Abstract Denitrification is a critically important biogeochemical pathway that removes fixed nitrogen from ecosystems and thus ultimately controls the rate of primary production in nitrogen-limited systems. We examined the community structure of bacteria containing the \textit{nirS} gene, a signature gene in the denitrification pathway, from estuarine and salt marsh sediments and from the water column of two of the world’s largest marine oxygen-deficient zones (ODZs). We generated over 125,000 \textit{nirS} gene sequences, revealing a large degree of genetic diversity including 1,815 unique taxa, the vast majority of which formed clades that contain no cultured representatives. These results underscore how little we know about the genetic diversity of metabolisms underlying this critical biogeochemical pathway. Marine sediments yielded 1,776 unique taxa when clustered at 95 % sequence identity, and there was no single \textit{nirS} denitrifier that was a competitive dominant; different samples had different highly abundant taxa. By contrast, there were only 39 unique taxa identified in samples from the two ODZs, and 99 % of the sequences belonged to 5 or fewer taxa. The ODZ samples were often dominated by \textit{nirS} sequences that shared a 92 % sequence identity to a \textit{nirS} found in the anaerobic ammonium-oxidizing (anammox) genus \textit{Scalindua}. This sequence was abundant in both ODZs, accounting for 38 and 59 % of all sequences, but it was virtually absent in marine sediments. Our data indicate that ODZs are remarkably depauperate in \textit{nirS} genes compared to the remarkable genetic richness found in coastal sediments.

Keywords Anammox · Denitrification · Estuarine sediments · Nitrite reductase (\textit{nirS}) · Oxygen-deficient zones · Salt marshes

Introduction

Nitrogen, one of the essential macronutrients, has a complicated biogeochemical cycle because it connects multiple reservoirs (atmosphere, oceans, and soils) and multiple redox states [1, 2]. Thus the cycling of nitrogen is critically important to maintaining ecosystem function in both terrestrial and aquatic systems [3]. Nitrogen limits the rates of primary production in vast areas of the biosphere, such that perturbations to the supply of nitrogen can have far ranging repercussions [4]. Application of synthetic nitrogenous fertilizers to agricultural crops has increased food yield; today ~40 % of the world’s population depends on food made possible by the industrial fixation of nitrogen through the Haber Bosch process [5]. Excess fixed nitrogen, however, cascades through ecosystems [6], leading to a host of negative side effects including increased incidences of estuarine eutrophication [7] and expanding areas of coastal dead zones [8].
The industrialization of nitrogen fixation via the Haber Bosch process has increased the amount of fixed nitrogen in the biosphere, but there is no widely used industrial analog for the removal of this fixed nitrogen [2]. Instead, most fixed nitrogen is returned to the atmosphere via two microbially mediated processes in wastewater treatment and natural systems: denitrification and the anaerobic oxidation of ammonia (anammox).Canonical denitrification is a dissimilatory process in which nitrate (NO$_3^-$) is used as a respiratory substrate for the oxidation of organic matter [9]. Long thought to occur only among bacteria, there is considerable evidence that this process can also occur among archaea, protists, and fungi [2]. More recently, anammox, the anaerobic autotrophic process whereby nitrite (NO$_2^-$) is used to oxidize ammonium (NH$_4^+$), has been shown to be another quantitatively important mechanism for the loss of fixed nitrogen in some systems [10–13].

The removal of fixed nitrogen by denitrification and anammox is critical to the global cycling of nitrogen. Estimates of the loss of fixed nitrogen suggest that 65–75 % of oceanic nitrogen loss occurs in estuarine and continental shelf sediments, with the remainder occurring in the three major oxygen-deficient zones (ODZs), the Eastern Tropical North and South Pacific and the Arabian Sea [14, 15]. The extent to which anammox or denitrification dominates the loss of fixed nitrogen varies widely depending on location. In shallow coastal sediments, especially salt marsh sediments, where areal rates are among the highest reported [16], denitrification appears to be the dominant loss process [2, 17, 18], and rates of denitrification appear to be enhanced by increased nitrogen supply [18, 19]. Anammox, however, appears to increase in importance in offshore sediments, where overall losses of fixed nitrogen are considerably lower [2]. In ODZs, reports indicate that either canonical denitrification [20] or anammox [11–13] can be the dominant loss process, but the relative contribution of each overall is about 71:29 (denitrification:anammox), the ratio dictated by the C:N ratio of the organic matter being degraded under anoxic conditions [21, 22].

The enzymes that catalyze the key reactions in the nitrogen cycle, including denitrification and anammox, are fairly well understood [23]. The defining step in canonical denitrification, the conversion of dissolved NO$_3^-$ to gaseous nitric oxide (NO), is catalyzed by one of two different, though functionally redundant, nitrite reductases encoded by either the nirS or nirK gene [9]. The nirS-encoded cytochrome cd$_1$ nitrite reductase has received more attention [24], and in marine systems where comparisons have been done, the nirS gene tends to be more prevalent than nirK in both DNA [25] and cDNA [26]. The process of anammox, thus far, has been found only in a subset of bacteria within the Planctomycetes phylum and much of the work to quantify these organisms relied on the amplification of 16S rRNA genes specific to that cluster. There is a multitude of protein-coding genes involved in the anammox process [27] that also provide functional biomarkers for anammox, the most commonly used of which is the hzo gene encoding the hydrazine oxidoreductase enzyme [28, 29]. Recently, however, it has been proposed that the nirS gene from anammox bacteria may be a more useful biomarker because it is present in one copy per genome in both the prevalent anammox clusters, those related to the Scalindua and the Kuenenia clades [12, 30]. In anammox bacteria, the nitrite reductase produces nitric oxide, which then combines with NH$_4^+$ to produce N$_2$.

Recent advances in next-generation sequencing have led to an explosion of new information regarding the diversity and function of microbes in the environment. Much of this work has focused on deep sequencing of 16S rRNA gene amplicons (e.g., Refs. [31, 32]. High-throughput whole community metagenomic and metatranscriptomic sequencing is also revealing new insights into the distribution of protein-coding genes in the environment [33–35], although the sequencing depth needed to decipher statistically relevant results is considerable [36]. Instead, we use a targeted metagenomics approach, in which next-generation sequencing technologies are used to sequence amplicons of specific protein-coding genes to examine the distribution and diversity of microbes in the environment [37–39]. As with all amplicon-based approaches, there are biases inherent in both the choice of primers and in the amplification reaction itself. Furthermore, any unknown sequences that are highly divergent from those in the databases will be missed by this analysis. However, the primers used here are the most robust for marine systems and we are able to achieve statistically relevant sequencing depths. Thus, the relative abundance of the amplifiable sequences can be compared and patterns that underlie specific microbial functions that are geochemically relevant can be elucidated. Here we report patterns in the distribution and diversity of the nirS gene from two of the world’s ODZs and from coastal marine sediments via pyrosequencing. We show that coastal sediments harbor extensive novel diversity of the nirS gene. Further, we show that the most abundant nirS sequence in the ODZs is most similar to the nirS sequence found in the anammox clade Scalindua and that the nirS gene in the ODZs is remarkably depauperate in comparison to the richness uncovered in coastal sediments.

**Materials and Methods**

Sample Description

We sequenced nirS amplicons from DNA extracted from 21 samples that have all been previously described (Table 1;
ESM 1 and 2). Four samples were collected across a range of salinities from sediments in the Choptank River [40], a tributary of the Chesapeake Bay, and from along the length of the main stem of Chesapeake Bay [41]. Since NO$_3^-$ is a primary substrate for denitrification, we also collected eight samples from sediments underlying the tall ecotype of *Spartina alterniflora* at different levels of nutrient enrichment from the fertilized plots at Great Sippewissett Salt Marsh in Falmouth, MA [42–44]. Four samples were collected from locations within the Arabian Sea ODZ in 2004 [43] and in 2007 [20], and another five samples were collected from stations in the Eastern Tropical South Pacific (ETSP) ODZ [20]. Sediment samples were collected with syringe corers at a depth less than 1 cm. Seawater samples were directly filtered onto 0.2 μm pore size Sterivex filters (Millipore, Billerica, MA). Detailed descriptions of sites, sample collection and storage, and nucleic acid processing can be found in the associated references [20, 40–46]. Metadata for each sampling site can be found in ESM 1 and 2.

### PCR Amplification and Sequencing

DNA fragments for pyrosequencing were prepared using nested polymerase chain reaction (PCR). First, DNA from the samples were amplified using the *nir*S primer set *nir*S1F and *nir*S6R [47] in three independent PCR runs to obtain approximately 890 bp *nir*S gene amplicons. Although these primers have been shown to be suboptimal in amplifying *nir*S from some phyla (Chloroflexi, Bacteroidetes), they do perform well against α-, β-, and γ-Proteobacteria, which account for a large portion of known denitrifiers [37]. The PCR cocktail consisted of 0.4 μM of each primer, 0.02 Units μl$^{-1}$ Phusion® High-Fidelity DNA Polymerase (New England BioLabs, Ipswich, MA), 1× Phusion HF buffer, 1.5 mM MgCl$_2$, 3 % DMSO, 1.6 mM total dNTP (Roche Applied Science, Indianapolis, IN), and 400 μg ml$^{-1}$ nonacetylated BSA (Sigma-Aldrich, St. Louis, MO). The PCR reaction was conducted on a S1000™ Thermal Cycler (Bio-Rad, Hercules, CA) using an initial denaturing step at 98 °C for

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Station</th>
<th>Depth</th>
<th>Sequencing depth</th>
<th>$S_{obs}$</th>
<th>Chao1 (SE)</th>
<th>Ace (SE)</th>
<th>Shannon</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP1</td>
<td>Plot 1</td>
<td>Surface</td>
<td>7,988</td>
<td>153</td>
<td>337 (66)</td>
<td>294 (9)</td>
<td>2.12</td>
</tr>
<tr>
<td>MP2</td>
<td>Plot 2</td>
<td>Surface</td>
<td>4,418</td>
<td>354</td>
<td>865 (109)</td>
<td>945 (20)</td>
<td>4.16</td>
</tr>
<tr>
<td>MP3</td>
<td>Plot 3</td>
<td>Surface</td>
<td>8,435</td>
<td>210</td>
<td>485 (80)</td>
<td>448 (14)</td>
<td>2.15</td>
</tr>
<tr>
<td>MP5</td>
<td>Plot 5</td>
<td>Surface</td>
<td>9,880</td>
<td>282</td>
<td>552 (65)</td>
<td>588 (16)</td>
<td>3.28</td>
</tr>
<tr>
<td>MP6</td>
<td>Plot 6</td>
<td>Surface</td>
<td>10,396</td>
<td>529</td>
<td>1,239 (124)</td>
<td>1,209 (22)</td>
<td>4.60</td>
</tr>
<tr>
<td>MP7</td>
<td>Plot 7</td>
<td>Surface</td>
<td>8,650</td>
<td>280</td>
<td>640 (85)</td>
<td>711 (18)</td>
<td>2.72</td>
</tr>
<tr>
<td>MP8</td>
<td>Plot 8</td>
<td>Surface</td>
<td>5,075</td>
<td>172</td>
<td>294 (39)</td>
<td>312 (11)</td>
<td>3.97</td>
</tr>
<tr>
<td>MP9</td>
<td>Plot 9</td>
<td>Surface</td>
<td>11,870</td>
<td>330</td>
<td>665 (75)</td>
<td>677 (16)</td>
<td>3.16</td>
</tr>
</tbody>
</table>

Chesapeake Bay sediments

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Station</th>
<th>Depth</th>
<th>Sequencing depth</th>
<th>$S_{obs}$</th>
<th>Chao1 (SE)</th>
<th>Ace (SE)</th>
<th>Shannon</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB1</td>
<td>Oligohaline</td>
<td>Surface</td>
<td>2,050</td>
<td>92</td>
<td>130 (20)</td>
<td>135 (6)</td>
<td>3.86</td>
</tr>
<tr>
<td>CB2</td>
<td>Mesohaline</td>
<td>Surface</td>
<td>1,503</td>
<td>38</td>
<td>55 (16)</td>
<td>59 (4)</td>
<td>2.01</td>
</tr>
<tr>
<td>CB3</td>
<td>Polyhaline</td>
<td>Surface</td>
<td>4,061</td>
<td>45</td>
<td>97 (54)</td>
<td>82 (5)</td>
<td>2.21</td>
</tr>
<tr>
<td>CT2</td>
<td>Mesohaline</td>
<td>Surface</td>
<td>1,503</td>
<td>44</td>
<td>57 (12)</td>
<td>64 (4)</td>
<td>3.12</td>
</tr>
</tbody>
</table>

Arabian Sea (AS) water column

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Station</th>
<th>Depth</th>
<th>Sequencing depth</th>
<th>$S_{obs}$</th>
<th>Chao1 (SE)</th>
<th>Ace (SE)</th>
<th>Shannon</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS1*</td>
<td>1</td>
<td>230</td>
<td>8,650</td>
<td>11</td>
<td>21 (31)</td>
<td>33 (2)</td>
<td>0.08</td>
</tr>
<tr>
<td>AS2</td>
<td>1</td>
<td>150</td>
<td>3,494</td>
<td>4</td>
<td>4 (NaN)</td>
<td>4 (1)</td>
<td>1.08</td>
</tr>
<tr>
<td>AS3</td>
<td>1</td>
<td>175</td>
<td>5,052</td>
<td>8</td>
<td>9 (NaN)</td>
<td>10 (2)</td>
<td>0.90</td>
</tr>
<tr>
<td>AS4</td>
<td>2</td>
<td>150</td>
<td>3,243</td>
<td>9</td>
<td>9 (NaN)</td>
<td>9 (1)</td>
<td>1.16</td>
</tr>
</tbody>
</table>

Eastern Tropical South Pacific (ETSP) water column

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Station</th>
<th>Depth</th>
<th>Sequencing depth</th>
<th>$S_{obs}$</th>
<th>Chao1 (SE)</th>
<th>Ace (SE)</th>
<th>Shannon</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETSP1</td>
<td>10</td>
<td>80</td>
<td>5,766</td>
<td>8</td>
<td>8 (4)</td>
<td>10 (1)</td>
<td>1.53</td>
</tr>
<tr>
<td>ETSP2</td>
<td>10</td>
<td>150</td>
<td>4,977</td>
<td>11</td>
<td>21 (NaN)</td>
<td>21 (2)</td>
<td>0.96</td>
</tr>
<tr>
<td>ETSP3</td>
<td>10</td>
<td>260</td>
<td>2,290</td>
<td>8</td>
<td>14 (NaN)</td>
<td>23 (2)</td>
<td>1.13</td>
</tr>
<tr>
<td>ETSP4</td>
<td>14</td>
<td>160</td>
<td>12,928</td>
<td>8</td>
<td>18 (NaN)</td>
<td>NaN (NaN)</td>
<td>0.79</td>
</tr>
<tr>
<td>ETSP5</td>
<td>14</td>
<td>260</td>
<td>5,168</td>
<td>12</td>
<td>18 (NaN)</td>
<td>17 (2)</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 1 Sample location information, the number of observed OTUs ($S_{obs}$), two estimators of total taxonomic richness (Chao1 and Ace) and the Shannon Index for all samples sequenced.

* MP marsh plot, CB Chesapeake Bay, CT Choptank River, AS Arabian Sea, ETSP Eastern Tropical South Pacific, SE standard error, NaN not a number

* This Arabian Sea sample was collected in 2004, the remaining samples were collected in 2007
2 min, followed by 30 amplification cycles consisting of 10 s at 98 °C, 30 s at 61 °C, and 1 min at 72 °C, and by a final elongation step at 72 °C for 5 min. A nested PCR was performed in triplicate on 0.15 % (v/v in final PCR reaction mixture) of the pooled triplicate PCR products using nirS3F and nirS6R primers [47] constructed with the required sequence adaptors for the FLX sequencing protocol, along with a multiplex identifier (MID) that was unique to each sample. PCR conditions were as described above but with slightly higher MgCl₂ (1.75 mM) and with an annealing temperature of 56 °C. PCR fragments of approximately 650 bp, visualized on a 1.5 % agarose gel (1× TAE buffer), were excised and purified using Qiaquick Gel Extraction Kit (Qiagen, Valencia, CA). Gel purified PCR amplicons were quantified via QuantIT™ Picogreen® Reagent from Invitrogen (now Life Technologies, Grand Island, NY) on an Agilent MX3005p qPCR system (Santa Clara, CA), and the triplicate PCR products were normalized to a concentration of 10 ng μl⁻¹, pooled and eluted in 10 mM Tris–HCl buffer (pH 8) for pyrosequencing on a Roche FLX Genome Sequencer using titanium chemistry at the Josephine Bay Paul Center for Comparative Molecular Biology and Evolution at the Marine Biological Laboratory (Woods Hole, MA). All sequence data have been made publicly available through NCBI’s Sequence Read Archive (Accession numbers: SRP029151, SRP 030649, SRP030652, and SRP030654).

Data Analysis

Homopolymer sequencing errors are common in pyrosequencing data from the 454 FLX platform [48–51]. In protein-coding genes, the homopolymer errors create the appearance of frameshift mutations that we exploited to identify errors and de-noise the pyrosequences. We trimmed all sequences to 432 bp and then ran the HMM-FRAME algorithm [52] on the trimmed sequences using a hidden Markov model (HMM) for cytochrome D1 that was obtained from Pfam (accession PF02239.10). Chimera detection was performed by UCHIME (version 4.2.40) [53], running in de novo mode with default parameter settings. UCHIME was driven by custom software written in Python using the Biopython libraries (version 1.59) [54]. Further quality filtering was performed on the HMM-FRAME output using additional software written in R (http://www.R-project.org). These additional steps included removing sequences that had (1) ambiguous alignments to the HMM as indicated by using hmmsearch (HMMER version 3.0; http://hmmer.org/) along with the cytochrome D1 HMM, (2) more than one domain hit to the model, (3) hmmsearch scores lower than 95, (4) sequences with HMM-FRAME scores lower than 80, and (5) sequences with any internal stop codons. To facilitate comparison with previous clone library studies [55], the quality-filtered data were clustered at 95 % sequence identity using ESPRIT-Tree [56]. Alpha diversity calculations and UniFrac metrics were calculated using Qime [57] by subsampling the dataset down to the sample with the fewest operational taxonomic units (OTUs) [58]. Heat maps and principal coordinates analysis of UniFrac metrics, constrained correspondence analysis, and adonis (a permutation-based multivariate ANOVA) were performed on Hellinger-transformed data using the Vegan package for the R statistical programming language [59, 60]. Many of these analyses have since been fully integrated into an easily implementable pipeline explicitly designed for the analysis of functional gene pyrosequencing data [46]. To calculate UniFrac metrics, our sequences, along with a seed alignment [24], were aligned in PyNAST [61], and the phylogeny was inferred using FastTree [62]. This phylogeny was visualized using the Interactive Tree of Life (version 2.1) [63]. The multivariate bar chart feature was used to plot the relative abundances of the OTUs from each location.

Results and Discussion

NirS Alpha Diversity and Rarefaction

After spurious sequences were removed, approximately 125,000 nirS sequences were retained, which were then clustered into operational taxonomic units (OTUs) defined at 95 % sequence identity. In total, we observed 1,815 unique OTUs from all sites. Sippewissett marsh plot sediments (MP) had the highest number of unique OTUs, with up to 529 OTUs in a single marsh plot (Table 1). The marsh plots also had the highest estimated taxonomic richness (Chao1 and ACE estimators up to 1,239 and 1,209, respectively; Table 1). In contrast, the ODZ samples were much less rich, with no more than 12 OTUs in any one sample, and richness estimators two to three orders of magnitude lower than the MP samples. The Chesapeake Bay sediments were intermediate between the two (Table 1). Patterns in diversity were entirely consistent with the richness estimates (Table 1). When unique sequences from individual samples were compiled to derive a habitatwide estimate of diversity (Table 1), the Shannon Diversity Index values were high for the Sippewissett Marsh sediments (4.71), low for the ODZs (1.4 for the Eastern Tropical South Pacific and 1.87 for Arabian Sea), and intermediate for the Chesapeake Bay (3.99). These diversity values are similar to those previously reported in clone library studies of nirS from both the Arabian Sea [45] and the Chesapeake Bay [55].

Despite the remarkable diversity identified in marine sediments, the rarefaction curves for those samples did not plateau, suggesting that there is considerable diversity remaining to be sampled (Fig. 1). Rarefaction curves for the ODZ samples, by contrast, have much lower OTU numbers (note the difference in scale on the y axis), and despite the low number
of OTUs, rarefaction curves for many of the samples have already plateaued, indicating high sequence coverage. Although direct comparison of richness and diversity across different studies is a challenge due to differences in the depth of sequencing [58], the low levels of nirS genetic diversity in the ODZs has been observed in studies of other genes. Although clone library analysis of 16S rRNA gene sequences from the ETSP ODZ indicated that bacterial diversity in the ODZ was similar to diversity in the overlying surface waters [64], more recent pyrosequencing of the 16S rRNA gene demonstrated that the least diverse samples in the ETSP ODZ were derived from the oxygen-deficient core [65]. Samples from numerous depths in the ETSP showed that both taxonomic richness and phylogenetic diversity decrease with depth, reaching a minimum in the core of the ODZ [66]. Furthermore, metagenomic analysis of protein-coding genes also indicated that the lowest functional diversity occurred in the oxygen-deficient waters of the ETSP [65] although that analysis considered all functional genes in aggregate. Our results are the first to show that the genetic diversity underlying nirS denitrification in ODZs is also low, despite the overwhelming importance of denitrification as a metabolic strategy in these waters. The lack of diversity and the dominance of a single OTU that we identified in the ODZ nirS gene distribution is consistent with the late stage of succession in a denitrification bloom [45], in which environmental conditions (high NO₂⁻, decrease in NO₃⁻) correspond with a decrease in the diversity of denitrifiers [45]. The reduced diversity of denitrifiers under specific environmental conditions associated with high rates of NO₃⁻ removal suggests that only a handful of denitrifiers are responsible for much of the fixed N loss that happens from ODZs. This pattern appears to be absent in coastal sediments where taxonomic richness and evenness are considerably higher (Table 1).

By contrast, molecular analysis of genes in coastal sediments suggests that these habitats harbor among the highest levels of taxonomic richness and phylogenetic diversity [66, 67]. Among the functional groups present in coastal sediments, a high degree of diversity has been demonstrated for sulfate-reducing bacteria [68–70], ammonia oxidizers [71–73], denitrifiers [44, 55], and nitrogen fixers [74–76]. Several possible mechanisms could explain the much higher extent of diversity in coastal sediments. First, the structural properties of sediments [77] and the sharp redox gradients that characterize surface sediments [78] could result in much greater niche diversity than is present in ODZs. Additionally, coastal systems receive inputs from both marine and terrestrial end members and thus harbor microbes from both habitats that may promote a higher overall diversity of microorganisms [79]. The dramatically greater diversity of nirS in the coastal sediments in this study may result from these mechanisms.

NirS Phylogeny

The nirS phylogeny, inferred from OTUs clustered at 95 % sequence identity, resulted in 13 clades, with the majority of sequences belonging to four large clades (clades 5, 7, 8, and 13; Fig. 2). Only 35 of the 1,815 nirS OTUs had a cultured representative with greater than 85 % sequence identity (ESM 3). The average percent sequence identity to a cultured representative, across all 1,815 OTUs, was 78±3.2 %, a clear evidence that the sequences detected in this study diverge considerably from most model denitrifying organisms. By contrast, the environmental database yielded 708 OTUs within 85 % sequence identity (ESM 4). The average percent sequence identity to an environmental clone was 85±4 %, and the majority of the environmental sequences with high
similarity were derived from coastal and marine sediments around the world.

For each clade, we mapped the relative abundances of OTUs from each of the habitats (Fig. 2, insets). As suggested by the richness and diversity estimators, the majority of OTU sequences were found only in sediments (Fig. 2). In fact, 9 of the 13 clades identified contained OTUs drawn almost exclusively from the salt marsh plots. Of those, only clade 13 (Fig. 2, pink inset) had OTUs with cultured representatives within 85% sequence identity, all of which were members of the alpha proteobacteria (ESM 3). The other eight clades of denitrifiers found in the marsh sediments had no close cultured representatives, indicating that salt marshes harbor a community of unique denitrifiers that are considerably more divergent, at least at the nirS gene level, than the model organisms from which our understanding of denitrification is derived.

In addition to the alpha proteobacteria, further analysis of the nirS phylogeny indicates other classes of proteobacteria may also play key roles in denitrification in marsh sediments. Although the most abundant sequences from the marsh sediments were widely distributed, with the ten most abundant OTUs found in seven different clades (Fig. 2), the most abundant OTU (1,796, 15% of marsh nirS sequences) was found in clade 8, with closest matches to the gamma proteobacteria, and the second most abundant (1,800, 14% of marsh nirS sequences) was found in clade 11. Clade 11 had no matches within 85% sequence identity, but OTU 1,800 was an 80% match to the beta proteobacterium *Brachymonas denitrificans*, suggesting that taxa from this class of proteobacteria may also

---

**Fig. 2** Phylogenetic tree of nirS OTUs clustered at 95% sequence similarity. Inset boxes contain expanded views of selected clades. The relative abundances of each OTU from the four different sites are represented in the circular histograms. Subsets of clades that had matches to pure or enrichment culture representatives with >85% sequence similarities are depicted in the outer concentric histograms. All other OTUs did not have previously sequenced representatives within ~85% sequence similarity to known denitrifying taxa. Clades not shown as insets here are composed almost entirely of OTUs found only in salt marsh sediments.
play important roles in removing fixed nitrogen from marsh sediments. Proteobacteria were also shown to dominate the bacterial community structure of salt marsh sediments via analysis of the 16S rRNA gene, where abundances were approximately equally split between the alpha, delta, and gamma classes [67].

Surprisingly, even though the Chesapeake Bay samples were also collected from coastal sediments, the phylogenetic structure of the Chesapeake Bay community was considerably different than the Sippewissett Marsh microbial community (Fig. 2). Forty percent of the sequences from the Chesapeake Bay belonged to one subclade within clade 7 (Fig. 2, light blue inset) and had no close cultured representatives (ESM 3), though these sequences did have a >99 % match to other environmental sequences derived from marine or estuarine sediments (ESM 4). The second most abundant OTU in the Chesapeake Bay sediments (1,588; 17 % of all sequences) had no close cultured representative and was only a 72 % match to a previously sequenced environmental clone (Table S2). This high divergence from known denitrifiers underscores the paucity of data that exists on the microbes responsible for this critically important biogeochemical pathway.

To validate these Chesapeake Bay results and as a test of the biases associated with primer design, pyrosequencing methodology, and computational pipeline, we sequenced the Chesapeake Bay sediment nirS from the same samples as a recently published clone library study of over 500 nearly full-length nirS sequences [55]. The clone library sequences were amplified using the Braker et al. [47] 1F-6R primers and the pyrosequencing data were derived using the 3F-6R primers [47]. The results between the two analyses were largely consistent. The most abundant OTU in this study (OTU 1,748) was also the most abundant sequence found in the clone library (OTU1, [59]). The other highly abundant sequences in the clone library analysis also all mapped to highly abundant sequences from the pyrosequencing data, with the top 5 OTUs (~33 % of the clones), present in the 20 most abundant pyrosequences. Notably, the highly divergent OTU 1,588 mentioned above was not detected in the clone library, suggesting some degree of bias, at least among more divergent sequences.

The phylogenetic distribution of nirS sequences from the ODZs was dramatically narrower than the coastal sediment sequences. In the Arabian Sea, 99 % of the sequences belonged to 5 OTUs (OTUs 1,650, 1,570, 1,483, 1,425, and 1,453). In the ETSP, 99 % of the sequences belong to 4 OTUs (1,570, 1,520, 1,487, and 1,333). OTU 1,570, the most abundant nirS sequence identified in this study, was the only dominant OTU present in both ODZs, but was identified only once in the estuarine sediments, and not at all in the marsh sediments. This OTU accounted for 38 % of all Arabian Sea sequences and 59 % of all ETSP sequences. It belongs to a deeply branching cluster in clade 12 (Fig. 2, purple inset) with no close cultured representative (ESM 2). It was a 92 % match to nirS sequences from marine sediments that were derived using primers specific for nirS from anammox bacteria belonging to the genus Scalindua (HQ666310, [30]). Analysis of 16S rRNA genes for anammox bacteria also indicates a relatively low diversity of organisms capable of this metabolism in the Arabian Sea ODZ [80, 81]. This provides support for the genetic capacity of anammox as a key nitrogen loss process in the world’s ODZs [11–13], though globally, the process can only account for approximately 30 % of fixed nitrogen loss [22]. These results further imply that the most abundant microbes present might not be the ones responsible for much of the fixed nitrogen loss in the system.

Aside from the Scalindua-related nirS, the vast majority of remaining sequences in the ODZs was found in clades 7 and 8. OTU 1,650 (42 % of the Arabian Sea sequences) and OTUs 1520 and 1333 (36 % of ETSP sequences) belong to one cluster in clade 7. Members of this clade were a close match to sequences previously described from the ODZs [12, 45, 82]. Two other abundant ODZ sequences (1583 and 1487) were present in another deeply branching cluster in clade 7. These sequences had no match within 75 % to a cultured representative and were only a 77 % match to an environmental clone from Changjiang Estuary sediments (EU235813). The remaining abundant ODZ OTUs belong to another deeply branching cluster in clade 8 and were also only distant matches to other nirS sequences (ESM 3 and 4). These phylogenetic insights provide evidence that the ODZ denitrifying community is genetically depauperate compared to denitrifying communities in coastal sediments.

NirS Beta Diversity

A heat map of weighted UniFrac [83, 84] similarity metrics shows the formation of two clusters of samples, one containing the majority of the ODZ samples, and the second containing all the sediment samples and two samples, ETSP3 and AS1, from the ODZs (Fig. 3). These two samples cluster separately from the remaining ODZ samples primarily as a result of the lack of the nirS gene that was similar to the Scalindua nirS. The relatively large degree of similarity among the remaining ODZ samples is likely a result of the high abundance of the Scalindua-like nirS sequences from those locations. The second cluster, containing all sediment samples and the two remaining ODZ samples, is separated into two subclusters, with one sample, CB1, from the oligohaline region of the Chesapeake Bay, not identifying with either subcluster (Fig. 3). One subcluster contains all eight salt marsh samples and one more loosely associated cluster contains the remaining Chesapeake Bay and ODZ samples. These UniFrac results indicate that there are distinct differences among the denitrifier communities from each habitat, results that are supported by
The lower the UniFrac metric, the more similar the community structure of the samples being compared. Numbers located within the boxes correspond to the UniFrac metric. ETSP within the boxes is the UniFrac metric derived from the sequences. The numbers located in the boxes represent the similarity of the samples being compared.

To further explore controls on the community structure of the ODZ denitrifiers, we ran a constrained correspondence analysis (Fig. 5) on Hellinger transformed counts from the ODZ samples, along with basic oceanographic parameters from each site (ESM 1). The Arabian Sea samples and the ETSP samples differentiated themselves along the primary axis (explaining 35% of the variance in the data) though there was a considerable spread in the Arabian Sea data along the secondary axis (explaining 21% of the variance). ETSP samples clustered much more tightly together. Of the environmental variables measured, the primary feature differentiating the two ODZs was temperature. ETSP and Arabian Sea samples also clustered by region on the basis of nirS community composition as analyzed by a functional gene microarray [85].

To further explore controls on the community structure of the ODZ denitrifiers, we ran a constrained correspondence analysis (Fig. 5) on Hellinger transformed counts from the ODZ samples, along with basic oceanographic parameters from each site (ESM 1). The Arabian Sea samples and the ETSP samples differentiated themselves along the primary axis (explaining 35% of the variance in the data) though there was a considerable spread in the Arabian Sea data along the secondary axis (explaining 21% of the variance). ETSP samples clustered much more tightly together. Of the environmental variables measured, the primary feature differentiating the two ODZs was temperature. ETSP and Arabian Sea samples also clustered by region on the basis of nirS community composition as analyzed by a functional gene microarray [85].
Kuenenia nirS was not detected by the microarray and Scalindua nirS was not included on the array so the microarray cluster reflects primarily canonical denitrifier community composition. OTU 1,570, the highly abundant Scalindua-like nirS sequence (relative abundances appear parenthetically in Fig. 5), also played a role in the similarity of the ODZ samples. This OTU accounted for from 0 % of all sequences at ETSP2 to 81 % of the sequences at ETSP5. Those samples with low numbers of OTU 1,570, however, were differentiated toward the top of the secondary axis compared to samples where this sequence was a greater portion of the whole community. Nitrite concentration and depth were positively correlated along the same axis (Fig. 5), suggesting that supply of nitrite from either nitrate reduction or ammonia oxidation influences whether anammox or denitrification is the nitrogen loss process active at any given point in time. These results are consistent with the apparent lack of a relationship between nitrite concentrations and the anammox-derived loss of N2 that has been previously reported [13].

The distribution of the nirS gene among canonical denitrifiers and the Scalindua-like bacteria varied dramatically at both ODZs. The Scalindua-like nirS gene ranged in relative abundance from barely detected in the Arabian Sea ODZ in the 2004 sample to 71 % of the community in one of the 2007 samples. Similarly, it ranged from undetected to 81 % of the nirS sequences in the ETSP. This large variation in Scalindua-like nirS gene abundance is mirrored in measured rates of these processes in both ODZs [12, 21, 86, 87]. Nitrogen-labeled incubation experiments have shown that both anammox [12, 86] and denitrification [20, 87] can be the dominant nitrogen loss mechanism in the ODZs, depending on the time and location of sampling. In a comprehensive survey of the two processes in the ETSP, Dalsgaard et al. [21] demonstrated that although anammox rates were detected in a greater number of locations, the overall rates of anammox accounted for approximately 30 % of the fixed nitrogen loss from the system, in keeping with the stoichiometric constraints on the process [22]. Further work is needed to explicitly link the absolute abundances of this gene and its transcripts to the measured rates of anammox and denitrification to better understand what factors control these critical processes.

Next-generation sequencing of functional gene amplicons provides an exciting opportunity to explore the extent of diversity of key genes that govern biogeochemical pathways. The removal of fixed nitrogen via denitrification and anammox is a critically important biogeochemical process that occurs in marine oxygen-deficient zones and in anoxic marine sediments. Our next-generation sequencing data show that the genetic richness of the nirS gene in coastal marine sediments is dramatically greater than in the ODZs. We observed a total of 1815 unique nirS OTUs, but only 39 of these OTUs were uncovered in the ODZs. The only OTU that was widely distributed in both the Arabian Sea and the ETSP was a close match to a nirS gene identified from the anammox bacterial genus Scalindua, providing robust genetic support for the importance of this pathway in both the Arabian Sea and the Eastern Tropical South Pacific. The widely varying relative abundance of this OTU in the ODZ samples further demonstrates the boom and bust nature of these fixed nitrogen loss processes, as has been demonstrated by geochemical rate measurements. The novel genetic richness of the nirS gene in coastal marine sediments coupled with the lack of closely related sequences in culture collections underscores how little we know about what controls the distribution of these critical microorganisms in the environment. It also serves as a reminder that new efforts to culture biogeochemically important microbes should be undertaken if we are to promote the microbial removal of increasing supplies of anthropogenically fixed nitrogen.

Acknowledgments Funding for this work came from multiple NSF grants to BBW and from the following NSF grants: DEB-0717155 (to John Hobbie) and DBI-0400819 (to Jennifer L. Bowen). Additional support was provided via start-up funds from University of Massachusetts Boston to JLB. We would like to acknowledge the Salt Pond Sanctuaries and Dr. E. F. X. Hughes and family for allowing us to have access to the Great Sippewissett marsh plots, and Ivan Valiela for providing support and maintenance of the plots. Christopher Jones and Sara Hallin provided us with their curated nirS alignment. Comments from anonymous reviewers greatly improved the quality of this manuscript.

References