Diversity of ammonia monooxygenase (amoA) genes across environmental gradients in Chesapeake Bay sediments

CHRIS A. FRANCIS,1 GREGORY D. O’MULLAN2 AND BESS B. WARD1,2
1Department of Geosciences, and 2Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

ABSTRACT

Chesapeake Bay, the largest estuary in North America, encompasses a wide range of nutrient loading and trophic levels from the rivers and upper Bay to the sea, providing an ideal natural environment in which to explore relationships between functional diversity, physical/chemical complexity and ecosystem function (e.g. nitrification). In this study, amoA gene fragments (encoding subunit A of the key nitrification enzyme, ammonia monooxygenase) were PCR-amplified from DNA extracted from sediment cores collected at five stations spanning gradients of salinity, ammonium, nitrate, oxygen and organic carbon along the Bay and Choptank River, a subestuary of the Bay. Phylogenetic analysis of ∼30 amoA clones from each station revealed extensive diversity within the β-Proteobacteria group of ammonia-oxidizing bacteria (AOB), with the vast majority of sequences falling into coherent phylogenetic clusters distinct from sequences of cultivated AOB. Over 70% of the clones fell into two major phylogenetic clusters that appear to represent novel groups of Nitrosomonas-like and Nitrosospira-like amoA sequences that may be specific to estuarine and marine environments. Rarefaction analysis, estimators of genetic variation and dissimilarity indices all revealed differences in the relative amoA-based diversity and/or richness among most of the stations, with the highest diversity at the North Bay station and the lowest at the mesohaline stations. Although salinity appears to play a role, no single physical or chemical parameter entirely explains the pattern of diversity along the estuary, suggesting that a complex combination of environmental factors may shape the overall level of AOB diversity in this dynamic environment.

INTRODUCTION

The oxidation of ammonium (NH₄⁺) to nitrite (NO₂⁻) and ultimately nitrate (NO₃⁻) by chemolithoautotrophic nitrifying bacteria is a critical branch of the nitrogen cycle in soil, sedimentary, marine, freshwater and estuarine environments, where this process provides a key link between the mineralization of organic nitrogen and the subsequent loss of fixed nitrogen via denitrification. In many aquatic systems, particularly shallow systems and sediment–water interfaces, denitrification (the dissimilatory reduction of N oxides to gaseous products such as N₂O and N₂) depends directly on NO₂⁻ supplied by nitrification across the oxic/anoxic interface (Jensen et al., 1993, 1994). These coupled processes are quantitatively important in the nitrogen budgets of continental shelf sediments and of estuaries (Christensen et al., 1987; Seitzinger, 1988).

Sediment nitrification rates are generally regulated by availabilities of O₂ or NH₄⁺ (Henriksen & Kemp, 1988), but the influence of these and other environmental gradients on the distributions and diversity of nitrifying bacteria is not well understood.

Nitrification is a two-step process carried out by distinct functional groups of chemolithotrophic bacteria, the ammonia-oxidizing bacteria (AOB) and the nitrite-oxidizing bacteria. The oxidation of ammonia to nitrite by AOB is also a two-step process, whereby the first and rate-limiting step, the oxidation of ammonia to hydroxylamine, is catalysed by the membrane-bound protein ammonia monooxygenase (AMO), and the second step is catalysed by hydroxylamine oxidoreductase (Hooper et al., 1997). Chemolithoautotrophic AOB are phylogenetically restricted to two different monophyletic lineages within the Proteobacteria (Head et al., 1993). Most known...
AOB belong to a closely related group of β-Proteobacteria (Purkhold et al., 2000), which includes the genera Nitrosomonas and Nitrospira (which now encompasses the former genera Nitrosovibrio and Nitrosolobus); the widely distributed marine ammonia-oxidizer, Nitrosovibrio oceanus (Ward & O’Mullan, 2002), and the lesser known Nitrospira halophila comprise the only known AOB species within the γ-Proteobacteria (Head et al., 1993). Taking advantage of the restricted phylogenetic distribution of AOB, most molecular studies of these organisms in natural environments have focused upon the retrieval and analysis of 16S ribosomal RNA (rRNA) sequences for characterizing the diversity and structure of nitrifying communities (recently reviewed by Kowalchuk & Stephen, 2001). A number of these studies have attempted to link the presence of different phylogenetic clusters with key environmental parameters (Bruns et al., 1999; McCaig et al., 1999; de Bie et al., 2001). Of particular interest to our study is the work of de Bie et al. (2001), who reported distinct geographical patterns in Nitrosomonas-like AOB sequence distributions that appeared to be linked to steep gradients of salinity, ammonium and oxygen concentrations in the Scheldt estuary. Despite these insights, 16S rRNA analyses are somewhat limited by the specificity of frequently used primers and the low resolving power of rRNA sequences for distinguishing closely related AOB species (Rothhauwe et al., 1997).

An alternative approach to 16S rRNA analysis for investigating AOB diversity in the environment is to target the functional gene, amoA, encoding subunit A of ammonia monooxygenase (McTavish et al., 1993; Rothhauwe et al., 1997). This gene is a particularly useful molecular marker for studying AOB because, unlike many other biogeochemically important functional genes (e.g. nirS of denitrification) that have clearly undergone significant horizontal gene transfer (Zumft, 1997), amoA phylogeny appears to be coherent with that based on 16S rRNA sequences (Purkhold et al., 2000). As with most protein-encoding genes, amoA provides greater phylogenetic resolution than the non-coding rRNA genes. In addition, sequence divergence in the amoA gene is more likely to be of functional importance to the process of nitrification than is variation in the 16S rRNA gene. For example, differences in amoA sequences correlated with differences in isotopic discrimination for ammonia oxidation (εAMO) in various β-Proteobacterial AOB (Casciotti et al., 2003). The combination of phylogenetic reliability, adequate levels of sequence divergence and functional importance to the process of interest makes the amoA gene a powerful tool for the study of environmental samples. In recent years, amoA diversity has been investigated in a wide variety of natural environments, including soils, sediments, plant roots, groundwater, marine and fresh waters, and estuaries (Ivanova et al., 2000; Nold et al., 2000; Kowalchuk & Stephen, 2001; Oved et al., 2001; Avrahami et al., 2002; Nicolaisen & Ramsing, 2002; Caffrey et al., 2003). In the present study, we explore the distribution and diversity of β-Proteobacterial amoA sequences in sediments of Chesapeake Bay, the largest estuary in North America, where nitrification is a pivotal component of the N cycle at certain times of the year (Kemp et al., 1990; Cowan & Boynton, 1996). The sequence diversity at five stations was analysed quantitatively in terms of correlations with the large environmental gradients in the system (e.g. salinity, inorganic nitrogen, etc.), in order to explore relationships between physical/chemical complexity and AOB diversity.

**MATERIALS AND METHODS**

**Site description**

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 roughly 1% of the total freshwater to the bay. The five sampling stations were chosen to span a range of environmental conditions, with each having the following general characteristics: upper Choptank (CT1), high nitrate, low productivity, tidal freshwater/oligohaline, mud; lower Choptank (CT2), moderate nitrate, high productivity, mesohaline, mud; north Bay (CB1), high nitrate, low productivity, oligohaline, mud; mid-Bay (CB2), seasonally anoxic, moderate nitrate, high productivity, mesohaline, mud; and south Bay (CB3), low nitrate, moderate productivity, polyhaline, sand/mud. Hydrographic and nutrient data were obtained from the Biocomplexity project website (http://snow.tamu.edu) and from J. Cornwell (pers. commun.).

**Sample collection and DNA extraction**

Sediments were collected from upper (CT1) and lower (CT2) Choptank River stations using a box core deployed from a small boat in July 2000 and April 2001, respectively. Sediments from the north, middle and south Chesapeake Bay stations (CB1, CB2, CB3, respectively; see Fig. 1), were collected using a box core deployed off the R. V. Cape Henlopen in April 2001. Sediment subcores from cores incubated for nitrogen flux measurements (J. Cornwell, unpubl. data) using cut-off 5-mL syringes, dropped immediately into liquid N$\_2$, and stored at −80 °C until DNA extraction. DNA was extracted from replicate portions (∼0.25 g) of sediments (0–0.5 cm depth interval) using the FastDNA SPIN kit for soil (BIO 101, Inc.).

**PCR amplification and cloning of amoA gene fragments**

amoA gene fragments (∼490 bp) were amplified from sediment DNA extracts using the PCR primers (AmoA-1F and AmoA-2R) and conditions described by Rothhauwe et al. (1997). Products were visualized by electrophoresis in 1.2% agarose gels followed by ethidium bromide staining. Triplicate PCR reactions were pooled, gel purified using the QIAquick
gel extraction kit (Qiagen) and cloned using the TOPO-TA cloning kit (Invitrogen). Insert-containing white colonies were transferred to 96-well plates containing LB broth (with 50 µg mL⁻¹ kanamycin) and grown overnight at 37 °C. Clones were PCR screened directly for the presence of inserts using T7 and M13R vector primers.

Sequencing and phylogenetic analysis of amoA sequences

Sequencing of both strands of T7/M13 PCR products was performed using BigDye™ V.3.0 terminator chemistry (PE Applied Biosystems) and ABI 3100 and 310 capillary sequencers (PE Applied Biosystems). Nucleotide sequences were assembled, edited, aligned and translated using Sequencher™ v.4.1 (GeneCodes Corp.). A 450-bp region (corresponding to 150 amino acids) of 225 amoA sequences was chosen for phylogenetic analysis. Neighbour-joining phylogenetic trees were constructed based on alignments of DNA sequences (using Kimura two-parameter corrected distances) and amino acid sequences within the PAUP software package (version 4.0b10). Bootstrap analysis was used to estimate the reliability of phylogenetic reconstructions (100 replicates). To compare the amoA-based ‘species’ richness within each clone library, rarefaction analysis (Heck et al., 1975) was performed using the ANALYTICAL RAREFACTION 1.3 program (www.uga.edu.strata/software/) (Holland, 2001). For this analysis, operational taxonomic units (OTUs) were defined as amoA sequence groups in which sequences differed by ≤5%.

Statistical analyses

A multivariate correlation analysis of environmental variables (NH₄⁺, NO₃⁻, salinity) was performed in JMP (SAS Institute, 2002). The Arlequin population genetics software package (Schneider et al., 2000) was used to compare the levels of genetic variation within and between sampled stations. The mean number of pairwise genetic differences (π) (Nei & Li, 1979) within (πw) and between (πB) sampled populations was calculated from sequence data to estimate the level of genetic
variation within and between stations. An Analysis of Molecular Variance (Excoffier et al., 1992) was used to examine the partitioning of variation within stations vs. among stations in the entire data set. Dissimilarity indexes (Fst) (Reynolds et al., 1983; Martin, 2002) were calculated for pairs of stations. \( Fst = \frac{\pi_b - \frac{1}{2}\pi_v}{(\pi_v - \pi_a) - \frac{1}{2}\pi_v} \) and \( \frac{\pi_b}{\bar{\pi}_b} \) and ranges from a value of 1 if all the variation is found between stations to a value of 0 if the variation within stations is equal to the variation between stations. One hundred permutations with random assemblages of these sequences were compared with observed distributions to test for the statistical significance of non-random distributions. A Mantel test (Smouse et al., 1986) was performed to examine the correlations between the pairwise Fst matrix and matrices of the pairwise differences in both salinity and ammonia concentration between stations. Significance of correlations was calculated by randomly switching values between cells of the matrices and comparing the generated matrices to the observed matrix (1000 permutations).

**Exact test with OTUs**

The frequency of OTUs, defined by the >95% identity threshold, from each sampling site was recorded and a deviation from a random distribution of OTUs among stations was investigated by an exact test (Raymond & Rousson, 1995) for population differentiation with 10 000 Markov steps.

**Nucleotide sequence accession numbers**

The GenBank accession numbers of the \( amoA \) sequences from cultivated ammonia-oxidizers used for comparison are as follows: Nitrosonomas aquaturnii (AF272400), Nitrosonomas communis (AF272399), Nitrosonomas cryotolerans (AF272402), Nitrosonomas europaee (AJ298710), Nitrosonomas eutropha (AJ298713), Nitrosonomas halophilae (AF272398), Nitrosonomas marina (AF272405), Nitrosonomas nitrosa (AF272404), Nitrosonomas oligotroph (AF272406), Nitrosonomas ureae (AF272403), Nitrosonomas sp. Nm143 (AY123816), Nitrospira briensis (AY123821), Nitrospira multiformis (AF042171), Nitrospira tenuis (U76552) and Nitrospira sp. NpAV (U92432). The \( amoA \) sequences from the following environmental clones were also used: Agb35c8 (AY098886), AZPa2 (AY186228), AZPa5 (AY186229), AZPa6 (AY186230), AZPa8 (AY186232), AZPa9 (AY186233), AZPa10 (AY186234), AZPa11 (AY186235), AZPa12 (AY186236), AZPa13 (AY186237), AZPa14 (AY186238), B3-6 (AF293072), E-207M (AJ317947), K1 to K13 (AF489630 to AF489642), pDtA.2 (AJ388573), pDpA.3 (AJ388574), pDcC.1 (AJ388575), pgA.3 (AJ388567), pgMpa.1 (AJ388576), pgMpa.2 (AJ388577), RI-27 (AF532311), RR90-4 (AJ28546), RR90-6 (Z97839), RR90-8 (Z97841), Pflücker (Z97850), Schönsee (Z97851), T1 to T15 (AF489644 to AF489658), TUK18 (AJ285448) and w1710 (AF353262). The \( amoA \) sequences reported in this study have been deposited in GenBank under accession numbers AY352899 to AY535054.

**RESULTS AND DISCUSSION**

**Environmental gradients along the estuary**

Along the longitudinal transect from the North Bay (CB1) to South Bay (CB3) in April 2001 (Fig. 1), the bottom water salinity increased from 4.4 to 23.6 PSU, and \( \text{NH}_4^+ \) concentration decreased gradually from 10 to 4 \( \mu \text{M} \) (Table 1). Concentrations of \( \text{NO}_3^- \), the ultimate product of nitrification, but more often an indicator of agricultural runoff in estuarine systems, exhibited a much steeper gradient along this same transect, decreasing from 83 to 3 \( \mu \text{M} \). Similar opposing gradients of salinity and inorganic nitrogen were also observed in April 2001, from the freshwater upper station of the Choptank River (CT1) to the mesohaline lower Choptank station (CT2) (Table 1). The overall physical/chemical conditions at the two mesohaline stations, CT2 and CB2, were quite similar, with identical levels of \( \text{NH}_4^+ \) and \( \text{NO}_3^- \) and respective salinities of 14.5 and 18.7 PSU. Because sediment samples for molecular analyses were only collected from the three main Bay stations and the lower Choptank River station, CT2, in April 2001, a sediment sample collected from CT1 in July 2000 was also analysed in this study to provide some basis for comparison with the other stations. The key differences between the environmental conditions at CT1 in July 2000 and April 2001 were temperature (27 °C and 7 °C, respectively), \( \text{NH}_4^+ \) concentration (∼two-fold higher in April 2001), and \( \text{NO}_3^- \) concentration (44 \( \mu \text{M} \) and

<table>
<thead>
<tr>
<th>Station</th>
<th>Sampling date</th>
<th>Water depth (m)</th>
<th>Temperature (°C)</th>
<th>Salinity (PSU)</th>
<th>( \text{NH}_4^+ ) (( \mu \text{M} ))</th>
<th>( \text{NO}_3^- ) (( \mu \text{M} ))</th>
<th>( \text{O}_2 ) (( \mu \text{M} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper Choptank (CT1)</td>
<td>July 2000</td>
<td>5.5</td>
<td>27</td>
<td>0.3</td>
<td>5</td>
<td>44</td>
<td>NA</td>
</tr>
<tr>
<td>Lower Choptank (CT2)</td>
<td>April 2001</td>
<td>7</td>
<td>8</td>
<td>14.5</td>
<td>7</td>
<td>22</td>
<td>NA</td>
</tr>
<tr>
<td>North Bay (CB1)</td>
<td>April 2001</td>
<td>10</td>
<td>6.7</td>
<td>4.4</td>
<td>10</td>
<td>83</td>
<td>278</td>
</tr>
<tr>
<td>Mid Bay (CB2)</td>
<td>April 2001</td>
<td>17.5</td>
<td>7.2</td>
<td>18.7</td>
<td>7</td>
<td>22</td>
<td>247</td>
</tr>
<tr>
<td>South Bay (CB3)</td>
<td>April 2001</td>
<td>11</td>
<td>8.7</td>
<td>23.6</td>
<td>4</td>
<td>306</td>
<td></td>
</tr>
</tbody>
</table>

NA = not available.
identity thresholds for defining OTUs for the more divergent rRNA genes (Hughes et al., 2001), appropriate sequence identity thresholds for defining OTUs for the more divergent

functional genes have yet to be carefully considered or implemented (Ward, 2002). These thresholds will most likely vary, depending on the evolutionary history (i.e. conservation) of the particular gene in question. For the highly conserved amoA gene, thresholds ranging from 2 to 5% appear to be adequate for grouping closely related amoA phylotypes, without losing potentially valuable diversity information via the inclusion of phylogenetically distinct but closely related sequences. Because greater divergence at the DNA level will more likely result in key structural differences at the amino acid level, the 5% threshold applied in this study should be of greater functional relevance than the 2% threshold. It is important to note that although rarefaction analysis is useful for comparing the relative richness among clone libraries, it cannot predict the actual community richness (i.e. total number of OTUs) in the original samples.

Rarefaction analysis indicated by far the greatest amoA diversity/richness in the North Bay (CB1), and the lack of curvature after analysis of over 30 clones suggests that the diversity of distinct amoA sequences has yet to be saturated in this clone library. Interestingly, the lowest and essentially identical levels of richness were observed at the two mesohaline stations, CT2 and CB2, whereas intermediate levels were observed at CT1 and CB3. When the analysis was performed with a 2% cutoff to define an OTU (data not shown), almost identical trends were observed but with the striking exception of CB3, which exhibited much higher richness – almost comparable with that of CB1. The primary reason for this disparity between the two different OTU cutoffs can be seen by close examination of the novel Nitrosomonas-like cluster in the phylogenetic tree (Fig. 3). This cluster contains seven CB3 AmoA sequences that fall into four different positions. Whereas these sequences constitute five OTUs at the 2% OTU threshold, they comprise only two OTUs at the 5% DNA level, which significantly impacts the rarefaction analysis based on 31 clones. Therefore, arbitrary definitions of OTUs must be carefully considered for functional genes and analysis based on several different thresholds may be warranted.

Estimators of genetic variation

Rarefaction analysis can be used to compare the richness/diversity between clone libraries but it does not compare compositional overlap. In order to examine the distribution of OTUs between sites, a frequency-based ‘exact test’ for population differentiation (Raymond & Rousset, 1995) was applied to OTUs defined by the 5% cutoff. This test showed the distribution of OTUs among libraries to be non-random ($P < 0.002$). Pairwise station comparisons also indicated non-random distributions of OTUs ($P < 0.002$) in all cases.

The information content of the sequences themselves offers greater resolution than simply defining OTUs, for examining patterns of genetic variation and composition within and between stations. An Analysis of Molecular Variance (Excoffier et al., 1992) indicated that 74.75% of the genetic variation

Quantitative analyses of amoA sequences from Chesapeake Bay sediments

PCR amplification of a 490-bp amoA fragment was consistently obtained from sediment DNA extracts from all five stations (data not shown). Clone libraries containing 95 positive clones were subsequently generated and archived for each station, of which 30–32 were sequenced for a total of 156 amoA sequences. Before discussing the diversity and phylogenetic relationships of these sequences and their relationships to sequences from cultivated strains and other environments, some quantitative analyses of similarity among nucleotide sequences and community composition among the stations will be presented.

Rarefaction analysis

To compare the relative amoA-based OTU richness between stations, rarefaction analysis was performed on the ~30 amoA sequences from each station using a 5% cutoff at the DNA level to define an OTU (Fig. 2). Although identity cutoffs ranging from 1 to 3% are often used to define OTUs for 16S rRNA genes (Hughes et al., 2001), appropriate sequence identity thresholds for defining OTUs for the more divergent

© 2003 Blackwell Publishing Ltd, Geobiology, 1, 129–140
Fig. 3 Neighbour-joining phylogenetic trees of amoA gene products and gene sequences from Chesapeake Bay sediments. The amino acid tree (a) contains 156 sequences from the five Bay stations (see colour-coded key), together with sequences from cultivated ammonia-oxidizers and closely related environmental clones (black). Brackets highlight the different phylogenetic clusters referred to in the text. The DNA-based tree (b), corresponding only to *Nitrosospira*-like Cluster B in (a), illustrates the relative differences in topology and clustering between the DNA and protein-based trees. The * highlights the subcluster of amoA sequences clearly distinct at the DNA level, but that are identical to most of the sequences in the other major subcluster (**) at the amino acid level. Bootstrap values (>50%) are indicated at the branch points of both trees.
represented in the 156 amoA sequences from this study was found within populations, implying that the genetic composition found within a station was widely distributed relative to the total variation observed among sequences (Table 2). The mean number of pairwise differences (π) within a sample of DNA sequences (πω) is an estimator of the amount of genetic variation within the sample. The sequences from CT1 contain the highest πω value (76.14), and CB3 yielded the lowest πω value (61.80) (Table 2). The π value between stations (πB) is a meaningful estimator of the genetic variation between two stations. The πB value for the CB2 and CB3 pair (66.03) indicates that the communities at these two stations are the most similar, i.e. exhibit the least genetic variation between any two stations.

All of the comparisons between mesohaline (CT2, CB2) or polyhaline (CB3) stations have similarly small πB values (<75.87). By contrast, comparisons of πB values between mesohaline or polyhaline stations and the oligohaline stations (CT1 and CB1) have larger values (>89.57). The greatest genetic variation was found between CT1 and CT2 (106.62) but, because these samples were not collected during the same season, temporal differences may account for some of this variation.

Dissimilarity (Fst) indices (Reynolds et al., 1983; Martin, 2002) can be used to compare the average genetic variation within a group with the genetic variation between groups. This analysis evaluates the degree to which randomly generated groups of the sequences partition variation compared with the observed sequence distribution. Fst values of all pairwise comparisons indicated that all populations were significantly different (P < 0.05), except for the CB2/CB3 pair (Fst = 0.04) (Table 2). Again, this suggests that the amoA populations at the mesohaline to polyhaline stations are more similar to each other than to those at the other stations. These measures are calculated based upon DNA differences, and the estimators would differ if based upon amino acid data. In particular, the mesohaline and polyhaline stations (CT2, CB2, and CB3) would be even more similar to one another if amino acid data were analysed (see discussion of Fig. 3b below).

A Mantel test (Smouse et al., 1986) was used to examine the correlations between the pairwise Fst matrix (Table 2) and matrices of the pairwise differences in both salinity and ammonia concentration between stations (not shown). The Mantel test evaluates the variation in the Fst matrix due to differences in both environmental variables, providing a better fit than either variable alone. Salinity alone accounted for 62.0% of the variation in pairwise Fst, whereas ammonia accounted for only 2.7% of the variation. Together, the differences in these variables accounted for 64.7% of the variation in pairwise Fst (P = 0.013). However, caution must be used in the interpretation of correlation analyses, because the variables examined may significantly co-correlate with other variables. Although salinity may play an important role in the transition of amoA diversity and population structuring, it does not appear to explain the pattern of genetic variation completely. For example, CT1 and CB1 are both oligohaline (i.e. salinity <5 PSU), and are the two stations with the highest levels of within-sample genetic variation, but they are significantly different from each other according to Fst estimators (Table 2). This might be due to the salinity difference of ~4 PSU between these stations (Table 1) but is more likely due to other factors such as temporal variation and/or allochthonous inputs.

**Phylogenetic analysis of Chesapeake Bay amoA sequences**

Comparison of the 156 amoA sequences from Chesapeake Bay sediments to the most closely related sequences in the databases revealed that only a small fraction were phylogenetically closely related to known Nitrosomonas and Nitrosospira sequences. Instead, the majority of the sequences fell into distinct phylogenetic clusters comprised primarily of Chesapeake Bay sequences and, in some cases, closely related environmental clones (Fig. 3). In fact, over 70% of the Chesapeake Bay amoA sequences (112/156) fell into two major clusters (Nitrosospira-like Cluster B and the Nitrosomonas-like cluster; Fig. 3), with the remaining ~30% comprised of all 30 CT1 sequences, 13 CB1 sequences and a single CB3 sequence. Interestingly, the only environmental clones from the database that fell into the two major clusters were from either estuarine or marine sediments (see below for details). All of the clones within Nitrosospira-like Cluster B were derived from mesohaline (CT2, CB2) or polyhaline (CB3) environments, and this overall clade is clearly distinct (>22% divergence at the DNA level) from all other known amoA sequences. Therefore, it appears that this may represent a novel group of AOB found

---

**Table 2** Genetic variation within and between sites

<table>
<thead>
<tr>
<th>Station</th>
<th>CT1</th>
<th>CT2</th>
<th>CB1</th>
<th>CB2</th>
<th>CB3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT1</td>
<td>76.14</td>
<td>0.33</td>
<td>0.25</td>
<td>0.32</td>
<td>0.35</td>
</tr>
<tr>
<td>CT2</td>
<td>106.62</td>
<td>67.28</td>
<td>0.25</td>
<td>0.10</td>
<td>0.15</td>
</tr>
<tr>
<td>CB1</td>
<td>96.95</td>
<td>91.33</td>
<td>68.85</td>
<td>0.26</td>
<td>0.33</td>
</tr>
<tr>
<td>CB2</td>
<td>103.34</td>
<td>73.46</td>
<td>89.58</td>
<td>64.54</td>
<td>0.04*</td>
</tr>
<tr>
<td>CB3</td>
<td>106.13</td>
<td>75.86</td>
<td>97.86</td>
<td>66.03</td>
<td>61.80</td>
</tr>
</tbody>
</table>

Mean pairwise sequence differences within sites (πω) are on the diagonal in bold type. Mean pairwise sequence differences between sites (πB) are below the diagonal in italics. Pairwise population Fst values are above the diagonal. The asterisk (*) indicates an Fst value that is not significantly different at the 0.05 level from Fst values generated from 100 random distributions created by shuffling of the sequences between sites.

predominantly in brackish to marine environments. This is further supported by the fact that the majority of \textit{amoA} clones recently obtained from the Monterey Bay water column (O’Mullan & Ward, 2002) also fall into this phylogenetic cluster (data not shown). By contrast, the major \textit{Nitrosomonas}-like cluster (which is considerably more distinct at the DNA level) contains not only sequences from the mesohaline and polyhaline stations but also a large number of sequences from the oligohaline station, CB1. This may suggest a wider range of physiological types of AOB in this cluster or, alternatively, organisms capable of withstanding a wide range of environmental conditions.

In addition to comparing the relative diversity/richness of \textit{amoA} sequences between the different Chesapeake Bay stations, it is also useful to analyse the distribution of sequences throughout the phylogenetic tree and to compare the types of sequences found at these five stations with sequences from other environments. As suggested by the rarefaction analysis, the CB1 sequences were widely dispersed throughout the \textit{amoA} phylogenetic tree, but the majority fell into the lower half of the major \textit{Nitrosomonas}-like cluster, which also included a number of closely related and identical clones from CT2, CB2 and CB3. Considering the relative salinities of these different stations, it is possible that these CB1 clones correspond to moderately halotolerant AOB that can survive under mesohaline conditions, but prefer and hence show greater dominance under oligohaline conditions. Conversely, the fact that a single North Bay clone, CB1-13, is essentially identical to the 11 dominant CT2 (and six CB2) sequences in the upper half of this cluster might suggest a preference for mesohaline conditions, which rarely occur at CB1 but are characteristic of stations CT2 and CB2. In contrast to the other four stations, five CB1 sequences grouped within the \textit{Nitrosospira} cluster, comprised of a number of well-known \textit{Nitrosospira} species and various environmental clones, including three \textit{amoA} sequences recently reported from Azevedo Pond, located in Elkhorn Slough, a small California estuary (Caffrey et al., 2003). Azevedo Pond is characterized by large salinity, oxygen and temperature variations on both diel and seasonal scales. Clones CB1-6 and CB-19 fell just outside of the major \textit{Nitrosomonas}-like cluster and the \textit{Nitrosomonas marina/aestuarii} cluster, which also included seven additional Azevedo Pond sequences most closely affiliated with \textit{N. aestuarii}. Several of the CB1 \textit{amoA} clones were also strikingly similar to sequences derived from terrestrial environments, including soils and rice roots, possibly indicative of the impact of agricultural runoff at this station. CB1 was the only station which contained \textit{amoA} sequences that were identical to those of cultivated strains, namely \textit{Nitrospira briensis} and \textit{Nitrosomonas nitrosa}. Owing to the oligohaline character of both CB1 and CT1, overlap might be expected between sequences from these two stations. However, the only striking overlap occurred within a subcluster of the \textit{Nitrosomonas oligotropha}-like cluster, which contains two CB1 and five CT1 sequences that are not only essentially identical to one another, but also to \textit{amoA} clones (pDtA.2, pGtA.3 and pMPA.1) from three different lakes in The Netherlands. The only other overlap between CB1 and CT1 sequences occurred within \textit{Nitrospira-like} Cluster A, which contains a bootstrap-supported group of more divergent CB1 and CT1 sequences as well as lake and soil clones. Owing to the high \textit{amoA} sequence diversity/richness within the CB1 clone library, it is quite possible that sequencing of additional clones would reveal more overlap with CT1 sequences, which fall into only a few clusters.

Other than the minor overlap with the CB1 \textit{amoA} sequences, the CT1 sequences were totally distinct from those of other stations, including the lower Choptank River station, CT2. Using 16S rRNA probes to enumerate the main groups of \textit{Proteobacteria} in bacterioplankton of the Choptank River by fluorescence \textit{in situ} hybridization, Bovier & del Giorgio (2002) found consistent community shifts between the upper and lower CT regions. \textit{β-Proteobacteria} were abundant in the freshwater stations but were rare in the lower river, and the opposite pattern was observed for \textit{α-Proteobacteria}. The switch occurred at approximately the location of our station CT2, suggesting that a shift in the community structure of AOB might also be expected between the two river stations. Despite the fact that our study focused on sediment (rather than planktonic) communities and our primers were specific for AOB within the \textit{β-Proteobacteria}, a clear shift was also observed at the level of AOB molecular diversity.

Several CT1 sequences were closely related to \textit{amoA} clones obtained from other estuarine and freshwater environments. For example, within the \textit{Nitrosomonas oligotropha}-like cluster, clone CT-29 clustered tightly with a collection of \textit{amoA} sequences (designated T1 to T4, and T7) recently reported from surface sediments of an intertidal mudflat in the Tagus Estuary, Portugal (Nicolaisen & Ramsing, 2002), as well as clones obtained from various freshwater lakes. CT-1 and CT-22 were also 98–100% identical to two additional Tagus Estuary clones, a Lake Drontenmeer clone (pDTA3) and a soil clone (E207M). The remaining 19 CT1 clones fell into two distinct clusters, including one group of highly redundant sequences located between the \textit{Nitrosomonas ureae} and \textit{Nitrosomonas marina/aestuarii} clusters, and the other within the \textit{Nitrosomonas europaica}-like cluster in the phylogenetically distinct uppermost region of the tree. The latter CT1 cluster also comprises somewhat divergent sequences from Lake Drontenmeer (pDTC.1) and nitrifying bioreactor-associated biofilms. At present, it cannot be determined whether the major differences between the distribution and diversity of CB1 and CT1 sequences are due to temporal differences or actual physical/chemical differences between the two habitats. Interestingly, an exhaustive comparison of >100 \textit{nirS} nitrite reductase sequences obtained from these same sediment DNA extracts showed comparably high levels of diversity/richness between the two oligohaline sites, as well as greater overlap in
sequence types than that observed for the amoA sequences in this study (Francis & Ward, 2002). However, it is likely that denitrifying and ammonia-oxidizing bacteria respond differently to environmental fluctuations. To address these issues, future studies will address temporal/seasonal variability in both the ammonia-oxidizing and denitrifying communities at all of the stations, using amoA and nirB-based clone library and oligonucleotide microarray approaches (Taroncher-Oldenburg et al., 2003).

The low levels of amoA diversity/richness indicated by the rarefaction analysis of clone libraries from the two mesohaline stations, CT2 and CB2, are also clearly reflected in the distribution of sequences in the phylogenetic tree. The CT2 sequences were evenly divided between the two major Nitrospira-like and Nitrosomonas-like phylogenetic clusters. Within Nitrospira-like Cluster B, there is a striking lower subcluster containing 17 primarily redundant CT2 sequences, 17 CB2 clones, five CB3 clones and two environmental clones (K1 and K2) from sediments of Kysing Fjord, a shallow Danish estuary with virtually no tidal flow, a stable salinity of 17 PSU and a constantly high N content (Nielsen et al., 2003). A phylogenetic tree based on amoA DNA sequences (data not shown) revealed very similar overall topology and clustering as well as greater bootstrap support, with the exception of this novel subcluster containing 17 primarily redundant CT2 sequences, as well as Kysing Fjord clone K1) were absolutely identical (K1 and K2) from sediments of Kysing Fjord, a shallow Danish estuary with virtually no tidal flow, a stable salinity of 17 PSU and a constantly high N content (Nielsen et al., 1995; Nicolaisen & Ramsing, 2002). A phylogenetic tree based on amoA DNA sequences (data not shown) revealed very similar overall topology and clustering as well as greater bootstrap support, with the exception of this novel Nitrospira-like cluster (see Fig. 3b). At the DNA level, 16 CT2 clones (and CB3-27) formed a coherent cluster of nearly identical sequences clearly distinct (~8% DNA divergence) from the large subcluster of 20 mostly redundant CB2 and CB3 sequences. Surprisingly, however, the translated CT2 sequences (as well as Kysing Fjord clone K1) were absolutely identical to these same sequences. Detailed analysis of an alignment of these CT2 and CB2 amoA gene sequences revealed that the ~8% divergence occurred not only in the third ‘wobble’ position but also in the first position in a number of cases. This degree of variation in DNA sequence without corresponding variation in the amino acid sequence is strongly indicative of selection acting on AmoA. This is even more impressive because the DNA sequences largely cluster by station. This evolutionary signal of 3rd position (and in some cases 1st position) divergence may indicate that the AOB at the two stations differ substantially at non-amoA loci, with possible differences in physiology, while selection is holding the amino acid sequence of AmoA constant. It should be noted that strong selection would complicate the functional inference of DNA-based analyses (such as those presented above) by causing non-linear/inconsistent relationships between DNA sequence divergence and amino acid changes that are likely to be of functional importance. This illustrates the usefulness of combined DNA and amino acid analyses. Within the major Nitrospira-like cluster, 11 essentially identical CT2 sequences fell into the distinct upper subcluster, and the remaining four CT2 sequences grouped with closely related sequences in the lower subcluster.

As described above, all but one of the 18 CB2 sequences within Nitrospira-like Cluster B fell into a large subcluster, along with virtually identical amino acid sequences from CT2, CB3, as well as Kysing Fjord clones K1 and K2. The remaining 13 CB2 sequences fell into several subclusters of closely related sequences within the major Nitrospira-like cluster. Despite the similar environmental conditions at the two mesohaline stations, CT2 and CB2, in April 2001, nitrification is probably more important in CT2 sediments, where it is thought to be closely coupled to denitrification. By contrast, the sediments at the deeper CB2 station (~18 m water column) are exposed to seasonally anoxic conditions. When conditions become sulphidic, both nitrification and denitrification are generally inhibited. This inhibition of nitrification can lead to enhanced NH4+ fluxes out of these sediments (Kemp et al., 1990), and CB2 consistently exhibits the highest sediment NH4+ fluxes of the Bay stations (Cowan & Boynton, 1996; J. Cornwell, unpubl. observ.). Nevertheless, during April of 2001 when the CB2 station was still oxic, there was clearly striking overlap between the community compositions of CT2 and CB2. One rather noteworthy CB2 clone is the singleton, CB2-12, which is somewhat distinct (>5% DNA divergence) from the other Chesapeake Bay sequences in the major Nitrospira-like cluster, and appears to group with (and shares ~96–98% DNA identity to) seven clones (K7 to K13) from Kysing Fjord. In addition, CB2-12 also shows 97–99% DNA identity (over a shorter region of 332 bp obtained with different primers) to amoA sequences found to be dominant in sediments of Puget Sound (clones PS-5, -8 and -58) and the Washington continental margin (clone WC306) (Nold et al., 2000). Owing to the significantly shorter region of overlap with our Bay sequences, these Pacific Northwest (PNW) sequences were not included in the phylogenetic tree in Fig. 3. However, a DNA phylogenetic tree generated based on this smaller region confirmed that these sequences do in fact group extremely closely with the CB2-12 sequence (data not shown), whereas the corresponding AmoA amino acid tree revealed that many of the sequences within this cluster were closely related, or identical, to these PNW sequences. In light of our sequence data, it is somewhat surprising that all of the PNW amoA sequences reported by Nold et al. (2000) fell into the major Nitrospira-like cluster, and not the major Nitrospira-like cluster. The complete lack of Nitrospira-like sequences in that study could reflect a major difference in the overall AOB community composition between the respective sedimentary environments or, alternatively, could be due to differences in the relative specificities of the different primer sets used in the two studies. Incidentally, the primers used in our study designed by Rothhauwe et al. (1997) were also used in the Kysing Fjord study (Nicolaisen & Ramsing, 2002), which yielded sequences in both major clusters, with the majority falling into the Nitrospira-like cluster. Overall, the similarity between the mid-Chesapeake Bay sequence (CB2-12), several Kysing Fjord sediment sequences and
dominant PNW marine sediment sequences implies that this amoA phylotype (and other closely related representatives of this cluster) may be ubiquitous in brackish and marine sediments.

The majority (23 of 31) of the amoA sequences obtained from CB3, the most ‘marine’ or ‘polyhaline’ station along the main Bay transect, were located within *Nitrosospira*-like Cluster B. Although a number of CB3 amoA sequences were extremely closely related, if not identical, to the highly redundant CT2 and CB2 amoA sequences in this cluster, most formed distinct subclusters, suggesting greater ‘microdiversity’ at this station relative to the mesohaline stations. A distinctive feature of CB3 sediments that might potentially favour such AOB microdiversity is the extensive presence of large worms and their oxygenated burrows, which can serve to increase both small-scale physical variability and the potential for nitrification (Henriksen et al., 1983; Aller, 1988; Cowan & Boynton, 1996). In contrast to the CT2 sequences within this same *Nitrosospira*-like cluster, many of the CB3 nucleic acid differences apparently correspond to actual amino acid changes (see Fig. 3a,b). As discussed earlier with regard to the rarefaction analysis, there were seven CB3 sequences dispersed throughout the major *Nitrosomonas*-like cluster, with the greatest overlap occurring with closely related CB1 and CB2 sequences. The only CB3 sequence located outside the two major phylogenetic clusters was CB3-6, the only Bay sequence in the entire tree that grouped near the *Nitrosomonas marina/aestuarii* cluster. This is in contrast to the distribution of amoA sequences reported from Azvedo Pond sediments (Caffrey et al., 2003), in which the majority of cloned sequences were closest to this group. This may reflect the fact that the Azvedo Pond sequences were derived from sediments collected in August 1998, when the salinity and temperature were much higher than that of CB3 in April 2001. To determine the extent to which CB3 amoA sequences represent true ‘marine’ AOB, it will be useful to compare CB3 sequences with sequences obtained from the Chesapeake Bay plume as well as offshore stations.

**CONCLUSIONS**

Overall, this study has revealed extensive and unprecedented amoA diversity within these estuarine sediments, with over 70% of the sequences falling into two novel phylogenetic clusters that likely represent estuarine- and marine-specific sequences. Based on the distribution and diversity of sequences across the estuarine salinity gradient, it is tempting to speculate that many of the amoA sequences within these two major clusters correspond to the novel *Nitrosomonas*-like (e.g. group 5) or *Nitrosospira*-like (group 1) 16S rRNA environmental clones that have recently been reported in marine, coastal and estuarine environments (Stephen et al., 1996; McCaig et al., 1999; Phillips et al., 1999; de Bie et al., 2001; Hollibaugh et al., 2002). Although cultivation of environmentally representative AOB or an environmental genomics approach from such environments is ultimately necessary to link 16S rRNA and amoA phylogenies definitively, exhaustive clone library studies (such as this one) of both molecular markers across key environmental gradients (e.g. salinity) should also help resolve this issue. In terms of the community structure (inferred from clone libraries) of AOB along the Bay, a clear shift (i.e. decrease) in amoA diversity/richness occurred between the oligohaline and mesohaline stations, where the steepest environmental gradients were also observed. By contrast, the transition from the mesohaline stations to the poyhaline/marine station was much less obvious, with significant overlap observed in amoA sequence types, richness and genetic variation. In summary, this study represents the first characterization of amoA gene diversity across broad environmental gradients in sediments of a single estuary. Although salinity was the main environmental variable that was quantifiably related to the observed diversity patterns, covariation of other key parameters makes it difficult to discern a single driving factor. Future studies will take advantage of this new sequence database by combining innovative molecular approaches (e.g. oligonucleotide microarrays) with analyses of physical/chemical data and in situ nitrification rate measurements to address which environmental parameters are most important for determining the distribution, diversity and expression of amoA genes in the Chesapeake Bay.

**ACKNOWLEDGEMENTS**

We thank Jeffrey Cornwell for collecting the box core samples and providing ancillary chemical data for the sediment samples, and Jeff Alexander for assistance with sampling. This research was supported by a Biocomplexity research grant to B.B.W. (NSF grant OCE-9981482) and an NSF Postdoctoral Research Fellowship in Microbial Biology to C.A.F.

**REFERENCES**


© 2003 Blackwell Publishing Ltd, Geobiology, 1, 129–140


