OTU-1, assigned to the genus *Cenarchaeum*, was closely related to organisms detected in *Axinella carteri* (Y. Schechtman, E. T. Sieradzki & M. Ilan, unpublished) and *Phakellia fusca* (Han *et al.* 2012). Margot *et al.* (2002) and Holmes & Blanch (2007) previously suggested the existence of a symbiotic association between sponges belonging to the genus *Axinella* and an archaeon closely related to *Cenarchaeum symbiosum*. Here, due to the close taxonomic relationship of these *Axinella* and *Phakellia* sponges with the studied *Stylissa* (all belong to the order Halichondrida), we suggest the existence of a possible order-specific symbiosis between Halichondrida and the *C. symbiosum* phylotype (OTU-1) found in this study. This also suggests that, despite the high percentage of shared symbionts between both *Stylissa* species and seawater and the consequent putatively horizontal transmission of these OTUs (thus acquired from the surrounding environment), this *C. symbiosum* phylotype is transmitted vertically, that is, from parent to offspring. This suggests vertical transmission as an important component of microbial transmission for LMA sponges and not only for HMA sponges as suggested by Gloeckner *et al.* (2014). OTU-1 was also the most dominant OTU in *S. massa* in the Kepulauan Seribu reefs. In contrast to the Spermonde reef system, however, this OTU was also found in two seawater samples from Kepulauan Seribu reefs. This could indicate that OTU-1 is in fact horizontally transmitted and is a very rare component of the seawater community. However, it may also be true that *S. massa* is seeding seawater and that most transmission still occurs vertically.

The PICRUSt method used in the present study is only an estimate of microbiome function and not an actual measurement of such function. PICRUSt, however, includes various methods of quality control to test the reliability of the predictions. One of these entails the use of weighted NSTI scores. NSTI calculates the dissimilarity between reference genomes and the metagenome and was developed to evaluate the prediction accuracy of PICRUSt, particularly in poorly characterized environments with relatively few reference genome sequences available. NSTI scores in the present story were generally low, suggesting an accurate prediction of microbiome function. Scores were particularly low for the sponge *X. testudinaria* (mean NSTI: 0.035) but higher for sediment (mean NSTI: 0.108) and seawater (mean NSTI: 0.125). Langille *et al.* (2013) showed that the accuracy of PICRUSt decreased with increasing NSTI scores. In their study, the best results were obtained with well-covered Human Microbiome Project samples (mean NSTI: 0.03), mid-range for a data set of soil samples (mean NSTI: 0.17) and highest for the hyperdiverse and under-explored Guerrero Negro microbial mat (mean NSTI: 0.23). PICRUSt still produced accurate results for the soil samples despite the relatively high NSTI score. The accuracy for the hypersaline microbial mat community was, however, markedly lower although the Langille *et al.* (2013) noted that this was also related to shallow sequencing at a depth that was insufficient to fully sample the community's genomic composition.

Nevertheless, PICRUSt predictions must be treated with caution, particularly for novel communities with relatively high NSTI scores. In our study, the relatively low NSTI scores for sponges indicate that the PICRUSt results are reasonably accurate. The large range of NSTI values of sediment, however, is reason to be cautious with the obtained estimates of sediment microbiome function. This is not surprising given the pronounced diversity of the sediment microbiome and the presence of numerous taxonomically poorly resolved OTUs.

The nitrogen metabolism of Archaea assessed in the present study was dominated by KOs related to ammonia. Ammonium and ammonia incorporation can occur via two distinct pathways: glutamine synthetase/glutamate synthase and glutamate dehydrogenase (Harper *et al.* 2008). The Glutamate dehydrogenase pathway is responsible for the catalysis of the glutamate catabolism, that is for the breakdown of glutamate into ammonium and α-ketoglutarate, and thus is also responsible for feeding the tricarboxylic acid pathway (TCA) (Peterson & Smith 1999). The glutamate metabolism is, in this way, an important link between the carbon and nitrogen metabolisms (Belitsky & Sonenshein 1998).

All the biotopes had relatively high abundances of glutamine synthetase (K01915) although abundance was highest for both *Stylissa* species and water and lowest for sediment, *X. testudinaria* and *H. erectus*. Due to high ammonium affinity, the glutamate synthase pathway is used under restricted nitrogen availability while the glutamate dehydrogenase pathway, due to its low ammonium affinity, requires higher nitrogen availability (Harper *et al.* 2008). Seawater was enriched for both glutamate dehydrogenase ($NADP^+$; K00261) and glutamate synthase (NADPH/NADH; K00266). Sponges and sediment, in contrast, were enriched for glutamate dehydrogenase ($NAD(P)^+$; K00261) in comparison to seawater. These results indicate that seawater, and par-

ticularly Spermonde seawater, is a relatively nitrogen poor environment when compared to sponges and sediment (Hentschel *et al.* 2012). The presence of some KOs related to urea (UreA, UreB, UreC genes) and their enrichment in sponge biotopes suggests that urea might be an important source of carbon and energy for sponge archaeons, similar to results obtained by Alonso-Sáez *et al.* (2012).

The reversible oxidative deamination of glutamate to α-ketoglutarate and ammonia is catalysed by glutamate dehydrogenases. These enzymes can act with only one coenzyme (NAD^+ or $NADP^+$) or with two coenzymes ($NAD(P)^+$). The first case occurs normally in lower eukaryotes or in prokaryotes, while the second case occurs mainly in higher eukaryotes (Miñambres *et al.* 2000). The distinct differences in enrichment of glutamate dehydrogenase ($NADP^+$) and of glutamate dehydrogenase ($NAD(P)^+$) among biotopes suggest the existence of different degrees of specialization in the oxidative deamination of glutamate to α-ketoglutarate and ammonia.

Seawater, and particularly Spermonde reef system seawater, had the highest relative abundance of the coenzyme glutamate dehydrogenase ($NADP^+$), which participates in the glutamate anabolism, that is the transformation of ammonium and α-ketoglutarate into glutamate. This enrichment of a coenzyme promoting ammonium assimilation is consistent with the above suggestion of seawater as a nitrogen poor environment. Seawater also had a high relative abundance of carbamate kinase (K00926), an enzyme that catalyses the transformation of carbamoyl phosphate to carbamate with the concomitant production of ATP. This may be an important source of energy to planktonic Archaea (Uriarte *et al.* 1999).

Sediment and *X. testudinaria* from the Kepulauan Seribu and Spermonde reefs, in turn, had the highest relative abundances of nitrate reductase beta subunit (K00371) a possible indication that sediment and sponge archaeal communities use nitrate as a nitrogen source or electron acceptor (Tang *et al.* 2013), suggesting relatively high rates of reduction of nitrate to nitrite (the first step of denitrification) in these biotopes (Liu *et al.* 2012). To the best of our knowledge, this enzyme has not been previously associated with Cenarchaeales-related sequences that were responsible for almost all of the nitrate reductase beta subunit (K00371) counts in this study. This and other observations obtained using PICRUSt should be confirmed using actual transcriptomic or metagenomic methods. An advantage of the predictive approach used here is that it provides a rapid estimation of a wide range of gene families that contribute to the metagenomes assessed in this study thereby enabling future hypothesis testing and specific testing for the expression of genes of interest.

Cenarchaeum symbiosum (OTU-1) comprised more than 60% on average of the *Stylissa* archaeal community, while unclassified members of the Cenarchaeaceae family comprised 96% (OTU-2; OTU-11 and OTU-154) and 85% (OTU-3 and OTU-315) on average of *H. erectus* and *X. testudinaria* archaeal communities, respectively. Similarly, these dominant OTUs were responsible for almost all of the sponge KO counts and thus clearly play a dominant role in the Archaea-mediated nitrogen metabolism.

Our study provides novel insight into the function and distribution of Archaea in coral reefs. We observed that a higher proportion of Euryarchaeota in seawater appears to influence the proportion of this phylum in sponge hosts including both LMA and HMA sponges. In the Spermonde, this resulted in greater compositional similarity between the archaeal communities inhabiting *Stylissa* host species (LMA) and seawater. This effect was less pronounced in *X. testudinaria* and *H. erectus* (HMA). Our results also showed significant differences among biotopes with respect to functional gene content in archaeal assemblages. These differences were accentuated between host and nonhost biotopes and resulted in clear differences in dominant functions. In seawater, ammonium assimilation appears to be performed preferentially through the expression of $NADP^+$-specific glutamate dehydrogenase (typical for prokaryotes) and glutamate synthase (NADPH/NADH; K00266), whereas in all sponges and sediment, ammonium assimilation appears to be performed preferentially through the expression of glutamate dehydrogenase with mixed specificity ($NAD(P)^+$). The degree of enrichment for glutamate dehydrogenase with mixed specificity ($NAD(P)^+$) was also most pronounced for *X. testudinaria* and sediment. Both *Stylissa* species also had moderate enrichment of $NADP^+$-specific glutamate dehydrogenase and glutamate synthase (NADPH/NADH; K00266). Our results indicate that archaeal communities in host and nonhost biotopes have distinct ecophysiological roles and may thus provide complementary nitrogen cycling functions to coral reef ecosystems.

Acknowledgements

This research was supported by the Portuguese Foundation for Science and Technology (FCT) under grant PTDC/AAC-AMB/115304/2009 (LESS CORAL) and a PhD Fellowship SFRH/BD/33391/2008. Samples were collected under a research Permit issued by the Indonesian State Ministry for Research and Technology (Kementerian Riset Dan Teknologi Republik Indonesia—RISTEK). We thank the Indonesian Institute of Sciences (PPO—LIPI) for their support and especially Drs Yos Tuti.

References

Alonso-Sáez L, Waller AS, Mende DR *et al.* (2012) Role for urea in nitrification by polar marine Archaea. *Proceedings of the National Academy of Sciences of the USA*, **109**, 17989–17994.

Belitsky BR, Sonenshein AL (1998) Role and regulation of *Bacillus subtilis* glutamate dehydrogenase genes. *Journal of Bacteriology*, **180**, 6298–6305.

Bowen JL, Morrison HG, Hobbie JE, Sogin ML (2012) Salt marsh sediment diversity: a test of the variability of the rare biosphere among environmental replicates. *ISME Journal*, **6**, 2014–2023.

Capone DG, Dunham SE, Horrigan SG, Duguay LE (1992) Microbial nitrogen transformations in unconsolidated coral reef sediments. *Marine Ecology Progress Series*, **80**, 75–88.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Chan XY, Arumugam R, Choo SW, Yin WF, Chan KG (2013) Metagenomic sequencing of prokaryotic microbiota from tropical surface seawater. *Genome Announcements*, **1**, e00540–13.

Cleary DFR (2003) An examination of scale of assessment, logging and ENSO-induced fires on butterfly diversity in Borneo. *Oecologia*, **135**, 313–321.

Cleary DFR, Becking LE, de Voogd NJ *et al.* (2005) Variation in the diversity and composition of benthic taxa as a function of distance offshore, depth and exposure in the Spermonde Archipelago, Indonesia. *Estuarine, Coastal and Shelf Science*, **65**, 557–570.

Cleary DFR, Becking LE, Voogd NJ *et al.* (2013) Habitat-and host-related variation in sponge bacterial symbiont communities in Indonesian waters. *FEMS Microbiology Ecology*, **85**, 465–482.

Cleary DFR, Polónia ARM, Renema W *et al.* (2014) Coral reefs next to a major conurbation: a study of temporal change (1985–2011) in coral cover and composition in the reefs of Jakarta Indonesia. *Marine Ecology Progress Series*, **501**, 89–98.

Costa R, Keller-Costa T, Gomes NCM *et al.* (2013) Evidence for selective bacterial community structuring in the freshwater sponge *Ephydatia fluviatilis*. *Microbial Ecology*, **65**, 232–244.

Dang H, Zhou H, Yang J *et al.* (2013) Thaumarchaeotal signature gene distribution in sediments of the Northern South China Sea: an indicator of the metabolic intersection of the marine carbon, nitrogen, and phosphorus cycles? *Applied and Environmental Microbiology*, **79**, 2137–2147.

Desqueyroux-Faúndez and Valentine (2002) Family Petrosiidae. In: *Systema Porifera: A Guide to the Classification of Sponges* (eds Hooper JNA, Van Soest RWM), pp. 906–917. Kluwer Academic/Plenum Publishers, New York.

Diaz MC, Rützler K (2001) Sponges: an essential component of Caribbean coral reefs. *Bulletin of Marine Science*, **69**, 535–546.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996–998.

Edgar RC, Haas B, Clemente J, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Fan L, Reynolds D, Liu M *et al.* (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences of the USA*, **109**, E1878–E1887.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

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 Journal*, **1**, 19–27.

Gloeckner V, Wehrl M, Moitinho-Silva L *et al.* (2014) The HMA-LMA dichotomy revisited: an electron microscopical survey of 56 sponge species. *The Biological Bulletin*, **227**, 78–88.

Han M, Liu F, Zhang F, Li Z, Lin H (2012) Bacterial and Archaeal symbionts in the South China Sea sponge *Phakellia fusca*: community structure, relative abundance, and ammonia-oxidizing populations. *Marine Biotechnology*, **14**, 701–713.

Harper C, Hayward D, Wiid I, van Helden P (2008) Regulation of nitrogen metabolism in *Mycobacterium tuberculosis*: a comparison with mechanisms in *Corynebacterium glutamicum* and *Streptomyces coelicolor*. *IUBMB Life*, **60**, 643–650.

Hentschel U, Hopke J, Horn M *et al.* (2002) Molecular evidence for a uniform microbial community in sponges from different oceans. *Applied and Environmental Microbiology*, **68**, 4431–4440.

Hentschel U, Piel J, Degnan SM, Taylor MW (2012) Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology*, **10**, 641–654.

Holmes B, Blanch H (2007) Genus-specific associations of marine sponge with group I Crenarchaeotes. *Marine Biology*, **150**, 759–772.

Hoppert M, Krüger M, Reitner J, Cockell C (2013) Archaea in past and present geobiochemical processes and elemental cycles. *Archaea*, **2013**, 930493.

Jackson SA, Flemer B, McCann A *et al.* (2013) Archaea appear to dominate the microbiome of *Inflatella pellicula* deep sea sponges. *PLoS ONE*, **8**, e84438.

Kamke J, Taylor MW, Schmitt S (2010) Activity profiles for marine sponge-associated bacteria obtained by 16S rRNA vs 16S rRNA gene comparisons. *The ISME Journal*, **4**, 498–508.

Könneke M, Bernhard AE, José R, Walker CB, Waterbury JB, Stahl DA (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, **437**, 543–546.

Langille MGI, Zaneveld J, Caporaso JG *et al.* (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, **31**, 814–821.

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, **129**, 271–280.

Liu M, Fan L, Zhong L, Kjelleberg S, Thomas T (2012) Metaproteogenomic analysis of a community of sponge symbionts. *The ISME Journal*, **6**, 1515–1525.

Margot H, Acebal C, Toril E, Amils R, Puentes JF (2002) Consistent association of crenarchaeal Archaea with sponges of the genus *Axinella*. *Marine Biology*, **140**, 739–745.

Miñambres B, Olivera ER, Jensen RA, Luengo JM (2000) A new class of glutamate dehydrogenases (GDH) biochemical and genetic characterization of the first member, the AMP-requiring nad-specific GDH of *Streptomyces Clavuligerus*. *Journal of Biological Chemistry*, **275**, 39529–39542.

Moitinho-Silva L, Bayer K, Cannistraci CV *et al.* (2013) Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea. *Molecular Ecology*, **23**, 1348–1363.

Offre P, Spang A, Schleper C (2013) Archaea in biogeochemical cycles. *Annual Review of Microbiology*, **67**, 437–457.

Oksanen J, Kindt R, Legendre P *et al.* (2009) vegan: Community ecology package. R package version, 1,15-2.

Peterson PE, Smith TJ (1999) The structure of bovine glutamate dehydrogenase provides insights into the mechanism of allostery. *Structure*, **7**, 769–782.

Pires ACC, Cleary DFR, Almeida A *et al.* (2012) Denaturing gradient gel electrophoresis and barcoded pyrosequencing reveal unprecedented Archaeal diversity in mangrove sediment and rhizosphere samples. *Applied and Environmental Microbiology*, **78**, 5520–5528.

Polónia ARM, Cleary DRF, Duarte LN, de Voogd NJ, Gomes NCM (2013) Composition of Archaea in seawater, sediment and sponges in the Kepulauan Seribu reef system, Indonesia. *Microbial Ecology*, **67**, 553–567.

Preston CM, Wu KY, Molinski TF, DeLong EF (1996) A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proceedings of the National Academy of Sciences of the USA*, **93**, 6241–6246.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from http://www.R-project.org/.

Radax R, Rattei T, Lanzen A *et al.* (2012) Metatranscriptomics of the marine sponge *Geodia barretti*: tackling phylogeny and function of its microbial community. *Environmental Microbiology*, **14**, 1308–1324.

Reigstad LJ, Richter A, Daims H, Urich T, Schwark L, Schleper C (2008) Nitrification in terrestrial hot springs of Iceland and Kamchatka. *FEMS Microbiology Ecology*, **64**, 167–174.

Renema W (2010) Is increased calcarinid (foraminifera) abundance indicating a larger role for macro-algae in Indonesian Plio-Pleistocene coral reefs? *Coral Reefs*, **29**, 165–173.

Rinke C, Schwientek P, Sczyrba A *et al.* (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature*, **499**, 431–437.

Schöttner S, Pfitzner B, Grünke S, Rasheed M, Wild C, Ramette A (2011) Drivers of bacterial diversity dynamics in permeable carbonate and silicate coral reef sands from the Red Sea. *Environmental Microbiology*, **13**, 1815–1826.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tang K, Liu K, Jiao N, Zhang Y, Chen CTA (2013) Functional metagenomic investigations of microbial communities in a shallow-sea hydrothermal system. *PLoS ONE*, **8**, e72958.

Tavaré S (1986) Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences*, **17**, 57–86.

Taylor MW, Radax R, Steger D, Wagner M (2007) Sponge associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews*, **71**, 295–347.

Thacker RW, Freeman CJ (2012) Sponge-microbe symbioses: recent advances and new directions. *Advances in Marine Biology*, **62**, 57–111.

Tourna M, Stieglmeier M, Spang A *et al.* (2011) *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. *Proceedings of the National Academy of Sciences of the USA*, **108**, 8420–8425.

Uriarte M, Marina A, Ramón-Maiques S, Fita I, Rubio V (1999) The carbamoyl-phosphate synthetase of *Pyrococcus furiosus* is enzymologically and structurally a carbamate kinase. *Journal of Biological Chemistry*, **274**, 16295–16303.

Van Soest RWM, Erpenbeck D, Alvarez B (2002) Family Dictyonellidae Van Soest, Diaz & Pomponi, 1990. In: *Systema Porifera* (eds Hooper JNA, Van Soest RWM, Willenz P), pp. 773–786. Kluwer Academic/Plenum Publishers, New York, New York.

de Voogd NJ, Cleary DFR (2007) Relating species traits to environmental variables in Indonesian coral reef sponge assemblages. *Marine and Freshwater Research*, **58**, 240–249.

de Voogd NJ, Cleary DFR, Hoeksema BW, Noor A, van Soest RWM (2006) Sponge beta diversity in the Spermonde Archipelago, SW Sulawesi, Indonesia. *Marine Ecology Progress Series*, **309**, 131–142.

Wang Q, Garrity G, Tiedje J, Cole J (2007) Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.

Wang S, Xiao X, Jiang L *et al.* (2009) Diversity and abundance of ammonia-oxidizing Archaea in hydrothermal vent chimneys of the juan de fuca ridge. *Applied and Environmental Microbiology*, **75**, 4216–4220.

Weisz JB, Hentschel U, Lindquist N, Martens CS (2007) Linking abundance and diversity of sponge-associated microbial communities to metabolic differences in host sponges. *Marine Biology*, **152**, 475–483.

Wemheuer B, Wemheuer F, Daniel R (2012) RNA based assessment of diversity and composition of active archaeal communities in the German bight. *Archaea-An International Microbiological Journal*, **2012**, 695826.

Yin Q, Fu B, Li B, Shi X, Inagaki F, Zhang XH (2013) Spatial variations in microbial community composition in surface seawater from the ultra- oligotrophic center to rim of the South Pacific Gyre. *PLoS ONE*, **8**, e55148.

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, **7**, 203–214.

D.F.R.C. designed the study; R.F. and N.J.d.V. collected the samples; A.R.M.P. and R.F. performed the laboratory work under the guidance of N.C.M.G.; D.F.R.C. and A.R.M.P. performed the data analysis; A.R.M.P., D.F.R.C., N.J.d.V. and N.C.M.G. wrote the manuscript.

Data accessibility

The DNA sequences generated in this study can be downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA): Accession no. SRP047468.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Supplementary methods.

Appendix S2 OTU tables.

Table S1 List of most abundant archaeal OTUs (≥200 sequences).

Table S2 List of all archaeal OTUs.

Appendix S3 Sequence alignment file.

Appendix S4 The predicted contribution of sequences assigned to *Cenarchaeales* and E2 and their relative contribution to KOs involved in the Nitrogen metabolism from each biotope: *S. carteri* (Sc), *S. massa* (Sm), *X. testudinaria* (Xt), *H. erectus* (He), sediment (Sd) and seawater (Wt).

Appendix S5 Species accumulation curves

Fig. S1 Species accumulation curves as a function of the number of sequences using resampling of archaeal 16S rRNA gene sequences from *S. massa* (Sm), *S. carteri* (Sc), *H. erectus* (He), *X. testudinaria* (Xt), sediment (Sd) and seawater (Wt).

Fig. S2 Sponge species accumulation curves as a function of the number of sequences using resampling of archaeal 16S rRNA gene sequences from *S. massa* (Sm), *S. carteri* (Sc), *H. erectus* (He) and *X. testudinaria* (Xt).