environmental microbiology

Environmental Microbiology (2010) doi:10.1111/j.1462-2920.2010.02281.x

GeneFISH – an *in situ* technique for linking gene presence and cell identity in environmental microorganisms

Cristina Moraru, Phyllis Lam, Bernhard M. Fuchs, Marcel M. M. Kuypers and Rudolf Amann*
Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany.

Summary

Our knowledge concerning the metabolic potentials of as yet to be cultured microorganisms has increased tremendously with the advance of sequencing technologies and the consequent discoveries of novel genes. On the other hand, it is often difficult to reliably assign a particular gene to a phylogenetic clade, because these sequences are usually found on genomic fragments that carry no direct marker of cell identity, such as rRNA genes. Therefore, the aim of the present study was to develop geneFISH – a protocol for linking gene presence with cell identity in environmental samples, the signals of which can be visualized at a single cell level. This protocol combines rRNA-targeted catalysed reporter deposition – fluorescence *in situ* hybridization and *in situ* gene detection. To test the protocol, it was applied to seawater samples from the Benguela upwelling system. For gene detection, a polynucleotide probe mix was used, which was designed based on crenarchaeotal *amo*A clone libraries prepared from each seawater sample. Each probe in the mix was selected to bind to targets with up to 5% mismatches. To determine the hybridization parameters, the T_m of probes, targets and hybrids was estimated based on theoretical calculations and *in vitro* measurements. It was shown that at least 30%, but potentially the majority of the *Crenarchaeota* present in these samples harboured the *amo*A gene and were therefore likely to be catalysing the oxidation of ammonia.

Received 1 February, 2010; accepted 7 May, 2010. *For correspondence. E-mail ramann@mpi-bremen.de; Tel. (+49) 421 2028 930; Fax (+49) 421 2028 790.

Introduction

The rapid advance of sequencing technologies has enabled the retrieval of many sequences from environmental samples, mostly as PCR-based or metagenomic clone libraries. Hence, our knowledge concerning the metabolic potentials of yet to be cultured microorganisms has increased tremendously. However, it is difficult to identify from which microorganism a particular sequence originates, because it is usually found on a genomic fragment that carries no phylogenetic marker, such as rRNA genes. A promising technique to resolve this is fluorescence *in situ* hybridization (FISH). Pernthaler and Amann (2004) combined catalysed reporter deposition – fluorescence *in situ* hybridization (CARD-FISH), using rRNA-targeted oligonucleotide probes, with mRNA FISH, using polynucleotide probes. Thereby, information on *in situ* transcription of a gene was linked to the identity of a cell. When present, mRNAs of a certain gene usually occur in multiple copies, thus improving, on the one hand, the detectability of the gene of interest. On the other hand, mRNAs might only be expressed when the cells need that particular function, and they are often short lived (Jain, 2002; Condon, 2003; Deutscher, 2006). Alternatively, another way to link a function to cell identity would be a FISH-based identification of single genes. Although, as opposed to mRNA, DNA is always present in the cell, it is a challenge to detect genes, because they have a low number of copies per cell, sometimes no more than a single copy.

There are already several FISH protocols for visualization of genes in microorganisms, and they can be classified into two categories. First, there are methods that use oligonucleotide probes for detection by or upon amplification of the target sequence. These methods include *in situ* PCR (Hodson *et al.*, 1995; Tani *et al.*, 1998), cycling primed *in situ* amplification (CPRINS) (Kenzaka *et al.*, 2005), *in situ* loop-mediated isothermal amplification (Maruyama *et al.*, 2003), *in situ* rolling circle amplification (Maruyama *et al.*, 2005), and peptide nucleic acid-assisted rolling circle amplification FISH (Smolina *et al.*, 2007). Second, there are methods that use polynucleotide probes, such as the FISH-based detection of DNA fragments > 5 kb (Niki and Hiraga, 1997; Jensen and Shapiro,

1999), and RING-FISH, which targets single genes (Zwirglmaier *et al*., 2004). The methods in the first category amplify the target DNA either by PCR-based technologies (e.g. *in situ* PCR, CPRINS) or by isothermal amplification-based technologies (e.g. *in situ* loop-mediated isothermal amplification, *in situ* rolling circle amplification and peptide nucleic acid-assisted rolling circle amplification FISH). The disadvantage of using PCR-based technologies is the need for high temperatures, which often damages the cells. Although the methods using isothermal amplification technologies do not need high temperatures, these protocols require many short reporter probes and/or primers, which in complex environmental samples might raise probe specificity issues. The methods in the second category use polynucleotide probes with multiple labels. In RING-FISH, the signal has the shape of a halo around the cell, as a result of the formation of a molecular network by the probe. Since the formation of the probe network must be considered in probe design, it is difficult to predict the specificity of RING-FISH, which has hindered its application in complex environments. In addition, although the FISH method of Niki and Hiraga (1997) is robust for genomic fragments > 5 kb, this represents a size much larger than the average size of bacterial and archaeal genes. To our knowledge, there is therefore no robust protocol for the identification of gene fragments with a size of < 0.5 kb in complex environmental samples.

Such a FISH technique linking gene presence with cell identity could be applied to address some lingering ecological questions. One example would be whether the many putative ammonia monooxygenase sequences recently retrieved from various marine environments truly belong to *Crenarchaeota. Crenarchaeota*, especially those in Marine Group I, are widespread in marine environments, and they dominate microbial communities in the mesopelagic ocean with relative abundances of up to 40% of total microbial communities (Karner *et al*., 2001; Teira *et al*., 2006; Schattenhofer *et al*., 2009). Recent studies suggest that they might be capable of ammonia oxidation, a trait that has been shown previously exclusive to two specialized groups of *Proteobacteria* (Prosser and Nicol, 2008). Ammonia oxidation is mediated by the enzyme ammonia monooxygenase (AMO) and *amo*A, the gene encoding the alpha subunit of AMO, is often used as a functional gene marker. It shows congruent phylogeny to that of 16S rRNA (Purkhold *et al*., 2000). A first indication that *Crenarchaeota* possess ammonia oxidation potential came from the assembly of *amo*-like genes on archaeal scaffolds from the Sargasso Sea metagenomic dataset (Venter *et al*., 2004; Schleper *et al*., 2005). Later, both the *amo* genes and the crenarchaeotal 16S rRNA were found in the same metagenomic fragment from a soil clone library (Treusch *et al*., 2005). However, the first direct evidence came from the isolation of the crenarchaeon *Candidatus* 'Nitrosopumilus maritimus' from a marine aquarium. This crenarchaeon has been shown to oxidize ammonia to nitrite, via hydroxylamine, and it carries the archaeal type of putative *amo*A gene (Könneke *et al*., 2005). Based on the similarity between the 16S rRNA and *amo*A genes from *Candidatus* 'Nitrosopumilus maritimus' with sequences retrieved from different seawater samples, it has been suggested that Marine Group I *Crenarchaeota* present in marine waters also possess *amo*A genes and thus might play an important role in the marine nitrogen cycle (Prosser and Nicol, 2008). Additional evidence for this is the resemblance of the kinetics of ammonia oxidation by *Candidatus* 'Nitrosopumilus maritimus' to that of *in situ* nitrification in marine systems (Martens-Habbena *et al*., 2009). A number of studies have quantified both the gene copy number of archaeal 16S rRNA and putative *amo*A genes, although the relationships between these two in environmental samples were only inferred indirectly (Lam *et al*., 2007; Mincer *et al*., 2007; Agogue *et al*., 2008; Park *et al*., 2008). To date, there has been no *in situ* localization of these environmental *amo*A sequences in single crenarchaeotal cells.

The aim of the present study was thus to develop a protocol to directly link gene presence with cell identity in environmental samples, namely geneFISH. This protocol combined CARD-FISH, using rRNA-targeted oligonucleotide probes for single cell identification, with the detection of gene fragments using short polynucleotide probes. The method of rRNA-targeted CARD-FISH was previously developed (Pernthaler *et al*., 2002) and has since been used frequently in molecular ecology to study the composition of microbial communities. The gene detection method was adapted from Pernthaler and Amann (2004) and Wagner and colleagues (1998). It uses multiple digoxigenin (Dig)-labelled polynucleotide probes to target genes, followed by the binding of horseradish peroxidase (HRP)-conjugated antibodies and catalysed reporter deposition (CARD) to amplify and visualize the signal. This geneFISH protocol was first developed and tested in *Escherichia coli* clones with copy control plasmids, which were obtained from an *amo*A clone library. In a second phase, it was applied to seawater samples from the Benguela upwelling system on the Namibian shelf, in which the presence of the putative *amo*A gene was directly visualized in crenarchaeotal cells that were identified by 16S rRNA-targeted probes. This involved a specially designed polynucleotide probe mix (*amo*A-Nam) that targeted the crenarchaeotal putative *amo*A alleles present in these environmental samples.

Results

Polynucleotide probe design for the crenarchaeotal amo*A gene*

The design of appropriate polynucleotide probes targeting crenarchaeotal putative *amo*A genes started with the analyses of clone libraries of PCR-amplified *amo*A gene fragments. Three samples were collected from two stations in the Benguela upwelling system on the Namibian shelf, onboard the R/V Meteor in May/June 2008 during the M76/2 cruise: station 249, 128 m depth, station 213, 65 m depth and station 213, 76 m depth. For each *amo*A clone library, 163–198 clones were selected for sequencing. The resulting sequences were aligned and a region of 351 bp (position 95–445 of *amo*A, *Cenarchaeum symbiosum* numbering) was selected for probe design. For protocol development, we first tried to detect the *amo*A gene fragment in three *E. coli* clones. A single polynucleotide probe was designed, *amo*A-1E3, based on clone 1E3 (clone library from station 213, 76 m). This 351 bp probe had the following percentage of mismatches with the *amo*A clone inserts: 0% with the 1E3 clone, 0.6% with the 1E7 clone and 5% with the 3G4 clone.

A maximum likelihood tree was constructed that included all *amo*A sequences from this study (Fig. S1). Most of the sequences grouped within the previously described marine clusters A (here denoted A1) and B, with a lower number grouping close to *Nitrosopumilus maritimus* and in a new marine cluster, A2, which was well separated from cluster A1 (Fig. S1). For detection of the *amo*A gene in the Namibian seawater samples, the diversity of the 351 bp *amo*A region used for probe design needed to be analysed. More than 100 unique sequences were retrieved from each clone library. Operational taxonomic unit (OTU) grouping reduced the number of sequences to ~50 (99% identity cut-off) and to ~14 OTUs (95% identity cut-off) (Table 1). The two diversity estimators used (Chao and Jackknife, see Table 1) and the rarefaction curves (Fig. S2) indicated that, while the clone libraries did not capture the full *amo*A diversity at a 99% identity cut-off, most OTUs were retrieved if the 95% identity cut-off was used. At a 99% identity cut-off, none of the OTUs represented more than 10% of the sequences. At a 95% cut-off, between 70% and 80% of the sequences were found in only three OTUs (Fig. S3). The three clone libraries were examined using Venn diagrams and were found to be similar, sharing 49% to 58% of the sequences between them (Fig. S4).

The sequences from the three clone libraries were pooled for probe design. This was carried out with PolyPro software, using a mismatch threshold 1 (Th1) of 5% (Moraru *et al*., 2010). Polynucleotide probes bind to mismatched targets, but further studies are necessary to establish the maximum mismatch percentage for hybridization. A Th1 of 5% mismatches guarantees that the probes will bind to the respective targets (Wetmur, 1991). For a detailed description of the probe design see Fig. S5 through Fig. S10 and Table S1. The resulting probe mix (*amo*A-Nam) contained 12 polynucleotides, with a ΔT_m of 1.7°C (Table S2). From the total of 552 *amo*A target sequences, 77% were covered by two of the polynucleotides, while the remaining 23% were covered by the other 10 polynucleotides (Fig. 1). From the 12 probes, three belonged to cluster A1, seven to cluster B, one to cluster A2 and one to the *Nitrosopumilus*-like cluster (see Table S1). At a Th1 of 5%, none of the probes had hits outside their own cluster (Fig. S11).

Determination of the stringency parameters for hybridization with polynucleotide probes

Determination of the stringency parameters to be used for hybridization requires the knowledge of the melting temperature (T_m) of the hybrid formed between probe and target. Two methods were used to determine T_m: theoretical calculations and *in vitro* measurements. For the former, a formula for double-stranded DNA (dsDNA) implemented in the PolyPro software (Moraru *et al*., 2010) was used to calculate the T_m. For *in vitro* measurements, a protocol for T_m determination with dsDNA binding dyes and real-time fluorescence detection (Monis *et al*., 2005; Gudnason *et al*., 2007) was modified and used (see the text in *Supporting information* and Figs S12 and S13).

In this *in vitro* method, the fluorescence conferred by the intercalating dye SYTO 9 is high when the DNA is double-stranded, while fluorescence level drops when the DNA melts and becomes single-stranded. The T_m

Table 1. Diversity of *amo*A gene (the 351 bp fragment used for probe design) in the three Namibian seawater samples – clone libraries results.

Clone library	Total sequences	Unique OTUs	99% identity OTUs			95% identity OTUs		
			Observed	Chao	Jackknife	Observed	Chao	Jackknife
Station 249, depth 128 m	191	126	58	103	110	14	17	18
Station 213, depth 65 m	198	135	52	75	77	13	13	15
Station 213, depth 76 m	163	105	44	65	64	13	13	15
All samples	552	286	96	135	136	19	20	22

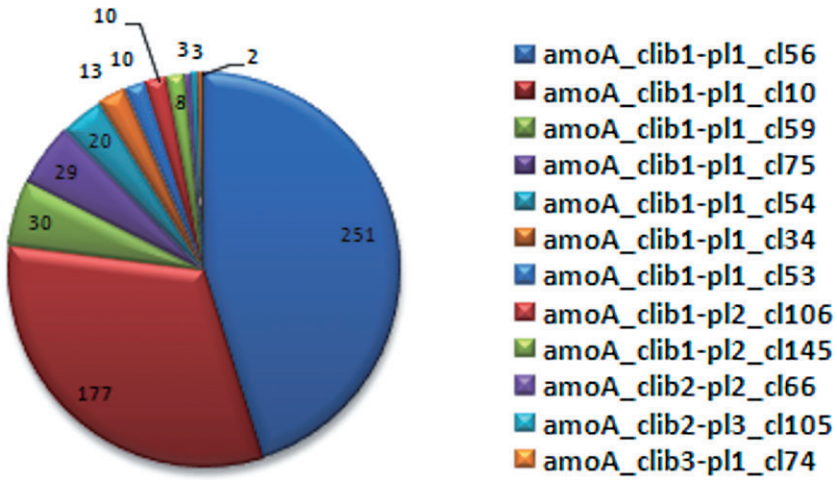

Fig. 1. Namibian seawater samples – coverage of *amo*A targets by the 12 polynucleotides in the *amo*A-Nam probe (number of targets hit by each polynucleotide). A sequence is a target for a certain probe when the percentage mismatch between it and the probe is between 0% and 5%. The legend shows the probe names.

appears as a distinct peak when the negative first derivative of the melting curve is plotted against temperature (Fig. 2). The original protocol was modified to measure the T_m in hybridization-like and washing-like buffers. These buffers had the exact composition of the hybridization and washing buffers used for gene detection, except that they did not contain the blocking reagents (see *Experimental procedures*). For the *amo*A-1E3 probe, we compared the T_m of (i) probe DNA, a dsDNA labelled with digoxigenin (Dig), (ii) target DNA, the same dsDNA, but without Dig, and (iii) probe–target hybrid, the same dsDNA, but with one strand with Dig but the other strand without. In the hybridization-like buffer, the measured T_m were 66.2 ± 0.1°C, 68.9 ± 0.0°C, 67.2 ± 0.1°C for probe, target and hybrid respectively (Fig. 2). In washing-like buffer, the measured T_m were higher: 71.9 ± 0.4°C and 75.2 ± 0.1°C for probe and target respectively.

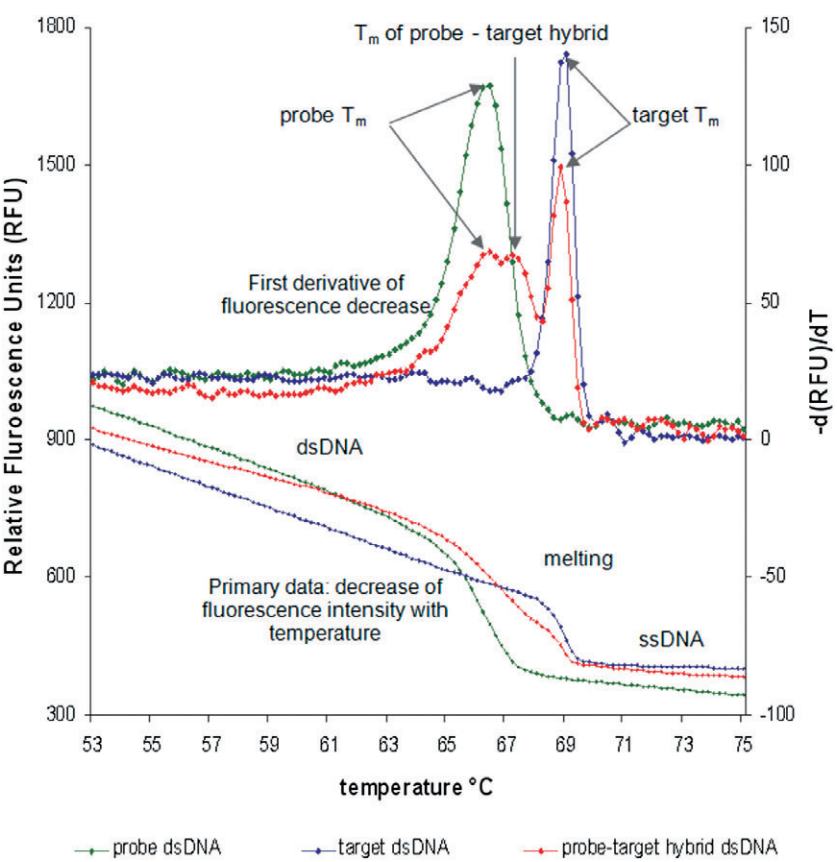

Fig. 2. *In vitro* measurements of the melting temperature (T_m) for the *amo*A-1E3 probe (green) (both strands with Dig), target (blue) (neither of the strands with Dig) and hybrid (red) (one strand with Dig and one without). The measurements were performed in hybridization-like buffer, with 35% formamide and 1.718 M Na^+. The primary data represent the fluorescence (expressed as relative fluorescence units – RFU) decrease with increasing temperature. The first derivative shows the T_m as a peak. In the hybrid reaction, three T_m peaks appear: one peak for the probe duplex, one peak for the hybrid duplex and one peak for the target. To identify the peaks, comparison with the individual reactions for probe and target was necessary.

In comparison, the T_m calculated using the PolyPro software (Moraru *et al*., 2010) for the target DNA were 73.4°C in hybridization-like buffer and 69.9°C in washing-like buffer. For the *amo*A-Nam probe mix, the mean values for the calculated T_m of the perfectly matched targets were 73.4°C in hybridization-like buffer and 69.9°C in washing-like buffer (Table S2), which were similar to those calculated for *amo*A-1E3. To calculate the T_m of the mismatched hybrids, the same formula as for perfectly matched hybrids was used, with the addition of the term for mismatches. It was considered that a 1% mismatch gives a 0.5–1.5°C decrease in the T_m, compared with the perfectly matched hybrid (Anderson and Young, 1985). Therefore, the calculated T_m for 5% mismatched hybrids were in the range of 65.2–71.9°C for hybridization-like buffer and 61.7–68.4°C for washing-like buffer.

Based on the above T_m values, the denaturation temperature, hybridization temperature and washing temperature of the *amo*A-1E3 probe were determined. The denaturation temperature was 75°C, which was chosen to be above the T_m of the probe and target duplexes in hybridization-like buffer. The hybridization temperature was calculated as follows: 67.2 $\pm$ 0.04°C (T_m of the hybrid in hybridization-like buffer) – 25°C (for optimum hybridization rate) = 42°C. The washing temperature was 42°C, which was chosen to be below the T_m in washing-like buffer. For the *amo*A-Nam probe mix, the same parameters as for the *amo*A-1E3 probe were used.

Development of the geneFISH protocol – E. coli *clones experiment*

The protocol was developed on pure cultures of *E. coli* EPI300 clones with crenarchaeotal *amo*A-containing plasmids, including clones 1E3, 1E7 and 3G4, with 0%, 0.6% and 5% mismatches, respectively, with the *amo*A-1E3 probe. A culture of *E. coli* K12 strain, without *amo*A inserts, was included as a negative control. The copy number of *amo*A genes in 1E3 and 3G4 clones was between 1 and 2 per cell, and the 1E7 clone was induced to a high copy number (10–200 per cell – according to kit manual and Wild *et al*., 2002).

The main steps of the geneFISH protocol are illustrated in Fig. 3. They consisted of two major parts: rRNA CARD-FISH, followed by gene detection with a dsDNA polynucleotide probe. The rRNA CARD-FISH part was performed as detailed elsewhere (Pernthaler *et al*., 2002). This step allowed identification of single cells via hybridization with an rRNA-targeted oligonucleotide probe. The gene detection part was based on polynucleotide hybridization and included two steps of signal amplification (Fig. S14). It started with the binding of the polynucleotide probe, which had multiple Dig labels. Antibodies conjugated with HRP were then added to bind to the Dig molecules. This was followed by a final amplification step, where the antibody-bound HRP catalytically deposited many fluorescently labelled tyramides.

A critical parameter for a good signal-to-noise ratio was the gene probe concentration. Different probe concentrations were tested – 0.25 pg μl^{-1}, 2.5 pg μl^{-1}, 25 pg μl^{-1} and 250 pg μl^{-1}. The hybridization time was 18–22 h or 41 h. The results are summarized in Table 2. The detection efficiency increased with probe concentration from 24.0 $\pm$ 0.1% at 0.25 pg μl^{-1} to 44.3 $\pm$ 3.5% at 250 pg μl^{-1} (clone 1E3 – 0% mismatches – 18–22 h). On the other hand, the false positive level also increased, from 2.5 $\pm$ 3.4% at 0.25 pg μl^{-1} to 19.0 $\pm$ 2.4% at 250 pg μl^{-1} (*E. coli* K12 – negative control – 18–22 h). The probe concentration which gave both a high detection efficiency and a low background was 2.5 pg μl^{-1}, with 42.8 $\pm$ 2.4% detection efficiency and 1.4 $\pm$ 1.1% false positives, followed close by 25 pg μl^{-1}, with 41.0 $\pm$ 6.2% and 4.4 $\pm$ 3.0% respectively. Longer incubation times did not result in a significant increase in the detection efficiency. At optimum probe concentrations of 2.5 pg μl^{-1} and 18–22 h hybridization, the detection efficiency was 34.2 $\pm$ 1.6% for the 3G4 clone (5% mismatches) and 92.0 $\pm$ 1.7% for the 1E7 clone (0.6% mismatches, high-copy-number clone). The geneFISH signals obtained were dot-like (Fig. 4), with most of the positive cells having one dot per cell. An exception was the 1E7 clone, which possessed a higher number of gene copies per cell and where, consequently, multiple dots were observed (Fig. S15).

Table 2. GeneFISH on *Escherichia coli* clones with *amo*A-1E3 probe – gene detection efficiency (percentage of cells showing a gene signal).

Probe concentration (pg μl^{-1})	*E. coli* cl1E3 (0% mismatches), 1–2 *amo*A copies per cell		*E. coli* cl3G4 (5% mismatches), 1–2 *amo*A copies per cell		*E. coli* 1E7 (0.6% mismatches), 10–200 *amo*A copies per cell		*E. coli* K12 (target not present)	
	18–22 h	41 h	18–22 h	41 h	18–22 h	41 h	18–22 h	41 h
0.25	24.0 $\pm$ 0.1%	–	–	–	–	–	2.5 $\pm$ 3.4%	–
2.5	42.8 $\pm$ 2.4%	41.4 $\pm$ 1.3%	34.2 $\pm$ 1.6%	29.9 $\pm$ 2.4%	92.0 $\pm$ 1.7%	96.9 $\pm$ 1.3%	1.4 $\pm$ 1.1%	2.6 $\pm$ 1.5%
25	41.0 $\pm$ 6.2%	40.0 $\pm$ 2.3%	34.0 $\pm$ 3.6%		99.0 $\pm$ 0.7%	98.1 $\pm$ 1.2%	4.4 $\pm$ 3.0%	6.0 $\pm$ 0.8%
250	44.3 $\pm$ 3.5%	–	–	–	–	–	19.0 $\pm$ 2.4%	–

–, not determined.

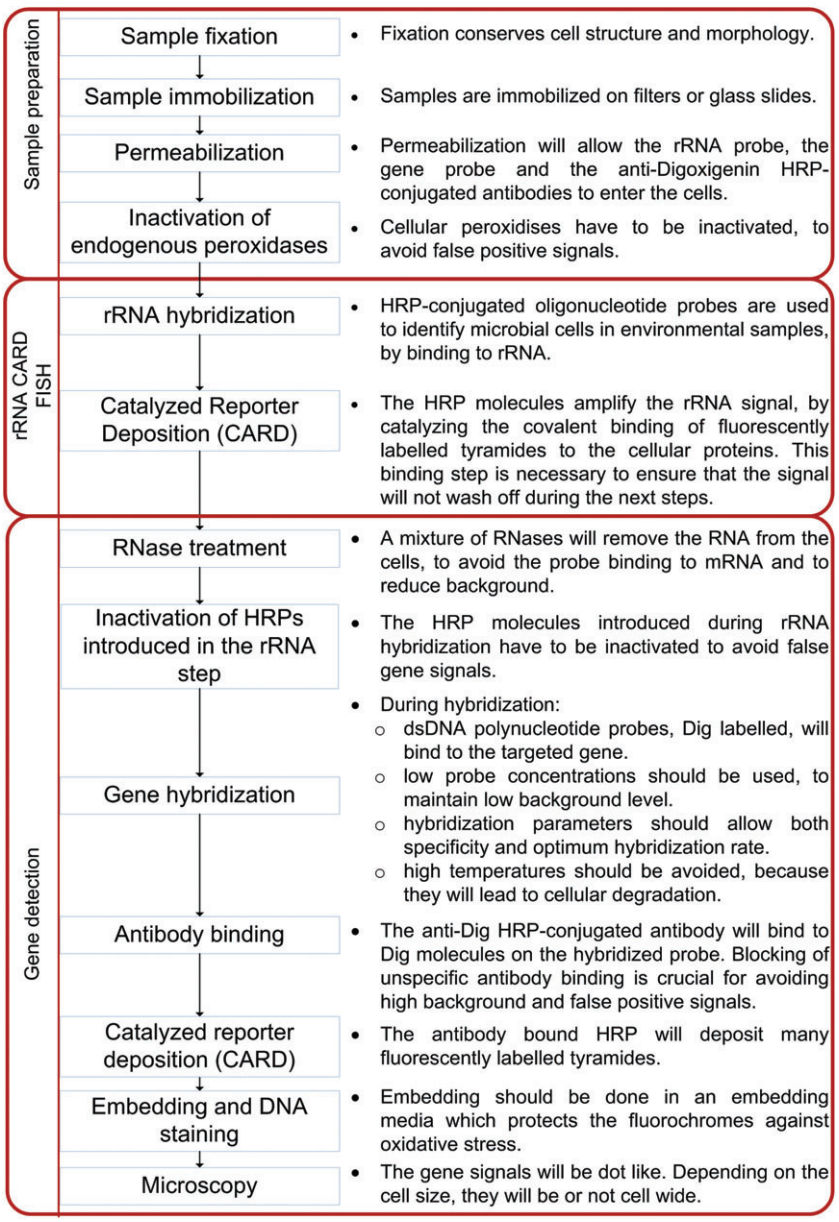

Fig. 3. GeneFISH protocol.

Detection of the crenarchaeotal amo*A gene in Namibian seawater samples*

The geneFISH protocol was applied for the Namibian seawater samples, with modifications in the preceding steps of permeabilization, peroxidase inactivation, rRNA hybridization and rRNA CARD steps, as previously found necessary for hybridizing *Crenarchaeota* (see *Experimental procedures*). The gene probe used was the *amo*A-Nam probe mix and the two negative controls were DNA digestion and the NonPolyPr350 probe. The NonPolyPr350 probe was synthesized from an artificial sequence, which gave no significant resemblance via BLASTN searches (Johnson *et al.*, 2008) with bacterial/archaeal sequences deposited in the public database GenBank (experimental procedures in *Supporting information* and Table S6). For the three samples hybridized, we obtained *amo*A gene signals which were colocalized with crenarchaeotal rRNA probe signals (Fig. 5). rRNA CARD-FISH with the Cren554 probe showed abundances of *Crenarchaeota* between 10% and 21%. From these crenarchaeotal cells, 30–34% showed positive *amo*A gene signals with the *amo*A-Nam probe, compared with ~0.7% for both DNA digestion with the *amo*A-Nam probe and NonPolyPr350 probe without DNA digestion (Table S3, Fig. S16). From the bacterial cells, only ~0.7% showed signals for the *amo*A-Nam probe, the same applying for the NonPolyPr350 probe.

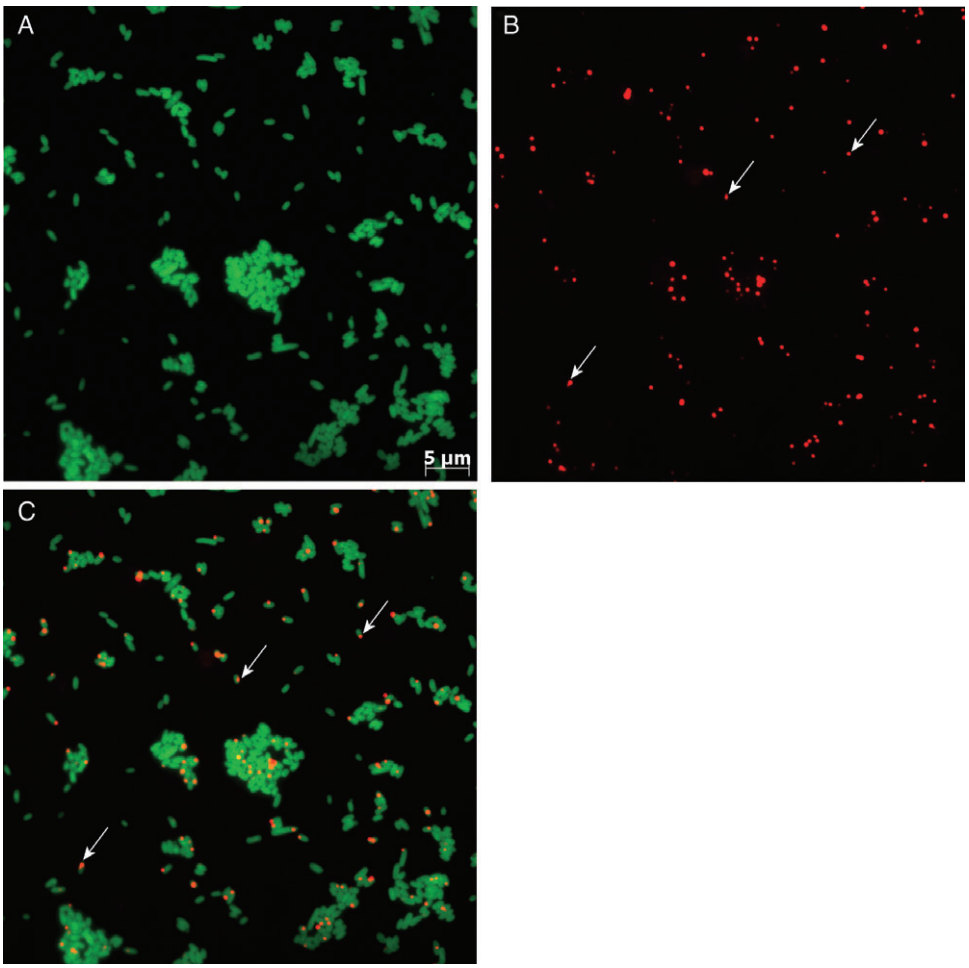

Fig. 4. GeneFISH on *Escherichia coli* clone 1E3 (one to two *amo*A copies per cell). Simultaneous 16S rRNA CARD-FISH with EUB338 (green), a general bacterial probe, (A) and gene detection with *amo*A-1E3 probe (red) targeting crenarchaeal ammonia monooxygenase subunit A (B). Arrows indicate geneFISH signals in overlay (C).

Discussions

Probe design

Previous work (Moraru *et al*., 2010) showed that the diversity of functional genes, as inferred from cultured microorganisms, is too high to allow the design of a single polynucleotide probe mix which targets all alleles of a functional gene. Therefore, we decided to design a specific probe mix for our particular environmental samples. An assessment of the *amo*A gene diversity in the three Namibian seawater samples showed the presence of a high number of alleles: 286 OTUs were retrieved when using a 99% identity cut-off to allow for *Taq* polymerase errors (Acinas *et al*., 2005). Despite this observed high diversity, it was possible to design a mix consisting of 12 probes, *amo*A-Nam, which covered all targets retrieved by the clone libraries and which, due to a low ΔT_m of 1.7°C, could be used as a single probe.

One question that might arise is whether the diversity had been sampled sufficiently by our cloning and sequencing efforts, such that the *amo*A-Nam probe mix truly covered the majority of the *amo*A alleles in these environmental samples. Since the mismatch threshold for probe design (Th1) was set to 5% mismatch, the coverage of the clone libraries should be estimated using the 95% identity cut-off for OTU grouping. Results (Table 1, Fig. S2) indicated that the 19 OTUs retrieved represented > 80% of the diversity. From the 12 polynucleotides of the *amo*A-Nam probe mix, two belonging to cluster A1 targeted 77% of all sequences. The other 10 probes covered only between 0.36% and 5.43% of the sequences (Fig. 1) but, on the other hand, they targeted cluster B, A2 and *Nitrosopumilus*-like sequences (Fig. S11). Although we took measures to reduce the PCR bias, we did not use the clone libraries to infer cellular abundances of the different clades. Some sequences present in small numbers in the clone libraries could be numerically abundant in environmental samples. Therefore, all 12 polynucleotides were used together. Although a Th1 of 5% was used for designing the *amo*A-Nam probe mix, we expect that the probes

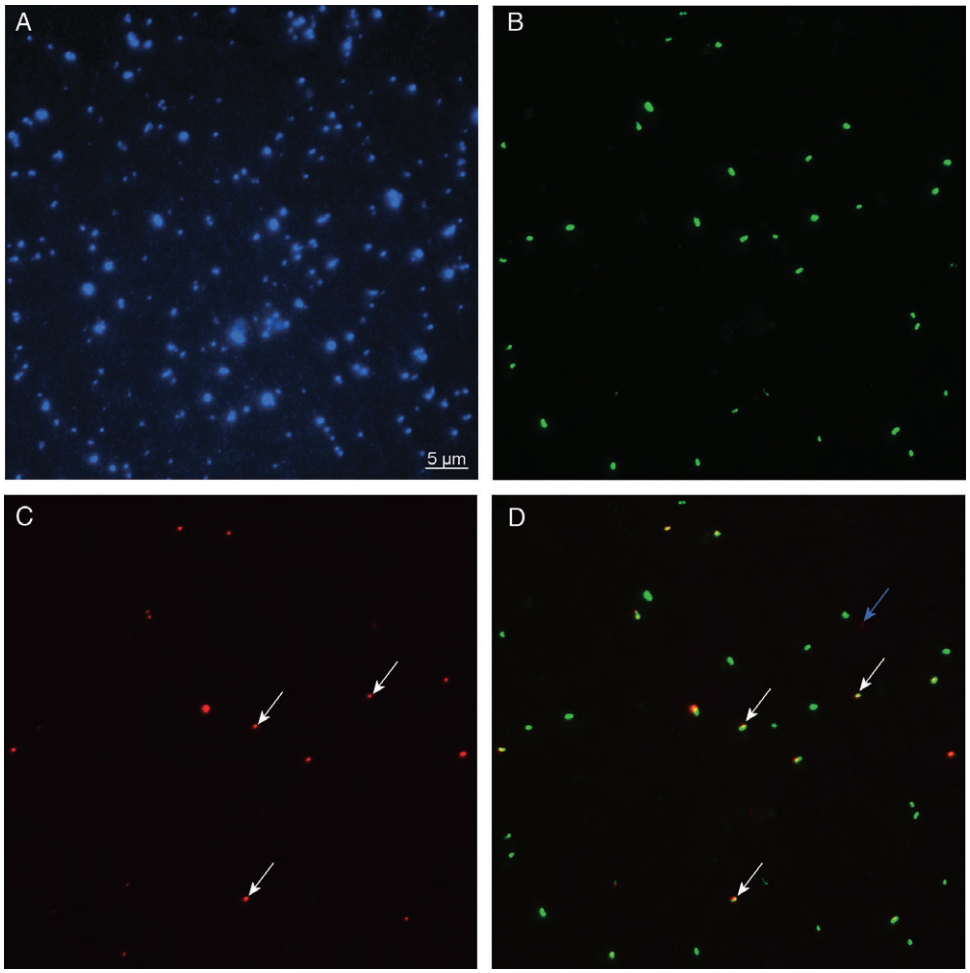

Fig. 5. GeneFISH on the Namibian seawater samples, station 249. All cells stained with DAPI (blue) (A). Simultaneous 16S rRNA CARD-FISH with Cren554 probe (green) for marine *Crenarchaeota* (B) and gene detection with *amo*A-Nam probe mix (red) for ammonia monooxygenase subunit A gene (C). White arrows indicate geneFISH signals in overlay and the light-blue arrow indicates non-specific signal (D).

will bind targets with higher percentage mismatches. However, as previous studies (Wetmur, 1991) and our results with clone 3G4 indicated (Table 2), the gene detection efficiency will decrease with the increasing percentage mismatch. Using only the two probes that covered 77% of the sequences would have resulted in low or no detection of targets from clusters B, A2 and *Nitrosopumilus*-like *amo*A. Therefore, it can be safely assumed that the *amo*A-Nam probe mix binds to the majority of the *amo*A alleles in the Namibian seawater samples.

It can even be possible that this probe mix is suitable for targeting crenarchaeotal *amo*A genes in habitats other than Namibian seawater samples. However, it is highly advisable to always check the specificity of the *amo*A-Nam mix against the allele diversity in the respective habitats and design additional probes or completely new probe mixes if necessary.

Determination of the stringency parameters for hybridization with polynucleotide probes

The stringency parameters for hybridization included the composition of hybridization and washing buffers (Na^+ and formamide concentration), as well as hybridization and washing temperatures. The ideal parameters should allow for: (i) specific hybridization, (ii) optimum hybridization rate and (iii) preservation of cellular morphology (low temperatures). The choice for all these parameters in turn depended on the T_m of the probe–target hybrids.

The crenarchaeotal *amo*A genes are so distant even from their closest bacterial homologue (Dunfield *et al*., 2007) that it is very difficult to obtain a good alignment at the nucleic acid level with the bacterial *amo*A. The number of mismatches to our crenarchaeotal probes was well beyond the 20–30% mismatches reported by Wetmur (1991) to be the limit for hybridization. Hence, no binding

was expected outside the crenarchaeotal *amo*A clade for both the *amo*A-1E3 and *amo*A-Nam probes. Therefore, the only criteria for choosing the hybridization temperature and composition of the hybridization buffer were maintenance of cellular morphology and an optimum hybridization rate.

To avoid degradation of the cellular morphology, the hybridization temperature had to be in the range 42–50°C (data not shown). To have an optimum hybridization rate, the hybridization temperature was set to ~25°C below the T_m of the perfectly matched hybrid, according to previous findings for DNA:DNA hybrids (Wetmur, 1991). As a result, the composition of the hybridization buffer was chosen to give a T_m in the range 67–75°C (see next section and *Experimental procedures*). For the *amo*A-Nam probe mix the same parameters as for *amo*A-1E3 probe were used. Because the *amo*A-Nam probe mix is targeting alleles up to 5% mismatches, it is obvious that the choice of the hybridization temperature cannot be the optimum value for all hybrids. Assuming a 0.5°C decrease in T_m for 1% mismatch, the deviation around the optimum would be between −1.95°C and +1°C, while assuming a 1.5°C decrease in T_m for 1% mismatch, the maximum deviation around the optimum would be between −4.45°C and +1°C (for mismatched hybrids, the reduction in the optimum hybridization temperature was considered to be half the reduction in T_m, according to Bonner *et al.* (1973)). This is placing many of the hybrids rather on the stringent side of the hybridization curve, but still close to their optimum hybridization rate.

The washing step is important to melt short hybrids that might form during hybridization (Wetmur, 1991) between parts of the probe and random genomic DNA fragments. For this, a low-salt buffer (washing buffer II) was used, as detailed in the next section and *Experimental procedures*. The choice for the washing temperature was made according to the T_m values for the washing-like buffer and it was lower than the T_m, in order to avoid melting of the *amo*A probe–target hybrids, both perfectly matched and mismatched.

The two methods used to estimate T_m gave slightly different results, which was expected, as the theoretical calculations will only provide an estimation (Moraru *et al*., 2010). Due to the saturation effect that high salt concentrations have on T_m (Hutton, 1977), it is expected that the *in vitro* measurements gave a good T_m in hybridization-like buffer. On the other hand, the low salt concentration in the washing-like buffer most probably led to a T_m overestimation due to the use of the dsDNA binding dyes, which are known to enhance the T_m (Gudnason *et al*., 2007). The enhancement depends on the type of dye and concentration used, and it is probably due to a 'salt-like' effect (Bjorndal and Fygenson, 2002). We used SYTO 9, which has been shown to give the lowest increase in T_m with increasing dye concentration (Monis *et al*., 2005; Gudnason *et al*., 2007). While bearing in mind that both the measured and the calculated values did not give the exact T_m, this range of values can still be used to determine the parameters for hybridization.

GeneFISH protocol

The protocol for gene detection was adapted from previously developed protocols for mRNA FISH (Wagner *et al*., 1998; Pernthaler and Amann, 2004). Our initial calculations indicated that the signal amplification system used in our protocol may deposit between 460 and 2000 fluorochromes per probe: a 300–400 nucleotide probe may have 6–16 Dig labels (Yu *et al*., 1994), each of which can bind an antibody conjugated with approximately three HRP molecules and each HRP molecule will in turn deposit between 26 and 41 fluorescently labelled tyramides (Hoshino *et al*., 2008). Previous studies using multiple Cy3-labelled probes (Niki and Hiraga, 1998; Viollier *et al*., 2004) indicated that around 200–400 fluorochromes in one localized spot were sufficient for the signal to be detected. Therefore, the signal amplification system used here should allow visualization of single targets.

The main difference between the geneFISH and the mRNA FISH protocols was the use of dsDNA probes in the former instead of ssRNA probes in the latter. Since the DNA:DNA hybrids have lower T_m and lower temperature for an optimum hybridization rate than the RNA:DNA hybrids (Wetmur and Davidson, 1968; Birnstiel *et al*., 1972; Hutton, 1977), the use of a dsDNA probe reduces the hybridization temperature. Therefore, the cellular morphology is better preserved. Moreover, DNA probes are less prone to degradation compared with RNA probes, which can be degraded by contaminating RNases, high temperatures or some chemicals (e.g. Mg^{2+}) (Watson *et al*., 1984; Kierzek, 1992). Due to the double-stranded nature of the DNA probes, an RNase treatment was introduced (discussion in *Supporting information*) to ensure the signal originated from the gene and not from the mRNA. This was necessary for method development. In future use of the geneFISH protocol, this step can be omitted, the omission potentially resulting in an increase in the detection efficiency.

Another difference from the mRNA FISH protocols was the composition of the hybridization and washing buffers. To increase hybridization kinetics, a higher Na^+ concentration was used in the hybridization buffer. It was 1718 mM (Table S4) compared with 1050 mM (Pernthaler and Amann, 2004) and 75 mM (Wagner *et al*., 1998) in earlier studies. The post-hybridization washes were similar to the ones used in Southern/Northern Blots (Sambrook, 2001) and included washing buffer I (WBI) with

390 mM Na^+ and 0.1% SDS, which was used to remove unbound probe molecules (Wetmur, 1991), and washing buffer II (WBII) with 23 mM Na^+ and 0.1% SDS (Table S5), used as a stringent washing. For probes with higher %GC (55–65%GC) or for cases when discrimination between similar alleles is required, the stringency of the WBII can be increased by formamide addition. Alternatively, the washing temperature can be increased.

As opposed to the mRNA FISH protocols, where probe concentrations of 250 pg μl^{-1} (Pernthaler and Amann, 2004) or 25 ng μl^{-1} (Wagner *et al*., 1998) were used, it was found that the best signal to noise ratio was obtained at probe concentrations of 2.5 pg μl^{-1} (Table 2). Higher probe concentrations increased the false positive signals ($19.0 \pm 2.4\%$ for 250 pg μl^{-1}), while lower probe concentrations decreased the detection efficiency (from $42.8 \pm 2.4\%$ cells hybridized at 2.5 pg μl^{-1} to $24.0 \pm 0.1\%$ cells hybridized at 0.25 pg μl^{-1}), with no decrease in the number of false positives. The optimum hybridization time was between 18 and 22 h. Longer times failed to increase the hybridization efficiency significantly.

The initial steps of sample preparation (fixation, permeabilization) were necessary to retain cell morphology and, in the same time, to allow large reagents to enter the cells. Together with the peroxidase inactivation steps, they are discussed in *Supporting information*. The antibody binding and the CARD for gene detection steps generally followed the protocol from Pernthaler and Amann (2004) and Pernthaler and Pernthaler (2005), with the main exception being the double dextran sulfate concentration in the CARD amplification buffer, to enhance the tyramide deposition (Kubota *et al*., 2006; see *Experimental procedures* and *Supporting information*).

In the case of the copy control clones, not all the cells carrying the *amo*A fragment had a gene signal. This can possibly be explained by the low number of targets per cell (one to two copies). Hybridizations are equilibrium reactions in which far less than 100% of the targets are bound by the probe. When the target number is 1–2 per cell, as in our study, this will result in part of the cells not having a gene signal at all. Moreover, different treatments necessary during the protocol are known to damage DNA: paraformaldehyde, acids, H_2O_2 and high temperatures (Raap *et al*., 1986). DNA degradation by any of these treatments can lead to a reduction in the number of targets, including a loss of target. This explanation for a hybridization efficiency of ~42% is supported by our experiments with the induced clone, where the increase in the number of targets per cell resulted in an increase of the hybridization efficiency to 90–100% (Table 2).

The presence of ~1.5% false positives for the *E. coli* experiment was most probably due to the use of antibodies and to the dot-like signals characteristic for geneFISH. Such a small dot would not be considered as a false positive for techniques where the whole cell is stained, such as rRNA FISH. Surprisingly, in the Namibia seawater samples the false positive signals were lower (~0.7%) than in the experiments with *E. coli*. The presence of false positives, combined with a lower detection efficiency, will make the application of geneFISH for the detection of genes present in rare populations more difficult.

In the case of the Namibian seawater samples, ~30% of the crenarchaeotal cells identified by 16S rRNA-targeted CARD-FISH were detected carrying an *amo*A gene signal. This provided the first direct visual link between single-celled *Crenarchaeota* in these seawater samples and the presence of putative *amo*A genes. In fact, the vast majority of *amo*A signals originated from cells identified as *Crenarchaeota*, the bacterial cells showing only background-level gene signals. Compared with the detection efficiency in *E. coli*, which had one to two target copies per cell (~42% for perfectly matched hybrids), the percentage of *amo*A signals (~30%) in these Namibia seawater samples suggested that likely the majority of crenarchaeotal cells in these seawater samples possessed the *amo*A gene. However, one has to be cautious with making quantitative assumptions starting from the detection efficiency in pure cultures, since the detection efficiency of geneFISH depends on several factors, like cell wall composition or the number of targets per cell (this can vary from microorganism to microorganism and with the cell cycle) (Kubitschek and Freedman, 1971; Breuert *et al*., 2006; Lundgren *et al*., 2008).

For use in quantitative investigations, the efficiency of geneFISH would have to be increased to 100% and this might be achieved by adding a target amplification step. Such a quantitative geneFISH protocol would be an even more useful tool for addressing certain ecological questions. For instance, a previously published study using quantitative PCR reported a discrepancy between the abundances of crenarchaeotal 16S rRNA and *amo*A genes in different oceanic regions (Agogue *et al*., 2008). These results suggested that not all crenarchaeotal cells carried the *amo*A gene. This conclusion has recently been challenged for an amoA primer bias (Konstantinidis *et al*., 2009). Hence, the use of an improved geneFISH protocol, with 100% detection efficiency, may help to resolve this important question.

In addition, the application of the geneFISH protocol may be further extended to examine the relative distribution of various subpopulations of functional gene variants. For instance, more than two major clusters of crenarchaeotal *amo*A genes have been identified in this study (Fig. S1). An interesting follow-up study would now be to use polynucleotide probes to discriminate between these

clusters identified to date. A plot of the percentage mismatches of each of the 12 probes with all *amo*A sequences (Fig. S11) shows that the *amo*A-Nam mix could be split into four probe mixes. By modifying the hybridization stringency, these four probe mixes could be used to discriminate between clusters A1, A2, B, and *Nitrosopumilus*-like *amo*A. Moreover, if halogen-containing tyramides are used in the last amplification step, the geneFISH protocol could be combined with NanoSIMS analyses (Musat *et al*., 2008), to determine whether the four clusters differ in their activity and substrate uptake rates.

Two negative controls were used for the environmental samples: DNA digestion and NonPolyPr350. The negative controls were necessary to confirm that the signals represented hybridization events and did not come from endogenous or introduced peroxidases, from unspecific probe or antibody binding to cellular matrixes, cell walls, etc. Both controls gave similar results, with the false positive signals being less than 1%. This confirmed that the NonPolyPr350 was a good negative control and therefore only one control was necessary. The NonPolyPr350 control is, however, preferred, because it did not introduce an extra step in the protocol, which increases both the hands-on time as well as potential cell damage. Nevertheless, if the template for NonPolyPr350 is not available, then DNA digestion can equally be used.

While this article was under review, another study regarding single gene detection in microorganisms was published (Kawakami *et al*., 2010). The authors used locked-nucleic-acid oligonucleotide probes labelled with Dig and two rounds of the antibody-CARD steps to detect single genes in a in pure culture. They reported a lower gene detection rate (15%) and simultaneous detection of the rRNA was not performed.

Summary and conclusions

The newly developed protocol for geneFISH was successfully applied to seawater samples from the Benguela upwelling system and it showed that at least 30%, possibly almost all, of the *Crenarchaeota* present in these samples harboured the *amo*A gene. Crenarchaeotal type *amo*A genes were only found in *Crenarcheota*. For gene detection, a polynucleotide probe mix was used, which was designed based on crenarchaeotal *amo*A clone libraries prepared for each seawater sample. Each probe in the mix was designed to bind to mismatched targets with at least up to 5% mismatches. To determine the hybridization parameters, the T_m of probes, targets and hybrids was estimated based on theoretical calculations and *in vitro* measurements.

Therefore, this geneFISH protocol can be used at least qualitatively, to link gene presence with cell identity in environmental samples, with potential extension to quantitative applications, should the detection efficiency be sufficiently improved.

Experimental procedures

Environmental samples

The environmental seawater samples were collected from the Benguela upwelling system on the Namibian shelf, onboard the R/V Meteor in May/June 2008 during the M76/2 cruise. Two stations were sampled: station 249 (22.99 S, 14.04 E), at a depth of 128 m, and station 213 (20.99 S, 13.36 E) at depths of 65 m and 76 m. Samples for DNA extraction were filtered through 0.2 μm Sterivex GS filters (Millipore), and stored at –80°C until extraction. Samples for FISH were fixed by adding paraformaldehyde at a final concentration of 1% to the freshly collected seawater and incubating at 4°C for 8–14 h. The fixation was stopped by filtration and the filters were stored at –20°C or –80°C.

Crenarchaeotal amo*A clone libraries*

DNA was extracted from the three environmental samples from the organic phases of RNA extraction via the Totally RNA Kit (Ambion). Cell lyses were carried out within a Sterivex filter cartridge according to Somerville and colleagues (1989), with an additional 10 units ml^{-1} of RNase inhibitor (SUPERaseIn, Ambion). The crenarchaeotal *amo*A gene was PCR amplified using Arch-*amo*AF (5′-STAATGGTCTGGCTTAGACG-3′) and Arch-*amo*AR (5′-GCGGCCATCCATCTGTATGT-3′) primers, as described by Francis and colleagues (2005). To minimize the PCR bias, the lowest cycle number at which a product remained detectable was used. Five replicates were prepared from each sample, and the replicates were pooled for cloning (Acinas *et al*., 2005). The PCR product was then purified with Gene Clean Turbo kit (MPBio), cloned in a pCR4 vector and transformed in *E. coli* TOP10 cells, using the TOPO TA cloning kit for sequencing (Invitrogen). Additionally, to obtain copy control clones, the PCR product was cloned in a pCC1 vector and transformed in *E. coli* Epi300 cells, using the Copy Control cDNA, Gene and PCR Cloning Kit (Epicentre). The clones were PCR-screened using M13F and M13R vector primers, or, in the case of copy control clones, by cell lysis and estimation of vector size. The positive clones were then sequenced from the screening PCR products (both directions) using T7 and T3 vector primers or, in the case of the copy control clones, from plasmid DNA and T7 primer, with the Big Dye Terminator Sequencing Kit (Applied Biosystems), and analysed on an ABI 3700 sequencer (Applied Biosystems). Nucleotide sequences were cleaned and assembled using DNA Baser software (HeracleSoftware, Germany, http://www.DnaBaser.com). The identity of the sequences was verified via BLASTN searches (Johnson *et al*., 2008). The sequences were aligned and phylogenetically analysed using ARB (Ludwig *et al*., 2004). The distance matrices generated therein were further used for statistical analysis using DOTUR (OTU grouping and distribution, Chao and Jackknife diversity estimators) (Schloss and Handelsman, 2005) and MOTHUR (Venn diagrams) (Schloss *et al*., 2009).

Nucleotide sequence accession numbers

All partial putative *amo*A gene sequences determined in this study were deposited in the GenBank database under the following accession numbers: GQ500142–GQ500577.

Cell cultures

Escherichia coli cultures used in this study were *E. coli* strain K12 and *E. coli* copy control clones 1E3, 3G4 and 1E7, which were *E. coli* EPI300 transformed with a pCC1 vector that contained three closely related crenarchaeotal *amo*A genes. The cultures were grown overnight starting from single colonies on LB media (1% Tryptone, 0.5% Yeast Extract and 1% NaCl) with 12.5 μg ml^{-1} chloramphenicol for the clones. In order to increase the number of vector copies per cell, the 1E7 culture was amended with Induction Solution (Epicentre) to a final concentration of 2×. The cells were harvested by centrifugation and then fixed in 1% paraformaldehyde in 1× PBS pH 7.4 (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4), for 1 h at room temperature. The paraformaldehyde solution was removed by centrifugation and aspiration, followed by one 1× PBS wash. The cells were resuspended in 1× PBS and 96% ethanol was added to a final concentration of 50%. The cells were stored at –20°C.

Probe design

For the tests in *E. coli* clones, a single polynucleotide based on a 351-bp-long region of clone 1E3 was used as a probe. For the Namibian seawater samples, a multi-polynucleotide probe mix was designed to target all *amo*A sequences retrieved from the three seawater samples, using the PolyPro software (probe sequences are given in *Experimental procedures* in *Supporting information*).

Probe synthesis

Probes were produced by incorporating Dig-dUTP into dsDNA via PCR (70 μM Dig-dUTP), using the PCR Dig Probe Synthesis Kit (Roche, cat. No. 11636090910), according to the manufacturer instructions. The PCR products were column purified using the Gene Clean Turbo kit. The concentration was determined spectrophotometrically, using a NanoDrop instrument (Fischer ThermoScientific). The probes were stored at –20°C, in water.

Determination of stringency conditions

First, the T_m of the probe–target hybrids was calculated using the formulae from Wetmur (1991), which were integrated in the PolyPro software. Second, the T_m of the perfectly matched probe–target hybrids was measured in a hybridization-like buffer and washing-like buffer II, *in vitro*, using a real-time PCR machine (IQ5, Bio-Rad) and SYTO 9 dye (Invitrogen, cat. No. S-34854). The hybridization-like buffer was composed of 3.5 ml of formamide (Sigma), 2.5 ml of 20× SSC (3 M NaCl, 0.3 M sodium citrate, from Ambion, AM9765), 1 g of dextran sulfate sodium salt (Sigma, cat. No. D8906), 50 μl of 20% sodium dodecyl sulfate (SDS), 0.4 ml of 0.5 M EDTA pH 8.0 and 2.6 ml of autoclaved MilliQ water. To 1.5 ml of hybridization-like buffer 3 μl of 5 mM SYTO 9 was added, to get a final concentration of 10 μM. To 100 μl of the latter mixture, 6 μl of dsDNA (230–350 ng) was added, and the resulting solution was aliquoted into 25 μl portions per well and used for T_m determinations. The T_m was measured for the probe dsDNA (with Dig-dUTP), for target dsDNA and for the hybrid dsDNA. The thermal protocol used for the determination in hybridization-like buffer was the following: denaturation at 80°C for 5 min, hybridization at 42°C for 25 min and melting from 50°C to 75°C, +0.2°C per 1.5 min, minimum ramp rate. The composition of the washing-like buffer II was 0.1× SSC, 0.1% SDS, 10 μM SYTO 9 dye and dsDNA (~240 per 25 μl reaction). The T_m was measured for the probe dsDNA (with Dig-dUTP) and for target dsDNA. The thermal protocol used for T_m in washing-like buffer was: from 50°C to 75°C, +0.2°C per 1.5 min, minimum ramp rate.

GeneFISH protocol

All the water used during the protocol was autoclaved 0.22 μm filtered MilliQ water. Unless stated otherwise, the incubations were performed at room temperature. All washing steps were carried in 50 ml volumes.

Sample immobilization. Different volumes of paraformaldehyde fixed *E. coli* cells were mixed with 10 ml of 1× PBS and filtered through 0.2 μm polycarbonate filters (GTTP, Millipore). The filters were previously coated with a Pd/Au alloy (Musat *et al*., 2008) to reduce fluorescence background, although this is not essential for the geneFISH protocol. The filters were then washed with 10 ml of 1× PBS, 15 ml of water, air dried and stored at –20°C. The Namibian seawater samples were directly filtered after paraformaldehyde fixation, washed once with water, air dried and stored at –80°C.

Inactivation of endogenous peroxidases. The inactivation was performed by overlaying the filter with 0.01 M HCl for 10 min, followed by washing with 1× PBS for 5 min and with water for 1 min.

Permeabilization. For the *E. coli* samples, permeabilization was undertaken in 0.5 mg ml^{-1} lysozyme (AppliChem, cat. No. A4972.0010), 1× PBS, 0.1 M Tris-HCl pH 8.0 and 0.05 M EDTA pH 8.0, for 1 h, on ice. The wash consisted of 5 min with 1× PBS, 1 min with water and 1 min with 96% ethanol, followed by air-drying. For the Namibian seawater samples the permeabilization was 1 min with 0.1 M HCl, followed by 1× PBS (1 and 5 min), 1 min with water, 1 min with 96% ethanol, air-drying.

rRNA hybridization. Escherichia coli samples were hybridized with the EUB338 probe (Amann *et al*., 1990), while the Namibian seawater samples were hybridized with the Cren554 probe (5′-TTAGGCCCAATAATCMTCCT-3′) (Massana *et al*., 1997), which targets Marine Group I *Crenarchaeota,* or with the EUB338 I–III probe mix (Amann *et al*., 1990; Daims *et al*., 1999), which targets *Bacteria.* The hybridization took place for 3 h or overnight at 46°C, followed by

15 min of washing at 48°C. The hybridization buffers and the corresponding washing buffers were prepared as described previously (Pernthaler *et al*., 2002). The formamide concentrations used in the hybridization buffers were 35% for the EUB338 probes and 0% for Cren554.

CARD for rRNA detection. All the samples were equilibrated for 20 min in 1× PBS. Next, the *E. coli* samples were incubated for 10 min at 37°C in a solution containing 1× PBS, 0.0015% H_2O_2 and 0.25 μg ml^{-1} Alexa$_{488}$-labelled tyramide (prepared as described by Pernthaler and Pernthaler, 2005). The wash consisted of 2 and 15 min with 1× PBS, 1 min with water and 1 min with 96% ethanol, followed by air-drying. The Namibian seawater samples were incubated for 40 min at 46°C in a solution containing 1× PBS, 10% dextran sulfate, 0.1% blocking reagent (Roche, Germany, cat. No. 11096176001), 2 M NaCl, 0.0015% H_2O_2 and 1 μg ml^{-1} Alexa$_{488}$-labelled tyramide. The wash steps were 1, 2 and 2 × 10 min with 1× PBS at 46°C, 1 min with water, 1 min with 96% ethanol, followed by air-drying. Starting with the CARD incubations, all steps were performed in the dark.

RNase treatment. Before RNase digestion, some of the samples were denatured as follows: 20 min at 60°C in denaturation mix (90% formamide, 0.1× SSC and 0.1% SDS), to melt any potential secondary structures in the RNA and thus, to facilitate digestion. No difference was observed between denatured and not-denatured samples. The wash consisted of 3 × 2 min ice-cold water, then 2 × 2 min ice-cold 96% ethanol, followed by air-drying. Samples were incubated at 37°C for 4–5 h, in RNase solution – 0.5 U μl^{-1} RNase I (Ambion, cat. No. AM 2295), 30 μg ml^{-1} RNase A (Sigma, cat. No. R4642-10), 0.1 M Tris-HCl pH 8.0. The wash consisted of 3 × 7 min with 1× PBS, 1 min with water and 1 min with 96% ethanol, followed by air-drying.

DNase treatment. One of the controls for the Namibian seawater samples consisted of DNase digestion. In the previous step, these samples had been treated only with RNase I and without RNase A. The RNase I was inactivated by incubating for 20 min in 1× PBS at 70°C, then washing for 1 min in water. The DNase treatment was performed at 37°C overnight, in a solution containing 0.2 U μl^{-1} DNase I (Epicentre, cat. No. D9910K), 20 mM $MgCl_2$, 4 mM $CaCl_2$ and 10 mM Tris-HCl pH 8.0. The washing and inactivation of DNase I consisted of 25 min at 75°C, 2 × 20 min with 1× PBS, 0.1 M EDTA pH 8.0 and 1 min with water.

Inactivation of HRP introduced with the rRNA probe. The inactivation of HRP consisted of 10 min with 0.2 M HCl (for *E. coli* samples) or 1 min with 0.1 M HCl (for Namibian seawater samples), followed by washing with 1× PBS for 1 and 5 min, then 1 min with water, 1 min with 96% ethanol and air-drying.

Gene hybridization. Pre-hybridization was carried out in the same buffer as hybridization, but without the probe, for 5 h at 42°C. The hybridization buffer had the following composition: 7 ml of formamide, 5 ml of 20× SSC, 2 g of dextran sulfate, 100 μl of 20% SDS, 0.8 ml of 0.5 M EDTA pH 8.0, 2.2 ml of water, 0.5 ml of sheared salmon sperm DNA (sssDNA) (Ambion, cat. No. AM 9680), 0.5 ml of yeast RNA (Ambion, cat. No. AM 7118) and 2 ml of 10% blocking reagent (Roche, Germany, cat. No. 11096176001). The sheared salmon sperm DNA and yeast RNA were added as further blocking reagents for unspecific signals. For preparation of the hybridization buffer, see experimental procedures in *Supporting information*. The polynucleotide probes used were *amo*A-1E3 (*E. coli* clones), *amo*A-Nam and Non-PolyPr350 (Namibian seawater samples), with the third one being a negative control. The probe concentrations tested with the *E. coli* clones were 0.25 pg μl^{-1}, 2.5 pg μl^{-1}, 25 pg μl^{-1} and 250 pg μl^{-1}. For the Namibian seawater samples, the probe concentration was 2.5 pg μl^{-1} for each polynucleotide in the mix. After adding the probes and the filters to the hybridization buffer, the samples were denatured for 25 min in a 75°C water bath, slow shaking. The tubes were transferred immediately to 42°C (slow shaking water bath) and hybridization took place for 18–22 h or 41 h. The washes were first performed with washing buffer I (WBI) (2× SSC, 0.1% SDS) for 5 min and 30 min at 42°C, followed by washing buffer II (WBII) (0.1× SSC, 0.1% SDS) for 3 × 1 min at room temperature and 1.5 h at 42°C, in a slow-shaking water bath, and finally 2 × 1 min with 1× PBS.

Antibody binding. The samples were blocked for 1 h in 1× PBS and 0.5% Western Blocking Reagent (WBR) (Roche, cat. No. 11921673001). The antibody binding took place for 1.5 h, in a solution containing 1× PBS, 1% WBR and 0.3 U μl^{-1} anti-Dig HRP-conjugated antibody (Fab fragments) (Roche, cat. No. 11207733910). The wash was carried out in a 1× PBS, 0.5% WBR solution for 1, 5 and 2 × 10 min. All steps were carried on a shaker at 50 rpm.

CARD for gene detection. The samples were equilibrated for 20 min in 1× PBS. Then, they were incubated for 40 min at 37°C in amplification buffer containing 1× PBS, 20% dextran sulfate, 0.1% blocking reagent, and 2 M NaCl with 0.0015% H_2O_2 and 2 μg ml^{-1} Alexa$_{594}$-labelled tyramide. For preparation of the amplification buffer see the experimental procedures in *Supporting information*, and for the labelled tyramides see Pernthaler and Pernthaler (2005). The samples were then washed for 1, 5 and 2 × 10 min with 1× PBS, in a 46°C oven, slow shaking, then 1 min with water, 1 min with 96% ethanol, followed by air-drying.

Embedding and counterstaining. The samples were embedded either in ProLongGold antifade reagent (Invitrogen, cat. No. P36930) or in SlowFadeGold antifade reagent (Invitrogen, cat. No. S36936), containing 1 μg ml^{-1} 4′,6-diamidino-2-phenylindole (DAPI).

Microscopy

Microscopy was performed on an Axio epifluorescence microscope (Carl Zeiss) equipped with the following fluorescence filters: DAPI (365/10 nm excitation, 420 LP emission, FT 395 Beam Splitter), Alexa$_{488}$ (472/30 excitation, 520/35 emission, 495 Beam Splitter) and Alexa$_{594}$ (562/40 excitation, 624/40 emission, 593 Beam Splitter). The photomicrographs

were taken with a black and white camera and the exposure times for $Alexa_{594}$ were 200 ms (*E. coli* samples) and 250 ms (Namibian seawater samples). The black and white photomicrographs were pseudocoloured with the Axio Visio software. Counting for the geneFISH hybridization efficiency was performed on the photomicrographs using the Axio Visio software. At least 450 cells were counted per sample and the standard deviation was calculated from the counts of 5–20 microscopic fields.

Acknowledgements

We are thankful to Daniela Franzke for performing the DNA extractions from the three seawater samples and to Jillian Petersen and Nicole Dubilier for providing us with the *hyn*L sequences used for T_m measurements. We wish to acknowledge the Max Planck Society and the EnviTools Project for providing financial support for this study.

References

Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M.F. (2005) PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71:** 8966–8969.

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

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., and Stahl, D.A. (1990) Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl Environ Microbiol* **56:** 1919–1925.

Anderson, M.L.M., and Young, B.D. (1985) Quantitative filter hybridisation. In *Nucleic Acid Hybridisation – A Practical Approach*. Hames, B.D., and Higgins, S.J. (eds). Oxford, UK: Oxford University Press, pp. 73–110.

Birnstiel, M.L., Sells, B.H., and Purdom, I.F. (1972) Kinetic complexity of RNA molecules. *J Mol Biol* **63:** 21–39.

Bjorndal, M.T., and Fygenson, D.K. (2002) DNA melting in the presence of fluorescent intercalating oxazole yellow dyes measured with a gel-based assay. *Biopolymers* **65:** 40–44.

Bonner, T.I., Brenner, D.J., Neufeld, B.R., and Britten, R.J. (1973) Reduction in the rate of DNA reassociation by sequence divergence. *J Mol Biol* **81:** 123–135.

Breuert, S., Allers, T., Spohn, G., and Soppa, J. (2006) Regulated polyploidy in halophilic archaea. *PLoS ONE* **1:** e92.

Condon, C. (2003) RNA processing and degradation in *Bacillus subtilis*. *Microbiol Mol Biol Rev* **67:** 157–174.

Daims, H., Brühl, A., Amann, R., Schleifer, K.-H., and Wagner, M. (1999) The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22:** 434–444.

Deutscher, M.P. (2006) Degradation of RNA in bacteria: comparison of mRNA and stable RNA. *Nucleic Acids Res* **34:** 659–666.

Dunfield, P.F., Yuryev, A., Senin, P., Smirnova, A.V., Stott, M.B., Hou, S., *et al*. (2007) Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450:** 879–882.

Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E., and Oakley, B.B. (2005) Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci USA* **102:** 14683–14688.

Gudnason, H., Dufva, M., Bang, D.D., and Wolff, A. (2007) Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature. *Nucleic Acids Res* **35:** e127.

Hodson, R.E., Dustman, W.A., Garg, R.P., and Moran, M.A. (1995) *In situ* PCR for visualization of microscale distribution of specific genes and gene products in prokaryotic communities. *Appl Environ Microbiol* **61:** 4074–4082.

Hoshino, T., Yilmaz, L.S., Noguera, D.R., Daims, H., and Wagner, M. (2008) Quantification of target molecules needed to detect microorganisms by fluorescence *in situ* hybridization (FISH) and catalyzed reporter deposition – FISH. *Appl Environ Microbiol* **74:** 5068–5077.

Hutton, J.R. (1977) Renaturation kinetics and thermal stability of DNA in aqueous solutions of formamide and urea. *Nucleic Acids Res* **4:** 3538–3555.

Jain, C. (2002) Degradation of mRNA in *Escherichia coli*. *IUBMB Life* **54:** 315–321.

Jensen, R.B., and Shapiro, L. (1999) The *Caulobacter crescentus smc* gene is required for cell cycle progression and chromosome segregation. *Proc Natl Acad Sci USA* **96:** 10661–10666.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* **36:** W5–W9.

Karner, M.B., DeLong, E.F., and Karl, D.M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409:** 507–510.

Kawakami, S., Kubota, K., Imachi, H., Yamaguchi, T., Harada, H., and Ohashi, A. (2010) Detection of single copy genes by two-pass tyramide signal amplification fluorescence in situ hybridization (Two-Pass TSA-FISH) with single oligonucleotide probes. *Microbes Environ* **25:** 15–21.

Kenzaka, T., Tamaki, S., Yamaguchi, N., Tani, K., and Nasu, M. (2005) Recognition of individual genes in diverse microorganisms by cycling primed *in situ* amplification. *Appl Environ Microbiol* **71:** 7236–7244.

Kierzek, R. (1992) Nonenzymatic hydrolysis of oligoribonucleotides. *Nucleic Acids Res* **20:** 5079–5084.

Könneke, M., Bernhard, A.E., Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437:** 543–546.

Konstantinidis, K.T., Braff, J., Karl, D.M., and Delong, E.F. (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.

Kubitschek, H.E., and Freedman, M.L. (1971) Chromosome replication and the division cycle of *Escherichia coli*. *J Bacteriol* **107:** 95–99.

Kubota, K., Ohashi, A., Imachi, H., and Harada, H. (2006) Visualization of mcr mRNA in a methanogen by fluorescence *in situ* hybridization with an oligonucleotide probe and two-pass tyramide signal amplification (two-pass TSA-FISH). *J Microbiol Methods* **66:** 521–528.

Lam, P., Jensen, M.M., Lavik, G., McGinnis, D.F., Müller, B., Schubert, C.J., *et al*. (2007) Linking crenarchaeal and bacterial nitrification to anammox in the Black Sea. *Proc Natl Acad Sci USA* **104:** 7104–7109.

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

Lundgren, M., Malandrin, L., Eriksson, S., Huber, H., and Bernander, R. (2008) Cell cycle characteristics of *Crenarchaeota*: unity among diversity. *J Bacteriol* **190:** 5362–5367.

Martens-Habbena, W., Berube, P.M., Urakawa, H., Torre, J.R., and Stahl, D.A. (2009) Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* **461:** 976–979.

Maruyama, F., Kenzaka, T., Yamaguchi, N., Tani, K., and Nasu, M. (2003) Detection of bacteria carrying the *stx2* gene by *in situ* loop-mediated isothermal amplification. *Appl Environ Microbiol* **69:** 5023–5028.

Maruyama, F., Kenzaka, T., Yamaguchi, N., Tani, K., and Nasu, M. (2005) Visualization and enumeration of bacteria carrying a specific gene sequence by *in situ* rolling circle amplification. *Appl Environ Microbiol* **71:** 7933–7940.

Massana, R., Murray, A.E., Preston, C.M., and DeLong, E.F. (1997) Vertical distribution and phylogenetic characterization of marine planktonic Archaea in the Santa Barbara Channel. *Appl Environ Microbiol* **63:** 50–56.

Mincer, T.J., Church, M.J., Taylor, L.T., Preston, C., Karl, D.M., and DeLong, E.F. (2007) Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. *Environ Microbiol* **9:** 1162–1175.

Monis, P.T., Giglio, S., and Saint, C.P. (2005) Comparison of SYT09 and SYBR Green I for real-time polymerase chain reaction and investigation of the effect of dye concentration on amplification and DNA melting curve analysis. *Anal Biochem* **340:** 24–34.

Moraru, C., Moraru, G., Fuchs, B.M., and Amann, R. (2010) Concepts and software for a rational design of polynucleotide probes. *Environ Microbiol Rep* (in press): doi:10.1111/j.1758-2229.2010.00189.x

Musat, N., Halm, H., Winterholler, B., Hoppe, P., Peduzzi, S., Hillion, F., *et al*. (2008) A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proc Natl Acad Sci USA* **105:** 17861–17866.

Niki, H., and Hiraga, S. (1997) Subcellular distribution of actively partitioning F plasmid during the cell division cycle in *E. coli*. *Cell* **90:** 951–957.

Niki, H., and Hiraga, S. (1998) Polar localization of the replication origin and terminus in *Escherichia coli* nucleoids during chromosome partitioning. *Genes Dev* **12:** 1036–1045.

Park, S.-J., Park, B.-J., and Rhee, S.-K. (2008) Comparative analysis of archaeal 16S rRNA and amoA genes to estimate the abundance and diversity of ammonia-oxidizing archaea in marine sediments. *Extremophiles* **12:** 605–615.

Pernthaler, A., and Amann, R. (2004) Simultaneous fluorescence *in situ* hybridization of mRNA and rRNA in environmental bacteria. *Appl Environ Microbiol* **70:** 5426–5433.

Pernthaler, A., and Pernthaler, J. (2005) Simultaneous fluorescence *in situ* hybridization of mRNA and rRNA for the detection of gene expression in environmental microbes. *Methods Enzymol* **397:** 351–371.

Pernthaler, A., Pernthaler, J., and Amann, R. (2002) Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microbiol* **68:** 3094–3101.

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

Purkhold, U., Pommerening-Röser, A., Juretschko, S., Schmid, M.C., Koops, H.-P., and Wagner, M. (2000) Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and *amo*A sequence analysis: implications for molecular diversity surveys. *Appl Environ Microbiol* **66:** 5368–5382.

Raap, A.K., Marijnen, J.G.J., Vrolijk, J., and Ploeg, M.v.d. (1986) Denaturation, renaturation, and loss of DNA during *in situ* hybridization procedures. *Cytometry* **7:** 235–242.

Sambrook, J. (2001) *Molecular Cloning Manual*. New York, USA: Cold Spring Harbor Laboratory.

Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A., and Pernthaler, J. (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ Microbiol* **11:** 2078–2093.

Schleper, C., Jurgens, G., and Jonuscheit, M. (2005) Genomic studies of uncultivated archaea. *Nat Rev Microbiol* **3:** 479–488.

Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71:** 1501–1506.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing MOTHUR: open source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Smolina, I., Lee, C., and Frank-Kamenetskii, M. (2007) Detection of low-copy-number genomic DNA sequences in individual bacterial cells by using peptide nucleic acid-assisted rolling-circle amplification and fluorescence *in situ* hybridization. *Appl Environ Microbiol* **73:** 2324–2328.

Somerville, C.C., Knight, I.T., Straube, W.L., and Colwell, R.R. (1989) Simple, rapid method for direct isolation of nucleic acids from aquatic environments. *Appl Environ Microbiol* **55:** 548–554.

Tani, K., Kurokawa, K., and Nasu, M. (1998) Development of a direct *in situ* PCR method for detection of specific bacteria in natural environments. *Appl Environ Microbiol* **64:** 1536–1540.

Teira, E., Lebaron, P., Aken, H.v., and Herndl, G.J. (2006) Distribution and activity of *Bacteria* and *Archaea* in the deep water masses of the North Atlantic. *Limnol Oceanogr* **51:** 2131–2144.

Treusch, A.H., Leininger, S., Kletzin, A., Schuster, S.C., Klenk, H.-P., and Schleper, C. (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ Microbiol* **7:** 1985–1995.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Viollier, P.H., Thanbichler, M., McGrath, P.T., West, L., Meewan, M., McAdams, H.H., and Shapiro, L. (2004) Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *Proc Natl Acad Sci USA* **101:** 9257–9262.

Wagner, M., Schmid, M., Juretschko, S., Trebesius, K.-H., Bubert, A., Goebel, W., and Schleifer, K.-H. (1998) *In situ* detection of a virulence factor mRNA and 16S rRNA in *Listeria monocytogenes*. *FEMS Microbiol Lett* **160:** 159–168.

Watson, N., Gurevitz, M., Ford, J., and Apirion, D. (1984) Self cleavage of a precursor RNA from bacteriophage T4. *J Mol Biol* **172:** 301–323.

Wetmur, J.G. (1991) DNA probes: applications of the principles of nucleic acid hybridization. *Crit Rev Biochem Mol Biol* **26:** 227–259.

Wetmur, J.G., and Davidson, N. (1968) Kinetics of renaturation of DNA. *J Mol Biol* **31:** 349–370.

Wild, J., Hradecna, Z., and Szybalski, W. (2002) Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res* **12:** 1434–1444.

Yu, H., Chao, J., Patek, D., Mujumdar, R., Mujumdar, S., and Waggoner, A.S. (1994) Cyanine dye dUTP analogs for enzymatic labeling of DNA probes. *Nucleic Acids Res* **22:** 3226–3232.

Zwirglmaier, K., Ludwig, W., and Schleifer, K.H. (2004) Recognition of individual genes in a single bacterial cell by fluorescence *in situ* hybridization – RING-FISH. *Mol Microbiol* **51:** 89–96.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. A maximum likelihood nucleotide tree of crenarchaeotal *amo*A obtained from the three Namibian seawater samples. Bootstrap values (> 50% from 1000 bootstraps) are denoted at branch points. Sequences are colour coded according to types of sample origins (terrestrial, marine sediments or seawater). The marine clusters A (here denoted A1), B and C, as defined by Francis and colleagues (2005), plus a new cluster, A2, are also indicated. The majority of the Namibia seawater *amo*A group in marine clusters A1 and B. The Namibian *amo*A are represented as groups labelled 'Namibia seawater', with the number in brackets representing the number of unique *amo*A sequences from each group.

Fig. S2. Namibian seawater samples – rarefaction curves for the *amo*A gene (the 351 bp fragment used for polynucleotide probe design) at a 99% identity cut-off for OTU grouping (A) and a 95% identity cut-off for OTU grouping (B). Shaded areas represent the confidence intervals.

Fig. S3. Namibian seawater samples – distribution of the *amo*A gene (the 351 bp fragment used for polynucleotide probe design) in OTUs, at 99% and 95% identity cut-off. At 99% identity cut-off, none of the OTUs represents more than 10% of the sequences. The majority of the OTUs are represented by a very low number or individual sequences. Pooling the sequences from the three samples results in an almost doubling of the OTU number, as compared with individual clone libraries, and none of the individual OTUs are represented by more than 8% of the sequences. At a 95% identity cut-off, between 70% and 80% of the sequences are found in three OTUs. Pooling the sequences from the three clone libraries results in an increase of the number of OTUs with less than 50%, as compared with individual clone libraries, and 68% of the sequences are found in three OTUs.

Fig. S4. Distribution of the *amo*A gene (the 351 bp fragment used for polynucleotide probe design) across the three Namibian seawater samples. The OTUs are grouped according to a 99% identity cut-off (A) and a 95% identity cut-off (B). At a cut-off of 99% identity, the three clone libraries shared amongst them between 49% and 58% of their sequences, while between 21% and 28% were unique sequences and the remaining 21–24% sequences were shared between two clone libraries. When using 95% identity as a cut-off for OTU grouping, between 75% and 85% of the sequences were shared among the three clone libraries.

Fig. S5. A nucleotide multiple alignment of *amo*A from Namibia seawater samples, performed with the integrated aligner from ARB. The target region for probe design is marked in orange (position 95–445 *Cenarchaeum symbiosum* numbering). The sequences are trimmed according to the marked region and exported in GenBank format, with phylogenetic information as inferred from the maximum likelihood tree (see Fig. S1).

Fig. S6. The GenBank files (A) containing the trimmed sequences are transformed into FASTA files (D) using the GTE module (B) from the PolyPro software. The phylogenetic information from the SOURCE field of the GenBank files is used to create the Taxonomy database (C).

Fig. S7. The PPD module from the PolyPro software receives the FASTA files generated by the GTE as input. All the *amo*A sequences are given both as probes and as targets. The hybridization type is set to DNA:DNA and a mismatch table is calculated between probes (horizontal header) and targets (vertical header). In the mismatch table, the cells with percentage mismatch (%MM) lower than mismatch threshold 1 (Th1) are marked in green.

Fig. S8. The threshold for a probe to hit a target (Th1) is set to $\leq$ 5% mismatch. No phylogenetic clade is selected. Therefore, the probes will be designed for all crenarchaeal *amo*A targets retrieved from the Namibian seawater samples. The probes with identical target groups and a T_m difference lower than 0.05°C are considered replicates and only one of them will be kept. As a consequence, the number of probes decreases, as reported in the LOG. The mismatch table is transformed into a hit matrix, which will be used in the next step to calculate the probe mixes.

Fig. S9. The probe mixes which hit all the targets are calculated. The combination algorithm is initiated by combinations of 2, followed by optimized combinations. Since the T_m toler-

ance parameter was set to 0, only the probe mixes with the lowest ΔT_m were selected during the combination algorithm. As a result, all the 600 probe mixes have the same ΔT_m (1.63°C), as reported in the LOG.

Fig. S10. Step 5 was used to select probe mixes with dsDNA probes less likely to cross-hybridize. From all probe mixes resulting in step 3, only the ones with the lowest similarity between the probes were selected.

Fig. S11. Hit map for the 12 polynucleotides in the *amo*A-Nam probe mix. For each probe the percentage mismatch with each sequence from the clone libraries is represented as a dot. The colour of the dot is green when the respective sequence belongs to the same phylogenetic clade as the probe and red when it belongs to a different clade. The threshold for a probe to target a sequence was set to $\leq$ 5 % mismatch (blue line in graph).

Fig. S12. The effect of Syto9 dye concentration on the T_m peak height and shape. These T_m curves have been measured for the *hyn*L dsDNA probe. It can be noticed that an increase in Syto9 concentration produces an increase in peak height, without a significant variation in T_m.

Fig. S13. Variation of the melting temperature with the concentration of Syto9 dye. The T_m of *hyn*L dsDNA probe, target and hybrid were measured in hybridization-like buffer (35% formamide). The T_m did not vary much with the dye concentration, whose increase to 50 μM lead to a decrease in the T_m of ~0.5°C.

Fig. S14. Schematic representation of the gene signal amplification system.

Fig. S15. GeneFISH for simultaneous detection of *amo*A gene (crenarchaeal ammonia monooxygenase subunit A) with *amo*A-1E3 probe (red, second column) and 16S rRNA CARD-FISH with *EUB338*, a general bacterial probe (green, first column), on *Escherichia coli* clones. The last column represents the overlap between the 16S and gene signals.

A. *Escherichia coli* clone 1E7, induced to contain 10–200 copies of the *amo*A gene fragment per cell, positive control.

B. *Escherichia coli* K12, no *amo*A copy, negative control.

Fig. S16. GeneFISH with the negative control probe for gene detection (NonPolyPr350) (red, third column) and 16S rRNA CARD-FISH with Cren554 probe for marine *Crenarchaeota* (green, second column) on Namibian seawater samples, station 249. All cells stained with DAPI: blue, first column. The last column represents the overlap between the 16S and gene signals.

Table S1. The *amo*A-Nam probe mix used for geneFISH on Namibian seawater samples (targeting all retrieved crenarchaeal *amo*A sequences). The parameters for this probe mix are: ΔT_m 1.7°C, average %MM between probes: 18.77 $\pm$ 6%. The probes have a length of 351 bp (position 95–445 *Cenarchaeum symbiosum* numbering).

Table S2. Calculated T_m for the polynucleotides forming the *amo*A probes. The T_m was calculated for the composition of the hybridization-like buffer (HB) (1718 mM Na^+ and 35% formamide) and the washing-like buffer II (WB) (23 mM Na^+).

Table S3. GeneFISH on the three Namibian seawater samples – percentage of crenarchaeotal cells showing a gene signal.

Table S4. Na^+ contribution of different components in the gene hybridization buffer.

Table S5. Na^+ contribution of different components in the washing buffer II.

Table S6. NonPolyPr template – restriction enzymes with the corresponding restriction sites and resulting fragments.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.