

Evaluating microbial community dynamics with regards to the metabolic denitrification pathway in the Saanich Inlet

Immanuel Abdi*, Melanie Law*, Arshia Shad, Jerry He, Aneesa Khan, Kristi Lichimo, Tamara Nichvolodoff, Johnston Wang, David Yeung, Mutlu Yilmaz

*These authors contributed equally

Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

SUMMARY The climate crisis has driven the expansion of oxygen minimum zones (OMZ) in oceans worldwide. OMZs disproportionately effect oceanic biological nitrogen loss, contributing to a global 50%, while only comprising of 7% of total volume. As such, it is becoming increasingly important to understand the effects of deoxygenation on marine microbial communities. These microorganisms play a major role in nutrient cycling by catalyzing geochemical reactions that maintain the levels of nutrients required to sustain marine life. However, the growing prevalence of OMZs is known to influence the microbial composition of marine environments, thereby causing an imbalance in the reactions such as ones pertaining to the nitrogen cycle. This study aims to address this concern by identifying microbial taxa at various depths of the Saanich Inlet, an OMZ in British Columbia, Canada that serves as a model ecosystem for global deoxygenation. Using the TreeSAPP metagenomic pipeline, microbial communities from the Saanich Inlet containing the denitrification genes *napA*, *narI*, *nirK*, *norB*, and *nosZ* were taxonomically classified at the phylum level. Gene abundance and alpha diversity was quantified and compared at the metagenomic and metatranscriptomic levels. Metagenomic and metatranscriptomic insights revealed that many species that were capable of executing reactions in the nitrogen cycle, often did not perform these reactions. Additionally, a regression model was created that depicts a high degree of association between diversity of taxa that contain a specific denitrification gene, differences in 'omic type, and energy of a denitrification metabolic reaction. This analysis provides insight into the roles of microbial communities involved in the denitrification pathway, and can be applied to broader OMZs to better understand the implications of deoxygenation on global nutrient cycling.

INTRODUCTION

The levels of dissolved oxygen and its dispersal patterns in the world's oceans have a direct impact on the biogeochemical flow of essential elements and nutrients in marine ecosystems, as well as on the aquatic life within them (1, 2). With global atmospheric temperatures rising due to climate change, the Earth's oceans are warming as well (3). Ongoing surveys show that this phenomenon is already beginning to lead to ocean deoxygenation events across the subarctic Pacific, an ocean basin that is bordered on one side by eastern Asia, and by western North America on the other (3). To study ocean deoxygenation and its potential effects on marine wildlife and nutrient flow, scientists often utilize oxygen minimum zones (OMZs), marine environments naturally low in dissolved oxygen levels often formed as a result of high oxygen consumption to ocean ventilation ratios (4). These OMZs provide a model ecosystem for the study of how marine microorganisms

Published Online: September 2022

Citation: Abdi I, Law M, Shad A, He J, Khan A, Lichimo K, Nichvolodoff T, Wang J, Yeung D, and Yilmaz M. 2022. Evaluating microbial community dynamics with regards to the metabolic denitrification pathway in the Saanich Inlet. UJEMI+ 8:1-19

Editor: Gara Dexter, University of British Columbia

Copyright: © 2022 Undergraduate Journal of Experimental Microbiology and Immunology.

All Rights Reserved.

Address correspondence to: Immanuel Abdi, immanuelazn@gmail.com

co-exist and evolve within deoxygenated ocean columns and the role they may play in the flow of essential elements, such as nitrogen, sulfur, and carbon in such environments (5).

One well-studied example of an OMZ is the Saanich Inlet, a seasonally anoxic fjord situated off the coast of Vancouver Island in British Columbia, Canada (2). During the spring and early summer months, the deep waters within the inlet become increasingly anoxic due to higher primary productivity in the surface waters (2). As the later summer months approach, the inlet is replenished with oxygenated water that displaces the anoxic deep waters upwards, causing changes in the redox chemistry of the environment (2). This seasonal stratification of oxygen within the Saanich Inlet makes an appropriate model ecosystem for the study of ocean deoxygenation (2). Under such low oxygen conditions in the ocean, nitrate (NO_3^-) often acts as a common electron acceptor in anaerobic microbial metabolism (6). In fact, OMZs provide 30-50% of the global fixed nitrogen (N) loss while only making up 7% of oceanic volume (6). The use of NO_3^- as an electron acceptor can lead to N_2 production and hence N-loss, via a microbe-catalyzed denitrification pathway and through intermediate compounds nitrite (NO_2^-), nitric oxide (NO), and nitrous oxide (N_2O) (6). The other potential pathway of N_2 generation in OMZs is anaerobic ammonium oxidation (anammox), a redox reaction coupling reduction of NO_2^- with the oxidation of ammonium (NH_4^+), once again catalyzed by marine microbes (6) (Figure 1).

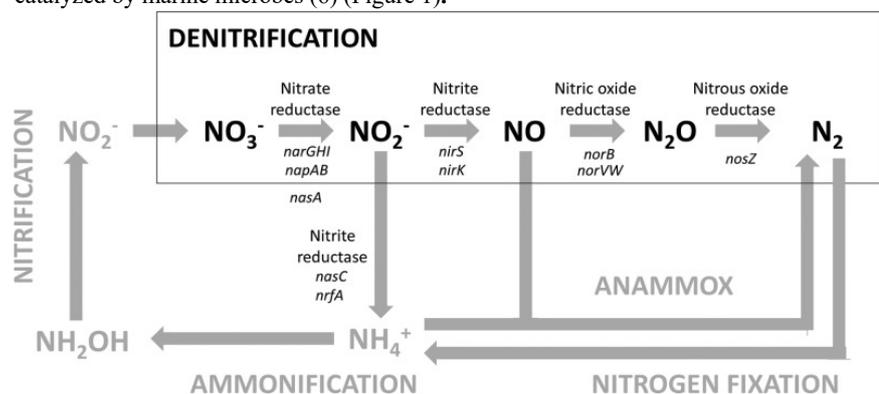


FIG. 1 Denitrification Pathway.

An illustration of the denitrification pathway as well as certain associated reaction pathways. Genes important to our analysis pipeline are labeled with their respective step in the pathway. Figure adapted from Alvarez *et al.* and with permission from the American Society of Microbiology (ASM) (43).

Both denitrification and anammox pathways have different outcomes on the carbon cycle, and the potential leakage of intermediate nitrogen species (6). Incomplete denitrification, for instance, can lead to the production and release of N_2O , a potent greenhouse gas, into the environment (6). It is becoming increasingly relevant to understand how climate change can result in ocean deoxygenation, and in turn, drive the release of greenhouse gasses into the environment. Moreover, the depth-based stratification of the microbes that catalyze these reactions may be closely linked with the distribution of molecules such as nitrate and ammonium throughout the water column (7). Hence, a major part of this understanding can be acquired through the study and analysis of such inter-microbial interactions, and intra-microbial gene abundances present in OMZs.

With these important considerations, this study aims to utilize a subset of metagenomic data collected from the Saanich Inlet to analyze the distribution of microbial organisms involved in the denitrification pathway across several depths. Each step of the denitrification pathway involves a variety of microbial genes in its catalysis. This paper will focus on *napA* and *narI* encoding the nitrate reductase enzyme, *nirK* for nitrite reductase, *norB* for nitric oxide reductase, and finally *nosZ* for nitrous oxide reductase (8). This analysis will provide insight into the distribution and potential compartmentalization of the reactions of the denitrification pathway within the context of the microbial species present in the Saanich Inlet dataset. This will be achieved utilizing taxonomic classification, phylogenetic analysis at the phylum level, quantification and comparison of gene abundance and expression, as well as analysis of alpha diversity metrics. Depth-based analysis could signify whether certain steps of the pathway are partitioned into different depth columns in the Saanich Inlet. Lastly, analysis on differential diversity in taxa that contain a denitrification gene, depending on the energy output of the metabolic reaction, will be explored. Overall, this study will provide insight into the roles of microbial groups in the nitrogen cycle within this OMZ model ecosystem, and the differences in insights given by differences in 'omics groups.

METHODS AND MATERIALS

Sample collection and sequencing overview. Water samples for the geochemical data were collected on February 18, 2006 from sample collection station S3 (48°35.500 N, 123°30.300 W) in Saanich Inlet during the 72 cruises aboard the *MSV John Strickland*, as previously described (5). To briefly reiterate, samples were collected from 7 different depths including 10, 100, 120, 135, 150, 165, and 200 m using Niskin or Go-Flow bottles for dissolved gasses and cell counts. Conductivity, temperature, and depth (CTD) instruments and the resulting data were used to determine salinity, density, and dissolved O₂. Additional geochemical data from the water samples were collected including CH₄, H₂S, N₂O, NH₄⁺, NO₂⁻, NO₃⁻ concentrations as previously described (5). Cell counts from 10 mL water were measured using flow cytometry (5).

Water for the metagenomic and metatranscriptomic data was collected at the same depths in Saanich Inlet on August 1, 2012. Biomass from the water samples was collected through a 0.22 µm Sterivex filter and total genomic DNA and RNA were extracted from the filters (9). Shotgun Illumina libraries were generated using the genomic DNA and cDNA and paired end sequenced on the Illumina HiSeq platform with 2x150bp technology (9).

Water for high resolution SSU rRNA was also filtered using 0.22 µm filters. Amplicon sequencing libraries were generated from genomic DNA targeting the V6-V8 region of the SSU rRNA gene or the V4-V5 region of the bacterial and archaeal SSU rRNA gene. Biomass samples were then further sequenced using the Illumina MiSeq platform (9). cDNA was prepared through reverse transcription of the RNA to reveal metatranscriptomic insights on our data set.

For single-cell amplified genomes (SAGs), water was collected on August 9, 2011 from Saanich Inlet at 100, 150, and 185 m depths. The water was filtered through a 40 µm mesh and sorted microbial single-cells were amplified using Multiple Displacement Amplification and sequenced using the Illumina HiSeq 2000 system (10). Following, SAGs were taxonomically characterized by screening against previously generated amplicon sequencing libraries.

Genomic assembly. Raw metagenomic and metatranscriptomic sequence reads for each depth were processed, quality controlled, and filtered using Trimmomatic (v.0.35) (11). The filtered reads were then assembled into contigs using MEGAHIT (v.1.1.3) (12). Binning and metagenome-assembled genomes (MAGs) generation was done using MetaWRAP (v.1.2.4) (13). Taxonomy was assigned using GTDB-TK v1.4.0 with the reference data version r95 (14). The 219 bins were subsequently updated with sample IDs and concatenated.

SAGs were assembled using a different methodology than MAGs. The single cell sequencing reads were filtered and trimmed using Trimmomatic v0.03 and assembled using SPAdes (v3.9.0) (15). Quality control steps were done using CheckM (v.1.0.5) and ProDeGe v2.3.0. 154 SAGs with > 50% completeness and < 10 % contamination were selected for further analysis. Taxonomic classification was done using GTDB-TK v1.4.0 and the SAGs were then updated with sample IDs and concatenated (14).

Phylogenetic analysis and abundance. Phylogenetic trees were created using each gene within the denitrification pathway selected as functional anchors. The TreeSAPP pipeline was executed to perform the analysis for both metatranscriptomic and metagenomic data in parallel (16). Firstly, reference packages from the command *treesapp create* were previously developed to form a multiple sequence alignment, hidden markov model, taxonomic lineages, and phylogenetic trees of reference amino acid sequences. Reference sequences were published and previously validated amino acid sequences to represent each denitrification gene (16). The *treesapp create* pipeline utilizes the RAXML-NG auto MRE for phylogenetic inference, with the default minimum bootstrap number of bootstraps necessary. Pre-created reference packages were compared against a small TIGRFAM seed sequence database using *treesapp purity*, to ensure that the reference packages accurately represent their targeted gene of interest (17).

To further increase the quality of our reference packages, we utilized publicly available sequences from UniProt specific to a gene of interest (18). These sequences were classified with *treesapp assign*, where homologous proteins are aligned in a profile HMM, and placed

into the reference phylogeny using EPA-NG (19). Here, sequences are then given a recommended taxonomic rank based on their lowest common ancestor of their descendants. Classified UniProt sequences were then used to update our gene's reference package. Purity was evaluated again for UniProt updates sequences.

Following, SAGs were classified using *treesepp assign*, and classified sequences were used to update our gene reference packages. Updating the gene reference packages with SAGs is vital to first introduce non-contaminated taxa to their respective phylogenetic trees. Nonetheless, the SAGs data was not investigated in downstream analysis in the context of this study. Another *assign* and *update* cycle was executed with the MAGs data. A final purity check was done to evaluate the reference package. However, updates and purity checks were unsuccessful in running *norB* files through treeSAPP due to issues with the original reference package and the original treeSAPP software. Lastly, non-binned metagenomic contigs and metatranscriptomic data were classified. Classified data was used as inputs for the *treesepp abundance* command to determine transcripts per million (TPM) abundances for each taxon and depth. Abundance data was then visualized with R (v.4.2.0), and the *dplyr* and *ggplot2* packages (20).

Alpha diversity analysis. TPM values for taxa at each gene and depth were used to represent abundance. This data was inputted to estimate Shannon alpha diversity (H') for denitrification genes at each depth (21). Shannon alpha diversity is calculated by the following equation:

$$\text{Shannon Index}(H') = - \sum_{i=1}^s p_i \ln_i$$

Here, i represents a taxa group containing a gene within a specific depth, while TPM ratio is represented by p_i . This measurement accounts for both richness and evenness within a specific sample. Diversity analysis was executed on both metagenomic contigs and metatranscriptomic data. Genes without any abundant taxa at a particular depth were filtered out, due to the redundancy to calculate diversity on these groups.

Modeling the association between Standard Gibbs free energy of a denitrification gene's metabolic reaction and the diversity of taxa that contain the gene. For each denitrification gene analyzed, standard Gibbs free energy of reaction values ($\Delta_r G^\circ$) of their gene products were obtained and used from literature (22) (Supplementary Table 1). $\Delta_r G^\circ$ values acted as an explanatory variable, with H' of taxa containing the gene acting as the response variable. To analyze variance due to the differences in abundances from metatranscriptomic and metagenomic data within a single model, a dummy variable was utilized with the baseline datatype being metagenomic data. This was used in a simple regression model minimizing for mean squared error (MSE), represented by the parametrized equation:

$$H'(\Delta_r G_f^\circ, D) = \beta_0 + \beta_1(\Delta_r G_f^\circ) + \beta_2(\Delta_r G_f^\circ)^2 + \beta_3(D) + \epsilon$$

Here, H' is the predicted Shannon diversity of the taxa that contain that gene, and D represents whether the data is from a metatranscriptomic sample. A quadratic term was utilized to lower heteroscedasticity and nonlinearity at higher $\Delta_r G^\circ$ values. Further polynomial terms were not added, since observed R-squared values did not positively change. Additionally, since our regression model does not consider any additional biological predictors, further polynomial terms would result in lowering residuals towards unexplained variance within a single sample, rather than providing true insight into the association of our predictors to H' in the population. Qualitative properties of the model were analyzed to validate quality using diagnostic plots (Figure S3).

RESULTS

***napA* expression is dominated primarily by Proteobacteria and SAR324, with expression peaking between 135 meters and 165 meters.** To determine the presence and abundance of denitrification genes across different taxonomic lineages at the phylum level,

metagenomic and metatranscriptomic samples collected at 7 different depths in the Saanich Inlet were analyzed using the TreeSAPP package for gene-centric analysis (Figure 1).

napA, responsible for the conversion of NO_3^- to NO_2^- , was detected in the metagenomes of 10 different phyla (Figure 2A). Proteobacteria represented the most abundant *napA*-containing phylum at all 7 depths, with it comprising a relative abundance minimum of 33.9% at 165m to 80.9% at 200m (Figure 2A). The next resolved phyla with the highest cumulative relative abundance were Bacteroidota and SAR324. While Proteobacteria consistently represented the most abundant *napA*-containing phyla across all depths, its relative abundance was notably lower at 100m and 165m (39.6% and 33.9%), with the Bacteroidota (27.8% and 31.7%) and SAR324 (4.6% and 10.7%) primarily compensating for the decrease in Proteobacterial abundance.

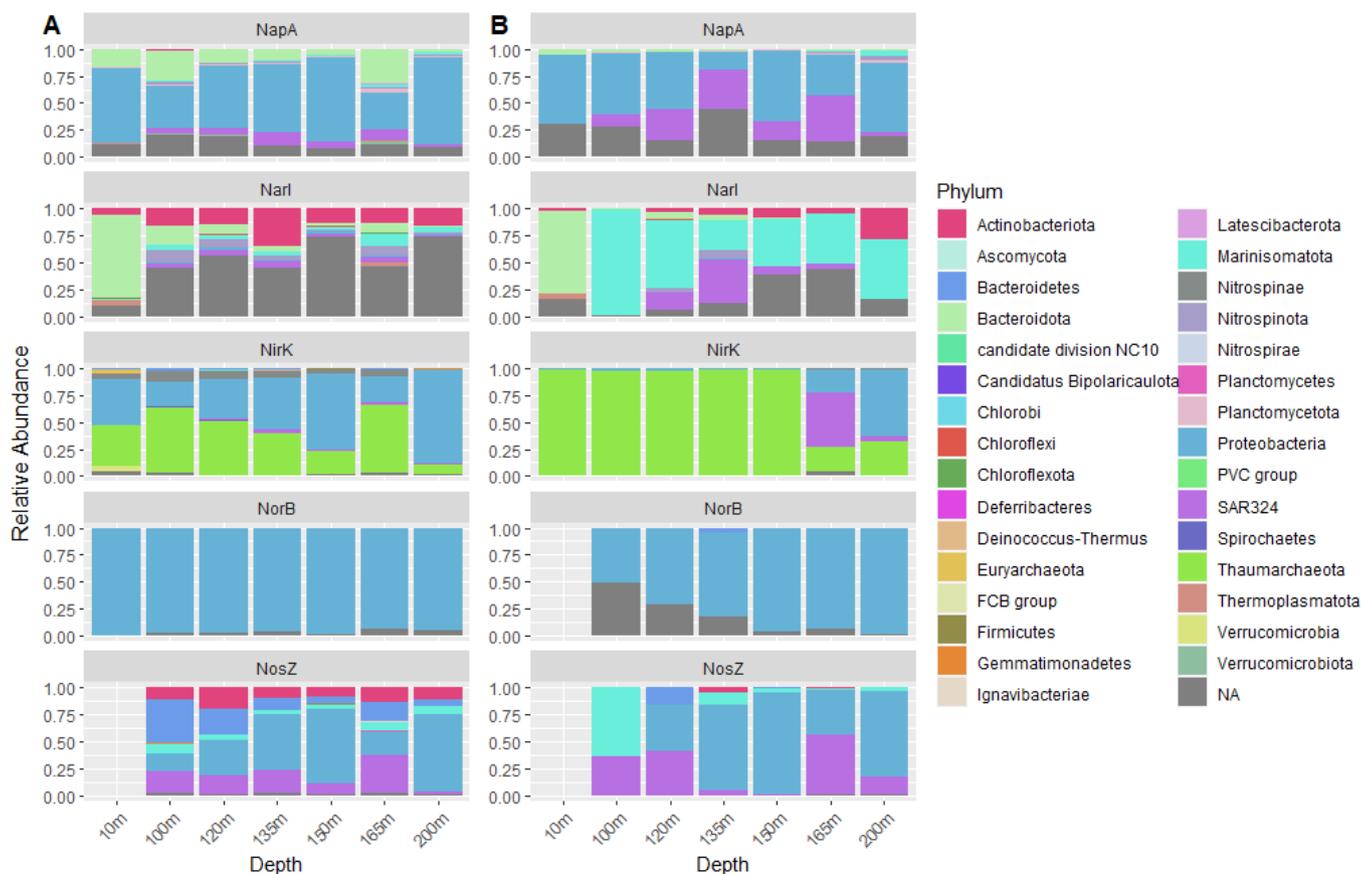


FIG. 2 Stacked bar graphs displaying the relative abundance of denitrification genes at different depths in Saanich Inlet. (A) Relative abundance of *napA*, *narI*, *nirK*, *norB*, and *nosZ* metagenomic reads, and (B) metatranscriptomic reads.

Interestingly, SAR324 *napA* expression was disproportionately greater at 135m and 165m, comprising a relatively low metagenomic abundance at these depths (Figure 2B). In fact, where SAR324 only encompassed 12.0% and 10.8% metagenomic relative abundance at 135m and 165m, SAR324 *napA* expression comprised 36.7% and 43.6% of all *napA* expression at these depths. This is in stark contrast to Proteobacteria, which despite representing the most abundant phylum at 135m (63.1%), only constituted 16.3% of the total relative *napA* expression. Additionally, while presence of Bacteroidota was generally high at most depths (as high as 31.7% at 165m), *napA* expression by this phylum was consistently low across all depths with maximum expression at 10m (5.75%) (Figure 3B).

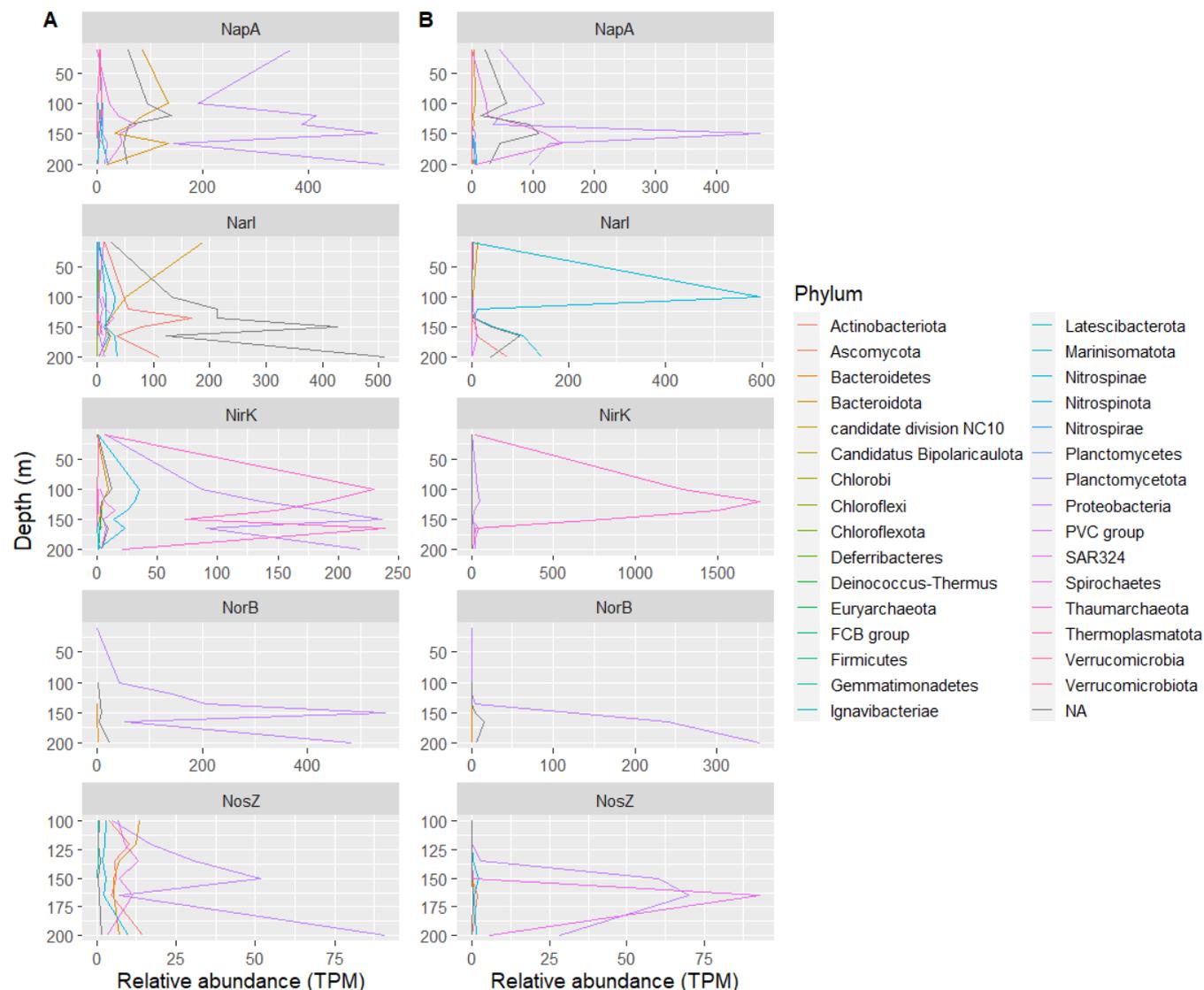


FIG. 3 Stacked line graph displaying the abundance of denitrification genes at different depths in Saanich Inlet. (A) Abundance of *napA*, *narI*, *nirK*, *norB*, and *nosZ* metagenomic reads, and (B) metatranscriptomic reads. Abundance is measured in transcripts per million (TPM), in which each metagenome sequence is normalized to the sum of abundances of all other sequences aligned to the metagenome.

Although Proteobacteria were consistently the most abundant *napA*-containing phylum at all depths, its expression of *napA* relative to other phyla varied across depths (Figure 3). Proteobacterial *napA* expression dominated from 10-120m, 150m, and 200m, with SAR324 and unresolved phyla compensating for *napA* expression at depths of 135m and 165m (Figure 3B). Curiously, while Proteobacterial *napA* expression was relatively lower at 135 and 165m, *napA* expression peaked at 150m (Figure 3B). This is in accordance with the high presence of Proteobacteria at this depth (Figure 3A). However, as Proteobacteria existed at similar levels at 200m, but did not express *napA* at nearly as high a level as at 150m, further investigation into this occurrence is needed.

***narI* expression is dominated by Marinisomatota, especially at 100 meters where *narI* is almost exclusively expressed by Marinisomatota.** Metagenomic and metatranscriptomic data revealed that 13 distinct phyla were identified to contain the *narI* gene in the Saanich Inlet dataset. In the metagenomic data, Bacteroidota and Actinobacteriota had high abundances at various depths relative to other phyla (Figure 2A). Bacteroidota had a relative abundance of 76.8% at 10 m while Actinobacteriota had a relative abundance of 34.9% at 135 m. Marinisomatota, Nitrospina, and Proteobacteria were detected at most

depth levels - however, their relative abundances are low compared to Bacteroidota and Actinobacteriota (Figure 2A). Chloroflexi, Chloroflexota, and Deferribacteres had even lower abundances across all sampled depth levels as nearly negligible components of the NarI-containing metagenome. However, a large proportion of the metagenomic phyla could not be resolved. While the phyla present in the metagenome at 10 m are for the most part known, many of the phyla at depths of 100-200m are unclassified.

In the metatranscriptomic data, Marinisomatota was abundant compared to all other phyla across all depth levels except for 10 m (Figure 2B). At 10 m, Bacteroidota was the dominant gene expressed at 76.6% relative abundance, while Marinisomatota was absent at this depth level. Descending in the water column, the metatranscriptomic data at 100 m consisted nearly entirely of Marinisomatota. The relative abundance of Marinisomatota diminishes to reach a minimum of 27.5% around 135 m before gradually increasing in abundance from 135 m to 200 m. This signifies that Marinisomatota plays a significant role in the reduction of nitrate at all depth levels, especially at 100 m.

***nirK* expression is dominated by ammonia-oxidizing Thaumarchaeota at depths above 150 meters, and denitrifying Proteobacteria and SAR324 below 150 meters.** The genomes of several taxonomic groups were found to harbor the *nirK* gene at all depths (Figure 2A). A total of 18 phyla were identified, with the two dominant phyla at all depths being Proteobacteria and Thaumarchaeota. At depths of 10 m, 100 m and 120 m, both of these *nirK*-containing phyla are almost equally as abundant. Interestingly, beyond these depths, Proteobacteria and Thaumarchaeota do not dominate equally, as shown by the alternating high peaks for these two phyla (Figure 3A). In other words, if one of these two phyla dominates at a given depth, the other phyla is less abundant at that depth, potentially indicating competition in energy resulting from utilizing the denitrification pathway. At 150 m and 200 m, Proteobacteria are 70.2% and 86.7% abundant, respectively, whereas at 165 m, Thaumarchaeota are dominant with an abundance of 63.5%. Thaumarchaeota abundance decreases dramatically at 200 m, the deepest depth sampled. At depths of 10 m to 165 m, Nitrospinae represent a small fraction, between 4.17- 9.35%, of *nirK*-containing phyla.

However, *nirK* abundance in the metagenome does not directly reflect its abundance in the metatranscriptome. At depths of 10 m to 150 m, Thaumarchaeota represents >95% of all *nirK* expression in the samples, while Proteobacteria expresses approximately the remainder (Figure 2B). In contrast, Proteobacteria represents 62% of *nirK* expression at 200 m. Interestingly, a SAR324 represented 49.8% of *nirK* expression at 165 m and 5.38% at 200 m, but is almost absent at all other depths. The dominance of *nirK* expression by Thaumarchaeota relative to all other phyla is clearly shown by the massive peak, with the greatest point of approximately 1750 TPM at a depth of 135 m (Figure 3B). Overall, although Proteobacteria and Thaumarchaeota appear to occupy the majority of *nirK*-encoding microorganisms in the metagenome together, Thaumarchaeota is the principal phylum in the metatranscriptome that expresses *nirK*.

Proteobacteria dominate in *norB* abundance at both metagenomic and metatranscriptomic levels. 7 distinct phyla were found to contain *norB* in our data across all depths (Figure 2). Proteobacteria accounted for the majority of the occurrence of *norB* for both the metagenomes and the metatranscriptomes (Figure 2). The next most abundant phyla were the accumulated unclassified phyla, followed by Bacteroidetes. Interestingly, Proteobacteria fully represented *norB* abundance at a depth of 10m. no metatranscriptomic *norB* was identified at 10 m in Proteobacteria or any other phylum. Additionally, *norB* was identified in Proteobacteria at the metagenome level at 100 m at 97% relative abundance, however the gene was only expressed at 51% relative abundance in Proteobacteria. The remaining 49% of *norB* expression occurred in unclassified phyla. The same general pattern was observed with increasing depth with slight increases in *norB* expression in Proteobacteria and decreases in expression in unclassified phyla until about 150 m depth where the expression of *norB* in Proteobacteria remained high (Figure 2). The occurrence of metagenomic *norB* remained relatively consistent in Proteobacteria with the least amount (93%) identified at 165 m. Metagenomic and metatranscriptomic *norB* was exceedingly low in the remaining classified phyla, with the most notable being 4% expression in Bacteroidetes at 135 m. The relative abundance for *norB* in TPM showed two peaks at the metagenomic level in Proteobacteria (Figure 3). The highest peak was at 150 m depth, followed by 200 m.

The metatranscriptomic relative abundance in TPM had a single peak in Proteobacteria at 200 m (Figure 3).

NosZ expression is highest in the Proteobacteria and SAR324 phyla at a depth of 165 meters.

At the phylum level, *nosZ* was detected in the metagenomes of 9 phyla (Figure 2). At a depth of 10 m, no sequences met the alignment quality cut off, hence relative abundance data were only available for depths of 100 m onward. A relative abundance value of 0 was used as a placeholder to maintain consistency across abundance plots for all genes.

The proportion of the metagenome that could not be resolved at the phylum level across all depths did not exceed 2.3 % of the metagenome and 0.3% of the metatranscriptome. At all depths except 10 m and 165m, Proteobacteria represented the phyla with the highest relative abundance for both the metatranscriptome and metagenome, with peaks at 150 m and 200m in the metagenome. At 150 m, Proteobacteria represented 93.9% of *nosZ* expression. However, at 165 m, SAR324 became the dominant phyla, representing 55.7% expression compared to 42.1% by the Proteobacteria (Figure 2). For these two phyla, trends in increases and decreases in relative abundance in the metagenome corresponded to those seen in the metatranscriptome.

At 100 m *nosZ* expression is dominated by the Marinisomatoa and SAR324, representing 64.3% and 35.7% of the metatranscriptome respectively, compared to 8.8% and 20.1% of the metagenome respectively. However, TPM counts in the metatranscriptome for all phyla were very close to or equal to 0 for all depths, except for those of the Proteobacteria and SAR324 from 120 m and greater. A maximum TPM of 91 at 165 m from the SAR324, whereas the TPM originating from the Proteobacteria was higher at the remaining depths (Figure 3B). Despite using N₂O as a substrate, *nosZ* expression appears to peak at depth of 165 m, whereas a maximum N₂O concentration of 18.1 μM is observed at 100 m (Figure S1). The concentration of N₂O decreases to 0 μM at 200 m, which was also found to correspond to a decrease in the metatranscriptomic TPM counts in both the Proteobacteria and SAR324 (Figure 3B). Of the Proteobacteria and SAR324, these peaks were attributed to one order for each phyla, SZUA-229 and SAR234, respectively (Figure S2). SAR324 could be resolved to the genus level where the sole contributing peak was from the UBA3442 genus.

Differing patterns exist in depth-based Shannon diversity values between the metagenomic and metatranscriptomic data for all five genes. As part of our overarching analysis pipeline, we looked into observing how Shannon diversity values for each gene changed as depth increased. We generated figures for both metagenomic and metatranscriptomic Shannon diversity for all five genes (Figure 4). In the metagenomic data, all five genes, *napA*, *narI*, *nirK*, *norB*, and *nosZ* peak in abundance at 165m, although to varying degrees. Furthermore, both *narI*, and *nirK* abundance experience a dip: *narI* from 10m to 150m, *nirK* from 120m to 150m, before this sudden increase in abundance at 165m. Meanwhile, both *norB* and *nosZ* experience a much smaller increase at this depth as compared to the other three genes.

In the metatranscriptomic data however, the patterns are quite different. While both *nirK* and *norB* still experience the peak in abundance at 165m, these peaks have shifted to 120m for *napA*, 10m for *narI*, and 200m for *nosZ*. Moreover, *nirK* experiences a major dip in abundance in surface waters at 10m, while a similar dip occurs for *norB* at 120m. These results signify that a considerable difference exists in depth-based changes in Shannon diversity values for all five genes when comparing metagenomic to metatranscriptomic data. Although at certain depths, a gene might be extremely abundant in the genome of the microbial community, it may not necessarily be expressed to nearly the same extent. Likewise, in certain depths, although the genomic abundance of the gene may be low, it could be expressed to a relatively high degree.

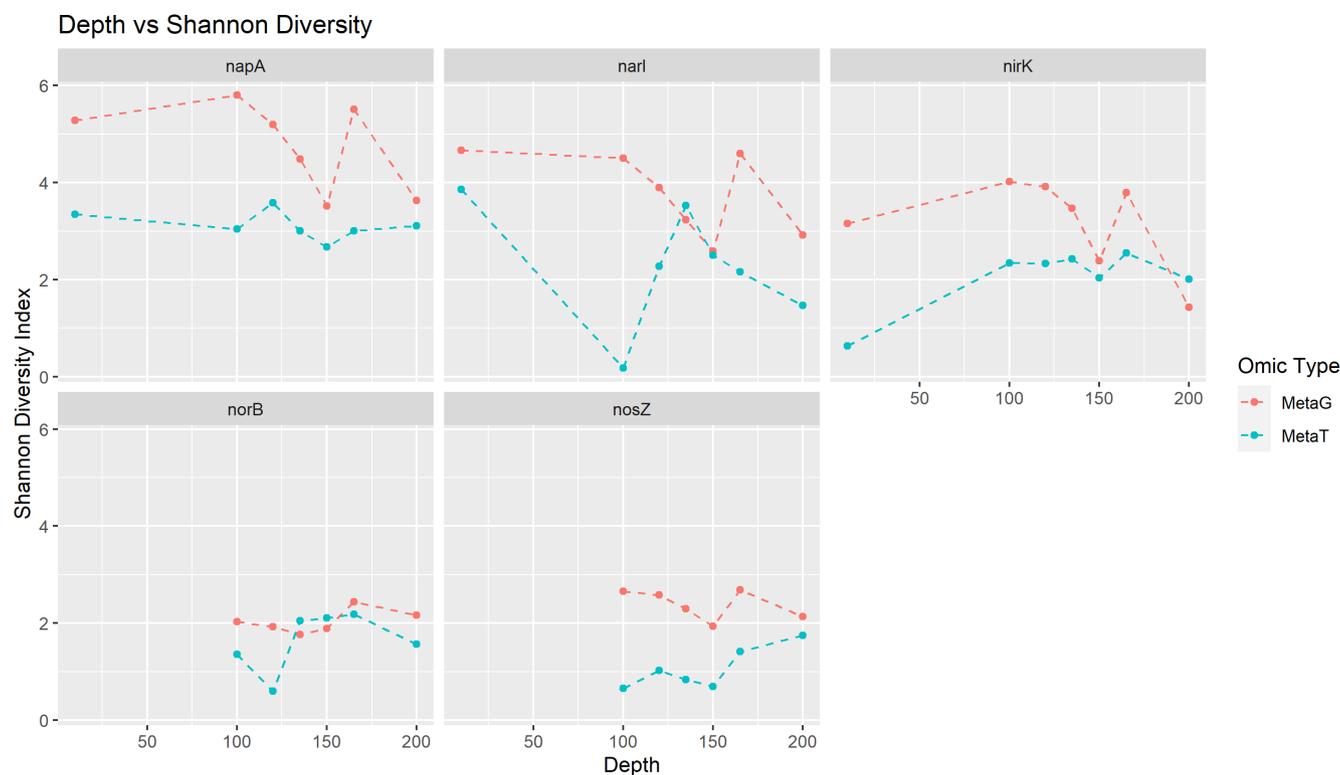


FIG. 4 Five graphs depicting metagenomic and metatranscriptomic Shannon diversity values at different depths in the Saanich Inlet water column. Data points depicted indicate sampling depths of 10m, 100m, 120m, 135m, 150m, 165m, and 200m from left to right on the x-axis. Shannon Diversity Index is depicted on the y-axis. Metagenomic and metatranscriptomic data are indicated by MetaG and MetaT respectively.

Association found between a denitrification gene product's standard Gibbs free energy of reaction and the diversity of taxa that contain the gene. Since abundance data and taxa were evaluated for each gene of the denitrification cycle, we sought to further our insights gained from the cycle by also analyzing alpha diversity and for taxa that contain the gene. Further, we sought to utilize the diversity metrics to see if they were correlated to the energy output of each gene product in the denitrification cycle. Energy output of a metabolic reaction can be simply quantified using standard Gibbs free energy of a reaction ($\Delta_r G^\circ$). Lastly, an additional interest was to find whether the data being metagenomic or metatranscriptomic had affected this relationship.

Therefore, a regression model was utilized to address all these interests all at once (Figure 5). This model considers both $\Delta_r G^\circ$ and whether the data is metatranscriptomic or metagenomic, as predictors against H' . Additionally, a quadratic $(\Delta_r G^\circ)^2$ term was included in this model, to better fit the dataset. Qualitative analysis of the residual plots and quantile-quantile plots for this model resulted in determining that there was a slight left skew of the predictor distribution, and nonmonotonic nonlinear heteroscedasticity (Figure S3). Nonetheless, these traits were negligible enough for us to continue analyzing our model.

An analysis of variance (ANOVA) for each predictor was done for this model (Table 1). Notably, both $\Delta_r G^\circ$ and $(\Delta_r G^\circ)^2$ were significant in their correlation with H' ($p=0.03321$, $p=0.00815$, respectively). $\Delta_r G^\circ$ exhibited a negative correlation to H' , while $(\Delta_r G^\circ)^2$ displayed a positive correlation to H' (-1.024×10^{-2} kcal/mol, 2.370×10^{-4} (kcal/mol) 2). Moreover, metatranscriptomic samples were predicted to result in an H' value that was lower than metagenomic samples by 1.001 ($p=2.23 \times 10^{-5}$) when holding $\Delta_r G^\circ$ constant. There was no interaction between the sample 'omic type and $\Delta_r G^\circ$ in their correlation with H' ($p=0.30228$), indicating that the curve slope between predictors and H' is due solely to the

TABLE. 1 ANOVA table to analyze the relationship between ‘omic data type, Shannon diversity, and standard Gibbs free energy of reaction for taxa that possess a denitrification gene.

Predictor	Value	Error	T-value	P-value
Y-intercept	2.285	0.1997	11.443	$<2*10^{-16}$
$\Delta_f G^\circ$	$-1.024*10^{-2}$	$4.698*10^{-3}$	-2.180	0.03321
$(\Delta_f G^\circ)^2$	$2.370*10^{-4}$	$8.659*10^{-5}$	2.737	0.00815
MetaT ‘Omic Type	$4.267*10^{-3}$	$4.101*10^{-3}$	1.04	$2.23*10^{-5}$
Interaction between ‘Omic type and $\Delta_f G^\circ$	$4.267*10^{-3}$	$4.101*10^{-3}$	1.04	0.3

gene’s value of $\Delta_f G^\circ$ and residual error. Therefore, with the ‘omic type denoted as D , the reduced model may be parameterized by the following equation:

$$H'(\Delta_f G^\circ, D) = 2.285 - 0.01024\Delta_f G^\circ + 0.000237(\Delta_f G^\circ)^2 - 1.001(D)$$

Additionally, the change in H' due to $\Delta_f G^\circ$ may be parameterized with the following differential equation:

$$\frac{\partial H}{\partial \Delta_f G^\circ} = -0.01024 + 0.000474\Delta_f G^\circ$$

Overall, this model exhibits an adjusted R-squared value of 0.6665, indicating a strong association of the predictors to H' ($p=1.535*10^{-14}$) (Figure 5). In other words, approximately 66.65% of the variability in H' of taxa who contain a denitrification gene in a sample can be predicted through knowing $\Delta_f G^\circ$ of that gene and what type of ‘omic data the sample is.

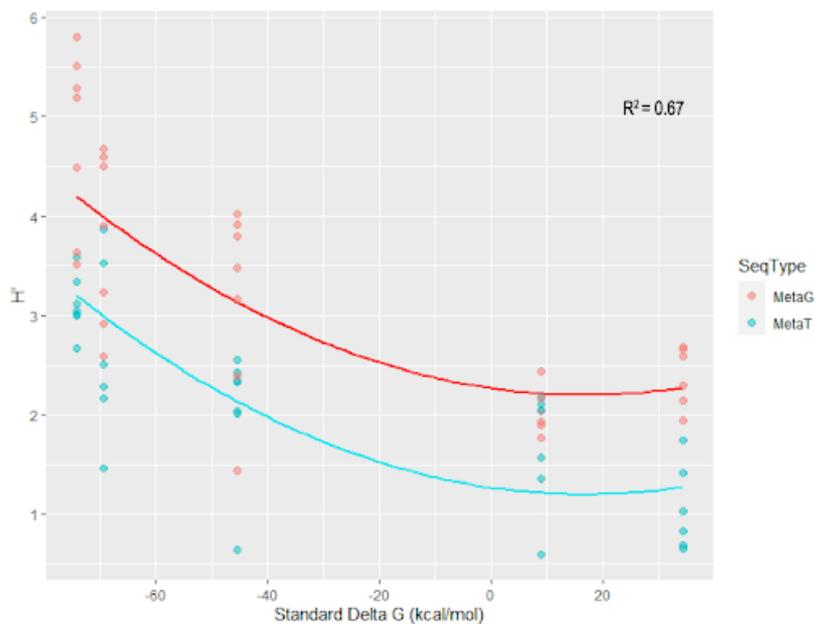


FIG. 5 Scatter plot depicting a strong association between standard Gibbs free energy of reaction of denitrification genes, and Shannon alpha diversity in taxa that contain that gene. H' here refers to the predicted Shannon diversity of the taxa that contain each gene. From left to right, the values refer to *napA*, *narI*, *nirK*, *norB*, and *nosZ*. Exact $\Delta_f G^\circ$ (kcal/mol) values for each gene can be found in Table S1. Metagenomic and metatranscriptomic data are indicated by MetaG and MetaT respectively.

DISCUSSION

***napA* analysis highlights the lack of a discernible expression pattern with regard to oxygen concentration.** NapA is a periplasmic nitrate reductase that contains a hexa-coordinated molybdopterin and iron-sulfur (4Fe-4S) cofactors (23). As its name suggests, NapA resides in the periplasmic space of a variety of microbes and is one of a plethora of enzymes responsible for the first step of denitrification, the reduction of NO_3^- to NO_2^- . Proteobacterial expression of *napA* displays an opposite trend with NO_3^- concentration patterns in the water column (Figure 3B). This observation is also supported by previous findings of Proteobacterial *napA* expression in stratified marine environments (24). Additionally, the increased Proteobacterial *napA* expression occurs at an anoxic depth in the

Saanich Inlet, the first sampling depth in which oxygen is not present at detectable levels (Figure S1). This finding is somewhat unexpected as *napA* is generally regarded as a marker for aerobic nitrate reduction, where other nitrate reductases such as NarI or NarG are considered anaerobic markers for this reaction. Indeed, one study observed a higher abundance of *napA* expression at oxic and hypoxic levels than anoxic regions in lake surface sediment (25). Curiously, different studies have not necessarily observed the same trend. Marchant *et al.* detected no correlation between oxygen concentration and *napA* expression in coastal sediments (26). Similarly, a study conducted investigating denitrifiers of the Saanich Inlet water column also did not observe *napA* expression as a function of oxygen concentration, which is precisely what is seen in this study (27).

Prevalent Proteobacterial *napA* expression is expected and may be linked to the presence of taxa such as SUP05. SUP05 is a largely uncultivated genus of Gammaproteobacteria and has been suggested to catalyze multiple denitrification reactions, ranging from nitrate reduction to complete denitrification (6). Moreover, SUP05 has frequently been described as an abundant member of the microbial community in oxygen minimum zones (28). Specific to nitrate reduction, evidence suggests that nitrate reduction is heavily conserved in this genus (28). Moreover, SUP05 has been described to be most abundant in areas of high NO_2^- , consistent with our metagenomic and metatranscriptomic data indicating high *napA* presence and expression at the second nitrite maximum (150m) (28). SAR324 was identified as the next highest *napA*-expressing phylum at 165m. SAR324 is an uncultivated phyla (previously a member of the Deltaproteobacteria) and is known to be ubiquitous in the oceanic water column, previously shown to possess denitrifying metabolic potential (29). Further investigation at lower taxonomic levels is required to elucidate the specific members of Proteobacteria and SAR324 responsible for *napA* expression.

At a depth of 200m, Proteobacteria are shown to express *napA* (Figure 3B). The presence of Proteobacterial *napA* expression at this depth is altogether quite curious. As noted above, *napA* expression commonly corresponds to regions of high NO_2^- and moderate NO_3^- concentrations. However, at 200m, there is little to no detectable amounts of nitric or nitrous oxide. A plausible explanation for this observation may be due to the processing workflow utilized for this gene. *napA* is highly homologous to other protein families and analysis with the TreeSAPP pipeline may result in the classification of false positives. As some genes present in the NapA reference package are classified as sulfur-oxide reductases, it remains possible that incorrect classification of sulfur oxide reductases as NapA homologs resulted in observed *napA* expression at 200 m, especially considering that H_2S concentrations are highest at 200 m. Overall, *napA* abundance and expression largely supports existing evidence and suggests that NapA may not strictly serve as an aerobic nitrate reductase. Regardless, deeper investigation into *napA*-expressing lineages and further modification of the NapA reference package to mitigate false positives is warranted.

***narI* metagenomic and metatranscriptomic abundance differ significantly in Marinisomatota** Fulfilling a functional niche similar to *napA*, *narI* encodes the nitrate reductase gamma subunit and is involved in the reduction of NO_3^- to NO_2^- . Our analysis indicated that the phylum Marinisomatota was a key contributor to *narI* expression at depths of 100 m and below (Figure 3B). Marinisomatota were formerly referred to as Marinimicrobia, SAR406, and Marine group A (30). They are bacteria that inhabit OMZs with a particularly high abundance above the mesopelagic zone at 150-200 m in both the Pacific and Atlantic Oceans (30, 31). This is consistent with our analysis of the Saanich Inlet depth column, which indicates a peak of Marinisomatota metatranscriptomic abundance for around 165 m (Figure 3B). Furthermore, literature suggests that Marinisomatota prefer environments with a low oxygen concentration (32–34). The Saanich Inlet data corresponds particularly well to this behavior. With increasing depth, the oxygen concentration decreases (Figure 2), and this has an inverse correlation to the relative abundance of Marinisomatota in the metatranscriptome (Figure 3B). Notably, this rise in relative abundance is not mirrored in the metagenome, and we expand on this further below. Marinisomatota is an important contributor to the production of OMZ-associated nitrite maxima, indirectly affecting the activity of other nitrogen cycling processes such as anammox (31). This can be seen in our analyses, which indicates a high relative contribution of Marinisomatota to nitrate reduction

through *narI* expression at all depths of the water column, but especially at 100 m. At 100 m, nearly all *narI* expression is accounted for by Marinisomatota (Figure 3B).

Interestingly, there is a notable discrepancy between the metagenomic and metatranscriptomic abundance of Marinisomatota. At 100 m, Marinisomatota has a relative abundance of 5.3% in the metagenome (Figure 2A), but a relative abundance of around 98.4% in the metatranscriptome (Figure 2B). This suggests that at this depth level, organisms in the Marinisomatota phylum may express a disproportionately large amount of *narI*. Conversely, organisms in other phyla may be expressing negligible levels of *narI*, causing low expression of *narI* by Marinisomatota to overshadow *narI* expression by other phyla. Examining the data, we find that the metatranscriptomic abundance of Marinisomatota is exceedingly high at 594.5 TPM, followed only by Bacteriodota at 4.0 TPM and SAR324 at 2.1 TPM, indicating that Marinisomatota indeed expresses a disproportionately large amount of *narI* at this depth. This provides insight into the high relative contribution of Marinisomatota to nitrate reduction in the nitrogen cycle, and sets the foundation for future analysis and experimentation to identify why Marinisomatota metagenomic and metatranscriptomic abundance differ so significantly.

***nirK* expression is dominated by ammonia-oxidizing Thaumarchaeota at depths above 150 meters, and denitrifying Proteobacteria and SAR324 below 150 meters.** The *nirK* gene encodes a copper-dependent nitrate reductase that catalyzes the conversion of NO_2^- to nitric oxide (NO) gas (35). The gene is often quantified in environmental samples to estimate the frequency of denitrification reactions (35). Microorganisms expressing *nirK* have been previously identified in OMZs owing to their metabolic diversity (36). The present study has identified the two predominant *nirK*-expressing phyla in the Saanich Inlet, Thaumarchaeota and Proteobacteria, which have been previously found to contain *nirK* in their genomes (37, 38). Thaumarchaeota has an important role in the nitrification cycle, in which these members oxidize ammonia (NH_3) to NO_2^- , and have been known to co-exist with denitrifying Proteobacteria in water columns with low O_2 concentration (39). Notably, the current study shows that *nirK*-containing Proteobacteria are highly abundant at a depth of 150 m (Figure 2, 3), where NO_2^- concentration is the highest in the Saanich Inlet (Figure 2), then abundance decreases sharply at 165 m, where NO_2^- concentration decreases. These findings are consistent with existing literature that determined that *nirK* gene abundance was positively correlated with NO_2^- concentration (40). These findings support the theory that Proteobacteria are most abundant in depths with greater NO_2^- , but not near the surface of the inlet where aerobic species dominate. Conversely, Thaumarchaeota appears to dominate at 165 m, where NO_2^- concentration decreases. Given that Thaumarchaeota is primarily known to express the ammonia monooxygenase operon *amoABC* for ammonia oxidation (41), it is surprising that this phylum contributes to >95% of *nirK* expression at depths above 150 m. Nonetheless, previous studies have found that *amoABC* transcription is highly correlated with *nirK* transcription (41). NH_4^+ concentration is the lowest at depths above 150 m in the Saanich Inlet (Figure 2), which could suggest that Thaumarchaeota expresses *nirK* at extremely high levels when NH_4^+ concentration is low. Here, instead of NH_4^+ oxidation, Thaumarchaeota could be using NO_2^- as an electron donor to compensate for the lack of NH_4^+ sources. Furthermore, *nirK* expression by Thaumarchaeota is almost completely decreased at depths greater than 150 m, which supports previous findings of its aerobic lifestyle (42). Lastly, SAR324 was found to account for approximately 50% of *nirK* expression at a depth of 165 m. Although not very abundant in the metagenome or expressed at very high levels, *nirK* expression by SAR324 is represented by a small peak at 165 m (Figure 3B) where *nirK* expression by Thaumarchaeota is almost non-existent. This provides insight into the metabolic versatility of SAR324, which has been found to carry out reactions to support alkane oxidation, heterotrophic and lithotrophic lifestyles in OMZs (29). Taken together, *nirK* abundance and expression varies with depth in the Saanich Inlet, and is carried out by metabolically diverse taxonomic groups.

***nosZ* expression patterns signify the potential role of other molecular factors beyond N_2O .** Despite the identification of *nosZ* in the metagenomes of 9 taxonomic groups resolved at the phylum level, its expression was found to peak at a depth of 165m in both the Proteobacteria and SAR324, with expression being highest in the latter (Figure 3B). Such findings suggest that peak denitrification occurs at this depth, reinforcing previous studies in

the Saanich inlet (6). Low TPM counts in the remaining phyla suggest that the conversion of N_2O to N_2 in the Saanich Inlet may be largely responsible from the metabolic activities of the Proteobacteria and SAR324. This could additionally suggest that these phyla are the primary specialists that control this niche. Despite remaining phyla having comparative relative abundance levels of *nosZ* in their metagenomes to the SAR324, their substantially lower expression levels may be indicative of tight regulation and lack of signaling to initiate *nosZ* transcription. Similarly, there is no discernible trend between relative abundance levels of metagenomic and metatranscriptomic data across the phyla, suggesting that *nosZ* expression is likely not correlated with its metagenomic abundance across depth. Further resolution of these dominant phyla at higher taxonomic levels revealed that a single group was contributing to the metatranscriptomic peaks produced at 165 m, which were SZUA-229 and SAR234, at the order and genus level for Proteobacteria and SAR234 phyla respectively. Lack of available literature for both indicates that these groups have only been identified metagenomically.

In the denitrification cycle, *nosZ* is responsible for the conversion of N_2O to N_2 (43). However, the depths at which N_2O concentration and *nosZ* expression is greatest differs by 65 m in the water column. One may initially suspect that trends in *nosZ* expression levels would overlap with those of N_2O concentration, however this offset could suggest that other molecular factors could have a larger influence on *nosZ* expression. Although N_2O is required for the activation of *nosZ*, previous studies have found that the presence of NO strongly amplifies *nosZ* activation signals (44). It is uncertain whether NO may be a requirement for the Proteobacteria and SAR324 bacteria that are responsible for the majority of *nosZ* expression found at 165 m. NO concentrations at various depths in the Saanich Inlet were not available to make a comparison between the depth at which maximum NO concentrations occur and that of *nosZ* expression. At a depth of 150 m, water conditions become anoxic, suggesting that despite a higher concentration of N_2O at shallower depths, anoxic conditions are required for the transformation of N_2O to N_2 . This is supported by previous work that found that SAR324 abundance is correlated, with greater abundance occurring in OMZs (34).

Evidence of high proteobacterial involvement in multiple denitrification steps at both genomic and transcriptomic levels. Our analysis of the relative abundance of the five genes of interest at different depths highlight the significant presence of these genes in the Proteobacteria phylum. More specifically, at the genome level, Proteobacteria appear to be the dominant phylum for *napA*, *norB*, and *nosZ* at all depth levels analyzed (Figure 3A). *nirK* is most abundantly present genomically in both Proteobacteria and Thaumarchaeota, although they never appear to dominate at the same depths (Figure 3A). When it comes to gene expression, the trends vary slightly. Most interestingly, Thaumarchaeota dominantly expresses *nirK*, with very little to none proteobacterial involvement in expression (Figure 2B). On the other hand, *nosZ* expression appears to be mediated by both Proteobacteria and the SAR324 clade, with the two co-dominating at 165 meters (Figure 3B). *napA* and *norB* expression seem to continue to be carried out primarily by proteobacterial species, somewhat mimicking the metagenomic data (Figure 3B).

As can be seen, proteobacteria seem to be heavily involved in most steps of the denitrification pathway. They carry all four *napA*, *nirK*, *norB*, and *nosZ* genes, and dominantly express *napA*, *norB*, and *nosZ* at most depths. Although the SAR324 clade co-dominates in *nosZ* expression, especially at 165m, it is important to note that this clade is often considered a delta-proteobacterial group, and hence a phylogenetically distinct proteobacteria lineage (45, 46). Furthermore, although Proteobacteria do not appear involved in *nirK* expression, previous studies of the Saanich Inlet point to dominant proteobacterial expression of *nirS*, a different nitrite reductase gene involved in the denitrification pathway (47). In the broader literature, Proteobacteria has often been a phylum associated with denitrification, in both aquatic and terrestrial ecosystems (48–50). Alphaproteobacteria isolated from OMZs within the Arabian Sea have been found to express *nosZ* and significantly participate in nitrous oxide reduction as a result (50). Other studies into the OMZs of the Indian Ocean, including the Arabian Sea, have found high representation of proteobacterial clades, such as delta and gammaproteobacteria, in these environments (51). Similarly, our own analysis could be underlying the substantial role of proteobacteria in the denitrification

pathways present within OMZs, and providing a potential avenue for further exploration of microbial community dynamics in the Saanich Inlet.

The lack of expression of *NorB* and *NosZ* at relatively low depths points to the potential stratification of the denitrification pathway. Clearly reflected in our data is the idea that the abundance of genes in the denitrification pathway varies with depth (Figure 2, 3). More specifically, from our analysis of the relative abundance of *NorB* and *NosZ*, we can see that both genes are undetectable at a depth of 10m for the metatranscriptomic (MetaT) condition, and *NosZ* is also undetectable at 10m in the metagenomic (MetaG) condition (Figure 2). This finding is also observed in our relative abundance (TPM) graphs, where we see a very low relative abundance value for both *NorB* and *NosZ* until approximately a depth of 125m, after which we observe a sharp increase in the relative abundance of both genes (Figure 3B). The observation that *NorB* and *NosZ* are expressed in greater concentrations at greater depths is further supported by the peak N_2O concentration seen at a depth of 100m, followed by a gradual decrease in concentration with increasing depth (Figure S1). The peak concentration at 100m suggests that *NorB* is actively catalyzing the production of N_2O , and the subsequent decrease could be attributed partially to the activity of *NosZ*, which uses N_2O as a substrate to produce N_2 .

One possible reason for this stratification of the genes in the denitrification pathway is the changes in O_2 concentrations with depth. O_2 was found to decrease in concentration until reaching a value of close to 0 μM at approximately 150m, which is also close to the depth at which *NorB* and *NosZ* relative abundance begins to increase (Figure S1, 3). This suggests that very low oxygen concentration is positively correlated with an increase in *norB* and *nosZ* expression. In their 1989 study, Bonin *et al.* found that nitrate reductase (like *NarI* and *NapA*) and nitrite reductase (*NirK*) are less sensitive toward oxygen than nitrous oxide reductase (*NosZ*) in *Pseudomonas nautica* (52). They found that nitrate and nitrite reductase activity was completely blocked at an oxygen concentration greater than 4.05 and 2.15 mg/L, respectively, while nitrous oxide reductase was blocked at only 0.25 mg/L (52). Nitric oxide reductase (*NorB*) was also found to only be active at very low oxygen concentrations (25, 53). The reduced oxygen sensitivity of nitrate and nitrite reductases would explain why we observe a greater relative abundance in *NarI*, *NapA* and *NirK* at higher depths as compared to the more oxygen sensitive *NorB* and *NosZ*.

The increase in denitrification gene diversity at 165m highlights the potential for a transition zone between 150 and 165 meters. Previous studies on the microbial community dynamics of the Saanich Inlet have denoted strong differences in community compositions between surface waters (~10m) and the deep waters below 100m (6). Interestingly, geochemical information collected from the Saanich Inlet often point to the 100m point as a sort of hypoxic boundary, below which O_2 concentrations rapidly decrease (5). Strikingly, our metagenomic data with regard to denitrification genes *napA*, *narI*, *nirK*, *norB*, and *nosZ* shows a noticeable increase in gene diversity at 165 meters (Figure 4). This increase could very well be connected to the aforementioned sharp decline in oxygen concentration often observed below 100m, leading to a need for alternative electron acceptors such as the intermediates involved in the denitrification pathway. However, given that our data does not include sampling between the depths of 150m and 165m, we may be missing an important transition point in the diversity of genes involved in denitrification. For future studies with access to samples collected from depths between 150m and 165m, it would be worthwhile to analyze whether this increase in genomic abundance is as sharp as it may seem, or if it is a more gradual increase depicting a transition point after 150m.

With this said, there is considerable difference between the metagenomic and metatranscriptomic data in our depth-based analysis of gene abundance. Although all genes experience an increase in abundance at 165m, the same can not be said of expression patterns for all genes (Figure 4). Metatranscriptomic abundance peaks at 120m for *napA*, 10m for *narI*, 165m for *nirK*, 165m for *norB*, and 200m for *nosZ*. It is interesting to point out that, as the denitrification pathway advances from left to right, and from nitrate oxide to nitrogen gas, it appears that the metatranscriptomic abundance of the genes associated with each step peaks at lower and lower depths. *napA* and *narI* are both involved in the nitrate oxide reductase step, and peak at higher depths, while *nirK* and *norB*, involved with the second and third step respectively, both peak at a deeper level, 165m. And finally, *nosZ*, involved with the final

step of denitrification, peaks at the deepest level of the Inlet that data was available for, 200m. These findings could signify the idea that the intermediate from each previous step of denitrification partitions deeper and deeper down the water column as the step is catalyzed, leading to the need for the next step of the pathway to be carried out at a deeper level than the one before it. Indeed, such a hypothesis could be rate-limited by diffusion of nutrients from one oceanic depth to another.

Energy yields of metabolic reactions available provide a driving force in determining community taxonomic abundances. In an intuitive sense, the number of different taxa that possess a gene should relate to how much energy is provided by said gene. The more energy a reaction provides, the higher propensity of a taxa to want to participate in that metabolic reaction as well. Indeed, this assumes that the metabolic products of that reaction are non-toxic to the taxa itself, as well as a non-zero concentration of the reactants at any point in time. We see this trend being accurately reflected within our model prediction for Shannon alpha diversity, with regards to the quadratic and linear terms of standard Gibbs free energy of reaction for each denitrification gene; the model predicts a strong negative association of H' associated with an increase of $\Delta_r G$ in a reaction. Interestingly however, once the $\Delta_r G$ of a reaction has required a sufficient input of energy, there is a gradual leveling in the predicted H' in taxa containing the gene.

One potential explanation as to why taxa possessing a non-energy yielding gene product continue to do so, is that once a reaction is no longer providing energy, different confounding variables on why that reaction needs to take place may have a bigger effect than with higher energy producing reactions. For instance, at least a couple representative taxa in a community may be required to execute a metabolic reaction to complete a pathway. Such has been well established in microbial-community spatial dynamics, where a single taxon acts for the greater good of continuing the flow of energy in an ecosystem (54, 55). End products of a pathway could potentially be required to drive another metabolic reaction to completion. In our context, metabolic pathways regarding nitrogen do not regularly begin at nitric oxide or nitrous oxide as a starting metabolic reactant (56). *NorB* and *NosZ* reactions cap off the denitrification pathway to completion, to yield biologically inert nitrogen gas (Table 1). Following, nitrogen gas is utilized in nitrogen fixation in anoxic environments to create ammonia. Lastly, ammonia is used in the nitrification pathway to create nitrite and nitrate. As previously explained, these are the initial reactants in the denitrification pathway, and yield high degrees of energy when acting as terminal electron acceptors (TEA) relative to compounds further down in the denitrification pathway.

Differential abundance between metatranscriptomic and metagenomic alpha diversity represents the dynamic nutritional conditions within the Saanich inlet. Our model signifies that there is a direct decrease in Shannon diversity when considering metatranscriptomic data in relation to metagenomic data. Indeed, this makes sense--a taxa is required to contain a gene to be able to transcribe it, indicating that the Shannon diversity in the metatranscriptomic data must be lesser than or equal to the value in the corresponding metagenomic data for a gene. However, it is notable how large the scale of decreases is in Shannon diversity between the two 'omic groups.

One explanation for this effect size is that gene products in the denitrification pathway are not constitutively expressed at all times, and some taxa may require distinct environmental signals to begin transcribing these respective genes (57). For instance, nitrate and nitric oxide are also signaling molecules to induct N oxide metabolizing enzymes (57). This notion is especially important in the context of the Saanich Inlet due to the seasonality of the geographical location. During the spring and summer months, there are high levels of primary productivity and limited vertical mixing below the glacial sill. In these conditions, there are clear minimal oxygen levels at deeper depths. However, neap tidal flows in the summer and fall are important for bringing in denser water, resulting in vertical mixing and creating oxic conditions at the deeper depths (5). As a result, microbial communities at all depths will never reach a long-term steady state equilibrium in nutrients within the Saanich inlet. This seasonality therefore has a potential to drive the creation of microbial redundancies in their genomes to be able to survive in various different nutritional conditions. This counteracts the microbial propensity to reduce their genome size for increasing growth rate. Such redundancies are particularly important within OMZs, and could represent similar microbial

trends in other global seasonality-driven OMZs. This signifies the importance of investigating both metatranscriptomic and metagenomic data in further oceanic gene-centric analysis.

Limitations There is a somewhat low statistical power in our analysis due to low sample size, as it focuses on data from a single cruise (SI072) representing one temporal level. As a result, our findings are a very small glimpse into the geochemical conditions and microbial community dynamics of the Saanich Inlet. Utilizing multiple time points would have allowed us to assess temporal correlations between nutrients, taxa, and diversity. Furthermore, the TreeSAPP pipeline has an inherent reliance on reference sequences of previously classified taxonomic groups. Hence, the reference packages constructed for each gene of interest may have a potential bias towards more well-studied microbial groups, such as the proteobacteria phylum. This bias could lead to a lack of identification of other denitrifying microorganisms present in the dataset—especially those less well-documented.

It is also important to mention that the *norB* reference package failed to update with SAGs and MAGs, which could be contributing to the larger number of unclassified phyla, labeled NA, present specifically in the metatranscriptomic relative abundance plot (Figure 3B). Additionally, purity analysis of NorB ran into an error, even with the non-updated reference package, indicating issues with the program or baseline reference package. Due to time limitations, we were unable to develop a new reference package to fix this issue. All other reference packages were successfully updated, and hence the issue most likely does not lie with the code, and rather with the *norB* reference sequences specifically.

Our model itself also suffers from various limitations. Firstly, our data was calculated using standard Gibbs free energy of reaction values, kept constant at every single depth and gene. However, this is evidently a large source of residual error, because pressure, and temperature vary extremely widely across different depths in the Saanich water column. Additionally, Shannon diversity does not utilize information on evolutionary distance between species. Unfortunately, with the way the analysis is structured, using denitrification genes as functional gene anchors within treesapp did not allow us to use evolutionary distances through an output tree network file. Additionally, work with quadratic terms prevents extrapolation in trends outside the range of our points. Lastly, prediction of Shannon diversity in the context of Δ_rG° should be done with every new metabolic pathway to ensure trends do not change.

Future Directions Future studies can attempt to carry out similar analysis with a much larger sample size, ideally from multiple cruises from the Saanich Inlet. An increase in sample size will allow for a study population more representative of the microbial community within the Saanich Inlet. In addition, the global generalizations of the results can be examined by replicating the study against other OMZs to correlate. One such OMZ is the Gotland Basin in between Sweden and the Baltic countries, which additionally has a depth similar to the Saanich Inlet with 249 meters (58). Moreover, assessing other similar genes in each step of the pathway can help further examine microbial organisms' involvement at different depths. Such as *narG*, *narH*, *narJ*, and *napB* for the nitrate reductase step, *nirS* for the nitrite reductase step, and *norV* and *norW* for the nitrous oxide reductase step (9). This information would provide more accurate data on the pathway as a whole at the differing depths while also identifying phylums with different genetic preference for each step.

Additionally, it would be very interesting to see whether the same trend between reaction energy and H° occurs in other metabolic pathways as well. In environments where TEAs yield much more energy (such as using oxygen gas), we would expect taxa that contain that pathway's genes to lack the high degree of evolutionary pressure for large changes, and therefore have a much less pronounced slope. Lastly, analysis should be done with Gibbs free energy taking into account temperature and pressure.

ACKNOWLEDGEMENTS

We would like to thank Dr. Steven Hallam, Dr. Connor Morgan-Lang, and Tony Tang for their continuous guidance and support throughout the duration of the project, from design, to execution, and to review. We would also like to thank the UBC Dept. of Microbiology and Immunology for access to space, guidance, funding, and technical resources. We would also like to thank two anonymous reviewers for constructive feedback on this manuscript.

CONTRIBUTIONS

I.A., and M.L. conceptualized the project. I.A., M.L. J.H., K.L., T.N., J.W., and D.Y. contributed to taxonomic abundance analysis. I.A. and M.L. executed diversity analysis. I.A. executed model creation. I.A., M.L., A.S., J.H., A.K., K.L., T.N., J.W., and D.Y. contributed to the writing of the manuscript. All authors contributed to editing the manuscript. All authors have read and agreed to the published version of the manuscript.

REFERENCES

1. **Heinze C, Blenckner T, Martins H, Rusiecka D, Döscher R, Gehlen M, Gruber N, Holland E, Hov Ø, Joos F, Matthews JBR, Rødven R, Wilson S.** 2021. The quiet crossing of ocean tipping points. *Proc Natl Acad Sci* **118**:e2008478118.
2. **Zaikova E, Walsh DA, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ.** 2010. Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* **12**:172–191.
3. **Whitney FA, Freeland HJ, Robert M.** 2007. Persistently declining oxygen levels in the interior waters of the eastern subarctic Pacific. *Prog Oceanogr* **75**:179–199.
4. **Garçon V, Karstensen J, Palacz A, Telszewski M, Aparco Lara T, Breitburg D, Chavez F, Coelho P, Cornejo-D’Ottone M, Santos C, Fiedler B, Gallo ND, Grégoire M, Gutierrez D, Hernandez-Ayon M, Isensee K, Koslow T, Levin L, Marsac F, Maske H, Mbaye BC, Montes I, Naqvi W, Pearlman J, Pinto E, Pitcher G, Pizarro O, Rose K, Shenoy D, Van der Plas A, Vito MR, Weng K.** 2019. Multidisciplinary Observing in the World Ocean’s Oxygen Minimum Zone Regions: From Climate to Fish — The VOICE Initiative. *Front Mar Sci* **6**.
5. **Torres-Beltrán M, Hawley AK, Capelle D, Zaikova E, Walsh DA, Mueller A, Scofield M, Payne C, Pakhomova L, Kheirandish S, Finke J, Bhatia M, Shevchuk O, Gies EA, Fairley D, Michiels C, Suttle CA, Whitney F, Crowe SA, Tortell PD, Hallam SJ.** 2017. A compendium of geochemical information from the Saanich Inlet water column. *Sci Data* **4**:170159.
6. **Michiels CC, Huggins JA, Giesbrecht KE, Spence JS, Simister RL, Varela DE, Hallam SJ, Crowe SA.** 2019. Rates and Pathways of N₂ Production in a Persistently Anoxic Fjord: Saanich Inlet, British Columbia. *Front Mar Sci* **6**.
7. **Tomasek A, Staley C, Wang P, Kaiser T, Lurndahl N, Kozarek JL, Hondzo M, Sadowsky MJ.** 2017. Increased Denitrification Rates Associated with Shifts in Prokaryotic Community Composition Caused by Varying Hydrologic Connectivity. *Front Microbiol* **8**.
8. **Bueno E, Mesa S, Bedmar EJ, Richardson DJ, Delgado MJ.** 2012. Bacterial adaptation of respiration from oxic to microoxic and anoxic conditions: redox control. *Antioxid Redox Signal* **16**:819–852.
9. **Hawley AK, Torres-Beltrán M, Zaikova E, Walsh DA, Mueller A, Scofield M, Kheirandish S, Payne C, Pakhomova L, Bhatia M, Shevchuk O, Gies EA, Fairley D, Malfatti SA, Norbeck AD, Brewer HM, Pasa-Tolic L, del Rio TG, Suttle CA, Tringe S, Hallam SJ.** 2017. A compendium of multi-omic sequence information from the Saanich Inlet water column. 1. *Sci Data* **4**:170160.
10. **Hallam SJ, Torres-Beltrán M, Hawley AK.** 2017. Monitoring microbial responses to ocean deoxygenation in a model oxygen minimum zone. 1. *Sci Data* **4**:170158.
11. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120.
12. **Li D, Liu C-M, Luo R, Sadakane K, Lam T-W.** 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**:1674–1676.
13. **Uritskiy GV, DiRuggiero J, Taylor J.** 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**:158.
14. **Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH.** 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**:1925–1927.
15. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol J Comput Mol Cell Biol* **19**:455–477.
16. **Morgan-Lang C, McLaughlin R, Armstrong Z, Zhang G, Chan K, Hallam SJ.** 2020. TreeSAPP: the Tree-based Sensitive and Accurate Phylogenetic Profiler. *Bioinformatics* **36**:4706–4713.
17. **Haft DH, Selengut JD, White O.** 2003. The TIGRFAMs database of protein families. *Nucleic Acids Res* **31**:371–373.
18. **The UniProt Consortium.** 2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**:D158–D169.

19. Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, Stamatakis A. 2019. EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Syst Biol* **68**:365–369.
20. Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer.
21. Shannon CE. 1948. A mathematical theory of communication. *Bell Syst Tech J* **27**:379–423.
22. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Karp PD. 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* **44**:D471–D480.
23. Kelley EE. 2017. Biochemistry of Molybdopterine Nitrate/Nitrite Reductases, p. 173–184. *In* Nitric Oxide. Elsevier.
24. He P, Xie L, Zhang X, Li J, Lin X, Pu X, Yuan C, Tian Z, Li J. 2020. Microbial Diversity and Metabolic Potential in the Stratified Sansha Yongle Blue Hole in the South China Sea. *Sci Rep* **10**:5949.
25. Hong P, Wu X, Shu Y, Wang C, Tian C, Gong S, Cai P, Donde OO, Xiao B. 2019. Denitrification characterization of dissolved oxygen microprofiles in lake surface sediment through analyzing abundance, expression, community composition and enzymatic activities of denitrifier functional genes. *AMB Express* **9**:129.
26. Marchant HK, Ahmerkamp S, Lavik G, Tegetmeyer HE, Graf J, Klatt JM, Holtappels M, Walpersdorf E, Kuypers MMM. 2017. Denitrifying community in coastal sediments performs aerobic and anaerobic respiration simultaneously. *ISME J* **11**:1799–1812.
27. Hawley AK, Brewer HM, Norbeck AD, Paša-Tolić L, Hallam SJ. 2014. Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc Natl Acad Sci* **111**:11395–11400.
28. Shah V, Chang BX, Morris RM. 2017. Cultivation of a chemoautotroph from the SUP05 clade of marine bacteria that produces nitrite and consumes ammonium. *ISME J* **11**:263–271.
29. Sheik CS, Jain S, Dick GJ. 2014. Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics: Disentangling the ecophysiological role of SAR324. *Environ Microbiol* **16**:304–317.
30. Pérez Castro S, Borton MA, Regan K, Hrabe de Angelis I, Wrighton KC, Teske AP, Strous M, Ruff SE. 2021. Degradation of biological macromolecules supports uncultured microbial populations in Guaymas Basin hydrothermal sediments. *ISME J* **15**:3480–3497.
31. Bertagnolli AD, Padilla CC, Glass JB, Thamdrup B, Stewart FJ. 2017. Metabolic potential and in situ activity of marine Marinimicrobia bacteria in an anoxic water column. *Environ Microbiol* **19**:4392–4416.
32. Fuchs B, Woebken D, Zubkov M, Burkill P, Amann R. 2005. Molecular identification of picoplankton populations in contrasting waters of the Arabian Sea. *Aquat Microb Ecol* **39**:145–157.
33. Allers E, Wright JJ, Konwar KM, Howes CG, Beneze E, Hallam SJ, Sullivan MB. 2013. Diversity and population structure of Marine Group A bacteria in the Northeast subarctic Pacific Ocean. *ISME J* **7**:256–268.
34. Wright JJ, Konwar KM, Hallam SJ. 2012. Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol* **10**:381–394.
35. Helen D, Kim H, Tytgat B, Anne W. 2016. Highly diverse nirK genes comprise two major clades that harbour ammonium-producing denitrifiers. *BMC Genomics* **17**:155.
36. Oakley BB, Francis CA, Roberts KJ, Fuchsman CA, Srinivasan S, Staley JT. 2007. Analysis of nitrite reductase (nirK and nirS) genes and cultivation reveal depauperate community of denitrifying bacteria in the Black Sea suboxic zone. *Environ Microbiol* **9**:118–130.
37. Kerou M, Ponce-Toledo RI, Zhao R, Abby SS, Hirai M, Nomaki H, Takaki Y, Nunoura T, Jørgensen SL, Schleper C. 2021. Genomes of Thaumarchaeota from deep sea sediments reveal specific adaptations of three independently evolved lineages. *ISME J* **15**:2792–2808.
38. Li F, Li M, Shi W, Li H, Sun Z, Gao Z. 2017. Distinct distribution patterns of proteobacterial *nirK* - and *nirS* -type denitrifiers in the Yellow River estuary, China. *Can J Microbiol* **63**:708–718.
39. Reji L, Tolar BB, Smith JM, Chavez FP, Francis CA. 2019. Differential co-occurrence relationships shaping ecotype diversification within Thaumarchaeota populations in the coastal ocean water column. *ISME J* **13**:1144–1158.
40. Saarenheimo J, Tiirola MA, Rissanen AJ. 2015. Functional gene pyrosequencing reveals core proteobacterial denitrifiers in boreal lakes. *Front Microbiol* **6**.
41. Klotz F, Kitzinger K, Ngugi DK, Büsing P, Littmann S, Kuypers MMM, Schink B, Pester M. 2022. Quantification of archaea-driven freshwater nitrification from single cell to ecosystem levels. *ISME J* <https://doi.org/10.1038/s41396-022-01216-9>.
42. Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, Swan BK, Fonseca MM, Posada D, Stepanauskas R, Hollibaugh JT, Foster PG, Woyke T, Luo H. 2019. Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. *ISME J* **13**:2150–2161.
43. Alvarez L, Bricio C, Blesa A, Hidalgo A, Berenguer J. 2014. Transferable Denitrification Capability of *Thermus thermophilus*. *Appl Environ Microbiol* **80**:19–28.
44. Vollack K-U, Zumft WG. 2001. Nitric Oxide Signaling and Transcriptional Control of Denitrification Genes in *Pseudomonas stutzeri*. *J Bacteriol* **183**:2516–2526.
45. Wright TD, Vergin KL, Boyd PW, Giovannoni SJ. 1997. A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Appl Environ Microbiol* **63**:1441–1448.
46. Cao H, Dong C, Bougouffa S, Li J, Zhang W, Shao Z, Bajic VB, Qian P-Y. 2016. Delta-proteobacterial SAR324 group in hydrothermal plumes on the South Mid-Atlantic Ridge. *Sci Rep* **6**:22842.

47. **Munzur A, Brown A.** 2021. Genome-resolved metagenomics of nitrogen cycling processes in Saanich Inlet. *UJEMI+* 7:13.
48. **Green SJ, Prakash O, Gihring TM, Akob DM, Jasrotia P, Jardine PM, Watson DB, Brown SD, Palumbo AV, Kostka JE.** 2010. Denitrifying bacteria isolated from terrestrial subsurface sediments exposed to mixed-waste contamination. *Appl Environ Microbiol*, 2010/03/19 ed. 76:3244–3254.
49. **Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, McLellan SL, Lückner S, Eren AM.** 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol* 3:804–813.
50. **Wyman M, Hodgson S, Bird C.** 2013. Denitrifying Alphaproteobacteria from the Arabian Sea That Express *nosZ*, the Gene Encoding Nitrous Oxide Reductase, in Oxidic and Suboxic Waters. *Appl Environ Microbiol* 79:2670–2681.
51. **Fernandes GL, Shenoy BD, Damare SR.** 2020. Diversity of Bacterial Community in the Oxygen Minimum Zones of Arabian Sea and Bay of Bengal as Deduced by Illumina Sequencing. *Front Microbiol* 10:3153–3153.
52. **Bonin P, Gilewicz M, Bertrand JC.** 1989. Effects of oxygen on each step of denitrification on *Pseudomonas nautica*. *Can J Microbiol* 35:1061–1064.
53. **Bergaust L, Shapleigh J, Frostegård Å, Bakken L.** 2008. Transcription and activities of NO_x reductases in *Agrobacterium tumefaciens*: the influence of nitrate, nitrite and oxygen availability. *Environ Microbiol* 10:3070–3081.
54. **Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, Bonilla G, Kar A, Leiby N, Mehta P, Marx CJ, Segrè D.** 2014. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep* 7:1104–1115.
55. **Mee MT, Collins JJ, Church GM, Wang HH.** 2014. Syntrophic exchange in synthetic microbial communities. *Proc Natl Acad Sci* 111:E2149–E2156.
56. **Canfield DE, Glazer AN, Falkowski PG.** 2010. The Evolution and Future of Earth's Nitrogen Cycle. *Science* 330:192–196.
57. **Zumft WG.** 1997. Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev* 61:533–616.
58. **Marzocchi U, Bonaglia S, van de Velde S, Hall POJ, Schramm A, Risgaard-Petersen N, Meysman FJR.** 2018. Transient bottom water oxygenation creates a niche for cable bacteria in long-term anoxic sediments of the Eastern Gotland Basin. *Environ Microbiol* 20:3031–3041.