Community composition of nitrous oxide reducing bacteria investigated using a functional gene microarray

A. Jayakumar, D. Balachandran, A.P. Rees, P.J. Kearns, J.L. Bowen, B.B. Ward

www.elsevier.com/locate/dsr2

PII: S0967-0645(17)30241-2
DOI: https://doi.org/10.1016/j.dsr2.2018.10.002
Reference: DSRII4507

To appear in: *Deep-Sea Research Part II*

Received date: 22 June 2017
Revised date: 23 April 2018
Accepted date: 4 October 2018

Cite this article as: A. Jayakumar, D. Balachandran, A.P. Rees, P.J. Kearns, J.L. Bowen and B.B. Ward, Community composition of nitrous oxide reducing bacteria investigated using a functional gene microarray, *Deep-Sea Research Part II,* https://doi.org/10.1016/j.dsr2.2018.10.002

Community composition of nitrous oxide reducing bacteria investigated using a functional gene microarray

A. Jayakumar[1*], D. Balachandran[1], A. P. Rees[3], P. J. Kearns[2], J. L. Bowen[2], B. B. Ward[1]

[1]Department of Geosciences, Princeton University, Princeton NJ 08540 USA

[2]Marine Science Center, Northeastern University, Nahant, MA 01908

[3]Plymouth Marine Laboratory, Plymouth, Devon, UK PL1 3DH

* Corresponding Author. Tel.: 609-258-6294. ajayakum@princeton.edu

Abstract

The diversity and environmental distribution of the *nosZ* gene, which encodes the enzyme responsible for the consumption of nitrous oxide, was investigated in marine and terrestrial environments using a functional gene microarray. The microbial communities represented by the *nosZ* gene probes showed strong biogeographical separation. Communities from surface ocean waters and agricultural soils differed significantly from each other and from those in oceanic oxygen minimum zones. Atypical *nosZ* genes, usually associated with incomplete denitrification pathways, were detected in all the environments, including surface ocean waters. The abundance of *nosZ* genes, as estimated by quantitative PCR, was highest in agricultural soils and lowest in surface ocean waters.

1. Introduction

Nitrous oxide is a trace component of the atmosphere, but is a potent greenhouse gas and an important ozone destroying agent. Oceans and agricultural soils are both significant sources of N_2O to the atmosphere, each contributing about 20% of the total flux (Ciais et al. 2013). Nitrous oxide is produced under both oxic and anoxic conditions during

nitrification (by oxidation of ammonia and reduction of nitrite) and denitrification (by reduction of nitrate and nitrite). The only known biological sink for N_2O is via reduction to N_2 by the enzyme nitrous oxide reductase (N_2OR), which is encoded by the *nosZ* gene cluster. N_2OR is typically associated with the canonical denitrification pathway, in which nitrogen oxides are sequentially reduced ($NO_3^- \rightarrow NO_2^- \rightarrow NO \rightarrow N_2O \rightarrow N_2$) as respiratory electron acceptors in anaerobic metabolism by bacteria. The pathway is somewhat modular, i.e., not all "denitrifiers" encode all the enzymes that catalyze the steps in the complete pathway (Graf et al. 2014), which may be part of the explanation for anomalous accumulations of the denitrification intermediates, nitrite and nitrous oxide, in environments such as the oxygen minimum zones (OMZs) of the ocean.

Canonical denitrification is usually restricted to anoxic or very low oxygen environments, and the enzymes involved in the pathway are inhibited by oxygen to varying degrees (Korner and Zumft 1989). Thus the consumption of N_2O is thought to be restricted to anoxic conditions. Conventional denitrifiers in the alpha-, beta- and gamma-proteobacteria, which are well characterized in both marine and terrestrial environments, possess what we shall refer to as conventional, or Clade I, *nosZ* genes. The recent discovery of novel *nosZ* sequences in bacterial groups, such as the Firmicutes, CFB (Cytophaga-Flavobacteria-Bacteroides) supergroup, and Verrumicrobia, not formerly associated with denitrification (Jones et al. 2013, Sanford et al. 2012) have extended the phylogenetic and physiological range of nitrous oxide consumption. Some of the atypical, or Clade II, *nosZ* genes occur in genomes that do not include the genes encoding nitrous oxide production from nitrite. For example, the Clade II genes are found in the Bacteroidetes and the Epsilon-proteobacteria, whereas the Clade I genes have not been reported in these

groups (Jones et al. 2013). Atypical *nosZ* sequences have been found widely in diverse environments, including arable soils, freshwater sediments and hydrocarbon contaminated soils (Jones et al 2013, Sanford et al. 2012). *nosZ* genes, diverse but with closer identity to Clade II, have recently been reported in two clades of Marinimicrobia, a phylum formerly known as Marine Group A and SAR406, and which has no cultured representatives but is widespread in the ocean (Hawley et al. 2017). Clearly N_2O reduction is not limited to the activity of complete denitrifiers and the two Clades of *nosZ* can be equally abundant in some environments.

The purpose of the present study was to investigate the community composition of N_2O consuming bacteria on the basis of their *nosZ* genes by screening a range of marine and terrestrial environments using a functional gene microarray that contains probes for both major clades of bacterial *nosZ* genes. We hypothesized that the NosZ community would vary among environments, showing regional biogeographical patterns, and that the *nosZ* genes will be more abundant under low oxygen conditions.

2. Materials and Methods

2.1 Nucleic Acid Manipulations and Quantitative PCR Analysis

The source of samples and physical/chemical data for each are listed in Table 1. Particulate material from 5-10 L of seawater was collected onto Sterivex capsules (0.2 µm filter, Millipore, Inc., Bedford, MA) with a peristaltic pump. Immediately after filtration the capsule filters were quick frozen in liquid nitrogen and stored at –80°C until the DNA could be extracted. Sediment samples were obtained from agricultural areas in Mississippi in July and August 2013 from homogenized 10 cm deep cores and frozen at -80°C until analysis.

DNA was extracted from 0.25-0.35 g of sediment samples using the UltraClean Soil DNA Kit (Mobio Laboratories, Inc., Carlsbad, CA). DNA was extracted from Sterivex capsule filters from the Arabian Sea using the PUREGENE Genomic DNA Isolation Kit (Gentra, Minneapolis, MN) following procedures recommended by the manufacturer. For all the other water samples DNA was extracted from Sterivex capsule filters using the ALLPrep DNA/RNA Mini Kit (Qiagen Sciences, Maryland, USA). PCR amplification and qPCR using SYBR Green and the Qiagen master mix (Qiagen Sciences, Maryland, USA) were described previously for seawater samples (Jayakumar et al. 2013). Those protocols were followed here for the *nosZ* gene using primers nosZ1F (WCSYTGTTCMTCGACAGCCAG) and nosZ1R (ATGTCGATCARCTGVKCRTTYTC) to amplify a 259 base-pair fragment (Henry et al. 2006) The locations of the PCR primers are shown in Figure S1. Standardization and verification of specificity for qPCR assays was performed as described (Jayakumar et al. 2009). Primers nosZ1F and nosZ1R targeted the Clade I genes; Clade II genes were not amplified separately. *Rhodopseudomonas palustris*, was used as a positive control to optimize the reaction and to construct a plasmid for use as a standard in the qPCR assays. The amplified products were visualized after electrophoresis in 1.0% agarose gels stained with ethidium bromide.

Assays for each sample group (Mississippi soil, seawater) were carried out within a single assay plate (Smith et al. 2006). Each assay included triplicates of the no template controls (NTC), no primer control (NPR), four or more standards, and triplicates of known quantity of the environmental DNA samples (20 – 25 ng). DNA was quantified using PicoGreen fluorescence (Molecular Probes, Eugene, OR) calibrated with several dilutions of phage lambda standards. Quantitative PCR was performed using a Stratagene MX3000P (Agilent

Technologies, La Jolla, CA, USA). Automatic analysis settings were used to determine the threshold cycle (Ct) values. The copy numbers were calculated according to:

$$copy\ number = \frac{ng\ x\ \frac{number}{mole}}{bp\ x\ \frac{ng}{g}\ x\ \frac{g}{mole\ bp}}$$

and then converted to copy number per ml seawater filtered, or per gram of sediment assuming 100% extraction efficiency.

To maintain continuity and consistency among qPCR assays a subset of samples from the first assay was included in subsequent assays, as well as fresh dilution series for standard curves on every assay. Template DNA and plasmid DNA were quantified prior to every assay as above using PicoGreen fluorescence to account for DNA loss that occurs upon repeated freeze-thaw cycles.

2.2 Microarray

The microarray (BC016) was developed following the archetype array approach described and employed previously (e.g., Ward and Bouskill 2011, Bulow et al. 2008) with 90-mer oligonucleotide probes. Each probe included a *nosZ*-specific 70-mer region and a 20-mer control region (5′-GTACTACTAGCCTAGGCTAG-3′) bound to a glass slide. The design and spotting of the probes has been described previously (Bulow et al. 2008).

Microarray BC016 contained two probe sets (NosZ and WNZ) for the *nosZ* gene. Clade I *nosZ*, most commonly found in marine and terrestrial heterotrophic denitrifying bacteria, was represented by 71 NosZ probes derived from whole genome sequences in public databases plus sequences obtained from clone libraries made using DNA extracts from the Great Sippewissett Marsh in Falmouth, MA, USA (Kearns et al. 2015). Clade I *nosZ* sequences were

associated with cultured strains representing alpha-, beta-, and gamma-Proteobacteria. An additional 43 WNZ probes were included to capture the atypical Clade II *nosZ* sequences (Sanford et al. 2012, Jones et al. 2013). Cultivated members of the atypical WNZ probe set include alpha-, beta- and delta-Proteobacteria, CFB supergroup, Firmicutes and Verrumicrobia. The probe accession numbers and sequences are listed in Table S1 and the phylogenetic trees of the probe sequences are found in Figure S2. The trees do not represent exactly the phylogeny of the complete genes because it is based only on the 70mer region of the probe. The 70mers were derived from the most variable part of the gene that allowed a robust alignment in order to maximize the resolution of the probe set (Bulow et al 2008). (The locations of the probe regions are shown in Figure S1.) Arrays were printed on glass slides (DeRisi et al. 1997) by Microarrays, Inc. (Huntsville, AL, USA).

Targets were prepared according to Ward and Bouskill (2011). PCR amplified DNA (from the qPCR assays above) was used for labeling with a BioPrime kit (Thermo Fisher Scientific, Cambridge, MA, USA) using random primers and a custom 1.2 mM dNTP mix with dUaa, followed by ethanol precipitation. The precipitated DNA was dissolved in 4.5 µl of 100 nM $NaCO_3$ (pH 9) before the addition of 4.5 µl of Cy3 dye and incubated for 3 h to overnight. Samples were then purified using a QIAquick PCR cleanup kit (Qiagen, Valencia, CA, USA) with buffer modifications described by Ward and Bouskill (2011). DNA concentration of the targets were measured on a Nanodrop spectrophotometer and the volume required for 200 ng of DNA was aliquoted into two separate tubes per sample, dried down under vacuum, and stored frozen until hybridized.

Samples were hybridized to the arrays overnight in sealed tubes and then washed according to Ward and Bouskill (2011). Arrays were scanned with an Agilent laser scanner

4300 (Agilent, Palo Alto, CA) and analyzed using GenePix 6.0 software. Replicate features on the same array were averaged to calculate the Cy3/Cy5 ratio for each probe. Relative fluorescence ratio (RFR, the fraction that each probe fluorescence (Cy3/Cy5 ratio) contributes to the total fluorescence of the probe group) and normalized fluorescence ratio (FRn, the Cy3/Cy5 ratio of each probe normalized to the maximum Cy3/Cy5 detected on that array for the probe group) were used for plotting and statistical analysis. The original array data are available at Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/projects/geo/) at the National Center for Biotechnology Information under **GEO Accession No. #####**

2.3 Statistical analysis.

The array data were analyzed using the vegan package in R (CRAN website; http://www.R-project.org) (Borcard et al. 2011). FRn values were transformed (Arcsin(Square root)) in order to normalize the proportional data. Environmental data were transformed (square root) and then standardized around zero (decostand in vegan). The transformed data were used in all diversity and ordination analyses according to Borcard et al (2011). The null hypothesis that the *nosZ* community composition did not differ between regions was tested in R using Multi-response Permutation Procedure (MRPP) (Zimmerman et al. 1985) with a significance level of 5%. The significance of the grouping of the stations by community composition was assessed using anosim in R.

3. Results

3.1 Abundance of nosZ genes

The abundance of *nosZ* genes was on the order of a few hundred to a few thousand copies ml^{-1} of seawater for the marine samples and much higher ($>10^7$ g^{-1}) in the soil samples (Table 1). Within the marine samples, the highest abundances ($\sim10^3$ ml^{-1}) occurred in the OMZ samples and the lowest in the polar samples.

3.2 Major Archetypes

The thirteen samples from four different environments showed regionally distinct hybridization patterns for Clade I *nosZ* archetypes (Figure 1). The two polar samples, both surface water, differed from each other but were even more different from the other 11 samples. Most samples had low evenness (Table 2) with one or a few archetypes dominating the total hybridization signal (Figure 1). The archetypes responsible for the major signals varied with habitat.

The major archetype in the Arctic sample was NosZ20, which represents *Pseudomonas stutzeri*, a commonly isolated marine and soil denitrifying strain. The second strongest signal in the Arctic sample was NosZ24, which represents several *Rhodopseudomonas palustris* strains derived from marine sediments. NosZ24 was the largest signal in the Antarctic sample, followed by NosZ64 (*Rhodanobacter denitrificans*, isolated from a contaminated aquifer, and representing a large number of uncultured soil *nosZ* sequences).

The three major Clade I signals in the ETNP and ETSP samples were NosZ65, NosZ14 and NosZ6. NosZ65 represents a number of cultured *Marinobacter sp*., denitrifiers

isolated from various marine environments and NosZ6 represents other *Marinobacter sp.* from sediments plus several salt marsh sequences. NosZ14 represents uncultured marine/salt marsh sediment sequences (Kearns et al. 2015), but was detected only in the OMZ samples, not in the soils.

NosZ15 (representing numerous uncultured soil sequences and *Achromobacter xylosoxidans*) was one of the largest signals in the Arabian Sea, followed by three archetypes with strong signals in the ETSP, NosZ6, NosZ14 and NosZ65. NosZ15 was the largest signal in the Mississippi soil samples, followed by NosZ60 (*Sinorhizobium meliloti*) and NosZ20 (*Pseudomonas fluorescence* and *P. stutzeri*), both associated with *nosZ* sequences derived from soils and coastal sediments.

Two archetypes were responsible for the major signals of the atypical *nosZ* in all samples across all locations (Figure 2). WNZ16 is an environmental sequence derived from agricultural soils while WNZ21 represents *Anaeromyxobacter delahlogenans*, which was isolated from soils contaminated with halogenated hydrocarbons.

One other WNZ archetype had a relatively large signal in most samples, WNZ43, an atypical *nosZ* sequence from activated sludge with no known cultured representatives. WNZ38, representing *Salinibacter ruber*, a halotolerant soil bacterium, was a large signal only in the Arabian Sea samples.

3.3 Community composition

The relative abundance of different archetypes, estimated from hybridization signal strength, varied biogeographically (Figures 1 and 2). MRPP analysis confirmed that the

community compositions in all four environments differed significantly from each other, both for NosZ and WNZ archetypes ($p < 0.001$ for both).

A Principal Components Analysis (PCA) for conventional NosZ archetypes (Figure 3A) revealed that all four environments clustered separately, although the two polar samples were quite distant from each other, and the Arctic sample was much more similar to the other marine samples than to the Antarctic sample. The PCA for WNZ archetypes (Figure 3B) also grouped the stations by environment. Both groupings were significant at $p < 0.001$ (anosim).

The archetypes that were correlated with different samples include both the major signals mentioned above and many more archetypes with minor signal strength. For example, both major signals NosZ15 and NosZ60 were correlated with the Mississippi soil samples. NosZ58 and NosZ59 are also strongly correlated with the Mississippi soil samples, where they represented only 5-6 % of the total signal, but they were much less than 1 % in all the marine samples, and thus they distinguish the soil samples from all others. Similarly, NosZ69 and NosZ70 were a small fraction of the total signal in the ETSP samples, but because they were absent in all the others, they distinguish the ETSP communities. NosZ29 distinguished the OMZ samples; it represents sequences derived from the salt marsh, but comprised ~3 – 9 % of the signal in the OMZ samples and was very minor or not detected in the soil or polar samples. *Pseudomonas stutzeri*, which has been cultivated from both soils and marine environments, is represented in NosZ20, and was detected at ~5% or more of the signal in all samples except the Pacific OMZ samples.

3.4 Environmental correlates

To investigate environmental drivers of *nosZ* communities, the array results for the marine samples were analyzed separately because analogous environmental data were not available for the soil samples. The three marine environments clustered separately when ordinated with key environmental parameters and *nosZ* archetype community composition (Figure 4). The polar samples were positively correlated with oxygen, in opposition to all the other samples, which were from OMZ environments. Salinity and temperature were both higher in the Arabian Sea than elsewhere, and these samples also had the highest abundance of *nosZ*, estimated by qPCR. NosZ15, NosZ20, NosZ14 and NosZ6 all correlated positively with the Arabian Sea communities, while NosZ65 was correlated with the ETSP and ETNP samples.

The clusters were less clear for WNZ community composition ordinated with environmental variables (Figure S3), but the general distributions were similar to the clusters obtained for *nosZ* archetype analysis. Only the archetypes that differentiated the samples, i.e., some of the smaller signals, were strongly correlated with community composition. WNZ38 and WNZ36 were distinguishing archetypes for the Arabian Sea communities and were highly correlated with the three environmental variables that distinguished the Arabian Sea (T, S and bottom depth). WNZ16, correlated with oxygen, was the strongest signal in the Antarctic sample (but was the second largest signal in all others)

4. Discussion

4.1 Abundance of nosZ genes

The highest abundance of *nosZ* genes was found in the Mississippi agricultural soils, on the order of 2 – 8 x 10^7 copies g^{-1} of soil. This is very similar to abundances reported for salt marsh sediments (0.45 – 5.2 x 10^7 copies g^{-1} of sediment; Kearns et al. 2016), agricultural soils under various fertilizer and crop regimes (~1 – 4 x 10^6 copies g^{-1} dry soil; Thompson et al. 2016) and from pasture and riparian soils (~5 x 10^5 – 1 x 10^7 copies g^{-1} dry soil; Deslippe et al. 2014). Deslippe et al. (2014) and Thompson et al. (2016) also quantified *nirS* or *nirS* and *nirK*, the genes that encode nitrite reductase, a step in the denitrification sequence prior to the reduction of nitrous oxide. In both cases, the abundance of the nitrite reductase genes exceeded that of *nosZ* by a factor of 10-fold or more.

The abundance of *nosZ* was lower in all of the marine samples than in the soil samples. The highest copy numbers (less than 2 x10^3 copies ml^{-1}) in OMZ samples were found in the three samples from the Arabian Sea OMZ. Similar abundances of *nosZ* genes (average 10^3, maximum of 5 x 10^3 copies ml^{-1}) and transcripts (up to 0.9 x 10^3 copies ml^{-1}) were reported previously from the Arabian Sea (Wyman et al. 2013). In the data of Wyman et al. (2013), expression of *nosZ* was highest at the depth of the primary nitrite maximum in well oxygenated waters, leading Wyman et al (2013) to suggest that facultative denitrifiers were associated with low oxygen microzones on the surface of *Trichodesmium* colonies. Abundances of *nirS* and *nirK* have been reported previously (Jayakumar et al. 2013) for some of the samples analyzed here for nosZ; *nirS* was present at much higher levels (~5 x 10^5 copies ml^{-1}) while *nirK* was present at only a few hundred copies ml^{-1}.

As far as is known, both *nirS* and *nosZ* are single copy genes, so their abundance should correlate with abundance of cells that contain the pathways. These gene abundances imply that cells containing these genes represent a small fraction of the total assemblage in both soil and seawater environments. Given typical bacterial abundances (10^9 cells g^{-1} for soils and maximum of 10^{-6} cells ml^{-1} for seawater at OMZ depths) *nosZ* bearing cells represent at most 8% and 2% of the total assemblage in agricultural soils and OMZ seawater, respectively. By contrast, *nirS* bearing cells are much more abundant and have even been estimated to be as much as 50% of the total assemblage in OMZ samples (Jayakumar et al. 2013).

The consistent finding that *nirS* abundance exceeds that of *nosZ* in both marine and soil environments may imply decoupling of the complete denitrification pathway. N_2O does not accumulate to very high levels except under certain environmental conditions. This might suggest that different parts of the *nirS* community are active under different conditions, while an effective N_2O scavenging assemblage, including microbes that do not produce N_2O themselves, is very efficient at removing N_2O.

It is also likely, however, that primer bias obscures the actual diversity and total abundances of the genes responsible for the multiple steps in the denitrification pathway. The *nirS* abundances for the marine samples reported here were obtained by qPCR using primers designed for marine denitrifiers (Braker et al. 1998). Using another set of primers (Throback et al. 2004), denitrifier *nirS* was detected at levels up to only 15 copies ml^{-1} (Lam et al. 2009) in the ETSP. Clearly the abundance estimates are heavily influenced by the choice of primers, and primer bias is probably also at work in the present estimates of *nosZ* abundance. The fact that *nosZ* abundances reported here were much lower in the Pacific OMZs than in the Arabian

Sea could be due to different community composition, which could influence both number and type of *nosZ* detected.

The *nosZ* primers used here were developed by Henry et al. (2006) at a time when both marine and terrestrial *nosZ* sequences were already well represented in the data base. The nosZ1F/nosZ1R primers of Henry et al. (2006) amplify a shorter region (259 bp) than that amplified by the primers developed by Scala and Kerkhof (1999; ~1100 bp) for marine targets, and so were more compatible with quantitative PCR amplification (Table S1). Nevertheless, we suspect that the qPCR primers used here may have underestimated the real abundance of *nosZ* genes in the marine samples

Both Jones et al. (2013) and Sanford et al (2012) developed new PCR primers for atypical *nosZ* genes, partly motivated by the clear separation between sequences of the two main clades of *nosZ* genes, and the finding that the existing primers (Scala and Kerkhof 1998, Henry et al. 2006) did not amplify some of the Clade II sequences. Although the primers used in the current study have been shown to underestimate the total *nosZ* abundance because they do not adequately amplify the Clade II sequences, it is clear from the array results that they do in fact amplify at least some of them. Many of the Clade II archetypes were detected on the array and two of them were among the highest FRn signals (Figure 2). That is, both clades are represented in the community composition detected on the arrays (the targets hybridized to the arrays were the very same PCR products that were produced in the qPCR assays), and the abundance estimates reported here include conventional as well as atypical *nosZ* genes.

All of the primers commonly used to amplify *nosZ* overlap in their target sequence range, such that most of them amplify the region represented by the NosZ and WNZ probes on the array (Figure S1). The primers used by Wyman et al. (2013) amplify a fragment that

overlaps only partially with the WNZ probe region and not at all with the NosZ probe region (Table S1). Therefore, it is not possible to know whether the *Labrenzia*-type *nosZ* genes from the Arabian Sea would have been detected by either qPCR or the array in the present study. The few *nosZ* sequences available from the recently described Marinimicrobia are not represented on the current array and would not have been amplified by the primers used in this study. These sequences are estimated to be present at the level of a few percent of the sequenced metagenome in some anoxic environments (Hawley et al. 2017), which would make them a significant component of the overall *nosZ* assemblage, but would not change the order of magnitude of the total *nosZ* abundance.

We conclude from this discussion that all the existing primers, including those used for this study, very likely underestimate both the abundance and the diversity of *nosZ* genes in terrestrial and marine environments. Therefore, it would be worthwhile to revisit primer design for the *nosZ* genes. It seems unlikely that truly universal primers are possible, but it may be possible to develop improved primers that explicitly target phylogenetically distinct sequences across the entire range of diversity, regardless of environment.

The modularity of the N cycle may also be a factor in understanding the relative abundance and phylogenetic affiliations of genes associated with different steps in the denitrification pathway (Graf et al. 2014, Stein and Klotz 2016). Stein and Klotz (2016) argue that the broad distribution of diverse nitrogen transformation genes emerged gradually over evolutionary time in response to changing environmental conditions. This resulted in a modular design, rather than a phylogenetically defined distribution of N cycle functions. Therefore, the abundance and phylogenetic relationships observed for these marker genes may reflect the growth and activity response of diverse microbes to the

continually changing conditions of their environment. These responses may lead to advantages in growth or transfer efficiency for individual members of the assemblage, but are not easily understood from the single gene perspective taken here with qPCR and community analysis based on the microarray.

4.2 Community composition

The most striking feature of the community composition as represented by the hybridization patterns is the biogeography of archetypes. Many of the same archetypes were important in many samples, but their relative contributions to the total signal varied significantly among habitats. The regional differences were stronger for the conventional NosZ archetypes than for the atypical WNZ archetypes. Clustering of community composition by habitat is clear whether the environmental variables are included in the analysis or not. This may imply that interactions among the assemblage represented by *nosZ* genes and with other components of the overall assemblage are important determinants of community composition, in addition to selection by environmental variables. Biogeographical patterns are often observed in microbial distributions, including ammonia-oxidizing archaea (Peng et al. 2013, Biller et al. 2012; Pester et al. 2014, Santoro et al. 2017) and *nirS*-defined denitrifying bacteria (Jayakumar et al. 2013) in OMZs and other environments, and *nifH* in sediment microbial communities (Zhou et al. 2016).

A second significant finding is that *nosZ* genes were detected in samples from the well oxygenated near surface waters of the polar regions. For the conventional *nosZ* archetypes, the two polar assemblages were different from each other and different from all the other samples as well, consistent with the idea that these environments are not likely to harbor conventional

denitrifying communities. By contrast, *nirS*, a key gene in the denitrifying pathway, is abundant within the anoxic nitrite containing waters but is generally undetectable in the oxic surface layer (Jayakumar et al 2009, Ward et al 2009). Despite the regional differences in atypical *nosZ* community composition, the same two atypical WNZ archetypes constituted the most intense signals in all 13 samples (Figure 2). All of the atypical *nosZ* sequences used on the microarray were obtained from soils (both agricultural soils and hydrocarbon contaminated sites; Sanford et al. 2012). The hybridization data make it clear that sequences closely related to those of soil microorganisms occur in seawater as well. Atypical *nosZ* genes are not generally associated with complete denitrification, but they may represent an important sink for N_2O that was previously not considered to be important in oxygenated waters. Wyman et al. (2013) suggested that facultative denitrifiers related to *Labrenzia* could be responsible for N_2O consumption in *Trichodesmium* aggregates in surface waters.

The microarray does not contain the entire breadth of *nosZ* diversity because the sequence database represented on the microarray is biased towards sequences retrieved from soil and marsh environments. While the large sequence database of marine *nosZ* sequences (e.g., Scala and Kerkhof 1999) is not fully represented on the microarray, *nosZ* probes based on sequences from cultivated marine denitrifiers produced major signals in the OMZ samples (e.g., *Marinobacter*, NosZ65 and NosZ6; *P. stutzeri*, NosZ20). On the basis of *nirS* clone libraries (Jayakumar et al 2009), *Marinobacter sp.* was identified as an important component of the denitrifying assemblage in the OMZs; *Marinobacter sp.* represented 33% of the sequences in a sample from the secondary nitrite maximum of the Arabian Sea OMZ.

As mentioned above, many of the Clade II *nosZ* genes are not amplified by the primers used in this study (Jones et al. 2013, Sanford et al. 2012), but our analysis indicates sufficient

identity with the primers to allow amplification of the atypical *nosZ* sequences represented by probes WNZ21 and WNZ16. Both of these probes represent *Anaeromyxobacter* species, a genus associated with diverse anaerobic metabolisms in soils. These two WNZ probes have very low identity with any of the conventional *nosZ* probes, and a recent BLAST search revealed no near identities with any sequence other than the WNZ sequences from which the probes were derived. Thus their signals truly represent the presence of atypical *nosZ* genes in the oxic surface ocean, as well as in the OMZs.

Within the constraints of primer selection, these data represent a snapshot of the presence of microbes that possess the genetic potential for N_2O reduction. It remains to be investigated which of these genes are being expressed under different environmental conditions. At some level, however, abundant archetypes must either be active or contained in cells that are active under those conditions, even if they are using some unrelated metabolism. Predation and degradation would eventually remove inactive cells and decrease the gene abundance in an inactive population. Hybridization to the microarray of targets derived from RNA can be used to address the question of activity vs. presence. Biogeochemical experiments would be useful to investigate whether N_2O is actually produced and consumed in surface waters (Rees et al. 2016).

Acknowledgements

We thank Ian Brown for collecting the polar samples. Funding for APR came from NERC UKOA grant NE/H017259/1. AJ, DB, PJK, JLB and BBW acknowledge funding from NSF award DEB-1019624

Figure S1. Map of the *nosZ* gene from *Pseudomonas fluorescens* showing location of commonly used PCR primers and 70-mer probe regions for both Clade I and Clade II *nosZ* gene amplification

Figure S2. Phylogenetic trees showing relationships among NosZ (A) and WNZ (B) probes

Figure S3. Redundancy analysis for marine samples, based on WNZ archetypes, overlain with environmental variables.

Table S1. NosZ Probe ID list

Table S2. WNZ Probe ID list

References

1. Biller, S. G., Mosier, A. C., Wells, G. F. and Francis, C. A., 2012. Global biodiversity of aquatic ammnia-oxidizing archaea is partitioned by habitat. Frontiers in Microbiology. 3, 252.
2. Borcard, D., Gillet, F. and Legendre, P., 2011. Numerical Ecology with R, Springer, New York.
3. Braker, G., Fesefeldt, A. and Witzel, K.-P., 1998. Development of PCR primer systems for amplification of nitrite reductase genes (nirK and nirS) to detect denitrifying bacteria in environmental samples. Applied and Environmental Microbiology. 64, 3769-3775.
4. Bulow, S. E., Francis, C. A., Jackson, G. A. and Ward, B. B., 2008. Sediment denitrifier community composition and nirS gene expression investigated with functional gene microarrays. Environmental Microbiology. 10, 3057-3069.
5. Ciais, P., Sabine, C., G, B., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R. B., Piao, S. and Thornton, P., 2013. Carbon and Other Biogeochemical Cycles. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. (T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauel, Y. Xia, V. Bex and P. M. Midgley). Cambridge University Press, Cambridge, UK and New York, NY, USA: 465-570.
6. DeRisi, J. L., Iyer, V. R. and Brown, P. O., 1997. Exploring the metabolic and genetic control of gene expression on the genomic scale. Science. 278, 680-686.
7. Deslippe, J. R., Jamali, H., Jha, N. and Saggar, S., 2014. Denitrifier community size, structure and activity along a gradient of pasture to riparian soils. Soil Biology & Biochemistry. 71, 48-60.
8. Graf, D. R. H., Jones, C. M. and Hallin, S., 2014. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N_2O emissions. PLOS One.

9. Hawley, A. K., Nobu, M. K., Wright, J. J., Durno, W. E., Morgan-Lang, C., Sage, B., Schwientek, P., Swan, B. K., Rinke, C., Torres-Beltran, M., Mewis, K., Liu, W. T., Stepanauskas, R., Woyke, T. and Hallam, S. J., 2017. Diverse Marinimicrobia bacteria may mediate coupled biogeochemical cycles along eco-thermodynamic gradients. Nature Communications. 8.
10. Henry, S., Bru, D., Stres, B., Hallet, S. and Philippot, L., 2006. Quantitative detection of the nosZ Gene, encoding nitrous oxide reductase, and comparison of the abundances of 16S rRNA, narG, nirK, and nosZ genes in soils Applied and Environmental Microbiology. 72, 5481-5189.
11. Jayakumar, A., Naqvi, S. W. A. and Ward, B. B., 2009. Distribution and relative quantification of key genes involved in fixed nitrogen loss from the Arabian Sea oxygen minimum zone. Indian Ocean Biogeochemical Processes and Ecological Variability. (J. D. Wiggert and R. R. Hood). American Geophysical Union, Washington, D. C.: 187-203.
12. Jayakumar, A., Peng, X. F. and Ward, B. B., 2013. Community composition of bacteria involved in fixed nitrogen loss in the water column of two major oxygen minimum zones in the ocean. Aquatic Microbial Ecology. 70, 245-259.
13. Jones, C. M., Graf, D. R. H., Bru, D., Philippot, L. and Hallin, S., 2013. The unaccounted yet abundant nitrous oxide-reducing microbial community: a potential nitrous oxide sink. Isme Journal. 7, 417-426.
14. Kearns, P. J., Angell III, J. H., Feinman, S. G. and Bowen, J. L., 2015. Long-term nutrient addition differentially alters community composition and diversity of genes that control nitrous oxide flux from salt marsh sediments. Estuarine Coastal and Shelf Science. 154, 39-47.
15. Korner, H. and Zumft, W. G., 1989. Expression of denitrification enzymes in response to the dissolved oxygen level and respiratory substrate in continuous culture of *Pseudomonas stutzeri*. Applied and Environmental Microbiology. 55, 1670-1676.
16. Lam, P., Lavik, G., Jensen, M. M., van de Vossenberg, J., Schmid, M., Woebken, D., Dimitri, G., Amann, R., Jetten, M. S. M. and Kuypers, M. M. M., 2009. Revising the nitrogen cycle in the Peruvian oxygen minimum zone. Proceedings of the National Academy of Sciences of the United States of America. 106, 4752-4757.
17. Peng, X., Jayakumar, A. and Ward, B. B., 2013. Community composition of ammonia-oxidizing archaea from surface and anoxic depths of oceanic oxygen minimum zones. Frontiers in microbiology. 4, 177-177.
18. Pester, M., Maixner, F., Berry, D., Rattei, T., Koch, H., Luecker, S., Nowka, B., Richter, A., Spieck, E., Lebedeva, E., Loy, A., Wagner, M. and Daims, H., 2014. NxrB encoding the beta subunit of nitrite oxidoreductase as functional and phylogenetic marker for nitrite-oxidizing Nitrospira. Environmental Microbiology. 16, 3055-3071.
19. Rees, A. P., Brown, I. J., Jayakumar, A. and Ward, B. B., 2016. The inhibition of N2O production by ocean acidification in cold temperate and polar waters. Deep-Sea Research Part Ii-Topical Studies in Oceanography. 127, 93-101.
20. Sanford, R. A., Wagner, D. D., Wu, Q., Chee-Sanford, J. C., Thomas, S. H., Cruz-Garcia, C., Rodriguez, G., Massol-Deya, A., Krishnani, K. K., Ritalahti, K. M., Nissen, S., Konstantinidis, K. T. and Loeffler, F. E., 2012. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. Proceedings of the National Academy of Sciences of the United States of America. 109, 19709-19714.
21. Santoro, A. E., Saito, M. A., Goepfert, T. J., Lamborg, C., Dupont, C. and DiTulliio, G. R., 2017. Thaumarchaeal ecotype distributions across the equatorial Pacific Ocean and their potential roles in nitrification and sinking flux attenuation. Limnology and Oceanography.
22. Scala, D. J. and Kerkhof, L. J., 1998. Nitrous oxide reductase (*nosZ*) gene-specific PCR primers for detection of denitrifiers and three *nosZ* genes from marine desiments. FEMS Microbiology Letters. 162, 61-68.

23. Scala, D. J. and Kerkhof, L. J., 1999. Diversity of nitrous oxide reductase (nosZ) genes in continental shelf sediments. Applied and Environmental Microbiology. 65, 11681-11687.
24. Smith, C. J., Nedwell, D. B., Dong, L. F. and Osborn, A. M., 2006. Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples. Environmental Microbiology. 8, 804-815.
25. Stein, L. Y. and Klotz, M. G., 2016. The nitrogen cycle. Current Biology. 26, R94-R98.
26. Thompson, K. A., Bent, E., Abalos, D., Wagner-Riddle, C. and Dunfield, K. E., 2016. Soil microbial communities as potential regulators of in situ N2O fluxes in annual and perennial cropping systems. Soil Biology & Biochemistry. 103, 262-273.
27. Throback, I., Enwall, K., Jarvis, A. and Hallin, S., 2004. Reassessing PCR primers targeting nirS, nirK and nosZ genes for community surveys of denitrifying bacteria with DGGE. FEMS Microbiology Ecology. 49, 401-417.
28. UNESCO, 1994. Protocols for the Joint Global Ocean Flux Study (JGOFS) Core Measurements,
29. Ward, B. B. and Bouskill, N. J., 2011. The Utility of Functional Gene Arrays for Assessing Community Composition, Relative Abundance, and Distribution of Ammonia-Oxidizing Bacteria and Archaea. Methods in Enzymology, Vol 46: Research on Nitrification and Related Processes, Pt B, 496: 373-396.
30. Wyman, M., Hodgson, S. and Bird, C., 2013. Denitrifying Alphaprotepbacteria from the Arabian Sea that express nosZ, the gene encoding nitrous oxide reductase, in oxic and suboxic waters. Applied and Environmental Microbiology. 79, 2670-2681.
31. Zhou, H. X., Dang, H. Y. and Klotz, M. G., 2016. Environmental Conditions Outweigh Geographical Contiguity in Determining the Similarity of nifH-Harboring Microbial Communities in Sediments of Two Disconnected Marginal Seas. Frontiers in Microbiology. 7.
32. Zimmerman, G. M., Goetz, H. and Mielke, P. W., 1985. Use of an Improved Statistical-Method for Group Comparisons to Study Effects of Prairie Fire. Ecology. 66, 606-611.

Figure 1. Stacked bar plot showing FRn for each conventional NosZ archetype at each station. Colors represent individual archetype signals (see legend).

Figure 2. Stacked bar plot showing FRn for each atypical WNZ archetype at each station. Colors represent individual archetype signals (see legend).

Figure 3A. Principal components analysis for all samples on the basis of community composition of NosZ archetypes.

Figure 3B. Principal components analysis for all samples on the basis of community composition of WNZ archetypes.

Figure 4. Redundancy analysis for marine samples, based on NosZ archetypes, overlain with environmental variables.

Table 1. Sample location and environmental data. NA = not applicable. ND = not determined. O_2 concentration was determined by Seabird SBE43 sensor mounted on the CTD; NO_2, NO_3 and PO_4 were determined using standard colorimetric protocols (UNESCO 1994).

Sample	Cruise	Station	Depth (m)	Longitude oE	Latitude oN	Bottom Depth [m]	To C	S	O2 (µM)	NO3 (µM)	NO2 (µM)	PO4 (µM)
Polar												
OA1_EO4	JR271	E04	40	3.66	78.35	100	-1.8	35.15	343	4.16	0.0	10.1
OA2_BO4	JR274	B04	60	25.93	-58.08	2896	-1.2	33.71	375	21.51	0.0	ND
Eastern Tropical Pacific OMZ regions												
ETNP_BB2_140m	TN278-ETNP 2012	BB2	140	-107.15	16.53	3600	13.0	34.81	1.70	19.79	6.1	2.6
ETSP_Stn55_175m	NBP13-05-2013	BB2	175	-70.66	-20.77	1782	12.6	34.87	2.32	11.6	7.1	2.8
ETSP_BB2_115m	NBP13-05-2013	BB2	115	-70.66	-20.77	1782	12.9	34.86	2.15	9.7	5.6	2.8
ETSP_BB2_300m	NBP13-05-2013	BB2	300	-70.71	-20.53	1714	10.9	34.76	2.38	21.2	6.7	2.8
Arabian Sea OMZ regions												
AS_Stn1_150m	KNOX-8	Stn1	150	66.38	19.22	3100	18.2	35.64	0.77	16.21	5.2	2.0
AS_Stn1_175m	KNOX-8	Stn1	175	66.38	19.22	3100	16.7	35.60	0.83	15.58	4.2	2.0

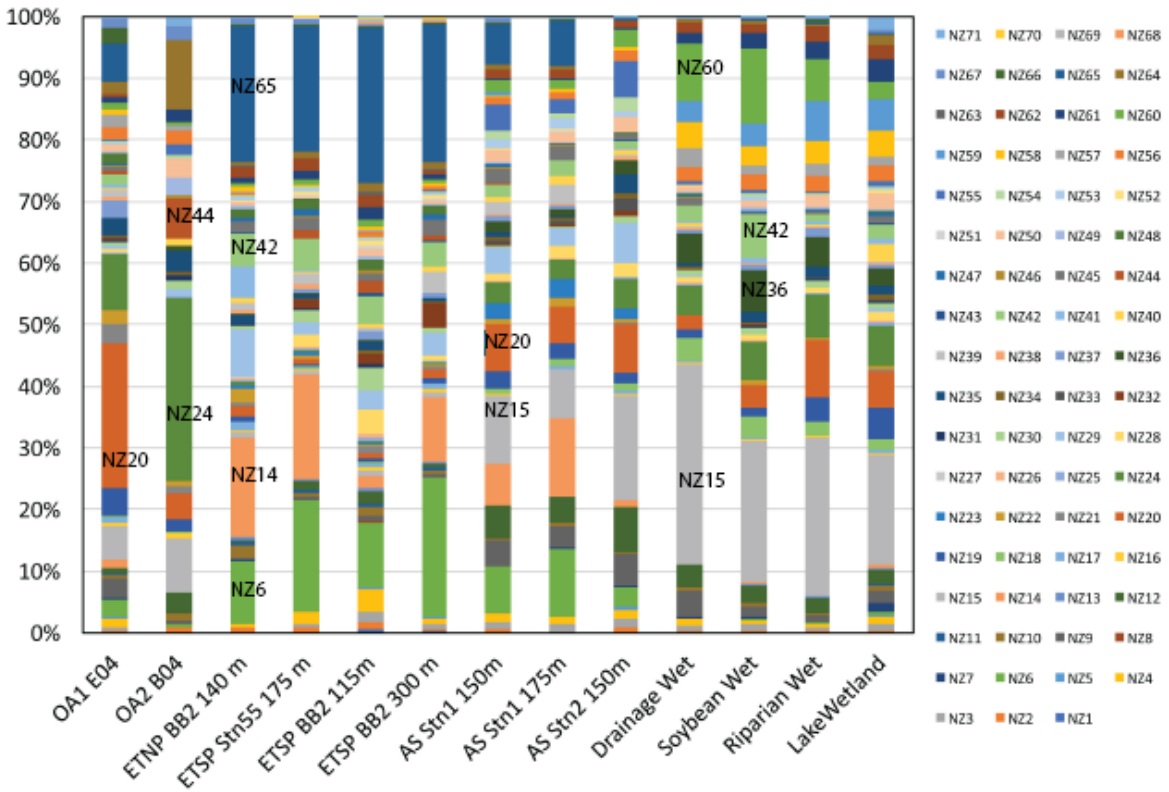
AS_Stn2_150m	KNOX-8	Stn2	150	69.00	15.00	3800	17.7	35.56	0.67	18.47	3.6	2.0
Agricultural soils												
Drainage_Wet				-90.68	33.40	NA	NA	NA	ND	3.20	ND	ND
Soybean_Wet				-90.66	33.40	NA	NA	NA	ND	25.1	ND	ND
Riparian_Wet				-90.66	33.40	NA	NA	NA	ND	ND	ND	ND
LakeWetland				-90.68	33.40	NA	NA	NA	ND	1.3	ND	ND

Table 2. Diversity measures and qPCR data for microarray samples. No = Richness (number of archetypes represented out of 71 total (NosZ) and 43 total (WNZ)); H = Shannon Entropy; E2 = Simpson Evenness. *nosZ* abundance in copies ml^{-1} (for seawater samples) or g^{-1} dry wt soil (for agricultural soil samples).

	NosZ			WNZ			Total nosZ genes	
	No	H	E2	No	H	E2	copies per ml	st. dev.
OA1_EO4	69	3.238	0.178	43	2.403	0.139	130.8	26.8
OA2_BO4	34	2.748	0.241	35	1.391	0.075	35.9	21.0
ETNP_BB2_140m	67	2.977	0.150	38	1.925	0.095	111.2	9.4
ETSP_Stn55_175m	66	2.909	0.137	34	1.959	0.112	457.4	28.9
ETSP_BB2_115m	68	3.299	0.168	42	2.343	0.117	636.4	49.0
ETSP_BB2_300m	62	2.881	0.135	36	2.535	0.173	174.9	10.4
AS_Stn1_150m	68	3.432	0.311	41	2.460	0.155	1181.7	180.1
AS_Stn1_175m	70	3.304	0.257	36	1.994	0.135	1702.6	10.8
AS_Stn2_150m	58	3.278	0.282	25	1.878	0.171	1206.4	60.0
Drainage_Wet	57	2.829	0.135	24	1.656	0.153	8.28E+07	2.90E+06

Soybean_Wet	67	3.033	0.165	36	2.116	0.143	2.02E+07	6.84E+06
Riparian_Wet	63	2.948	0.160	39	1.631	0.080	2.85E+07	5.18E+06
LakeWetland	67	3.442	0.263	36	1.942	0.122	7.98E+07	9.22E+06

Figure 1



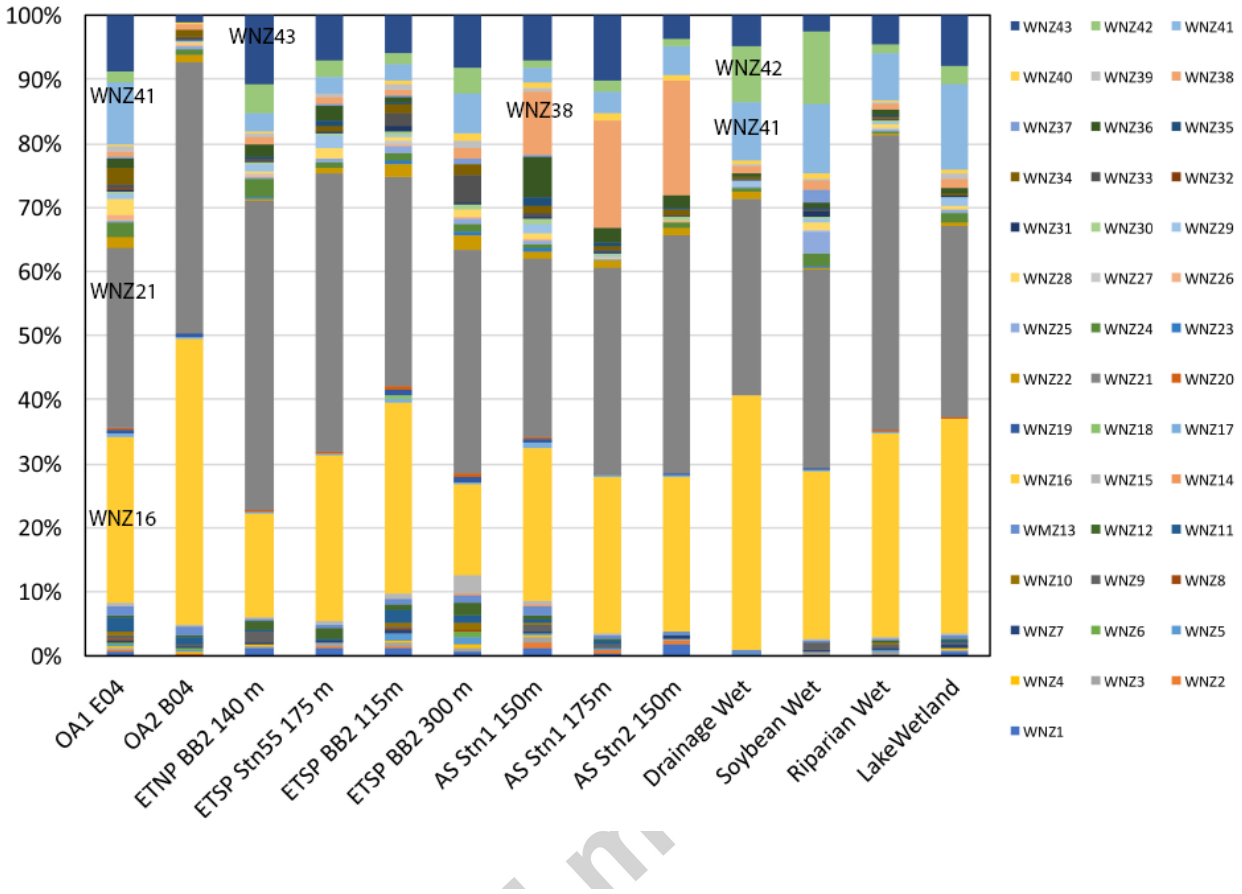

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
WNZ41
WNZ43
WNZ38
WNZ42
WNZ41
WNZ21
WNZ16
OA1 E04
OA2 B04
ETNP BB2 140 m
ETSP Stn55 175 m
ETSP BB2 115m
ETSP BB2 300 m
AS Stn1 150m
AS Stn1 175m
AS Stn2 150m
Drainage Wet
Soybean Wet
Riparian Wet
LakeWetland
WNZ43 WNZ42 WNZ41
WNZ40 WNZ39 WNZ38
WNZ37 WNZ36 WNZ35
WNZ34 WNZ33 WNZ32
WNZ31 WNZ30 WNZ29
WNZ28 WNZ27 WNZ26
WNZ25 WNZ24 WNZ23
WNZ22 WNZ21 WNZ20
WNZ19 WNZ18 WNZ17
WNZ16 WNZ15 WNZ14
WMZ13 WNZ12 WNZ11
WNZ10 WNZ9 WNZ8
WNZ7 WNZ6 WNZ5
WNZ4 WNZ3 WNZ2
WNZ1

Figure 3A

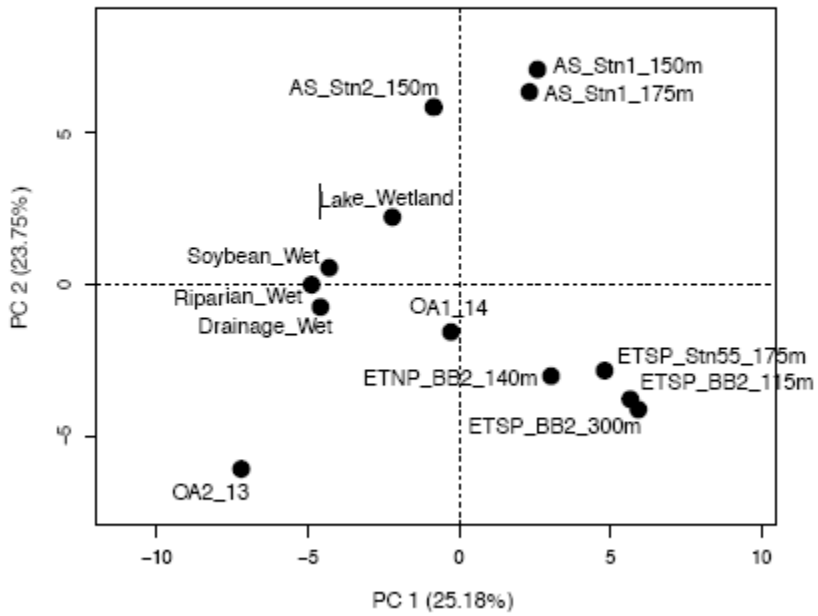

AS_Stn1_150m
AS_Stn2_150m
AS_Stn1_175m
Lake_Wetland
Soybean_Wet
Riparian_Wet
Drainage_Wet
OA1_14
ETSP_Stn55_175m
ETNP_BB2_140m
ETSP_BB2_115m
ETSP_BB2_300m
OA2_13
PC 2 (23.75%)
PC 1 (25.18%)
5
0
−5
−10
−5
0
5
10

Figure 3B

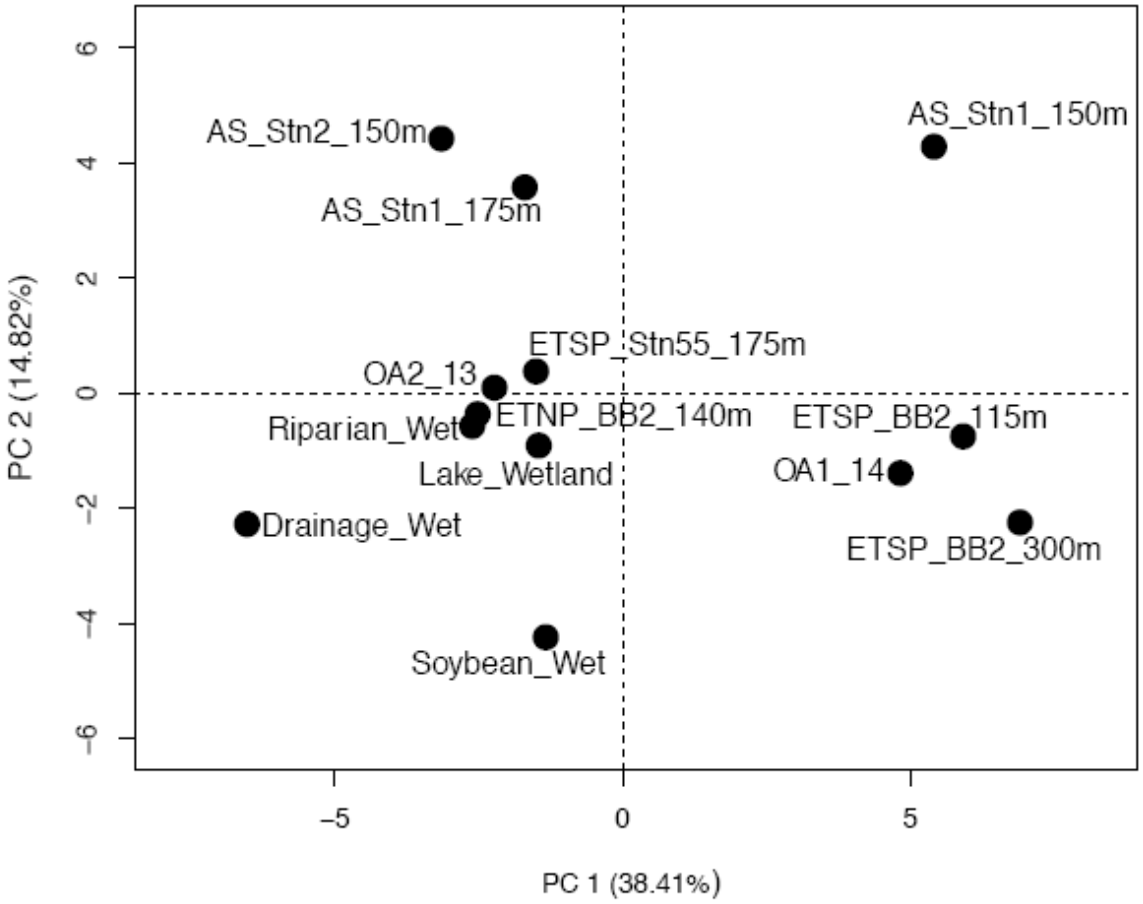


Figure 4