Environmental Microbiology (2007) doi:10.1111/j.1462-2920.2007.01371.x

Ammonia-oxidizing bacterial community composition in estuarine and oceanic environments assessed using a functional gene microarray

Bess B. Ward,[1]* Damien Eveillard,[2] Julie D. Kirshtein,[3] Joshua D. Nelson,[1] Mary A. Voytek[3] and George A. Jackson[2]
[1]*Department of Geosciences, Guyot Hall, Princeton University, Princeton, NJ, USA.*
[2]*Department of Oceanography, Texas A and M University, College Station, TX, USA.*
[3]*USGS, Reston, VA, USA.*

Summary

The relationship between environmental factors and functional gene diversity of ammonia-oxidizing bacteria (AOB) was investigated across a transect from the freshwater portions of the Chesapeake Bay and Choptank River out into the Sargasso Sea. Oligonucleotide probes (70-bp) designed to represent the diversity of ammonia monooxygenase (*amoA*) genes from Chesapeake Bay clone libraries and cultivated AOB were used to construct a glass slide microarray. Hybridization patterns among the probes in 14 samples along the transect showed clear variations in *amoA* community composition. Probes representing uncultivated members of the *Nitrosospira*-like AOB dominated the probe signal, especially in the more marine samples. Of the cultivated species, only *Nitrosospira briensis* was detected at appreciable levels. Discrimination analysis of hybridization signals detected two guilds. Guild 1 was dominated by the marine *Nitrosospira*-like probe signal, and Guild 2's largest contribution was from upper bay (freshwater) sediment probes. Principal components analysis showed that Guild 1 was positively correlated with salinity, temperature and chlorophyll a concentration, while Guild 2 was positively correlated with concentrations of oxygen, dissolved organic carbon, and particulate nitrogen and carbon, suggesting that different *amoA* sequences represent organisms that occupy different ecological niches within the estuarine/marine environment. The trend from most diversity of AOB in the upper estuary towards dominance of a single type in the polyhaline region of the Bay is consistent with the declining importance of AOB with increasing salinity, and with the idea that AO-Archaea are the more important ammonia oxidizers in the ocean.

Received 12 October, 2006; accepted 9 May, 2007. *For correspondence. E-mail bbw@princeton.edu; Tel. (+1) 609 258 5150; Fax (+1) 609 258 0796.

Introduction

Nitrification is an essential link in the nitrogen cycle between organic matter decomposition (ammonification) and denitrification and subsequent removal of fixed N from the system. The first step in the process is performed by ammonia-oxidizing bacteria (AOB) and archaea (AOA). Ammonia-oxidizing archaea are apparently the dominant group in the ocean (Wuchter *et al*., 2006) but the relative abundance of AOB and AOA in estuaries has not been investigated. The AOB are phylogenetically constrained to a few closely related groups in the gamma and beta classes of the *Proteobacteria*. Although negative results are rarely reported, attempts to detect the gammaproteobacterial AOB by polymerase chain reaction (PCR) are often not successful, implying that they are not very abundant in seawater, despite their apparent specialization on that environment.

At the 16S rRNA level, the betaproteobacterial AOB fall into approximately 10 clusters (Purkhold *et al*., 2000). The two largest clusters contain no closely related cultivated members and are commonly detected in seawater. One of these is the uncultivated *Nitrosospira* cluster, which appears to be the most important AOB in many marine environments; its members dominate clone libraries from the Mediterranean Sea (Phillips *et al*., 1999), polar seas (Bano and Hollibaugh, 2000; Hollibaugh *et al*., 2002) and Monterey Bay, CA (O'Mullan and Ward, 2005). The second uncultivated marine cluster is more closely related to *Nitrosomonas marina* and *Nitrosomonas oligotropha* strains and its members have been detected in the same environments, generally as a smaller component of those clone libraries. The *Nitrosomonas*-like cluster is more likely to be associated with particles (Phillips *et al*., 1999) and with higher ammonium concentrations (Kowalchuk and Stephen, 2001) and dominates in some estuaries (Bollman and Laanbroek, 2002).

Investigations of AOB diversity have focused on both the 16S rRNA gene and the signature functional gene, *amoA* (ammonia monooxygenase A) (Kowalchuk *et al*., 2000; Nicolaisen and Ramsing, 2002; Webster *et al*., 2002; Francis *et al*., 2003; O'Mullan and Ward, 2005; Freitag *et al*., 2006). O'Mullan and Ward (2005) detected much greater sequence diversity at the *amoA* gene level than for 16S rRNA genes from similar sized clone libraries obtained from Monterey Bay. Two main groups corresponding to the two uncultivated 16S rRNA clusters were identified. The same two groups were found in Ythan Estuary (Freitag *et al*., 2006) and in Chesapeake Bay sediments (Francis *et al*., 2003), where several subgroups were also detected.

The synthesis of previously published sequences from cultures and environmental clone libraries enabled Purkhold and colleagues (2000) to associate the main sequence clusters with gross environmental characteristics (e.g. seawater vs. soil), but this was a low resolution analysis, not based on quantitative analyses. Francis and colleagues (2003) compared the level of *amoA* sequence diversity in clone libraries from Chesapeake Bay sediments to the pairwise differences in environmental parameters (e.g. salinity, oxygen, temperature, nitrate and ammonium concentrations) between stations. Difference in salinity was the dominant variable and accounted for 62% of the variation in pairwise sequence dissimilarity between stations. Ammonia concentration, the next most important variable, explained only an additional 2.7%. Although AOA have not been investigated in estuaries, we expect that their relative importance and abundance will increase with salinity, opposite the trend for AOB.

Just as most investigations of diversity in the environment have used the 16S rRNA gene to identify species and phylotypes, assessment of natural samples using DNA microarrays has also exploited the 16S rRNA database. Peplies and colleagues (Peplies *et al*., 2004) represented 18 16S rRNA phylotypes using short oligonucleotides on a DNA microarray, but were able to detect only three groups by hybridizing with total RNA extracted from North Sea microbial communities. Functional groups have also been targeted using 16S rRNA genes. Loy and colleagues (Loy *et al*., 2004) used a hierarchical set of 16S rRNA oligonucleotide probes to identify the major sulfate reducing groups in acidic fen environments. Methanotrophs are analogous to the AOB in many ways, and they have been investigated in soils using short oligonucleotide DNA microarrays based on the particulate methane monooxygenase gene (Bodrossy *et al*., 2003; 2006; Stralis-Pavese *et al*., 2004). We previously demonstrated the approach with arrays containing multiple different functional genes (*nirS*, *nirK*, *nifH* and *amoA*) (Taroncher-Oldenburg *et al*., 2003) and developed criteria for probe design and data analysis.

The greater sequence divergence found in functional genes than in 16S rRNA genes from the same organisms (Purkhold *et al*., 2000) and the diversity patterns observed for *amoA* in the Chesapeake system drove our focus on functional genes for array development. Here we used DNA microarrays to investigate ecological signals in AOB *amoA* diversity in Chesapeake Bay (Fig. 1). By applying this approach in a system already characterized using the conventional clone library approach, we can compare the ecological conclusions obtained from each, as well as evaluate the resolution and resources required for both approaches. In order to relate the *amoA* phylotype composition to the environmental variables, we used a broadly applicable computational methodology for analysis of the diversity of microbial function in biogeochemical cycles using genetic data. This approach should also allow us to make new conjectures about the mechanisms that control the distribution of the functional gene and of the functional guilds that this diversity represents.

Results

Probe design

The phylogenetic relationships among the 24 probe 70-mer oligonucleotides are represented in the tree (Fig. 2). The probe region corresponds to 449–518 bp in the *Nitrosospira briensis* (NBU76553) sequence. The probes had an average melting temperature of 81.3 $\pm$ 2.7°C (Breslauer *et al*., 1986) and an average GC ratio of 51 $\pm$ 6 (Table S1, *Supplementary material*). The 24 oligos include several clusters of sequences that are very similar to each other (e.g. CT200d4 and CT200s1) as well as single probes that are quite distant from all others (e.g. CBsed37 and CBsed12) (Fig. 2; Table S2, *Supplementary material*). Assuming that sequences within 85% identity of a probe will hybridize to that probe (Taroncher-Oldenburg *et al*., 2003), this probe set should capture the entire range of sequence diversity represented in the Chesapeake Bay clone libraries. A gammaproteobacterial ammonia oxidizer sequence from culture (*Nitrosococcus oceani*) was included in the probe set as a negative control.

Array characterization

Probe resolution and specificity. Several experiments were performed in which two targets were competitively hybridized in a label inverse design as previously described (Taroncher-Oldenburg *et al*., 2003). In every case, the expected perfect match probe yielded the strongest hybridization signal, which in some cases was matched by the signal from probes with nearly identical sequences. Related probes also hybridized significantly

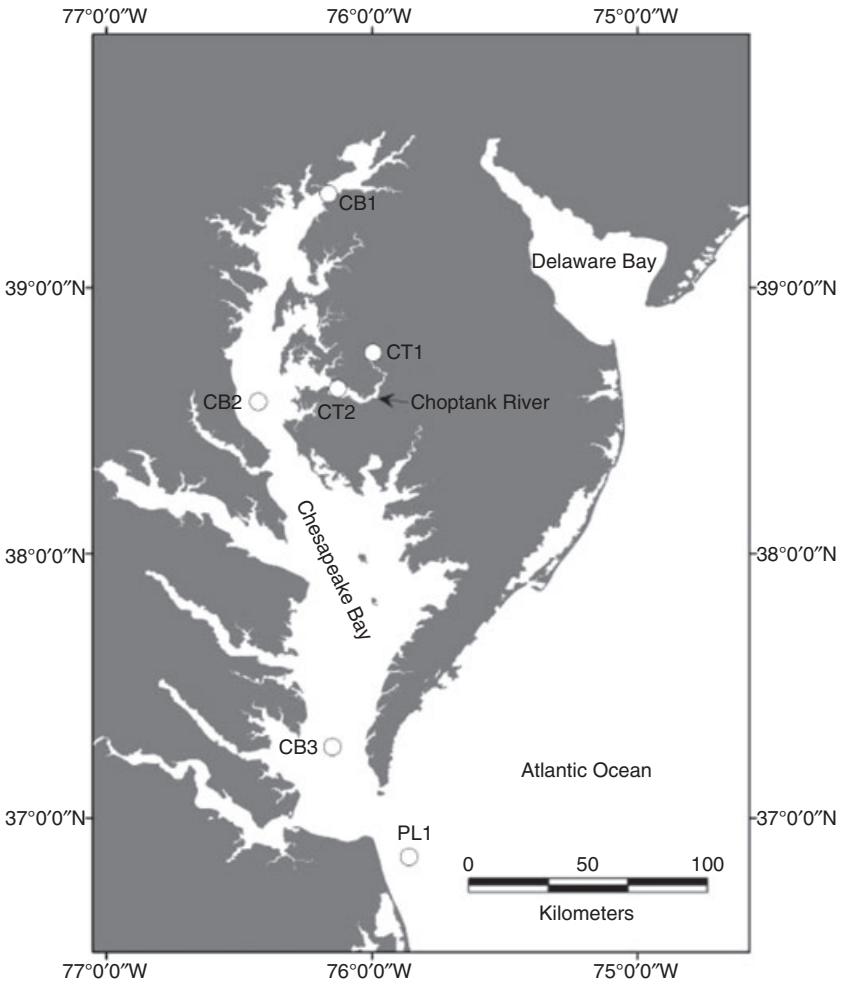


Fig. 1. Map of Chesapeake Bay showing station locations. Not shown is the Sargasso Sea station at 36°40 N, 71°60 W.

(as defined by the results of signal filtering; see *Experimental procedures* and Fig. S1, *Supplementary material*). At the relatively high concentrations of target hybridized in these experiments, some probes that were slightly lower in identity than 85–87% (the identity level predicted to allow hybridization) did in fact produce significant signal (see *Discussion*).

The pattern of hybridization from various groups of competitively hybridized probes was completely reproducible (2–4 replicate experiments for each probe mixture) and indicated an identify cut-off of 80–85% for independent probe resolution. Several 'cluster probes' were thus identified, as groups of probes that cross react, i.e. that each hybridize with the same set of target sequences. This analysis was consistent with the phylogenetic analysis (Fig. 2) and reduced the effective number of probes to those that are distinguishable by hybridization, i.e. have relatively low sequence identity. The total number of unique probes was thus reduced to 14, including single probes Noceani, Nmestuarii, Nseuro, Nmureae, Nmcryo, Nmmarina, CBsed1, CBsed12 and CBsed37, plus Cluster probes 1–5: Cluster probe 1 (i.e. the set of probes CBsed8, CBsed13, CBsed63); Cluster probe 2 (CBsed10, CBsed26, CBsed61); Cluster probe 3 (CT200s1, CB200d4, CB300s1, CB300s2); Cluster probe 4 (Nsbriensis, Nstenuis, Nsmultiform); Cluster probe 5 (Nmhalo, Nmoligo). The Nsbriensis probe sequence is in fact more distant from the other members of its cluster than the usual cut-off (78% vs. 80–85%); neither of the other two cultivated sequences in this cluster contributed significantly to the hybridization signal in environmental samples (see below), so it is convenient to group them with Nsbriensis.

Probe capacity. Equal amounts of some targets yielded different fluorescence intensities even when hybridized to their perfect match probes. Therefore, most of the probes were tested independently to ascertain the scale of this phenomenon, by comparing the fluorescence intensity of the perfect match hybridization (using the internal ratio

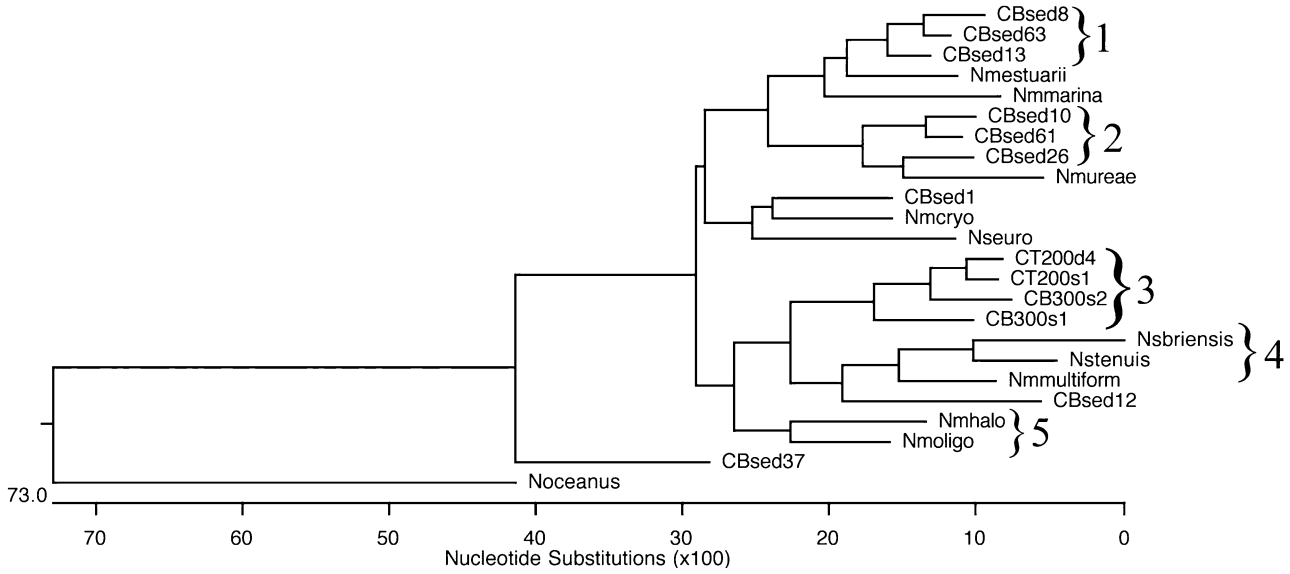


Fig. 2. Distance tree (Clustal V, Higgins *et al*., 1991) based on the alignment of 24 70-mer oligonucleotide sequences on the *amoA* array. The numbered brackets indicate the groups of probes that comprise the five cluster probes.

method, summing the individual components of the Cluster probes to treat each Cluster as one signal) with that of the CBsed12 perfect match. Three of the probes were not different from CBsed12, but four probes had significantly less fluorescence and two significantly greater (Fig. 3). Cluster probe 1 in particular had a fluorescence capacity 25-fold greater than that of CBsed12, i.e. 1 ng of one of the targets that hybridized with the four probes in Cluster probe 1 yielded 25 times greater fluorescence intensity than did 1 ng of target CBsed12 hybrid-

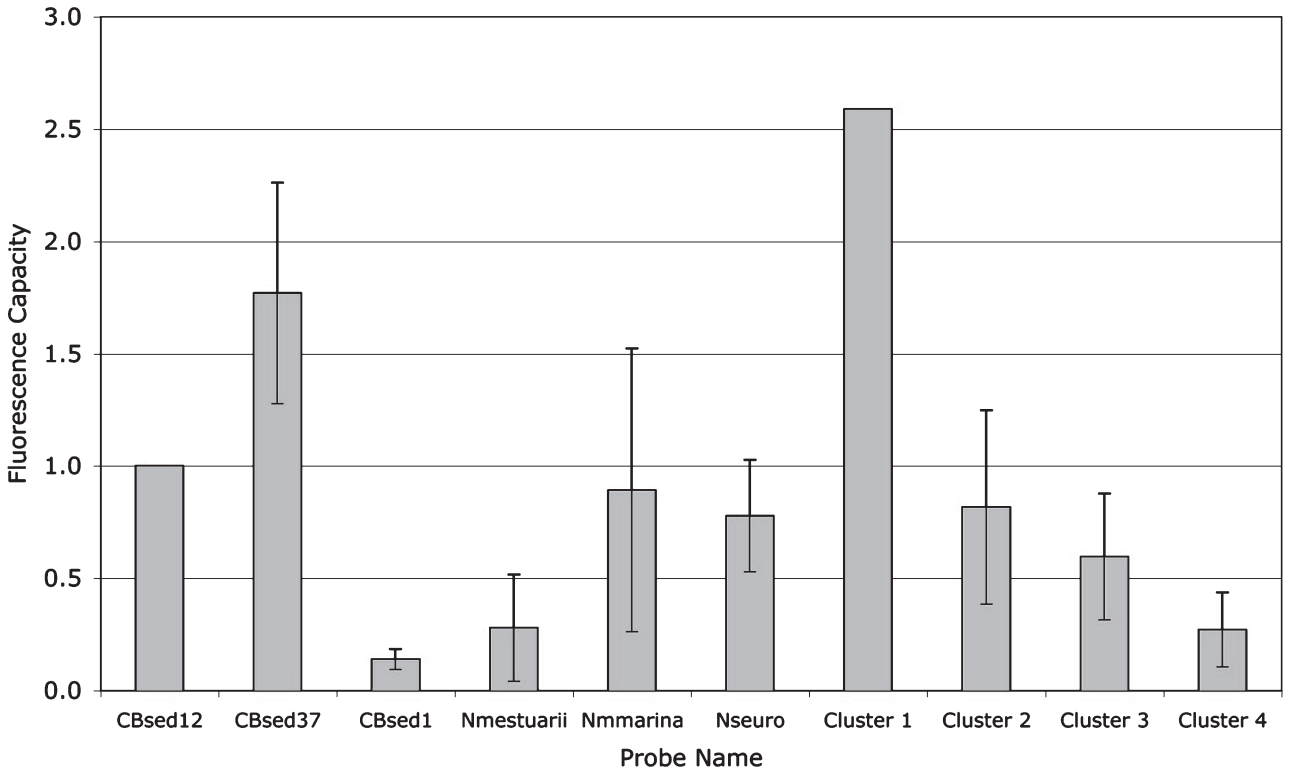


Fig. 3. Differential fluorescence capacity of 100% perfect match target/probe combinations normalized to CBsed12 as 100% signal. Error bars represent standard deviation of results from multiple separate arrays.

ized with probe CBsed12. This dramatic difference has serious implications for quantification of unknown mixtures (see below).

Results of array application to environmental samples

Hybridization data were determined for 14 water samples using the internal standard ratio method. Two examples, illustrating quite different hybridization patterns, are shown in Fig. S2 (*Supplementary material*). The coefficient of variation (CV) for significant fluorescence ratio (FR) values averaged 30% across all slides. The largest CVs were associated with very low (insignificant) signals, which were observed frequently for features representing cultivated AOB sequences. Because the FR scale is not calibrated and represents the average Cy3/Cy5 ratio for each probe, the absolute FR scales cannot be compared directly between arrays. To assess the robustness of the individual experiments, we analysed biological replicates for eight of the samples after completion of the original analyses. Duplicate filters that had remained frozen at −80°C since collection were extracted, labelled and hybridized using the same protocols as the original. Probes that yielded a hybridization signal $\geq 5\%$ of the total signal on each array were identified as present, and a similarity analysis using presence/absence data (Nei and Li, 1979) was performed. For the eight sets of duplicates, the similarity index $[F = 2n_{xy}(n_x + n_y)]$ ranged from 50% to 100% and averaged 72% (SD 16%). The main difference between replicates was that the first set of experiments detected more significant signals than the later ones; in only one sample did the later experiment detect one signal that was not present in the first set. This was attributed to degradation of the arrays – the replicate experiments were performed 2.5 years after the initial experiments.

Relative FR values were used to compare hybridization patterns among different samples (Fig. 4A). The same amount of total target (100 ng) was hybridized with each array, so the proportion of signal attributed to each probe is the appropriate comparison. The largest signals were observed for *amoA* probes derived from a group of closely related clones: CB300s2, CB300s1, CT200d4 and CT200s1. This group, identified as Cluster probe 3 above, constituted 4.6–93.7% of the total signal in the 14 samples. Its highest contributions (91.7%, 93.5% and 93.7%) occurred at CT200D, PL100D and CB200D, all deep water samples from the river, plume and bay respectively. Its smallest contributions (4.6%, 28.2% and 34.5%) occurred at CB100D, PL100M and SS100S, deep water from the upper bay and mid and surface water samples from the plume and Sargasso Sea respectively. Cluster probe 3 represented on the order of 50% of the total signal in all other samples.

Probes CBsed8, CBsed13 and CBsed63 (identified as Cluster probe 1 above) were the next most important group in magnitude of the hybridization signal, accounting for almost half the total signal at CB200S and close to 40% at CB300D.

The 11 individual probe sequences representing cultivated strains were consistently among the lowest signals, frequently being in the background (Fig. 4A). The individual probe representing *N. briensis* (Nsbriensis, identified as part of Cluster probe 4 above) produced the highest signal among the cultivated strains. Cluster probe 4 represented up to 28.9% of the total signal (sample CB100D), but was usually less than 5% and often less than 1% of the total. The only other cultivated sequence represented significantly in the hybridization signal was *Nitrosomonas cryotolerans* (Nmcryo), which reached 8.2% of the total signal in SS100M. It is clear from the capacity tests, however, that the fluorescence signals cannot be interpreted quantitatively in terms of target abundance in the sample (see *Discussion*).

Analysis of probe resolution

Relative FR results for all individual probes were subjected to discrimination to identify probes that tended to behave similarly in the hybridizations with environmental samples. This initial analysis identified groups of probes that did not behave independently because they hybridized to the same targets – the Cluster probes identified above. The phylogenetic analysis of sequence comparisons, the simple mixture experiments and the probe behaviour in hybridization experiments with environmental samples were all consistent with each other, and established a distance criterion slightly larger than $15 \pm 3\%$ criterion established with the earlier arrays (Taroncher-Oldenburg *et al*., 2003) for resolution of distinct genotypes.

In the experiments with environmental samples, the tendency of the probes to cross react cannot be distinguished from the potential co-occurrence of different targets (correlation in the signal intensity because the different targets have similar distributions in the samples). The fact that the hybridization results from environmental samples were consistent with the phylogenetic analysis and the simple mixtures (i.e. all probes within each Cluster showed similar hybridization patterns, essentially acting as a single probe), however, indicates that the components of the Cluster probes detected the same targets, rather than correlated occurrences of unrelated targets. For example, the four probes (CT200d4, CT200s1, CB300s1 and CB200s2), which consistently yielded the highest signal in most samples, were 82.9% – 95.7% identical to each other (Table S2, *Supplementary material*) and were identified as Cluster probe 3 above. Targets

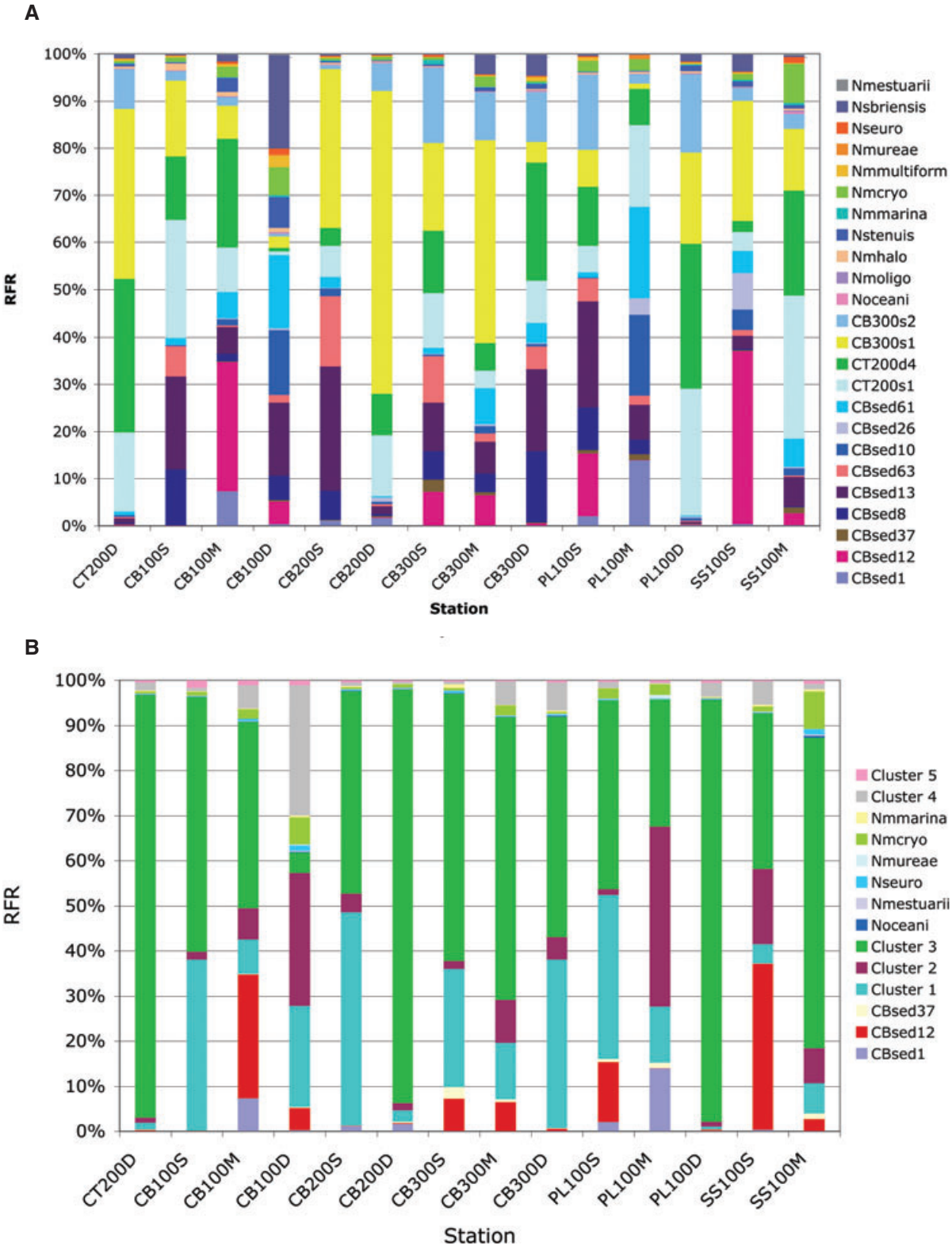


Fig. 4. Composite bar plot showing relative FR for all probes (A) and for the 14 independent probes (B) in all samples. The total FR signals were summed and the FR value for each probe is represented by its fraction of the total signal. The first station on the *x*-axis is the only Choptank River station, CT200, from a location with intermediate salinity. The other stations are arranged beginning with the freshwater upper station in the Bay (CB100) in order of increasing salinity towards the open ocean station, SS100.

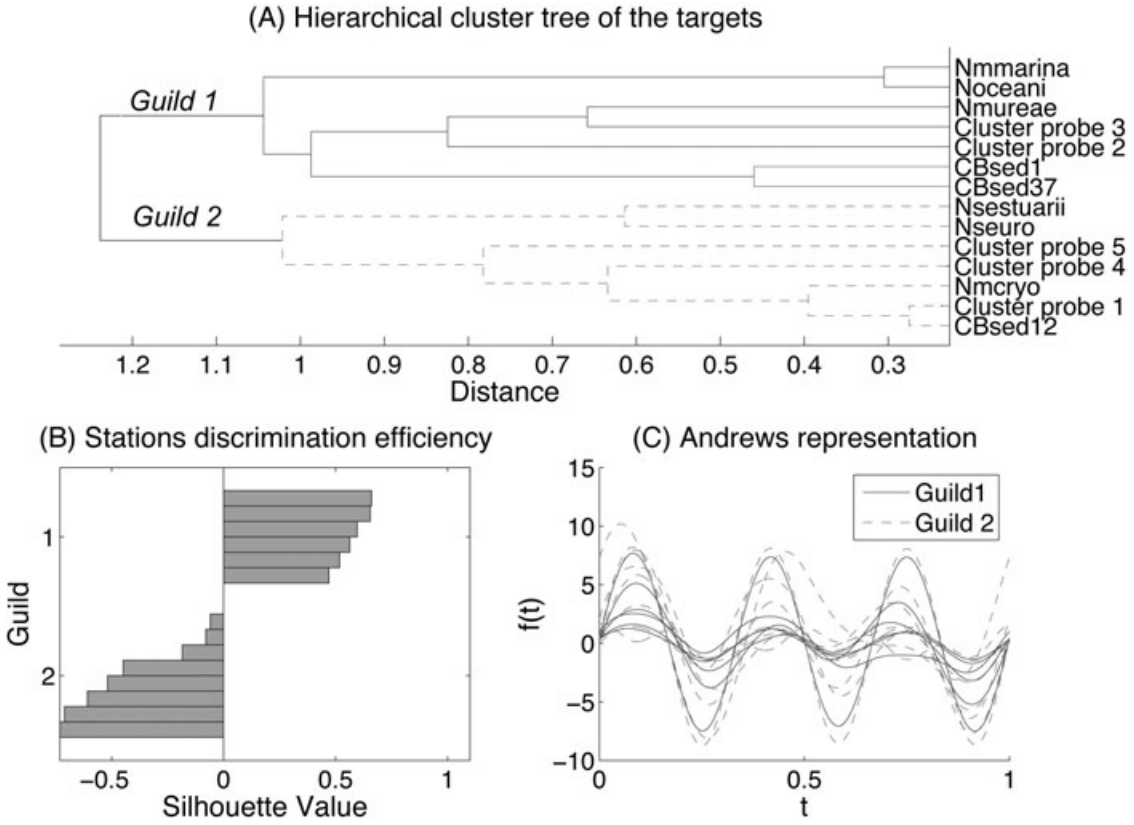


Fig. 5. Classification and discrimination analysis of the targets using hybridization results.
A. Classification results of the targets used in microarray experiments. Using a cut-off on the dendrogram, we are able to identify two groups.
B. Bars represent the sensitivity results after *K*-means discrimination algorithm application for two clusters. The classification (A) and the discrimination (B) approaches give the same groups.
C. An Andrews representation for each targets that highlights the efficiency of the discrimination analysis.

that hybridize with one of the four should in fact hybridize with all of them (Fig. S1, *Supplementary material*), and it is clear from the bar plot that these four probes did indeed behave similarly across the sample set (Fig. 4A). In contrast, the Nmestuarii probe was 88.6% identical to both CBsed13 and CBsed63, but only 82.9% identical to CBsed8. Although the three sediment clones exhibited strong hybridization signals, Nmestuarii never reached 1% of the total signal strength, so it was not considered a member of Cluster probe 1.

When the relative FR values for individual probes within each Cluster probe were summed, treating each Cluster probe as one signal, the patterns of the 14 independent probes (nine single probes and five Cluster probes) (Fig. 4B) emphasize the dominance of Cluster probes 1, 2 and 3 in the overall distribution.

Analyses of hybridization patterns in environmental context

Due to the large variability of the data and the variability in probe capacity, we did not use fluorescence intensity directly to analyse the hybridization patterns, but instead used a correlation matrix derived from the relative variation of each probe signal (FR), not its absolute value. Based on the relative FR values at different stations (Fig. 4B), the result identifies trends in the hybridization patterns among probes, for example, showing those that tend to occur together. The classification/discrimination approaches applied to this matrix represent the repartitioning of the probes; the classification results are illustrated in Fig. 5A. A distinction of two well-defined groups (guilds) was allowed by a distance of 1.1. The classification was confirmed by a *K*-means discrimination that matches significantly 71% of the distribution in two clusters, moreover with a good discriminant criterion as the silhouette score (Fig. 5B).

The convergence of these results allows us to distinguish two well-defined guilds of probes on the basis of their hybridization patterns: Guild 1 = Nmmarina, Noceani, Nmureae, Cluster probe 3, Cluster probe 2, CBsed1, CBsed37. Guild 2 = Nmestuarii, Nseuro, Cluster probe 5, Cluster probe 4, Nmcryo, Cluster probe 1, CBsed12.

The Andrews representation (Fig. 5C) summarizes these multivariate data and the related guilds. The representation indicates similar frequency for both guild curves, which means that both guilds exhibit the same qualitative variation in the function of the stations. However, the amplitudes of the guild signals are different which implies that the probes exhibit different hybridization intensities as a function of station (i.e. environmental conditions). We interpret these results to imply that the hybridization patterns reflect variation in the distribution of the functional gene *amoA*, which is related to the ecophysiology of ammonia-oxidizing microorganisms. Thus, relative FR patterns reflect the variation in biological function related to organisms that possess *amoA*.

The guilds defined on the basis of relative FR in an environmental context constitute functional groups, which suggests some ecological significance of the hybridization patterns. Guild 1 contains mostly marine sequences (e.g. Nmmarina, Nmureae), and its signal is dominated by Cluster probe 3, which contains sequences derived from the water column at stations CB200 and CB300. Guild 2 includes the obviously freshwater AOB sequences (e.g. Nseuro and Cluster probe 4) and Cluster probe 1 (composed of sequences from CB100 sediments) is a main component of its signal. The signal from the cultivated strains was so low that to base the guild characteristics on them is not warranted. The ecological and physiological characteristics associated with organisms from which the sequences in Cluster probes 1, 2 and 3 were derived cannot be known *a priori*, but phylogenetic affiliations and subsequent analysis suggests that Cluster probe 3 represents the main marine planktonic *Nitrosospira*-like group and Cluster probe 1 represents an important *Nitrosomonas*-like group from freshwater sediments. This observation suggests some relationship between the hybridization results and the physical characteristics of the environment at the various stations.

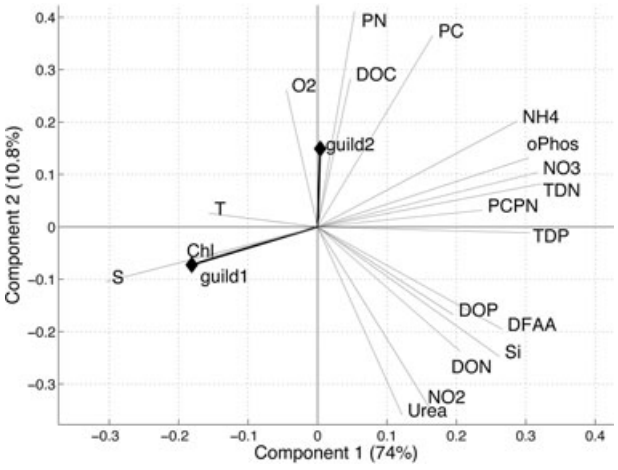


Fig. 6. Projection of the variables in the first component feature space in the principal components analysis. The first and the second component represent 74% and 10.8% of the global variability respectively. The bold variables correspond to the two guilds identified from the microarray hybridization patterns. Symbols represent salinity (S), temperature °C (T) and concentrations of oxygen (O_2), particulate nitrogen (PN), dissolved organic carbon (DOC), particulate carbon (PC), ammonium (NH_4), orthophosphate (oPhos), nitrate (NO_3), total dissolved nitrogen (TDN), ratio of particulate carbon/particulate nitrogen (PCPN), total dissolved phosphate (TCP), dissolved organic phosphate (DOP), dissolved free amino acids (DFAA), silicate (Si), dissolved organic nitrogen (DON), nitrite (NO_2), urea (Urea) and chlorophyll a (Chl).

Projection of the functional guilds into physical parameter space

These two guilds are composed of targets that tend to co-occur in the samples, and may represent two *amoA* communities with different ecophysiological characteristics and different geographical distributions, likely related to environmental conditions. The corresponding physical and chemical parameters for each sample were included in a correlation based principal components analysis (PCA) with the hybridization signals for the two guilds represented by the 14 independent probes (Fig. 6). The first and second components represent 74% and 10.8%, respectively, of the global variability in the data set. Guild 1 was positively correlated with salinity, temperature and chlorophyll concentration and negatively with most of the other variables. Guild 2 was positively correlated with oxygen, particulate nitrogen, particulate carbon and dissolved organic carbon. In the same feature space, the centroid of both guilds indicates an antagonistic relationship to urea, nitrite, dissolved organic nitrogen, silicate, dissolved organic phosphate and dissolved amino acids, which implies at least a non-linear effect of these parameters on the global *amoA* community.

Community composition as a function of station

One major goal of the microarray experiments is to allow comparison of microbial community composition among stations. Thus, a classification/discrimination approach similar to that described above was applied to the stations in order to classify or group them according to their hybridization patterns (Fig. 7A). A distinction of two clusters was allowed for 1.1 distance cut-off and a posteriori confirmed by a *K*-means approach for two clusters (cf. corresponding silhouette score for each station in Fig. 7B) and an Andrews representation (Fig. 7C). Thus, two main groups of stations were distinguished:

Station group 1: CB100S, CB200S, PL100S, CB300D, CB300S. Station group 2: SS100M, PL100M, PL100D, CB200D, CB100D, CB100M, SS100S, CB300M, CT200D.

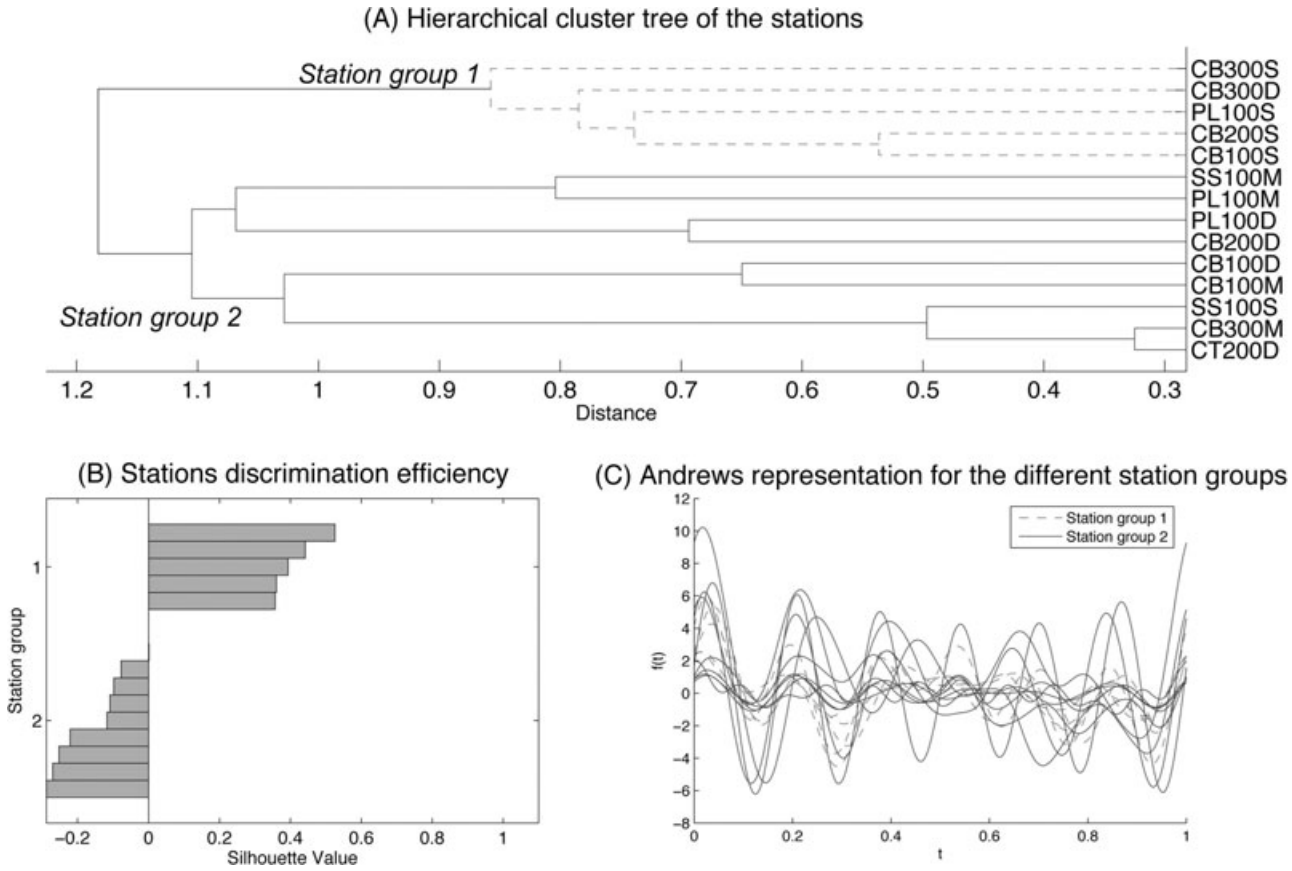


Fig. 7. Classification analysis of stations grouped by hybridization pattern.
A. Classification result of the stations based on hybridization results. Using a cut-off on the dendrogram, we are able to identify two groups.
B. Bars represent the sensitivity results after *K*-means discrimination algorithm application for two clusters. The classification and discrimination approaches give the same groups.
C. An Andrews representation for each station that highlights the efficiency of the discrimination analysis.

A classification of these stations according to physical and chemical parameters only was attempted. The discrimination efficiency for stations was less relevant and robust, however, than the guild discrimination. Moreover, using the available data, the physical and biological distributions are not significantly identical (55% identical), i.e. classifying the stations according to physical/chemical patterns results in different groups than when classifying them by hybridization pattern in 45% of the 10 000 iterations of the discrimination calculation.

Discussion

Interpretation of environmental microarray data

The microarray approach depends first on the development of an adequate sequence database, in order to represent on the array the full range of sequence diversity likely to be encountered in the environment (Francis *et al*., 2003; Stralis-Pavese *et al*., 2004), although comparisons among large numbers of samples on the basis of exhaustive clone libraries is prohibitive in time and expense. Due to their high throughput capabilities, microarrays have great potential for 'fingerprinting' microbial communities at both the DNA and RNA level for presence/absence and diversity information. There are two main issues, however, that preclude more quantitative interpretation of the array data at present.

Probe specificity. The ability of the individual probes to resolve closely related sequences is limited and variable (Marcelino *et al*., 2006), but for 70-mers and PCR products in various hybridization formats is about 15% sequence identity (Kane *et al*., 2000; Taroncher-Oldenburg *et al*., 2003; Steward *et al*., 2004). The two-colour competitive hybridization and the internal standard ratio hybridization methods showed comparable resolution in terms of target binding specificity. The heterologous probes most likely to hybridize at increased target concentration were robustly predictable on the basis of the phylogenetic tree shown in Fig. 1 – e.g. Nmestuarii is on

the outer edge of the cluster containing CBsed8, CBsed13 and CBsed63, which all cross react strongly with each other. At the target concentrations used for environmental samples, it is unlikely that any individual target reached levels (~25 ng) at which significant cross hybridization was observed in the specificity tests. Thus, cross reaction among probes that differ by ≥ 15% is unlikely in environmental experiments.

The degree of cross hybridization depends on the binding free energy of the probe/target hybrid molecule as well as the degree of sequence identity. Mismatched target/probe molecules generally hybridize at target concentrations less than 5 ng if the mismatched hybrid molecule has a calculated binding free energy ≥ 56% of the binding free energy of the perfect match target/probe hybrid (Taroncher-Oldenburg *et al*., 2003), although exceptions occur. For example, CB300s1 and CB200d4 are 84.3% identical and have binding free energy for the hybrid molecule of 45% of the CT200d4 perfect match. Nevertheless, CB300s1 hybridized consistently in simple mixtures (e.g. Fig. S1, *Supplementary material*) and in field samples as if it were indistinguishable from the other members of Cluster probe 3. Because the simple mixture results confirm that CB300s1 belongs in this cluster, we discount the possibility that the field data represent hybridization with an independent sequence that simply covaries with the other members of Cluster probe 3.

Nmestuarii, on the other hand, is 88.6% identical to CBsed8 and has a binding free energy equivalent to 49.9% of the CBsed8 perfect match. Although the Nmestuarii probe yielded significant signal in the simple mixture test such that it behaved like a member of Cluster probe 1, Nmestuarii never hybridized at a significant level in field samples. Apparently the targets from the natural samples that hybridized with the probes that comprised Cluster probe 1 were distinct from Nmestuarii.

Despite these constraints, the degree of resolution offered by the array hybridizations is still relatively high and is sufficient to be ecologically informative. Ammonia-oxidizing bacteria differ at the 16S rRNA gene level by less than 10% over the whole gene (Purkhold *et al*., 2003), and the amount of variation in the most commonly studied 490 bp fragment of the *amoA* gene is on the order of 25% (Francis *et al*., 2003). By selecting a highly variable 70-mer region from within the *amoA* fragment, the array probe sequences differ by as much as 70%, vastly increasing the power to distinguish among types. The minimum probe resolution for the 14 independent probes used here was about 15%, which corresponds roughly to the species cut off of 2–3% at the 16S rRNA level. Purkhold and colleagues (2000) compared the 16S rRNA and *amoA* similarities from cultured AOB strains, and found that a 3% 16S rRNA distance equates to a 15% *amoA* distance. Similar patterns were observed in correlated shifts of paired environmental samples using analyses of AOB-specific 16S rRNA similarity compared with *amoA* similarity (O'Mullan and Ward, 2005). Thus, the array resolves *amoA* community composition at an ecologically relevant level.

Probe capacity. Differential fluorescence intensity among perfect match probe/target combinations has not been widely considered as a factor in the interpretation of array hybridization data in environmental studies. Using short oligonucleotide probes on a tiling array, and hybridizing to rRNA targets from microbial eukaryotes, it was concluded that hybridization intensity cannot be predicted from thermodynamic calculations based on probe sequence and mismatch characteristics (Pozhitkov *et al*., 2006), and suggested this may be due to the inability of thermodynamic calculations to explain the behaviour of the surface-bound probe molecule interacting with the soluble target. While it is not practical to evaluate each individual probe on arrays containing hundreds or thousands of features, for the small probe set used here, it was practical to test most probes separately or in non-cross reacting mixtures (Fig. 3). If the extreme case of Cluster 1, for which only one experiment is available, is excluded, the range of relative fluorescence capacity for different perfect match probes is > 10-fold (0.14–1.8 for CBsed1 and CBsed37 respectively) relative to CBsed12. Full length (~1500 bp) 16S rDNA probes hybridized with 18–21-mer oligonucleotide probes exhibited capacity differences of at least sevenfold for perfect match probe/target hybrids and up to nearly 200-fold for probes with small but similar levels of mismatch (Sanguin *et al*., 2006). Pozhitkov and colleagues (2006) reported intensity variations among perfect match probes of 70-fold that could not be explained by thermodynamic or sequence characteristics. Some of the variability in capacity may be explained by difference in melting temperature; for the probes tested here, the fluorescence capacity generally decreased as melting temperature increased. Sanguin and colleagues (2006) reported the same tendency, but also noted unexplained examples of complete failure of hybridization for certain locations in the amplicon, even for probes targeting perfect match sequences. Steric hindrance of long PCR products or loss of target to competitive rehybridization with the complementary strand of the PCR product may explain part of capacity variability. However, these issues were minimized in our protocol because the target itself was a Klenow-labelled product, i.e. a mixture of fragments < 450 bp in length, and the mixture should have been random and similar among target preparations.

The net effect of both of these issues is to introduce unknown variation into the relationship between target abundance and hybridization fluorescence intensity. Both resolution and capacity can be quantified and corrected

for in known mixtures, but not for unknown mixtures, e.g. environmental samples. Because of the thermodynamic unpredictability of hybridization, we cannot assume that all targets that hybridize with a certain probe (perfect match targets plus mismatch targets with up to ~15% identity differences) will exhibit the same capacity, and thus we cannot simply apply the capacity correction that was determined for the perfect match hybridization to unknown mixtures. Therefore, the ecological analyses presented here were not based on absolute fluorescence intensity. *amoA* occurs in multiple copies in most AOB (up to three copies in cultivated strains; Norton *et al*., 2002), which introduces further variability in the relationship between hybridization intensity and AOB abundance. Even given these limitations for quantification, striking patterns emerged in the distribution of positive hybridization.

Ammonia-oxidizing bacteria community composition

The range of *amoA* diversity represented on the array focused on sequences represented in clone libraries from the water column and sediments of the Chesapeake system, but included sequences representing most, but not all, of the cultivated species/strains. The most striking result of the hybridization patterns is that the signal strength for the cultivated strains was much lower than for most of the environmental sequences. The target populations were produced by PCR amplification using *betaproteobacteria*-specific *amoA* primers. Therefore, Noceani should not have been represented in the target mixture and no signal was ever detected with the Noceani probe. All of the other cultivated strain probes, however, represent species that can be easily amplified with these primers. Nsbriensis was the only probe from a cultivated strain that hybridized to any extent and it was most important in the deep sample from the most freshwater station (Fig. 4A). This probably represents an influence from the surrounding terrestrial environment and indicates the organisms with *amoA* sequences very similar to *N. briensis* may be important in that environment. The capacity of Cluster 4, which included Nsbriensis, was only 27% of CBsed12, so its contribution is underestimated in Fig. 4B, but even if it were possible to correct for capacity, Cluster 4 would still be a minor component of the other samples.

The strongest hybridization was observed with probes representing Cluster probe 3, a group within the uncultivated marine *Nitrosospira*-like sequences in the overall *amoA* phylogeny. The reproducible and widespread strong detection of this group is consistent with the group's previously implied importance in the marine environment on the basis of clone libraries. The relative contribution of Cluster probe 3 was often more than 50%, of the total signal. The lowest value was observed in the deep water sample from the upper Chesapeake Bay, where Nsbriensis was an important signal. This upper bay sample was one of the least marine in character, and the deep water composition may reflect a community of freshwater sediments unlike that in the upper water column of the larger Bay environment. With a fluorescence capacity of 60% relative to CBsed12, Cluster 3 would be underrepresented in Fig. 4B, making its contribution even more overwhelming in many of the samples.

The high fluorescence capacity of Cluster 1 implies that its contribution to the target composition is overrepresented in Fig. 4B. The single probe CBsed12 exceeded 30% of the total signal in the Sargasso Sea surface sample. None of these sequences are associated with cultivated strains, but all represent members of the *Nitrosospira*-like and *Nitrosomonas*-like clades, respectively, that are frequently reported as major components of 16S rRNA and *amoA* clone libraries from marine sites (Bano and Hollibaugh, 2000; Freitag and Prosser, 2004; O'Mullan and Ward, 2005). Cluster probe 3 was identified as a component of Guild 1, while both Cluster probe 1 and CBsed12 were part of Guild 2, suggesting that the organisms represented by these sequences may represent ecologically distinct niches.

Ecological significance of guilds

The *amoA* array as described here was used to detect DNA directly. It thus provides an indication of the relative abundance and diversity of different *amoA* phylotypes, as reflected in pooled PCR products, at the level of genetic capability for ammonia oxidation. Visual inspection of the hybridization patterns (Fig. 4B) reveals obvious groups of stations: CT200D, CB200D, PL100D and SS100M were all dominated by Guild 1, especially Cluster probe 3. CB100S, CB100D, CB200S, CB300S, CB300M, CB300D, PL100S, PL100M and SS100S all showed an important contribution from Guild 2, mostly due to Cluster probe 1 (*Nitrosomonas*-like sequences) and CBsed12. Clear biogeographic patterns emerged in the identification of the two guilds and their relationship to environmental variables (Figs 5 and 6).

The classification of stations on the basis of phylotype patterns (Fig. 7) is, however, significantly different from the classification of stations on the basis of physical and chemical variables. That is, some stations with similar environmental characteristics differ in *amoA* community composition. Thus, it is not possible to identify completely the environmental variables that determine the biogeographical patterns evident in the hybridization data.

The next step in the application of microarray technology to understanding biogeochemical processes in the environment is to assess gene expression by hybridizing with mRNA or cDNA derived from environmental samples.

RNA/cDNA might reflect short time scale changes but due to the slow growth rate of AOB, DNA may be a better reflection of AOB response on the time scale of environmental variability in the bay. That is, highly abundant types must be successful in the environment as integrated over the timescale of the last few bacterial lifetimes. Methanotrophs are analogous to AOB in many ways, and similar methodology has recently been used to investigate the community composition of methanotrophs under various landfill and vegetation conditions (Stralis-Pavese *et al*., 2004). This study used a very different microarray approach (Bodrossy *et al*., 2003), in which short oligos (18–24-mers) and competitive two-colour methodology were employed. Stralis-Pavese and colleagues (2004) were able to identify the dominant methanotrophs at the DNA level and to identify different groups associated with different experimental conditions. Using RNA targets (Bodrossy *et al*., 2006), the same major groups were identified, but RNA was detected for some groups not detected at the DNA level – groups that must have responded particularly well to current conditions.

Ecology of AOB in estuarine environments

Salinity appears to be a major environmental selection factor for estuarine microbial communities, both for the total community (Bouvier and del Giorgio, 2002; del Giorgio and Bouvier, 2002) and for AOB assemblages in particular (Bianchi *et al*., 1999; de Bie *et al*., 2001; Bernhard *et al*., 2005). Microbial community composition (Bouvier and del Giorgio, 2002) and cellular activity (del Giorgio and Bouvier, 2002) both varied dramatically in the gradient from freshwater in the upper Choptank River to the moderately saline (14 psu) region where the Choptank joins Chesapeake Bay. *Betaproteobacteria* were the dominant bacterial clade in the freshwater environment and were replaced by *alphaproteobacteria* at the higher salinity stations. The greatest change occurred between 4 and 8 psu in the salinity gradient. Cottrell and Kirchman (Cottrell and Kirchman, 2004) documented the same community composition changes in the Delaware estuary.

Large shifts in the composition of 16S rRNA and *amoA* clone libraries representing AOB communities have been reported for the Schelde estuary in the Netherlands (de Bie *et al*., 2001; Bollman and Laanbroek, 2002), Waquoit Bay in the US (Bernhard *et al*., 2005), and Chesapeake Bay as an initial step in our study of *amoA* functional diversity in this system (Francis *et al*., 2003). *amoA* clone libraries (Francis *et al*., 2003) from the same CB and CT stations investigated here showed the greatest compositional differences between the upper bay (CB100) and upper river (CT100) stations compared with all the rest. On the basis of visual inspection of the hybridization patterns (Fig. 4B), the three freshwater samples (station CB100) were quite different from most other samples, likely representing the freshwater end member of the AOB community. The fact that CB200S (but not the other CB200 samples) was very similar to CB100S probably reflects the fact that surface water at CB200 originates in the freshwater region of CB100, while the deeper samples represent the deep water brought in from the mouth of the bay by the usual estuarine circulation. CT100 was not sampled for the present study, but CB100 was different from many of the other stations in its lower contribution by Guild 1 (Cluster probe 3) and the greater importance of members of Guild 2. At CB200 and CB300, the freshwater sediment Guild 2 was well represented in the surface samples, which was also important in the deep sample from CB300. CB200D and CB200D were both dominated by Guild 1, consistent with the seawater wedge of the estuary extending further north in the deep water than at the surface. Thus, the microarray results and clone library data are consistent in both their descriptions of the community and the ecological implications of variations in community composition, validating the microarray approach for high throughput investigations of community composition and diversity.

Salinity variations are not important in the ocean, at least nowhere near the scale of variability found in estuaries. Nevertheless, both major clades, *Nitrosospira*- and *Nitrosomonas*-like, are found in seawater, showing small-scale spatial and temporal variability (O'Mullan and Ward, 2005). It now appears that the AOA dominate the ammonia oxidizing community of the open ocean (Wuchter *et al*., 2006). The trend from high diversity to low in the AOB along the estuarine gradient is consistent with the idea that AOB are not very important in the ocean. Only a few types are capable of living in full strength seawater, and even those are not abundant. A similar study of AOA along the estuarine gradient might be expected to yield the opposite pattern.

Experimental procedures

Study site

The Chesapeake Bay drains a watershed of 166 000 km^2 and fills a dendritic river valley system consisting of a main channel and seven main rivers, including the Choptank River, a subestuary that contributes about 1% of the total freshwater to the bay. Six stations (Fig. 1) were chosen to span a range of environmental conditions; their characteristics have been described previously by Francis and colleagues (2003), who described the diversity of *amoA* clone libraries from this system. Characteristics of the depths sampled in April 2002 are provided in Table 1.

Sample collection and DNA extraction

The water column (two or three depths) was sampled in April 2001 and April 2002 using a pump in the river and Niskin

Table 1. Station locations and characteristics (data from April 2002).

Station	Depth (m)	T (°C)	Salinity	Nitrate (μM)	Chlorophyll a (μg l^{-1})	Sediment type
CT200D	7	ND	ND	0.03	4.79	Mud
CB100S	1.9	9.9	0.1	59.1	1.27	
CB100M	4	9.8	0.1	ND	ND	
CB100D	7.8	9.8	0.1	58.1	1.49	Mud
CB200S	2	9.9	15.1	4.5	8.53	
CB200D	19	9.0	21.1	0.4	8.81	Mud
CB300S	1.9	11.4	24.5	0.14	3.93	
CB300M	7.5	11.4	25.8	ND	ND	
CB300D	9.2	11.7	25.9	0.64	4.39	Sand/mud
PL100S	1	12.5	29.2	0.04	2.57	
PL100M	6.5	11.8	27.8	ND	ND	
PL100D	12.5	11.6	32.8	0.68	2.12	NA
SS100S	1	20.4	36.4	0.15	0.22	
SS100M	100	19.5	ND	0.10	0.23	

The temperature and salinity data were provided by W. Boicourt and the nutrient data were provided by P.M. Glibert, both of Horn Point Laboratory, University of Maryland Center for Environmental Sciences.
T, temperature; ND, not determined; NA, not applicable.

bottles on a CTD rosette in the bay, plume and Sargasso Sea. Samples collected in April 2001 were used to construct the clone library and samples collected in April 2002 were used in the array hybridization experiments. Water samples were collected from surface, mid depth and deep water, defined by the relative depth at each station: surface was generally 1 m, deep was 1 m off the bottom (except in the plume and the Sargasso Sea) and mid was in between, often chosen to sample a stratification feature such as the bottom of the thermocline or a particle maximum feature. Surface sediments were collected at the North Bay site (CB100) in April 2001 using a box core. A few grams of surface sediment were subcored using a cut-off plastic syringe, then scraped into a small tube and immediately frozen in liquid nitrogen. Samples were stored at −80°C until DNA extraction.

Particulate material from the water samples was collected by peristaltic pump filtration onto Sterivex GP 0.2 μm filter capsules (Millipore, Billerica, MA), which were drained, capped and frozen in liquid nitrogen immediately upon collection. DNA was extracted from the capsules using the PureGene DNA kit with slight modification of manufacturer's protocol for extraction of Gram negative bacteria (Gentra, Minneapolis, MN): lysis buffer (0.9 ml) was added to the filter and incubated with gentle rotation for 10 min at 80°C. Incubation times for all subsequent steps were the maximum recommended. Sediment DNA was extracted according to manufacturers instructions using the Ultraclean soil DNA kit (MoBio Laboratories, Carlsbad, CA).

Polymerase chain reaction, cloning, RFLP screening and sequencing

A fragment of the *amoA* gene was amplified from environmental samples and pure culture extracts using the betaproteobacterial *amoA* primers and PCR conditions of Rotthauwe and colleagues (Rotthauwe *et al*., 1997). Amplicons were purified by Wizard PCR prep (Promega, Madison, WI) and cloned into pCR2.1 vector using the Topo TA Cloning kit (Invitrogen, Carlsbad, CA). Transformants were screened for inserts by PCR using the *amoA* primers. At least 48 positive *amoA* amplicons from each sample were digested with Sau3A and HinfI for 3–6 h at 37°C to assess diversity of the cloned sequences. Digests were electrophoresed on a 3% Methaphor agarose gel (FMC BioProducts, Rockland, ME, USA) at 60 V, stained and photographed. Several representatives of each RFLP pattern type were then sequenced in both directions using *amoA* primers and Big Dye 3.1, according to manufacturers instructions (Applied Biosystems, Foster City, CA). Sequencing was performed on the ABI310 genetic analyzer.

Microarray oligonucleotide probe design

Rather than including every known sequence as a probe on the array, we chose a subset designed to represent the whole range of diversity with minimal cross reactivity among probes. Sequences from the April 2001 data set (surface and deep water samples from CB200, surface and deep water samples from CB300, and CB100 sediment) were locally aligned with each other and with other environmental and culture-based *amoA* sequences available in GenBank (NCBI). Phylogenetic trees were built using PAUP v 4.0 both for the total fragment (approximately 400 bp) and for several possible 70 bp stretches of sequence to find areas of the gene with the highest sequence dissimilarity that still represented the phylogenetic relationships visualized in the whole fragment phylogeny. Based on this comparison, 24 oligos were chosen, 13 from our environmental *amoA* sequences and 11 additional culture-based sequences (Table S1, *Supplementary material*). *Nitrosomonas eutropha* and *Nitrosomonas europaea* sequences are identical over the probe region, so the probe called Nseuro represents them both. Three cultivated sequences (*Nitrosococcus mobilis, Nitrosomonas communis* and *Nitrosomonas nitrosa*) have been detected rarely if at all in our clone libraries and were not available for testing so they were not included on the array. The original clone library survey was performed using *betaproteobacteria*-specific *amoA* primers and the same primers were used to produce target for the hybridization experiments (see next section). The 70-mer oligo probes (Operon Technologies)

were adjusted to a concentration of 0.05 μg μl^{-1} in 50% DMSO and were spotted on CMT-GAPS amino silane-coated glass slides (Corning, Corning, NY). After printing, the slides were baked at 80°C for 3 h and stored in the dark at room temperature under slight vacuum.

Hybridization experimental design

Two different protocols were used: (i) the standard two-colour competitive approach was used for the initial array testing and (ii) a new internal standard ratio method was used for the analysis of environmental samples and for capacity tests. The standard two-colour competitive hybridization method was used previously to establish hybridization criteria for functional gene oligonucleotide arrays (Taroncher-Oldenburg *et al.*, 2003). It was used here to characterize the behaviour of the *amoA* array in simple defined mixture experiments in order to verify the criteria established previously in a different gene system. The competitive hybridization approach was not appropriate for the actual analysis of field samples, however, because it is difficult to define a true 'control' sample or a 'normal' sample against which all others should be compared, as is the case for gene expression in diseased tissue versus healthy tissue, for example. One of the major advantages of the competitive approach, however, is the use of FRs, rather than simple hybridization intensity, to quantify the hybridization signal for each feature and to avoid artefacts associated with variations in hybridization patterns across the slide. Therefore, in the internal standard ratio method, we spotted a reference oligo (5′-GATCCCCGGGAATTGCCATG-3′) in several replicate features in each block so that an average FR could be computed for each *amoA* feature (*amoA*-specific fluorescence/reference fluorescence), allowing us to correct for across slide variability in hybridization intensity.

Hybridization experiments with known mixtures – probe resolution and cross reactivity

Several experiments with known mixtures of targets were used to investigate the resolution of the array using the two-colour competitive hybridization approach. Target was prepared by the incorporation of amino-allyl-dUTP (dUaa) into the PCR product during the *amoA*-specific PCR amplification step. The PCR protocol was the same as described for gene-specific *amoA* PCR above, except that dTTP was omitted from the dNTP mixture and was replaced with a combination of dTTP and dUaa-tagged dUTP (in a ratio of 1:2 at the same final concentration as the individual dNTP additions). Several duplicate dUaa-PCR products were pooled and cleaned using the Qiaquick spin column protocol (Qiagen, Valencia, CA, USA), and dried under vacuum. dUaa-labelled PCR products were dissolved in 4.5 μl of 100 mM Na_2CO_3 buffer (pH 9.5) and allowed to incubate at room temperature in the dark for 15 min. Then 4.5 μl of Cy3 or Cy5 dissolved in DMSO (1000-pmol μl^{-1}) were added and the incubation continued for 1 h in the dark, followed by the addition of 4.5 μl of 4 M hydroxylamine and an additional 15 min in the dark. The labelled target was once more cleaned using a Qiaquick spin column, and adjusted to the appropriate concentration, by drying under vacuum and resuspending in water, and stored frozen in the dark. Equal amounts of the two competitive target PCR products were heated to 95°C for 5 min, then cooled on ice and added to hybridization solution (Clontech, Franklin, NJ; prewarmed to 65°C). Prehybridization and hybridization at 65°C and the subsequent washes were performed as previously described (Taroncher-Oldenburg *et al.*, 2003). Experiments were performed in duplicate, one slide with the two targets labelled Cy3 and Cy5 and the second slide in the label inverse combination. The dried slides were stored at room temperature in the dark and scanned using a GenePix 4000 A scanner (Axon Instruments, Foster City, CA) and the GenePix Pro software provided with the scanner. The fluorescence data were transferred to Excel spreadsheets for manipulation and analysis.

Hybridization experiments with known mixtures – signal intensity

The fluorescence capacity of independent probes, i.e. the signal intensity produced by hybridization of a standard amount of target, was tested in experiments with single targets or mixtures of non-cross-reacting probes (established above).

Hybridization targets were produced from cloned *amoA* gene fragments using the standard *amoA* gene-specific PCR amplification protocol described above. Incorporation of label was accomplished using linear amplification with the Klenow enzyme, which is randomly primed, minimizing sequence-specific differential labelling. Products from parallel reactions were pooled and cleaned with Qiaquick PCR clean-up columns and then tagged with amino-allyl-dUTP by random priming using Klenow fragment and random octomers supplied in the BioPrime labelling kit (Invitrogen). The standard dNTP mixture was replaced with a mixture of 1.2 mM dACG plus a 1:2 mixture of 1.2 mM dTTP and 1.2 mM dUaa. Parallel reactions were pooled and again cleaned with Qiaquick columns, and the eluted dUaa-labelled DNA was dried under vacuum and stored as a dry pellet at –20°C.

dUaa-labelled fragments were coupled to Cy3 as described above. The fluorescently labelled target was then cleaned again with Qiaquick columns and dried to a pellet under vacuum. Immediately prior to hybridization, the target was dissolved in water using a volume calculated to provide a convenient addition to the hybridization mixture. The concentration of target was computed by measuring the DNA concentration (PicoGreen assay, Molecular Probes) and Cy3 concentration on 1 μl of the Qiaquick eluate (diluted to 500 μl) prior to the last drying step using a Perkin Elmer LS 55 Luminescence Spectrometer. Hybridization solution, prewarmed to 65°C, contained 100 ng of the Cy3-labelled target mixture plus 2 μl (200 pmol) Cy5-labelled reverse complement of the 20-mer reference in a total volume of 80 μl. Prehybridization, hybridization and post-hybridization washes were done as described previously (Taroncher-Oldenburg *et al.*, 2003).

Hybridization experiments with field samples

Ammonia-oxidizing bacteria are generally a minor component of the total microbial biomass, and thus it was not

possible to avoid the use of PCR to amplify the target gene sequence from natural water samples in order to detect those sequences on the *amoA* array. Polymerase chain reaction bias cannot be avoided, but we make the simplest assumption that the bias necessitated by our use of PCR to amplify *amoA* targets resulted in consistent bias across all our samples, and is the same bias introduced by use of the same primers to build the clone library from which the probes were derived (see probe design section). To maximize consistency, the simplest PCR protocol (i.e. without direct label incorporation) was employed in target preparations from environmental samples, and multiple replicate PCR products were pooled for the production of hybridization target. Target DNA from environmental samples was prepared using the Klenow labelling protocol and analysed using the hybridization protocols used in the capacity tests above.

Quantification of hybridization signals

Hybridization signals in the competitive two-colour experiments were filtered using the data filtering and processing criteria to ensure evenness and reproducibility and to normalize for dye intensity and labelling efficiency as described by Taroncher-Oldenburg and colleagues (2003). Only signals that were consistent with these criteria and with significant fluorescence in at least five of the eight replicate features for each probe on both of the label-inverse slides were considered positive. The results are expressed in terms of relative fluorescence units (RFU) to represent the labelling efficiency normalized data expressed as the $\log_2$ of the Cy5/Cy3 ratio (Taroncher-Oldenburg *et al*., 2003).

The hybridization experiments using the internal standard method employed a normalization procedure to correct for differences in hybridization strength across the array. The array contained two blocks, which contained identical patterns of features arranged in 16 subblocks. Each feature was represented in four replicates in each block (eight total replicates for the entire array). The array included features containing the reference 20-mer and features containing several gene-specific 70-mer *amoA* oligos in each subblock. The initial steps in data filtering were similar to that used for quality control in the two-colour competitive method: fluorescence data were screened to identify features with above background fluorescence. A significant signal was defined as one that was greater than the average signal of all empty features on the slide by at least 2 standard deviations. The ratio of Cy3 to Cy5 fluorescence was then computed for every *amoA* feature in each subblock using the Cy3 fluorescence for the *amoA* feature and the Cy5 fluorescence from the closest reference features within each subblock. The four replicates for each block were then averaged to obtain an upper and a lower block *amoA*/reference FR. The average *amoA*/reference FRs for the lower block were multiplied by the ratio (upper block reference Cy5 lower^{-1} block reference Cy5) to normalize the signal between blocks. The resulting value represents the final FR for each probe. Only if the average FR was $\geq 2\times$ the standard deviation for replicate features among blocks was the FR considered significantly greater than zero. The FR values for all features with significant signal were summed and the signal for each probe is presented as a fraction of the total (relative fluorescence ratio; relative FR). The relative FR data can be compared among slides despite variations in absolute signal strength and they were used in the computation analysis below to investigate clustering among probe signals and relationships to environmental variables.

Determination of ecological relationships

There were two goals of the analysis of microarray data: the grouping of probes with similar functionality and the analysis of the relationship between environmental properties and occurrence of probe groups. The probes chosen on the basis of sequence identity (described above) were evaluated a posteriori using the hybridization data from field samples. The probe grouping started with the calculation of a matrix describing the pairwise correlation between probes based on the microarray fluorescence data matrix. We calculated pairwise correlation distances between the targets rather than Euclidean-like distances because large variability of the RFR allows the use only of a semi-quantitative approach such as the correlation distance. For this comparison based on correlation coefficients, summing and averaging the component signals in the clusters probes produce mathematically equivalent results (whereas this would not be true if the absolute values of individual probe signals were used for the comparisons). The resulting correlations were visualized as a hierarchical cluster tree using unweighted average distance (UPGMA) (Legendre and Legendre, 1998 for data analysis overview). We used this classification to find the clusters of probes that were similar enough to be considered as effectively one probe for a given similarity level. I.e., this analysis tested whether the unknown mixtures (field samples) behaved as predicted on the basis of probe sequence similarity.

Two different discrimination methods were used to analyse the probe relationships. The first was the standard *K*-means algorithm (Hartigan, 1975) applied on the hybridization results. We estimated the discriminant efficiency using a silhouette score (Kaufman and Rousseeuw, 1990). The silhouette value for each observation is a measure of how similar a probe is relative to the other probes from the same cluster compared with its similarity to the probes from other clusters. The probes from the same cluster must possess similar silhouette scores for the cluster to provide a relevant discrimination.

The second test of discrimination was an Andrews representation (Andrews, 1972), in which each probe is represented by one curve. Each curve is a sum of sines and cosines of varying frequencies (i.e. similar to a Fourier transform for each target hybridization result) with the amplitude of each term determined by the value of the correlation distance with the appropriate other target. Similar probes have similar curves. If the probes within a cluster do not have similar curves, the classification is suspect.

To compare the microarray results to the environmental data, we performed a PCA on the physical parameters after calculating the matrix describing correlations between samples (Jackson, 1991). The resulting vectors describe eigenvectors (characteristic patterns of variation) as a function of station. The patterns of any environmental variable or probe can be described as a linear combination of these

patterns. The *amoA* hybridization results were included in the PCA by computing the representation of each guild in each sample and including guild representation along with the physical variables. The *amoA* Guilds were then projected into this physical feature space. This identifies the stations that are similar in terms of *amoA* community composition.

Data files

All MIAME-compliant data are posted on the project website: http://snow.tamu.edu/arrayms_data/index.htm.

Acknowledgements

This research was supported by grants from the US National Science Foundation to B.B.W. and G.A.J. The authors thank the members of their lab groups for help with field work and in critically reviewing previous drafts of the manuscript.

References

Andrews, D. (1972) Plots of high dimensional data. *Biometrics* **28:** 125–136.

Bano, N., and Hollibaugh, J.T. (2000) Diversity and distribution of DNA sequences with affinity to ammonia-oxidizing bacteria of the beta subdivision of the class *Proteobacteria* in the Arctic Ocean. *Appl Environ Microbiol* **66:** 1960–1969.

Bernhard, A.E., Donn, T., Giblin, A.E., and Stahl, D.A. (2005) Loss of diversity of ammonia-oxidizing bacteria correlates with increasing salinity in an estuary. *Environ Microbiol* **7:** 1289–1297.

Bianchi, M., and Feliatra, and Dominique, L. (1999) Regulation of nitrification in the land-ocean contact area of the Rhone River plume (NW Mediterranean). *Aquat Microb Ecol* **18:** 301–312.

Bodrossy, L., Stralis-Pavese, N., Murrell, J.C., Radajewski, S., Weilharter, A., and Sessitsch, A. (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ Microbiol* **5:** 566–582.

Bodrossy, L., Stralis-Pavese, N., Donrad-Koszler, M., Weilharter, A., Reichenauer, T.G., Schofer, D., and Sessitsch, A. (2006) mRNA-based parallel-detection of active methanotroph populations by use of a diagnostic microarray. *Appl Environ Microbiol* **72:** 1672–1676.

Bollman, A., and Laanbroek, H.J. (2002) Influence of oxygen partial pressure and salinity on the community composition of ammonia-oxidizing bacteria in the Schelde estuary. *Aquat Microb Ecol* **28:** 239–247.

Bouvier, T., and del Giorgio, P. (2002) Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. *Limnol Oceanogr* **47:** 453–470.

Breslauer, K.G., Frank, R., Blocker, H., and Markey, L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci USA* **83:** 3746–3750.

Cottrell, M., and Kirchman, D.L. (2004) Single-cell analysis of bacterial growth, cell size, and community structure in the Delaware estuary. *Aquat Microb Ecol* **34:** 139–149.

de Bie, M.J.M., Speksnijder, A.G.C.L., Kowalchuk, G.A., Schuurman, T., Zwart, G., Stephen, J.R., *et al.* (2001) Shifts in the dominant populations of ammonia-oxidizing beta-subclass *Proteobacteria* along the eutrophic Schelde estuary. *Aquat Microb Ecol* **23:** 225–236.

del Giorgio, P.A., and Bouvier, T.C. (2002) Linking the physiologic and phylogenetic successions in free-living bacterial communities along an estuarine salinity gradient. *Limnol Oceanogr* **47:** 471–486.

Francis, C.A., O'Mullan, G.D., and Ward, B.B. (2003) Diversity of ammonia monooxygenase (*amoA*) genes across environmental gradients in Chesapeake Bay sediments. *Geobiology* **1:** 129–140.

Freitag, T.E., and Prosser, J.I. (2004) Differences between betaproteobacterial ammonia-oxidizing communities in marine sediments and those in overlying water. *Appl Environ Microbiol* **70:** 3789–3793.

Freitag, T.E., Chang, L., and Prosser, J.I. (2006) Changes in the community structure and activity of betaproteobacterial ammonia-oxidizing sediment bacteria along a freshwater-marine gradient. *Environ Microbiol* **8:** 684–696.

Hartigan, J. (1975) *Clustering Algorithms*. New York, USA: John Wiley & Sons.

Higgins, D., Bleasby, A., and Fuchs, R. (1991) CLUSTAL V: improved software for multiple sequence alignment. *CABIOS* **8:** 189–191.

Hollibaugh, J.T., Bano, N., and Ducklow, H.W. (2002) Widespread distribution in polar oceans of a 16S rRNA gene sequence with affinity to Nitrosospira-like ammonia-oxidizing bacteria. *Appl Environ Microbiol* **68:** 1478–1484.

Jackson, J.E. (1991) *A User's Guide to Principal Components*. New York, USA: John Wiley & Sons.

Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu, J., Thomas, J.D., and Madore, S.J. (2000) Assessment of the sensitivity and specificity of oligonucleotide (50 mer) microarrays. *Nucleic Acids Res* **28:** 4552–4557.

Kaufman, L., and Rousseeuw, P. (1990) *Finding Groups in Data: An Introduction to Cluster Analysis*. New York, USA: John Wiley & Sons.

Kowalchuk, G.A., and Stephen, J.R. (2001) Ammonia-oxidizing bacteria: a model for molecular microbial ecology. *Annu Rev Microbiol* **55:** 485–529.

Kowalchuk, G.A., Stienstra, A.W., Heilig, G.H.J., Stephen, J.R., and Woldendorp, J.W. (2000) Changes in the community structure of ammonia-oxidizing bacteria during secondary succession of calcareous grasslands. *Environ Microbiol* **2:** 99–110.

Legendre, P., and Legendre, L. (1998) *Numerical Ecology*. Amsterdam, the Netherlands: Elsevier.

Loy, A., Kusel, K., Lehner, A., Drake, H.L., and Wagner, M. (2004) Microarray and functional gene analyses of sulfate-reducing prokaryotes in low-sulfate, acidic fens reveal cooccurrence of recognized genera and novel lineages. *Appl Environ Microbiol* **70:** 6998–7009.

Marcelino, L.A., Backman, V., Donaldson, A., Steadman, C., Thompson, R.J., Preheim, S.P., *et al.* (2006) Accurately quantifying low-abundant targets amid similar sequences by revealing hidden correlations in oligonucleotide microarray data. *Proc Natl Acad Sci USA* **103:** 13692–13634.

Nei, M., and Li, W.-H. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA* **76:** 5269–5473.

Nicolaisen, M.H., and Ramsing, N.B. (2002) Denaturing gradient gel electrophoresis (DGGE) approaches to study the diversity of ammonia-oxidizing bacteria. *J Microbiol Methods* **50:** 189–203.

Norton, J., Alzerreca, J., Suwa, Y., and Klotz, M. (2002) Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. *Arch Microbiol* **177:** 139–149.

O'Mullan, G.D., and Ward, B.B. (2005) Relationship of temporal and spatial variabilities of ammonia-oxidizing bacteria to nitrification rates in Monterey Bay, CA. *Appl Environ Microbiol* **71:** 697–705.

Peplies, J., Lau, S.C.K., Pernthaler, J., Amann, R., and Glockner, F.O. (2004) Application and validation of DNA microarrays for the 16S rRNA-based analysis of marine bacterioplankton. *Environ Microbiol* **6:** 638–645.

Phillips, C.G., Smith, Z., Embley, T.M., and Prosser, J.I. (1999) Phylogenetic differences between particle-associated and planktonic ammonia-oxidizing bacteria of the beta subdivision of the class *Proteobacteria* in the northwestern Mediterranean Sea. *Appl Environ Microbiol* **65:** 779–786.

Pozhitkov, A., Noble, P.A., Domazet-Loso, T., Nolte, A.W., Sonnenberg, R., Stehler, P., *et al.* (2006) Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted. *Nucleic Acids Res* **34:** e66.

Purkhold, U., Pommerening-Roser, 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 and amoA sequence analysis: implications for molecular diversity surveys. *Appl Environ Microbiol* **66:** 5368–5382.

Purkhold, U., Wagner, M., Timmermann, B., Pommerening-Roser, A., and Koops, H.-P. (2003) 16S rRNA and *amoA*-based phylogeny of 12 novel betaproteobacterial ammonia-oxidizing isolates: extension of the dataset and proposal of a new lineage within the nitrosomonads. *Int J Syst Evol Microbiol* **53:** 1485–1494.

Rotthauwe, J.-H., Witzel, K.-P., and Liesack, W. (1997) The ammonia monooxygenase structural gene amoA as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. *Appl Environ Microbiol* **63:** 4704–4712.

Sanguin, H., Herrera, A., Oer-Desfeux, C., Deschesne, A., Simonet, M., Navarro, E., *et al.* (2006) Development and validation of a prototype 16S rRNA-based taxonomic microarray for *Alphaproteobacteria*. *Environ Microbiol* **8:** 289–307.

Steward, G.F., Jenkins, B.D., Ward, B.B., and Zehr, J.P. (2004) Development and testing of a DNA macroarray to asess nitrogenase (*nifH*) gene diversity. *Appl Environ Microbiol* **70:** 1455–1465.

Stralis-Pavese, N., Sessitsch, A., Weilharter, A., Reichenauer, T., Riesing, J., Csontos, J., *et al.* (2004) Optimization of diagnostic microarray for application in analysing landfill methanotroph communities under different plant covers. *Environ Microbiol* **6:** 347–363.

Taroncher-Oldenburg, G., Griner, E., Francis, C.A., and Ward, B.B. (2003) Oligonucleotide microarray for the study of functional gene diversity of the nitrogen cycle in the environment. *Appl Environ Microbiol* **69:** 1159–1171.

Webster, G., Embley, T., and Prosser, J. (2002) Grassland management regimes reduce small-scale heterogeneity and species diversity of proteobacterial ammonia oxidizer populations. *Appl Environ Microbiol* **68:** 20–30.

Wuchter, C., Abbas, B., Coolen, M.J.L., Herfort, L., van Bleijswijk, J., Timmers, P., *et al.* (2006) Archaeal nitrification in the ocean. *Proc Natl Acad Sci USA* **103:** 12317–12322.

Supplementary material

The following supplementary material is available for this article online:

Fig. S1. The results of an example of a competitive two-colour hybridization with two inversely labelled targets are shown in Fig. S1. CT200s1 and CBsed8 PCR products were the two competitively hybridized targets (~25 ng each). In the pairwise experiment, one microarray (slide 1) was hybridized with each pair of targets added in equal concentrations (e.g. CT200s1-Cy3 and CBsed8-Cy5) and a second identical microarray (slide 2) was hybridized with the same targets at the same concentration in the opposite label combination (CT200s1-Cy5 and CBsed8-Cy3). The results are shown for all probes that yielded a significant hybridization signal, plotted as each probe's RFU on slide 1 versus its RFU on side 2 (Fig. S1). The only probes that exhibited significant RFU values on both label inverse slides were those that had high sequence identity (Table S1). All the ratios, normalized to account for labelling efficiency and between-slide differences in intensity, fall very close to the 1:1 line, indicating that each probe hybridized equally well to both members of the paired sets of targets.

Fig. S2. Examples of hybridization signals from two arrays. Bar height is the average fluorescence ratio (Cy3/Cy5) of the eight replicate features for each probe on the array. Error bars denote standard deviations of these replicates. At CT200D (Fig. S2A), the four individual probes comprising Cluster probe 3, representing environmental sequences, dominated the hybridization, with very little signal present for probes derived either from other environmental sequences or cultivated strains. At CB100D (Fig. S2B), the four probes that dominated at CT200D were not important; a strong signal was present for probes derived both from other environmental sequences as well as cultivated strains.

Table S1. Probe names and sequences for the 70-mer oligonucleotide probes representing *amoA* gene fragments on the array. Melting temperature was calculated by the nearest neighbour method (Breslauer *et al.*, 1986). Surface and deep refer to water column samples; accession numbers are provided as the source for sequences of cultivated strains.

Table S2. Per cent identity and for oligonucleotide probes on the *amoA* array. Grey boxes indicate pairwise sequence identifies > 85%, such that hybridization between these pairs is predicted under standard conditions. Outlined boxes indicate identities slightly less than 85%, for pairs in which hybridization was consistently observed.

This material is available as part of the online article from http://www.blackwell-synergy.com

Fig. S1. The results of an example of a competitive twocolour hybridization with two inversely labelled targets are shown in Fig. S1. CT200s1 and CBsed8 PCR products were the two competitively hybridized targets (~25 ng each). In the pairwise experiment, one microarray (slide 1) was hybridized with each pair of targets added in equal concentrations (e.g. CT200s1-Cy3 and CBsed8-Cy5) and a second identical microarray (slide 2) was hybridized with the same targets at the same concentration in the opposite label combination (CT200s1-Cy5 and CBsed8-Cy3). The results are shown for all probes that yielded a significant hybridization signal, plotted as each probeâ€™s RFU on slide 1 versus its RFU on side 2 (Fig. S1). The only probes that exhibited significant RFU values on both label inverse slides were those that had high sequence identity (Table S1). All the ratios, normalized to account for labelling efficiency and between-slide differences in intensity, fall very close to the 1:1 line, indicating that each probe hybridized equally well to both members of the paired sets of targets.

Fig. S2. Examples of hybridization signals from two arrays. Bar height is the average fluorescence ratio (Cy3/Cy5) of the eight replicate features for each probe on the array. Error bars denote standard deviations of these replicates. At CT200D (Fig. S2A), the four individual probes comprising Cluster probe 3, representing environmental sequences, dominated the hybridization, with very little signal present for probes derived either from other environmental sequences or cultivated strains. At CB100D (Fig. S2B), the four probes that dominated at CT200D were not important; a strong signal was present for probes derived both from other environmental sequences as well as cultivated strains.

Table S1. Probe names and sequences for the 70-mer oligonucleotide probes representing *amoA* gene fragments on the array. Melting temperature was calculated by the nearest neighbour method (Breslauer *et al.*, 1986). Surface and deep refer to water column samples; accession numbers are provided as the source for sequences of cultivated strains.

Table S2. Per cent identity and for oligonucleotide probes on the *amoA* array. Grey boxes indicate pairwise sequence identifies >?85%, such that hybridization between these pairs is predicted under standard conditions. Outlined boxes indicate identities slightly less than 85%, for pairs in which hybridization was consistently observed.

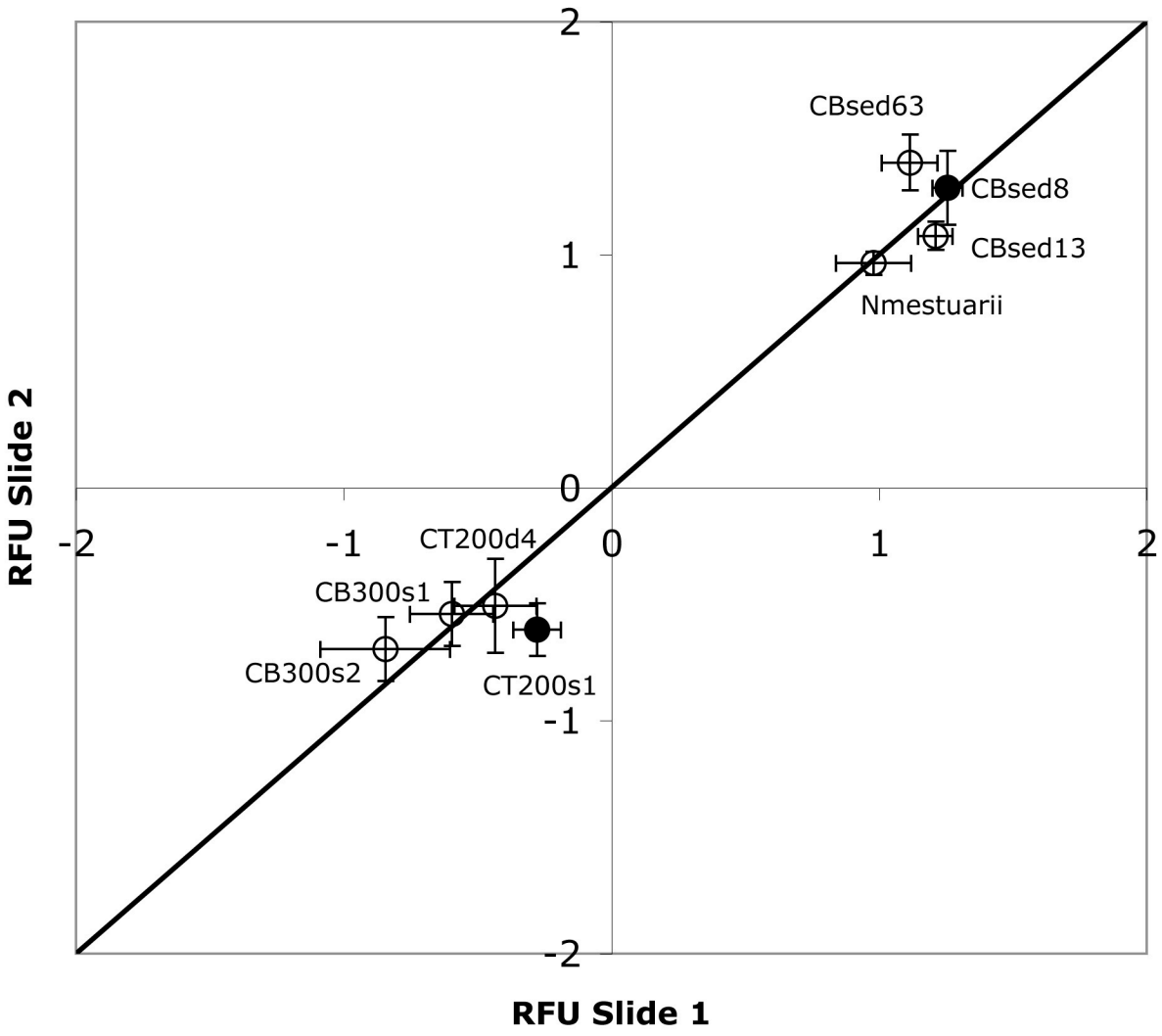

RFU Slide 2
RFU Slide 1
2
1
0
-1
-2
-2
-1
0
1
2
CBsed63
CBsed8
CBsed13
Nmestuarii
CT200d4
CB300s1
CB300s2
CT200s1

Figure S1

Figure S2A

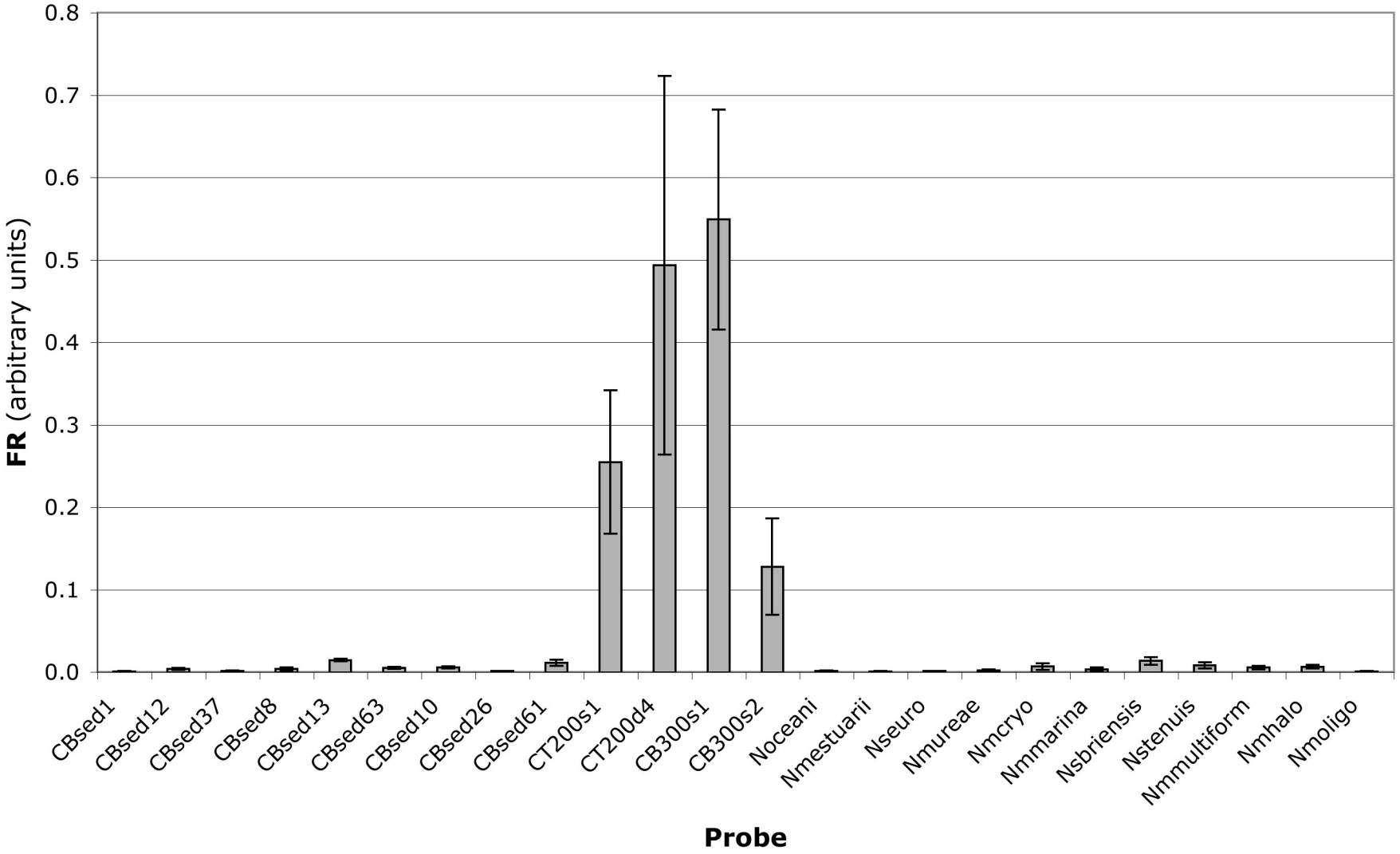
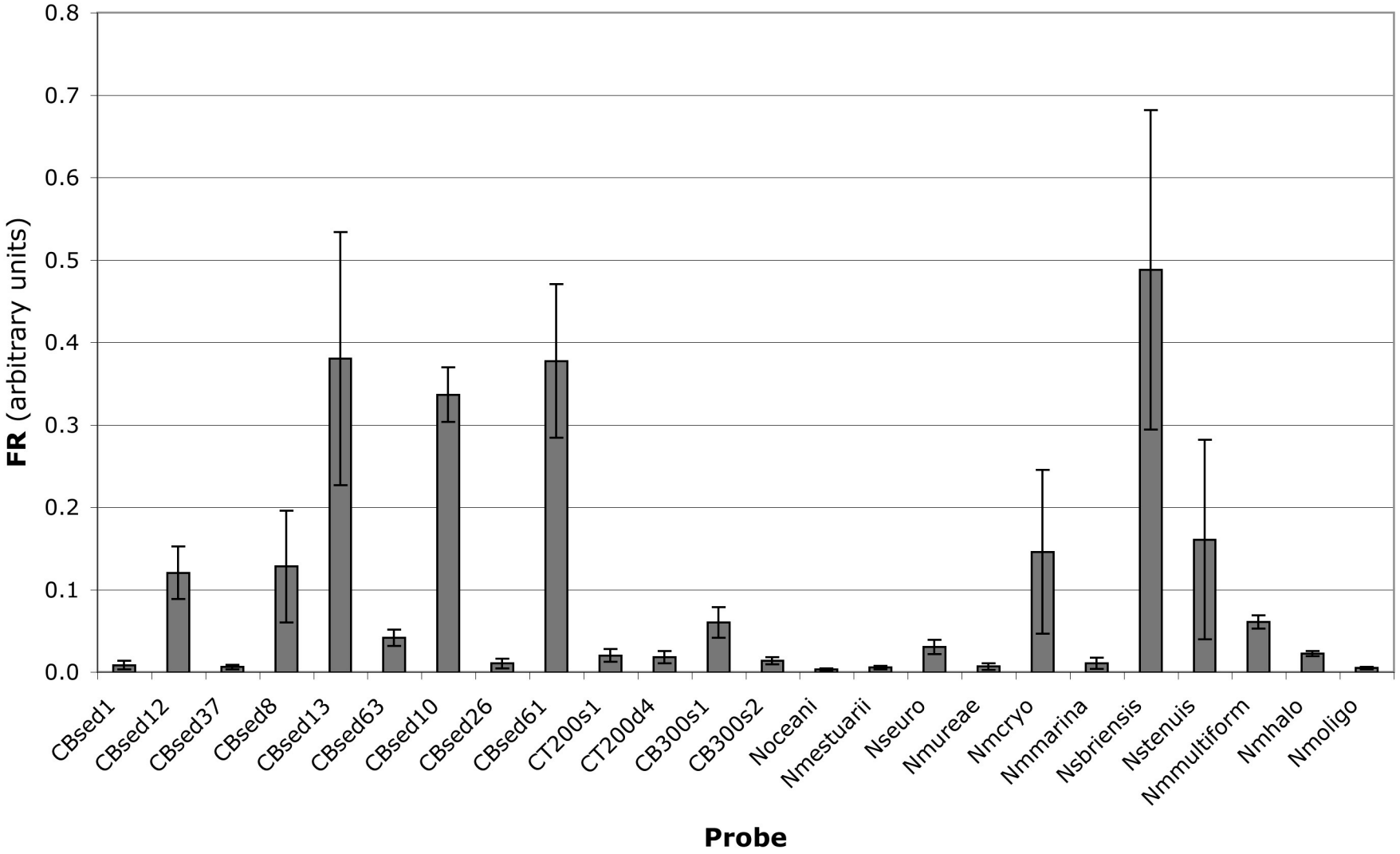

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
FR (arbitrary units)
CBsed1
CBsed12
CBsed37
CBsed8
CBsed13
CBsed63
CBsed10
CBsed26
CBsed61
CT200s1
CT200d4
CB300s1
CB300s2
Noceani
Nmestuarii
Nseuro
Nmureae
Nmcryo
Nmmarina
Nsbriensis
Nstenuis
Nmmultiform
Nmhalo
Nmoligo
Probe

Figure S2B


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
FR (arbitrary units)
CBsed1
CBsed12
CBsed37
CBsed8
CBsed13
CBsed63
CBsed10
CBsed26
CBsed61
CT200s1
CT200d4
CB300s1
CB300s2
Noceani
Nmestuarii
Nseuro
Nmureae
Nmcryo
Nmmarina
Nsbriensis
Nstenuis
Nmmultiform
Nmhalo
Nmoligo
Probe

Supplementary Materials

Table S1. Probe names and sequences for the 70-mer oligonucleotide probes representing *amoA* gene fragments on the array. Melting temperature was calculated by the nearest neighbor method (Breslauer et al., 1986). Surface and deep refer to water column samples; accession numbers are provided as the source for sequences of cultivated strains.

Probe name	DNA sequence 5' - 3"	Source	%GC	Tm
CBsed1	GTTACTGACACGTAACTGGTTGGTTACAGCATTGTTGGGTGGTGGCTTCTTTGGATTATTCTTCTACCCA	CB100 sediment	44.3	87.1
CBsed12	GTTGCTGACCCGCAGCTGGTTGATCACGGCCTTGGTAGGGGGTGGTGCATTCGGGCTGCTGTTCTATCCG	CB100 sediment	61.4	85.1
CBsed37m	ATATCTAACCCGTAGTTGGCTCGTGACAGCGTTAATTGGTGGTGGATTCTTTGGGTTATTCTTCTATCCA	CB100 sediment	42.9	77.5
CBsed8	GTTACTGACAGGTAACTGGTTGGTTACGGCATTGCTAGGTGGTGGGTTTTGGGGTCTGTTCTTTTACCCT	CB100 sediment	48.6	79.9
CBsed13	GTTACTGACAGGTAACTGGTTGGTTACGGCATTGCTAGGTGGTGGTTTCTGGGGCTTGTTCTTTTATCCT	CB100 sediment	47.1	79.3
CBsed63	GTTACTGACAGGTAACTGGTTGATTACGGCATTACTAGGTGGTGGATTCTGGGGTCTGTTCTTTTACCCT	CB100 sediment	45.7	78.7
CBsed10	ACTGTTGACGGGTAACTGGCTGGTTACAGCATTGTTAGGAGGTGGTTTCTGGGGCTTGTTTTTCTATCCA	CB100 sediment	47.1	79.3
CBsed26	ATTGTTGACGGGTAACTGGCTGATAACAGCACTGTTAGGTGGTGGATTCTGGGGACTATTTTTCTACCCA	CB100 sediment	45.7	78.7
CBsed61	ACTGTTGACGGGTAACTGGCTGGTTACAGCATTGTTAGGTGGTGGTTTCTGGGGATTGTTTTTTTACCCA	CB100 sediment	45.7	78.8
CT200s1	GTTGTTAACCCGTAACTGGATGATCACAGCACTGTTCGGTGGCGGAGCCTTTGGGTTACTGTTCTATCCG	CT200 surface	51.4	81.1
CT200d4	GTTGTTAACCCGTACCTGGATGATTACAGCACTGTTCGGTGGCGGAGCCTTTGGGTTGCTGTTCTATCCG	CT200 deep	52.9	81.6
CB300s1	GTTGTTGACCCGTAACTGGATGATTACAGCATTATTCGGTGGCGGAGCATTTGGACTATTGTTCTACCCG	CB300 surface	47.1	79.3
CB300s2	ATTGTTGACCCGTAACTGGATGATCACGGCACTGATCGGTGGCGGAGCCTTTGGGCTACTGTTCTATCCG	CB300 surface	54.3	82.2
Noceani	TGCGATTTCCAAGAGCTACGGCTTGACGGCGGTAGTGGGTGGATTAATGTACGGTTTGTTGATGTATCCC	AF509001	50.0	80.5
Nmestuari	GTTACTGACTGGTAACTGGTTGATTACAGCGTTACTGGGTGGTGGATTCTGGGGCTTGTTCTTTTATCCG	AF272400	47.1	79.3
Nseuro	GTATCTGACACGTAACTGGTTGGTGACTGCATTGGTTGGAGGTGGATTCTTTGGCCTGATGTTTTACCCG	NEZ97861	48.6	79.9
Nmureae	ATTATTGACGGGTAACTGGCTGGTAACCGCACTGTTAGGCGGTGGATTCTGGGGTTTATTTTTCTATCTG	AF272403	45.7	78.7
Nmcryo	GTTGCTGACACGCAGCTGGTTGGTTACTGCATTGGTTGGTGGTGGTTTCTTTGGGTTGTTTTTCTACCCA	AF272403	48.6	79.9
Nmmarina	GCTGTTGACGGGTAACTGGTTGATCACGGCCCTGCTGGGTGGTGGATTCTGGGGCTCGTTCTTTTACCCT	AF272405	57.1	83.4
Nsbriensis	GCTGCTCACCCGCAACTGGATGATCACAGCCCTGGTTGGCGGAGGCGCATTCGGACTCCTGTTCTACCCG	NEZ97859	62.9	85.7
Nstenuis	GCTGCTCACGCGCAACTGGATGATCACCGCCCTGGTAGGCGGGGGCGCCTTCGGGTTATTGTTCTACCCT	NBU76553	62.9	85.7
Nmmultiform	GCTTCTGACGCGCAACTGGATGATCACGGCACTGGTTGGCGGCGGCGCCTTTGGTTTGTTGTTCTATCCT	NMAMOABGN	57.1	83.4
Nmhalo	GTACCTGACCCGTAACTGGCTGATCACAGCACTGGTAGGCGGCGGCTTCTTTGGCTTGCTGTTCTACCCT	AF272398	57.1	83.4
Nmoligo	GTTGCTGACCGGTAACTGGCTGATCACAGCCCTGTTAGGCGGCGGATTCTTTGGCTTATTCTTCTACCCA	AF272408	52.9	81.6
		average	51.0	81.3
		st.dev.	6.0	2.7

Table S2. Percent identity and for oligonucleotide probes on the *amoA* array. Gray boxes indicate pairwise sequence identifies >85%, such that hybridization between these pairs is predicted under standard conditions. Outlined boxes indicate identities slightly less than 85%, for pairs in which hybridization was consistently observed

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16		Probe
	84.3	78.6	65.7	65.7	71.4	61.4	74.3	60.0	68.6	70.0	82.9	85.7	67.1	67.1	68.6	1	CB300s1
		65.7	58.6	60.0	75.7	60.0	72.9	61.4	58.6	62.9	87.1	92.9	61.4	60.0	71.4	2	CB300s2
			77.1	75.7	60.0	81.4	77.1	71.4	78.6	77.1	70.0	70.0	82.9	80.0	72.9	3	CBsed1
				74.3	61.4	92.9	74.3	55.7	78.6	94.3	57.1	52.9	71.4	82.9	65.7	4	CBsed8
					58.6	81.4	81.4	64.3	94.3	72.9	68.6	65.7	72.9	74.3	68.6	5	CBsed10
						62.9	58.6	54.3	55.7	62.9	71.4	70.0	74.3	60.0	71.4	6	CBsed12
							71.4	62.9	81.4	91.4	62.9	60.0	75.7	88.6	70.0	7	CBsed13
								62.9	84.3	77.1	68.6	72.9	65.7	70.0	68.6	8	CBsed26
									62.9	58.6	60.0	61.4	71.4	65.7	62.9	9	CBsed37
										77.1	67.1	64.3	74.3	74.3	67.1	10	CBsed61
											60.0	57.4	71.4	88.6	68.6	11	CBsed63
												95.7	67.1	64.3	74.3	12	CT200d4
													62.9	62.9	75.7	13	CT200s1
														68.6	70.0	14	Nmcryo
															65.7	15	Nmestuari

Table S2 cont.

17	18	19	20	21	22	23	24		Probe
60.0	65.7	75.7	60.0	34.3	71.4	68.6	65.7	1	CB300s1
65.7	74.3	72.9	65.7	32.9	72.9	65.7	71.4	2	CB300s2
67.1	65.7	80.0	72.9	32.9	58.6	74.3	64.3	3	CBsed1
75.7	58.6	67.1	71.4	32.9	50.0	72.9	54.3	4	CBsed8
74.3	64.3	75.7	81.4	31.4	50.0	67.1	61.4	5	CBsed10
55.7	72.9	64.3	51.4	41.4	80.0	68.6	74.3	6	CBsed12
78.6	62.9	72.9	77.1	34.3	45.7	74.3	58.8	7	CBsed13
77.1	60.0	82.9	85.7	30.0	58.6	62.9	65.7	8	CBsed26
51.4	57.1	67.1	62.9	41.4	48.6	67.1	52.9	9	CBsed37
75.7	61.4	74.3	77.1	31.4	51.4	67.1	61.4	10	CBsed61
78.6	61.4	71.4	68.6	35.7	50.0	72.9	57.1	11	CBsed63
60.0	70.0	72.9	64.3	35.7	67.1	64.3	67.1	12	CT200d4
62.9	72.9	77.1	67.1	34.3	71.4	62.9	72.9	13	CT200s1
61.4	70.0	68.6	62.9	37.1	64.1	78.6	68.6	14	Nmcryo
78.6	58.6	72.9	68.6	40.0	48.6	70.0	54.3	15	Nmestuari
65.7	80.0	84.3	64.3	35.7	72.9	77.1	75.7	16	Nmhalo
	67.1	75.7	65.7	32.9	57.1	62.9	65.7	17	Nmmarina
		68.6	62.9	40.0	78.6	68.6	85.7	18	Nmmultiform
			74.3	31.4	65.7	65.7	74.3	19	Nmoligo
				30.0	44.3	58.6	60.0	20	Nmurea
					34.3	34.3	35.7	21	Noceani
						64.3	84.3	22	Nsbriensis
							60.0	23	Nseuro
								24	Nstenu