Volume 7: 1-13

UJEMI+

The Undergraduate Journal of Experimental Microbiology & Immunology (+Peer Reviewed)

Genome-resolved metagenomics of nitrogen cycling processes in Saanich Inlet

Asli Munzur, Andrew Brown, Nicholas deGoutiere, Daniela Garcia, Lindsey Gross, Amy Lin
Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

SUMMARY In the oxygen minimum zones of the Saanich Inlet, genes encoding nitrogen loss processes e.g. denitrification were more abundant than those encoding nitrification or nitrogen fixation pathways. The most abundant denitrification genes recovered along the Saanich Inlet oxygen gradient spanning 100 and 200 meter depth intervals were the nitrate reductase subunits, *napA* and *narG* . Minor levels of nitrification genes such as *nxrA* and *nxrB* were identified by PROKKA, while the denitrification genes such as *napA*, *narG* and *nirS* were more abundant, implying denitrification pathways are expressed at a higher level than nitrification pathways. Similarly, the increased denitrification gene abundances can be related to the accumulation of products typical in the denitrification pathway. For example, nitrate (NO_3^-), a reactant in the denitrification pathway, is completely consumed at 200 meters, whereas hydrogen sulfide (H_2S) concentration sharply increases. Taxonomic grouping of these denitrification genes, identified by GTDB-Tk, indicate that they are primarily encoded by the Gammaproteobacteria class. Together, these results support a global trend of community metabolism shifting from utilizing oxygen to nitrogen as a terminal electron acceptor in response to oxygen minimum zones, in addition to identifying the key community members involved. This shift towards the denitrification pathways may potentially aggravate climate change in coastal marine environments through the formation of potent greenhouse gases, mainly nitrous oxide (N_2O), in oxygen minimum zones.

IMPACT SUMMARY The global trend of oceanic warming and deoxygenation poses a potential threat to the balance of nutrient cycling in oxygen minimum zones. Under increasingly anoxic conditions, nitrogen cycling processes may become permanently skewed towards denitrification. The results of this study show that the occurrence of denitrification pathways in the Saanich Inlet is more predominant than nitrification pathways. Oceanic denitrification is already a globally significant contributor of nitrous oxide greenhouse gas; the additional nitrous oxide emissions from an even greater shift to denitrification could further accelerate climate change. These findings provide a methodology that can be used in modelling the shift in key nitrogen gene abundances and their implications for climate change.

INTRODUCTION

One of the most drastic effects of climate change is the rising ocean temperatures [1]. The warming of marine ecosystems results in a decrease in dissolved oxygen concentration due to decreased oxygen solubility and thermal stratification of the water column, which hinders gas exchange used to replace consumed oxygen [2,3]. This in turn contributes to the expansion of oxygen minimum zones (OMZs). OMZs occur naturally any time the demands of oxygen respiration metabolism exceed available oxygen [4], but similar phenomena can also be precipitated by human activity (ex. creation of dead zones due to coastal eutrophication [5]). Due to limited oxygen availability, life in higher trophic levels (ex. fish) is drastically reduced in OMZs, thus shifting the available energy in the ecosystem towards

Published Online: September 2021

Citation: Asli Munzur, Andrew Brown, Nicholas deGoutiere, Daniela Garcia, Lindsey Gross, Amy Lin. 2021. Genome-resolved metagenomics of nitrogen cycling processes in Saanich Inlet. UJEMI+ 7:1-13

Editor: Daniela Morales, Stefanie Sternagel and Brianne Newman, University of British Columbia

Address correspondence to:
https://jemi.microbiology.ubc.ca/

microbial activity (Figure 1A, adapted from [5]). These microbial communities play an important role in nutrient cycling in marine ecosystems. Specifically, microbes are capable of numerous forms of anaerobic respiration using alternative terminal electron acceptors such as nitrate and sulfate (Figure 1B, adapted from [4]). The microbial contribution to the nitrogen cycle in OMZs is of particular interest as it may further drive climate change. Microbial respiration of nitrate (NO_3^-) in OMZs results in nitrous oxide (N_2O) production, a powerful greenhouse with 300x the radiative forcing effect relative to CO_2, on a scale of global significance [6]. Furthermore, NO_3^- respiration results in the loss of fixed nitrogen from the marine environment [7]. The loss of this limiting nutrient can be disruptive to carbon transport processes and threaten the health to the whole ecosystem. These implications of changing molecular balances motivate an investigation into the microbial communities and microbial-driven nutrient cycling that pervade oxygen minimum water columns.

In marine ecosystems nitrogen exists in a number of molecular states. The most abundant is nitrogen gas (N_2), but also includes organic (e.g. ammonia NH_3) and inorganic forms (e.g. nitrate NO_3^-). The nitrogen cycle is the process by which nitrogen is converted between these forms [8], typically through enzymatic reactions carried out by microbes. This set of processes is of paramount importance to all life because nitrogen gas is not a usable source of elemental nitrogen for all but a select few microorganisms. The nitrogen cycle is composed of the following processes: nitrogen fixation, nitrification, anammox, and denitrification. Nitrogen fixation is the conversion of nitrogen gas to ammonia ($N_2 + 8H^+ + 8e^- \rightarrow 2\ NH_3 + H_2$) and is facilitated by nitrogenase encoded by the *nif* family of genes. Importantly, nitrogenase is deactivated by the presence of oxygen [8] and therefore nitrogen fixation is favoured under suboxic ($0 < [O_2] < 50$ μM) to anoxic (0μM) conditions in marine environments. Next, the oxidation of ammonia to nitrate occurs in several steps in a process called nitrification. Ammonia is first converted to hydroxylamine (NH_2OH) and then nitrite (NO_2^-) by *amo* and *hao* family gene products. Then, the *nxr* family of nitrification gene products can convert nitrite to nitrate ($NO_2^- + H_2O \rightarrow NO_3^- + 2H^+ + 2e^-$). Nitrification processes are favoured under oxic ($[O_2] > 100$ μM) to dysoxic ($50 < [O_2] < 100$ μM) conditions in marine environments. Together, nitrogen fixation and nitrification are the processes by which nitrogen can be retained in an ecosystem as biomass. On the other hand, the two processes which result in loss of nitrogen in an ecosystem are anaerobic ammonium oxidation (anammox) and denitrification as both form nitrogen gas as products which cannot be retained as biomass. Denitrification consists of the following subsequent reduction steps: nitrate to nitrite ($NO_3^- + 2H^+ + 2e^- \rightarrow NO_2^- + H_2O$) facilitated by nitrate reductases of the *NapA/NarG* gene families, nitrite to nitric oxide ($NO_2^- + 2H^+ + 2e^- \rightarrow NO + H_2O$) facilitated by nitrite reductases of the *Nir* gene family, nitric oxide to nitrous oxide ($2NO + 2H^+ + 2e^- \rightarrow N_2O + H_2O$) facilitated by nitric oxide reductases of the *Nor* gene family, and finally, nitrous oxide to nitrogen ($N_2O + 2H^+ + 2e^- \rightarrow N_2 + H_2O$) facilitated by nitrous oxide reductases of the *Nos* gene family. Meanwhile, anammox converts NO_2^- to N_2, after passing through the toxic intermediate products NO and N_2H_4 [9]. Lastly, the *nrf* family of gene products may alternatively convert NO_2^- to NH_3 instead of N_2. The aforementioned processes are represented in Figure 1C [adapted from 10].

Major denitrification genes present in the the marine microbiome include the nitrate reductases *napA* and *narG*, nitrite reductases *nirS* and *nirk,* heterodimeric nitric oxide reductases *norB* and *norC,* and finally a nitrous oxide reductase *norZ* which produces N_2 [11]. Sequentially these genes turn NO_3^- into N_2, which is critical for biological denitrification. Nitrate reductases require NO_3^- as a terminal electron acceptor, demonstrated by the redox reactions presented in Figure S5, therefore necessitating many bacteria in anoxic conditions to encode either *napA* or *narG* proteins for catabolic processes [12]. For these reasons, it is expected to find an abundance of *napA* and *narG* proteins in anoxic marine ecosystems. Examining the remaining denitrification gene abundances can be used as a proxy for the shift towards denitrification itself, which is established to have critical consequences on the marine ecosystem [15]. The aforementioned complete redox reactions involving the mentioned genes are summarized in Figure S6.

The Saanich Inlet, a seasonally anoxic fjord, is an ideal model ecosystem for studying microbial responses to ocean deoxygenation. The Saanich Inlet water column provides an

opportunity for rich analysis of the microbial community, owing to the compendium of both geochemical and multi-omic sequence data collected over 10 years [13,14]. Together, the data enable a mechanistic analysis of the interplay between biogeochemical transformations and microbial diversity. Thus, this study will first investigate the composition of dissolved gases and nutrients present in the Saanich Inlet across different depths. Secondly, the dominant nitrogen cycle processes will be identified and quantified in those areas, followed by the characterization of the diversity of the microbial community driving these processes. Lastly, the study will probe the relationship between the found gene abundances and taxonomic groups present in the Saanich Inlet water column.

METHODS AND MATERIALS

Water Sample Collection. To characterize the microbial diversity of Saanich Inlet, water samples were collected from various depths spanning across the oxic, hypoxic transition, and anoxic basin waters according to the protocol outlined in Zaikova et al., 2009 [15]. Data was provided by the teaching team of MICB 405 for the purposes of this bioinformatics study. An outline of the methods provided in this section have been summarized in Figure 2. For the purposes of this study, sequencing data from exclusively 100 meters, 120 meters and 200 meters from Cruise 72 (August 1, 2012) were considered in downstream metagenomics analyses.

Read Processing and Binning. The reads were trimmed and filtered using the default settings for Trimmomatic [16]. The reads were assembled, mapped into a contigs database, and the resulting binary alignment mapped (BAM) files were sorted, indexed, initialized, profiled, and merged with Anvi'o [17]. In order to recover individual genomes, two binning algorithms were used: MaxBin2 [18] and MetaBAT [19], to choose the highest quality metagenome assembled genomes (MAGs). For quality assessment of MAGs, completion - contamination plots based on CheckM results for MetaBAT and MaxBin2 were generated (Figure S1) [20]. The low quality MAGs were disregarded, while high and medium quality MAGs with minimum 50% completion and maximum 10% contamination were utilized for analysis (Figure 1). DASTool was used to integrate results from both MaxBin2 and MetaBAT in order to identify the highest quality bins based on the frequency of single copy marker genes [21]. DASTool's identification of high and medium quality bins was identical to the 18 bins selected earlier based on completion - contamination plots. The bioinformatics workflow was continued with only MAGs generated with MetaBAT as MaxBin2 did not generate any high or medium quality MAGs. Contigs in the high and medium quality bins were used as inputs to count the number of occurrences for nitrogen cycle genes. Further processing was done in linux to quantify the number of contigs retained after initial quality control on reads. The resulting combined file was visualized using R.

Taxonomic and Functional Annotations using PROKKA. PROKKA was used to locate open reading frames and annotate the genomes separately for each depth [22]. To get gene abundance information for the nitrogen cycle genes and their variants (*napA, nifD, nirK, nirS, norB, norC, nosZ, nxrA, nxrB, narG, nrfA,* and *nasA*) for each depth, PROKKA outputs that contained the gene names and the contigs that can be traced to a certain depth were considered. Furthermore, information about non-coding sequences such as transfer ribonucleic acid (t-RNA) was removed from PROKKA annotations to retain the genes of interest mentioned above.

TreeSAPP and Phylogenetic Tree Construction. TreeSAPP was used for functional and taxonomic annotation of amino acid sequences for the nitrogen cycle genes mentioned above [23]. TreeSAPP commands were run on the chosen MAGs with default parameters, and the resulting data files were provided by the teaching team of MICB 405 course.

For each nitrogen cycle gene that was included in the queries, TreeSAPP outputs abundance and taxonomic classification information. Taxonomic identifications were

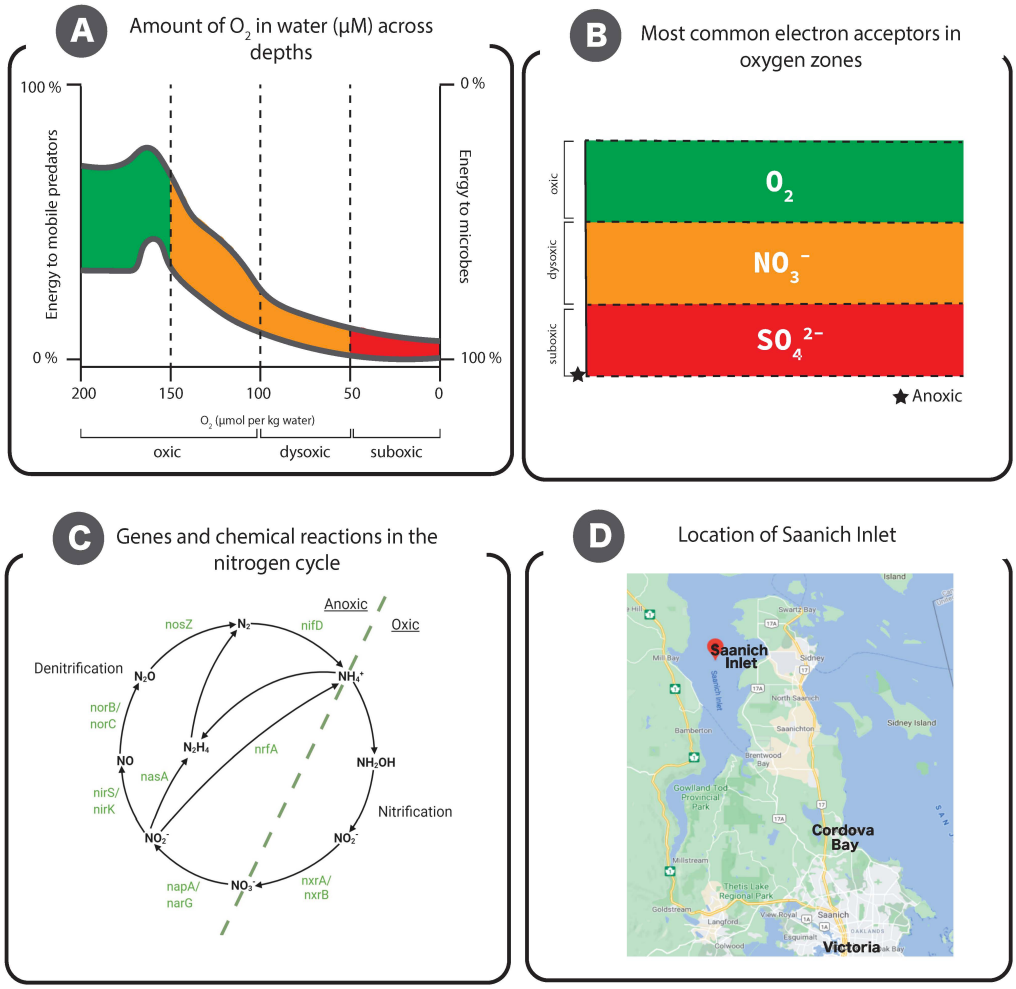


FIG. 1 Overview of the environmental background in Saanich Inlet. In Panel (A) the relationship of hypoxia and ecosystem energy flow are shown. The energy generated by microbial respiration uses different primary terminal electron acceptors where green represents O_2, orange represents NO_3^-, and red represents SO_4^{-2} . Panel (A) is modified from Ref. 5. In Panel (B), the trends of terminal electron acceptors in decreasing oxygen levels are shown. Green represents oxic conditions where the O_2 is the primary electron acceptor, orange represents dysoxic conditions where $NO_3^-$is prominent, and red represents suboxic regions where SO_4^{-2} is the main electron acceptor. Anoxic conditions are shown with the black star. Panel (B) is modified from Ref. 4. In Panel (C), the overall nitrogen cycle with key functional genes for each reaction are shown. The cycle is divided into denitrification in anoxic conditions and nitrification in oxic conditions with a green dashed line. Panel (C) is modified from Ref. 10. In Panel (D), the location of Saanich Inlet, a model oxygen minimum zone of the coast of Vancouver Island, British Columbia, Canada, is shown. Image has been dynamically created using Google Maps.

consolidated from TreeSAPP outputs for the top five species with the highest Reads Per Kilobase of transcript per Million mapped reads (RPKM) values. The phylum each species belongs to was identified using the List of Prokaryotic names with Standing in Nomenclature (LPSN) [24-27] and the NCBI Taxonomy Browser [28]. Gene abundance information and taxonomic classifications (at species and phylum levels) are found in Figure S3. Primary X axis shows species information, and secondary X axis indicates which phylum each species belongs to. Chemical reactions the genes are involved with are indicated in figure S6.

The top five classes were then organized into a phylogenetic tree in based on the phylum for which each class is a member using manual curation [Figure 5A]. Nitrogen cycle genes were then matched to the class that contained the species with the highest RPKM values to visualize the taxonomic relationships between the present species.

Technical Considerations Regarding Compositions of MAGs. To investigate the composition of MAGs across depths and assess how representative the bins are in the Saanich Inlet, the chosen MAGs were assessed for containing equal reads from all 3 depths. The mean

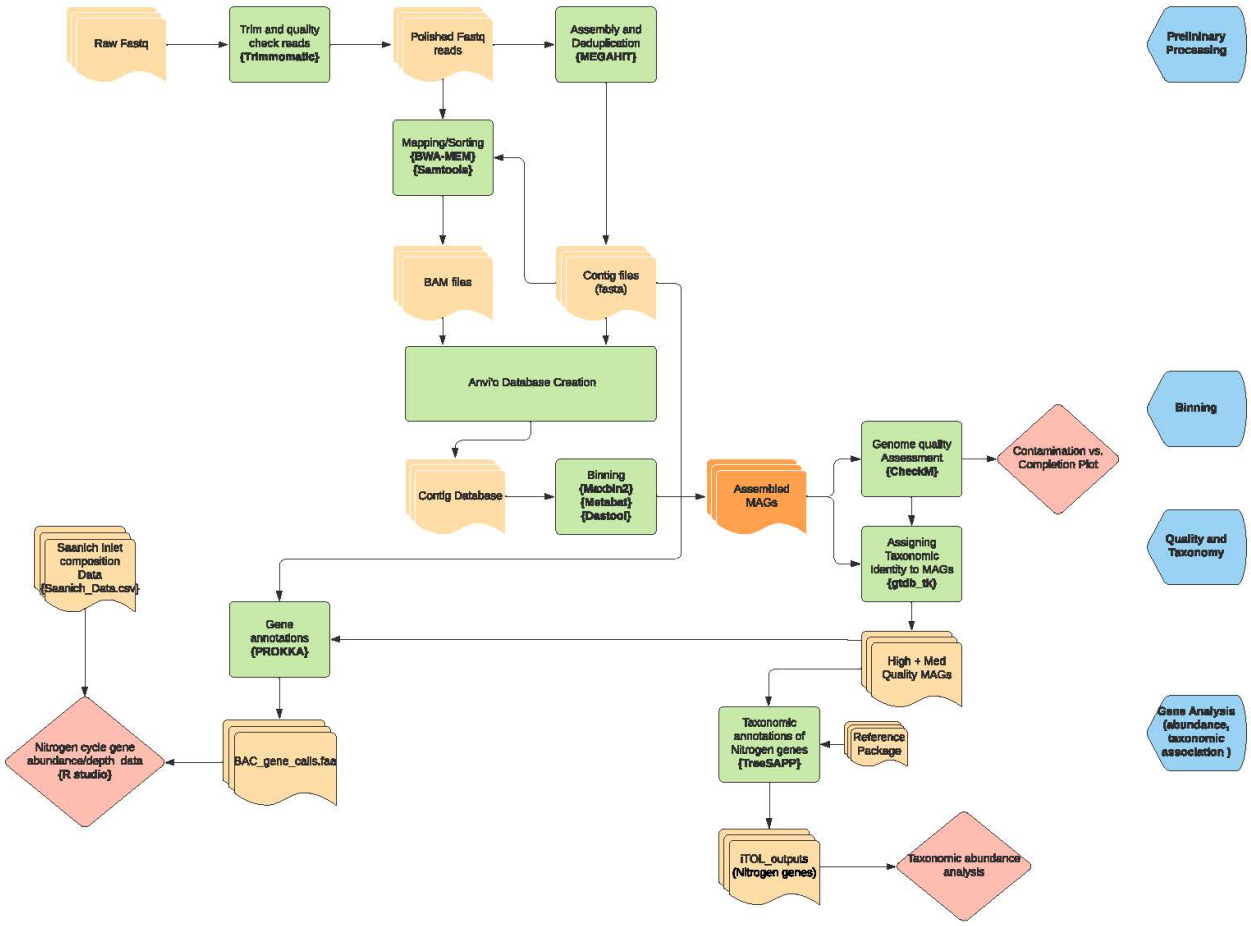


FIG. 2 Overview of the bioinformatics workflow. Water samples were sequenced and the resulting FASTA files were trimmed with Trimmomatic. For contig database creation and MAG assembly, Anvi'o was used. Genome quality assembly of the MAGS were done with CheckM and DASTool. For taxonomic annotations of bins GTDB-Tk was used. Functional gene annotations and taxonomic annotations of nitrogen cycle genes were completed with PROKKA and TreeSAPP, respectively. The figure was generated with Lucid charts.

coverage of reads for each high and medium quality MAGs across 100 meters, 120 meters and 200 meters depths were calculated using DASTool by mapping the reads onto MAGs. Coverage information was extracted from the "mean_coverage.txt" files for each bin that DASTool outputs. Differential coverage information is visualized in Figure 6A for each high and medium quality bin that was included in the workflow.

Analysis of Geochemical Data. To visualize the chemical composition of Saanich Inlet from 2006 to 2014 across depths and seasons, the water temperature (Figure S2C), density (Figure S2A), salinity (Figure S2B), as well as oxygen (Figure S6A, Figure 4A), and nitrate (Figure S6B, Figure 4B) were all plotted. Concentrations from the data provided by the teaching team of MICB 405 course at UBC. All plots to show changes in salinity, temperature, density, and molecule concentration were prepared as previously stated by filtering the output data accordingly. The ORCA bioinformatics Docker image, provided by the teaching team of MICB 405, was used for running all metagenomics related tools.

Figure Generation and Code availability. R, Adobe Illustrator, and Biorender were used to generate all the figures in the paper. All data files used in the bioinformatics workflow are present in the ORCA server. All R scripts developed in this study are provided in the following GitHub link:
https://github.com/amunzur/Microbial_Diversity_of_the_Saanich_Inlet/tree/master

More information regarding the custom code written for the study is available with no restrictions from authors.

RESULTS

Geochemical Data Defines OMZ and nitrogen cycling. It is well known that chemical reactions can be limited by the available reagents in the environment. As the purpose of this study is to assess biological aspects contributing to the nitrogen cycle, it is important to consider the concentration levels of the necessary compounds for the cycle. Thus, NO_3^- and O_2 concentrations were plotted to understand nitrogen cycle activity at different depths.

NO_3^- concentrations were approximately close to 0µM at shallow depths (<20 meters), but rose rapidly to 30µM at 80 meters (Figure 3A). Below this level in the water column, NO_3^- levels steeply diminished to minimal levels at 160 meters. O_2 concentrations were the highest at the water surface, with measurements being over 200µM (Figure 3B). The concentrations appeared to decrease until 80 meters, where dissolved O_2 levels fluctuate around 30µM until decreasing again at 140 meters. O_2 concentrations remained near zero from 150 meters onwards.

It may be important to notice the highest concentration of NO_3^- (Figure 3A) is concurrent with the decline in O_2 concentrations (Figure 3B) at 80 meters. The concentrations of both molecules proceed to decline at independent rates at subsequent depths.

Positive correlation between geochemical presence and gene abundances. The large reduction of nitrogen across depths incentivised the examination of gene abundances to observe any possible biotic factors which may contribute to this change or be influenced by it. Minimal concentrations of H_2S and NH_4 were measured at 100 meters and 120 meters, while relatively larger concentrations were measured at 200 meters (Figure 4A). Conversely, N_2O abundances appeared to be the opposite as high N_2O concentrations were seen at 100 meters and 120 meters, while absent at 200 meters (Figure 4A). The abundance of PO_4 in all three depths remained approximately the same (Figure 4A). Most pertinent to this research, O_2 concentrations at 100 meters and 120 meters appeared to be very low while there was a high relative abundance of NO_3^- (Figure 4A). These are in consensus with previously obtained geochemical data shown in Figure 3.

The top two most abundant genes in PROKKA annotations are known NO_3^- reductase genes, whose products are utilized in the denitrification pathway: *napA* and *narG* (Figure 4B). Both genes exhibited the greatest abundance at 120 meters with 189 and 64 variants respectively (Figure 4B). Other gene products involved with denitrification that are less

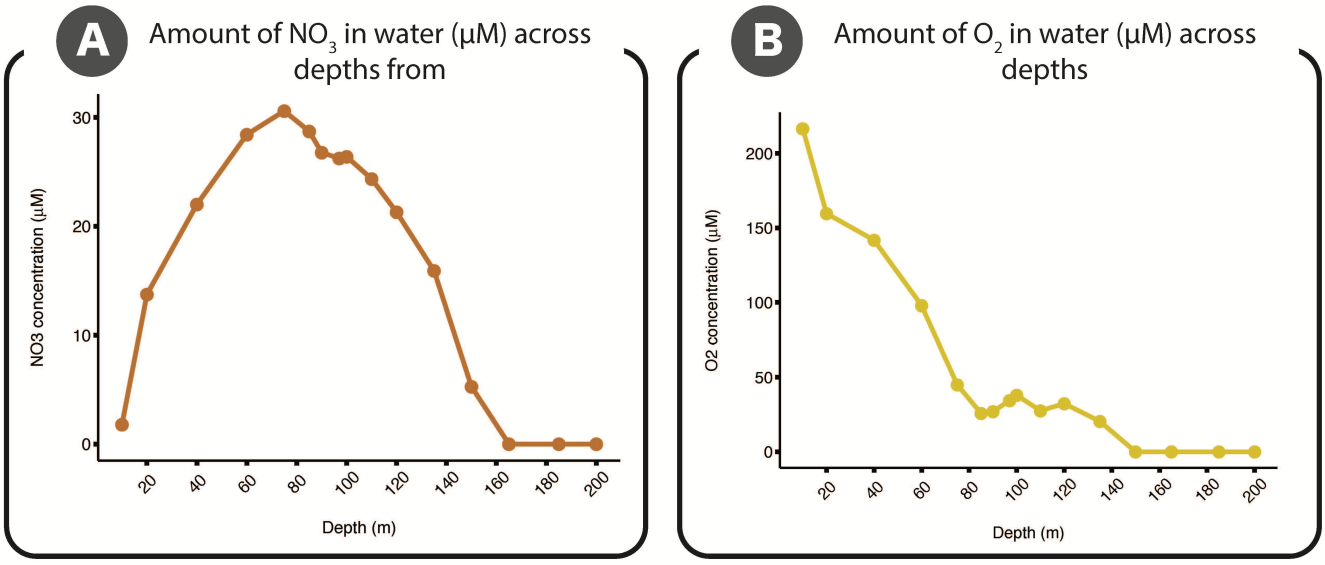


FIG. 3 Dissolved O_2 (dioxide) and NO_3^- (nitrate) concentrations measured across varying depths from cruise 72. In Panel (A) NO_3^- concentration is shown in micromolar (µM). In Panel (B) O_2 concentration is shown in micromolar (µM). In both panels, depths across 10 m - 200 m are included.

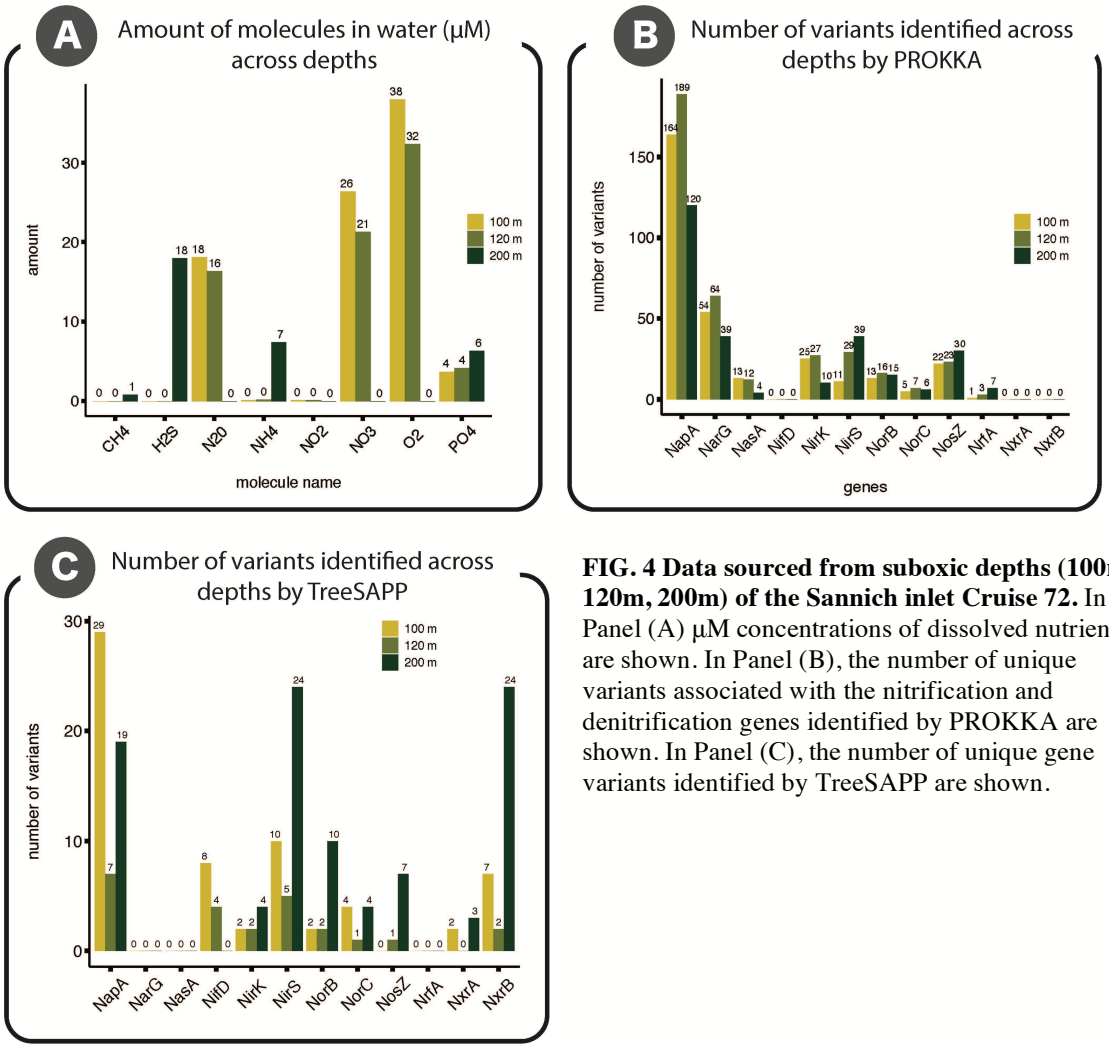


FIG. 4 Data sourced from suboxic depths (100m, 120m, 200m) of the Sannich inlet Cruise 72. In Panel (A) μM concentrations of dissolved nutrients are shown. In Panel (B), the number of unique variants associated with the nitrification and denitrification genes identified by PROKKA are shown. In Panel (C), the number of unique gene variants identified by TreeSAPP are shown.

abundant but noticeably present at all depths are: *nasA, nirK, nirS, norB, norC, nosZ* , and *nrfA* (Figure 4B). In contrast, the genes, *nifD, nxrA,* and *nxrB,* which produce proteins involved with nitrification, have little to no presence in these depths (Figure 4B).

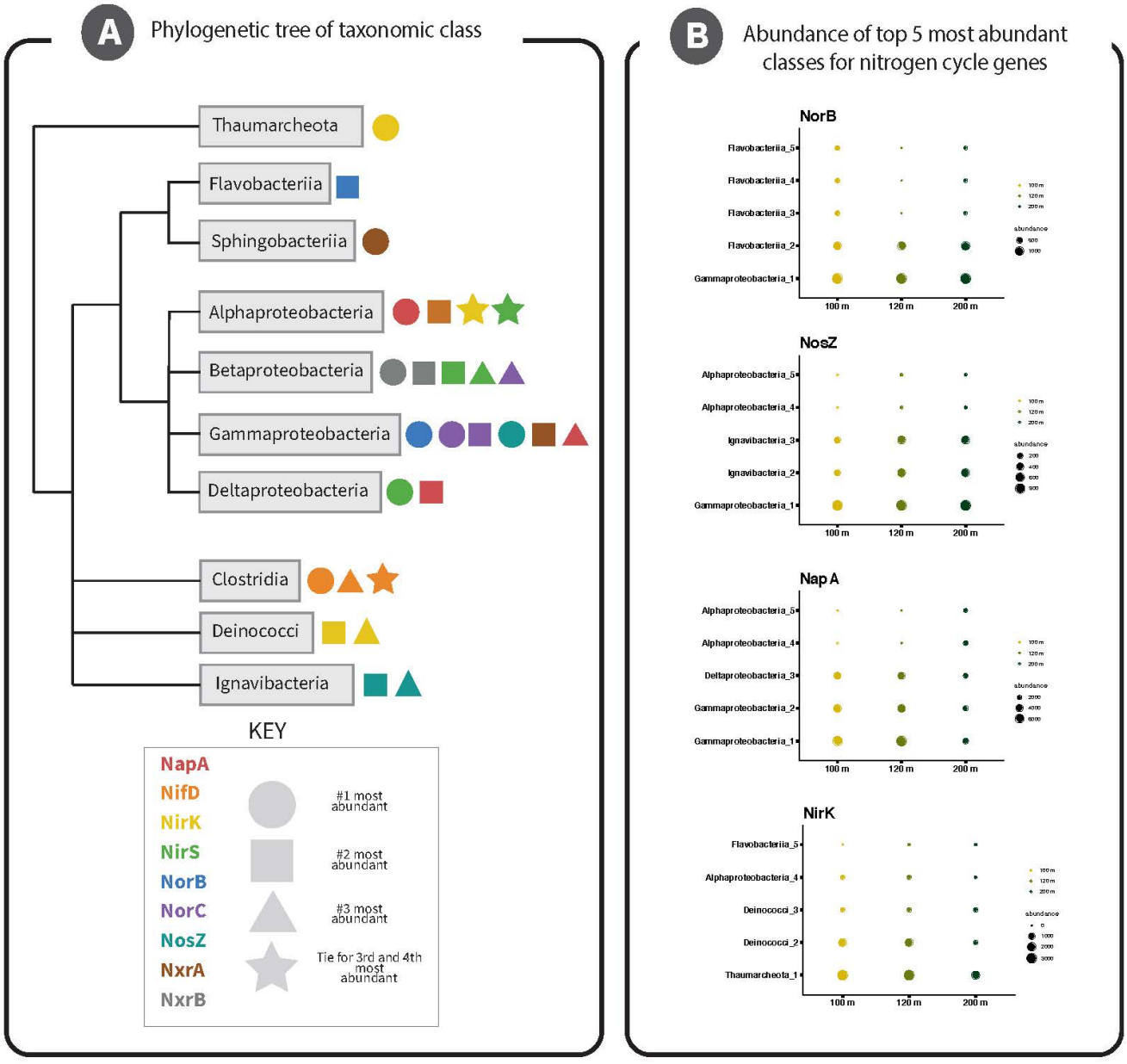
The outputs provided by TreeSAPP produced a different assessment compared to PROKKA results. Figure 4C shows *napA* having the highest abundance of all genes at 100 meters with 29 variants, followed by a lower level of 7 variants at 120 meters and then another spike to 19 variants at 200 meters, though it does not reach the 100 meter level. Other highly abundant genes include *nirS* and *nxrB*, both spiking to 24 variants at 200 meters (Figure 4C). Genes like *nirK, norB,* and *norC* have moderate to low abundance (between 1 and 11 variants) at all measured depths, whereas *nosZ* is seen at 120 meters and 200 meters, with 1 and 7 variants respectively, but not at 100 meters (Figure 4C). Also, contrasting to the lack of appearance in the PROKKA output, *nifD* is present at 100 meters and 120 meters, and *nxrA* is seen at 100 meters and 200 meters (Figure 4C). Finally, *narG, nasA,* and *nrfA* show no abundance at any depth from the TreeSAPP analysis (Figure 4C).

Taxonomic classification of microbes observed in the Saanich Inlet. The phylogenetic tree illustrates a diverse group of microbes, from archaea to five different bacterial phyla, carrying a high abundance of genes related to the nitrogen cycle (Figure 5A). In general, the Proteobacteria phylum contains the highest abundance of genes involved with nitrogen fixation (Figure 5A). Specifically, the class Gammaproteobacteria has the largest distribution of genes concerned with both nitrification and denitrification (Figure 5A). The presence of Sphingobacteriia may be indicative of low levels of nitrification, given the prevalence of nitrification genes, such as *nxrA,* in this class (Figure 5A). The denitrification genes *nirK* and

nosZ, as well as the gene *nifD* were found most prevalent in the following classes, although not exclusively: Clostridia, Deinococci, and Ignavibacteria (Figure 5A).

Of the selected genes, *nirK* did not have the Gammaproteobacteria class in the highest ranking spot, rather Thaumarchaeota appeared to be the most abundant class. *nirK* had Deinococci as the second and third ranking classes, Alphaproteobacteria as the fourth and Flavobacteriia in fifth. The gene distribution was more diverse across the second through fifth ranking classes as *napA* was present in different classes of Proteobacteria, while *norB* was featured almost exclusively in the class Flavobacteriia. Ignavibacteria placed second and third

FIG. 5 Phylogeny and bar graphs showing taxonomic abundances of nitrogen cycle genes. In Panel (A), the phylogenetic tree showing the classes top 3 most abundant species of each nitrogen cycle gene across 100 m, 120 m and 200 m are shown. Each color represents a different nitrogen cycle gene and each shape represents its abundance ranking as shown in the figure key. In Panel (B), bubble plots for the top 5 most abundant phyla for selected genes are shown. The abundance values are RPKM values computed by TreeSAPP. Y axis shows the phylum names; the suffix shows the ranking for abundance where "_1" indicates the highest abundance and "_5" indicates the lowest abundance. Both figures are based on the taxonomic annotations from TreeSAPP. One gene from each step in the nitrogen cycle was selected to further examine its abundance in certain classes at different depths (m).



with regards to *nosZ* gene abundance, and a very small amount of Alphaproteobacteria representing the fourth and fifth spots.

Methodological Limitations. To confirm the reliability of the results, DASTool was used to compute the coverage of reads gathered from various depths across all medium and high quality MAGs included in the study in order to assess the amount of data that was actually used to produce the referred outputs. In particular, these results describe if the reads from different depths were distributed evenly across all MAGs, or whether the composition of some MAGs were dominated by reads from a certain depth. Figure 6 shows that the data was made up from only medium to high quality MAGs, and did not include any input from low quality bins which might have led to a loss of information in downstream analyses. While results derived from a higher proportion of medium quality MAGs than high quality MAGs, the coverage of the bin should also be considered. The majority of the MAGs appear to have equal coverage across depths. However, MAGs numbered 165, 187 and 27 contain more reads from 200 meters compared to the rest of the MAGs (Figure 6). There are also several MAGs for which there was no noticeable read coverage in the 200 meter depth. These results indicate that the MAGs used for downstream analyses may not fully encompass the diversity present in the water column; however, there do appear to be depth-specific trends that may be biologically relevant.

DISCUSSION

Nutrient availability as rate-limiting step in the nitrogen cycle progression. The first step in understanding the Saanich Inlet as a model oxygen minimum zone is to evaluate the availability of geochemical nutrients and define the nitrogen cycle processes driven by the microbial community. For this purpose, the abundance of nitrogen cycle genes was derived from the metagenomic sequencing data and compared with the geochemical data to infer the prevalence of specific reactions in the nitrogen cycle (Figure 4). The presence of genes involved with the conversion of NO_3^- to N_2 gas, and the decreasing NO_3^- concentration, suggests denitrification occurs across the measured depths of the oxygen minimum zone.

Interestingly, NO_2^- found at all depths is nearly zero, yet NO_2^- reductase genes *nirK* and *nirS* are found at all depths (Figure 4B, 4C). This suggests that NO_2^- is possibly a limiting reagent or an intermediate consumed in downstream reduction steps as it is involved in multiple reaction pathways (Figure 1C). These include the reduction of NO_2^- to NO by products of *nir* gene variants and the direct conversion to NH_4^+ via *nrf* family enzymes. The first reduction processes proceed towards denitrification, generating the greenhouse gas N_2O, and loss of fixed nitrogen from the ecosystem (Figure 1C). Conversely, the conversion to NH_4^+ allows fixed nitrogen to be retained in downstream nitrification reactions (Figure 1C). However, the gene abundance and geochemical data suggest there is very little conversion to

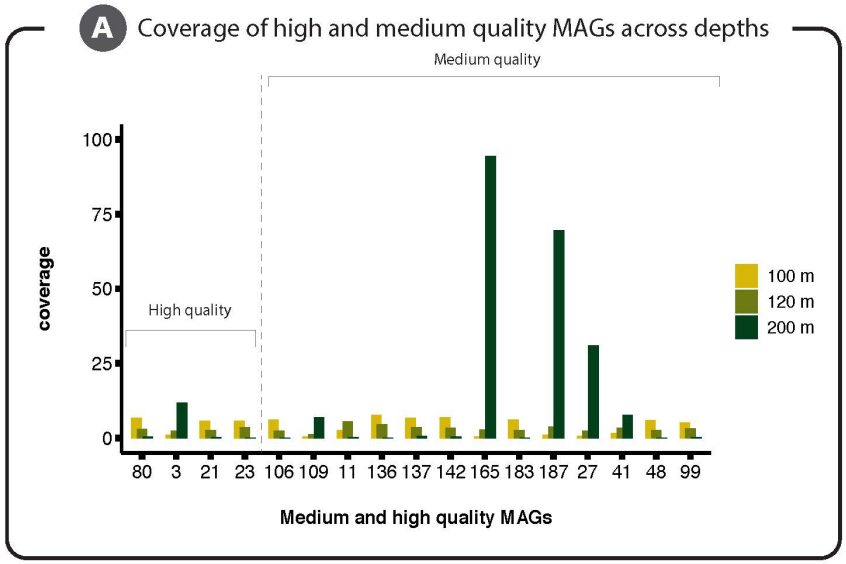


FIG. 6 Visual assessment of the read coverages of MAGs across different depths (m). DASTool was used to evaluate each bin to determine the distribution of reads in the chosen representative MAGs.

ammonium occurring in the Saanich Inlet water column (Figure 4). Only at a depth of 200 meters is there evidence of this conversion, where both NH_4^+ and *nrf* genes are found (Figure 4A, 4B). This observation is punctuated by the comparatively high concentrations of N_2O over NH_4^+ at 100 meter and 120 meter depths, which reinforces the prevalence of the denitrification reaction pathway. Competition for NO_2^- may also explain the abundance of *narG* and *napA*, the genes encoding proteins associated with the conversion from NO_3^- to NO_2^-, over all subsequent denitrifying genes. While all other denitrifying steps are limited by the low levels of NO_2^-, NO_3^- is comparatively abundant across most depths allowing NO_3^- consuming bacteria to flourish.

The apparent abundance of denitrification reactions contrasts sharply with the lack of evidence for nitrification reactions. Gene abundance analysis found minimal traces of *nif* or *nxr* family genes (Figure 4B, 4C), which encode key enzymes for the conversion of N_2 to NH_4^+ and subsequent conversion of NO_2^- to NO_3^- (Figure 1C). A similar pattern is also observed in decreasing NO_3^- concentration from 100 meter to 200 meter depths (Figure 4A). As microbial denitrification reactions consume NO_3^- without nitrification reactions to replace it, NO_3^- is completely consumed by 200 meters (Figure 4A). Thus, it is not surprising to observe a rise in H_2S at 200 meters (Figure 4A), as it could indicate a metabolic switch from using NO_3^- to SO_4^{2-} in anaerobic respiration.

Distribution of nitrogen cycle genes corresponds with environments of relevant microbes. A large variety of microbes across different phyla are found with diverse nitrogen cycle genes (Figure 5). While many species involved with these processes can be commonly found and are well-documented, there is evidence of extremophiles participating in the nitrogen cycle [29, 30]. This implies that the environment in which microbes are found can be correlated with their role in the nitrogen cycle. For instance, the Sphingobacteriia class is found with nitrification genes (Figure 4B, 4C), as it is composed of aerobic or facultatively anaerobic microbes and nitrification requires aerobic conditions [31]. Thaumarchaeota are one of the most abundant prokaryotes on Earth, and are the first Archaea identified to oxidize NH_4^+ as part of the nitrification cycle [32]. Although this class is not physiologically well characterized, they are presumed to use aerobic autotrophy and are expected to be found at 100 meters [32].

Flavobacteriia can be either aerobic or facultatively anaerobic, and as such are expected to be present in mainly the 100 meter and 120 meter depths. However, Flavobacteriia are recorded to be present in high abundance in marine biofilms [33]. This may result in decreased collection of this class of bacteria from a practical standpoint as biofilms tend to cling to surfaces, which are not typically found to be floating in the middle of the sampling area. This may explain why this class was less abundant in the collected samples (Figure 5, Figure S3).

Deinococci are potentially one of the more unique classes observed from this dataset, as they are primarily studied for their remarkable ability to survive vacuum conditions, ionizing radiation, desiccation, and oxidative stress [34]. This research however, tends to focus on the mechanisms of DNA repair and protection of proteins [34]. Future research could explore why *nirK* was so strongly observed in this class, especially given that the related gene, *nirS,* was found to be more widely distributed across different bacterial classes within the Proteobacteria (Figure 5, Figure S3).

Proteobacteria as a whole are one of the largest phyla of bacteria [35] and were originally called ‘the purple bacteria and their relatives’. This group of organisms contain 9 different classes including the Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, and Gammaproteobacteria [24-27]. Of these classes, Gammaproteobacteria have the highest number of officially recognized species, which can explain the wide diversity and high abundance of nitrogen cycle genes observed [24-27].

Collectively, these results indicate an imbalance of nitrogen cycle processes in the oxygen minimum zone of the Saanich Inlet during collection in August 2012, with denitrifying processes dominating over nitrification processes. Additionally, a large amount of organic loading from the surface and watershed provide ammonia to the water column, which is not atypical for coastal ocean conditions. Based on the abundance values of the nitrogen cycle genes at the various depths, it can be inferred that the microbes present in the samples from the Saanich Inlet as a whole, favor the denitrification process of the nitrogen cycle. This trend

towards denitrification processes is expected given the low oxygenation levels of Saanich Inlet at the sampled depths, as nitrification processes require the presence of oxygen.

Technical Limitations and Considerations. While the taxonomic and functional annotations of genes illustrate the variety of microbes involved in various nitrogen cycle pathways, it is important to acknowledge that the information presented in this study is only a glimpse into the true extent of microbial diversity in the Saanich Inlet. Nevertheless, conclusions about species diversity based on the annotations in various depths can be made. For example, the unequal distribution of read coverage across MAGs might be indicative of decreasing species diversity at 200 meters compared to shallow depths (Figure 6), as proven in a recent paper that species diversity declines sharply at deeper depths in the ocean [36]. From a bioinformatics perspective, binning algorithms are more likely to recover individual genomes from environmental samples with low species diversity since contamination is less probable. Lack of species diversity at deeper depths can also explain why reads from 200 meters are assembled into 3 MAGs instead of being distributed across all MAGs equally, as observed for most reads from 100 meters and 120 meters. In water samples with lower species diversity, reads are more likely to be assigned to MAGs accurately with higher coverage in those bins.

This can impact the ability to perform downstream metagenomic analyses, as metagenomics research relies heavily upon reference sequences of previously identified taxonomic groups. For instance, Gammaproteobacteria in particular are well studied and documented, which may have skewed the reference packages away from species that also have high abundance of nitrogen genes, but are not well identified or studied. In addition to this, the metagenomics tools used in the analyses may also have an impact on the significance of the research performed. It should also be noted that the reference packages and sequence analysis may not have the same level of annotation and research provided for all the genes. UniProtKB five-point annotation scoring system is a heuristic measure of the annotation content for a given entry. Thus, entries with a high annotation score do not necessarily confirm the accuracy of the annotations but rather imply the gene in question has been widely found and characterized compared to genes with lower annotation scores. Furthermore, gene annotations are not always experimentally obtained, and can be inferred based on homology from other orthologs. While most genes chosen in this study have five out of five annotation scores, *norC, nxrA,* and *nxrB* have decreasingly lower scores respectively. To overcome such limitations, constructing specific reference packages for the Saanich Inlet can allow annotation tools such as TreeSAPP to identify gene variants and abundances more accurately.

Furthermore, the distribution of contigs after read trimming, binning and removing low quality bins is visualized in Figure 6B. Only a small proportion of the total reads available in the FASTQ files were binned into individual medium to high quality MAGs. Therefore the MAGs included in the study do not fully represent the metagenomic diversity, and the scope of this research was likely unable to identify all of the species inhabiting the Saanich Inlet.

Lastly, a number of factors limit the conclusions drawn from this study. Even though metagenomics has the potential to identify novel genes crucial for various natural processes, it lacks the capacity to verify the gene expression in the bacterial and archeal hosts. Despite the development of various bioinformatics tool and pipelines, the field still lacks effective screening methods to connect enzyme expression with the related microbial hosts.

Conclusion and Future Directions

Climate change and deoxygenation pose a threat to the balance of the nitrogen cycle in OMZs. Under anoxic conditions, nitrogen cycling processes progressively shift towards denitrification. Our findings support this general global trend, showing that the nitrogen cycle is dominated by denitrification pathways in the Saanich Inlet. This imbalance may have feedforward repercussions for marine ecosystems and thus accelerate climate change. A shift towards denitrification would result in loss of fixed nitrogen, a key nutrient which could cause the ecosystem to collapse if removed entirely. As OMZs expand due to global climate change, it can be expected that the amount of denitrification occuring in aquatic environments will likewise expand. Meanwhile, cycling of the water column can replace the geochemical composition of the water and result in a shift in nitrifying processes. Such a shift from

denitrifying to nitrifying processes may be expected in a seasonally anoxic fjord such as the Saanich Inlet. The oscillating nature of the nitrogen cycle in the Saanich Inlet can be confirmed with similar analyses of the geochemical and nitrogen gene abundance across future seasons and years. Such a longitudinal study may elucidate the balance of nitrogen cycling in Saanich Inlet and serve to monitor the water column for evidence of OMZ expansion and nutrient cycle imbalances.

ACKNOWLEDGEMENTS

We thank Dr. Steven Hallam, Dr. Martin Hirst, Axel Hauduc, Dr. Stephan Koenig for giving valuable feedback and guidance during this project. We also thank Dr. Julia Anstett for providing us with the FASTA files and the preliminary bioinformatics analysis. We would also like to thank two anonymous reviewers for constructive feedback on this manuscript.

REFERENCES

1. **Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)** (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *IPCC*, 151 pp. From https://www.ipcc.ch/report/ar5/syr/
2. **Keeling, R. E., Körtzinger, A., & Gruber, N.** (2010). Ocean deoxygenation in a warming world. *Annual Review of Marine Science*, ***2***:199–229. https://doi.org/10.1146/annurev.marine.010908.163855
3. **Helm, K. P., Bindoff, N. L., & Church, J. A.** (2011). Observed decreases in oxygen content of the global ocean. *Geophysical Research Letters*, 38(23). https://doi.org/10.1029/2011GL049513
4. **Wright, J. J., Konwar, K. M., & Hallam, S. J.** (2012). Microbial ecology of expanding oxygen minimum zones. *Nature Reviews. Microbiology*, ***10***(6), 381–394. https://doi.org/10.1038/nrmicro2778
5. **Diaz, R. J., & Rosenberg, R.** (2008). Spreading dead zones and consequences for marine ecosystems. *Science (New York, N.Y.)*, ***321***(5891), 926–929. https://doi.org/10.1126/science.1156401
6. **Naqvi, S. W. A., Bange, H. W., Farías, L., Monteiro, P. M. S., Scranton, M. I., & Zhang, J.** (2010). Marine hypoxia/anoxia as a source of CH_4 and N_2O. *Biogeosciences*, ***7***(7), 2159–2190. https://doi.org/10.5194/bg-7-2159-2010
7. **Codispoti, L. A., Brandes, J. A., Christensen, J. P., Devol, A. H., Naqvi, S. W. A., Paerl, H. W., & Yoshinari, T.** (2001). The oceanic fixed nitrogen and nitrous oxide budgets: Moving targets as we enter the anthropocene? *Scientia Marina*, ***65***(S2), 85–105. https://doi.org/10.3989/scimar.2001.65s285
8. **Bernhard, A.** (2010) The Nitrogen Cycle: Processes, Players, and Human Impact. *Nature Education Knowledge* **3**(10):25
9. **Kartal, B., & Keltjens, J. T.** (2016). Anammox Biochemistry: A Tale of Heme c Proteins. *Trends in Biochemical Sciences*, ***41***(12), 998–1011. https://doi.org/10.1016/j.tibs.2016.08.015
10. **Smith, C. J., McKew, B. A., Coggan, A., & Whitby, C.** (2017). Primers: Functional Genes for Nitrogen-Cycling Microbes in Oil Reservoirs. In T. J. McGenity, K. N. Timmis, & B. Nogales (Eds.), *Hydrocarbon and Lipid Microbiology Protocols: Primers* (207–241). Springer. https://doi.org/10.1007/8623_2015_184
11. **Pajares, S., & Ramos, R.** (2019). Processes and Microorganisms Involved in the Marine Nitrogen Cycle: Knowledge and Gaps. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00739
12. **González, P. J., Correia, C., Moura, I., Brondino, C. D., & Moura, J. J. G.** (2006). Bacterial nitrate reductases: Molecular and biological aspects of nitrate reduction. *Journal of Inorganic Biochemistry*, ***100***(5), 1015–1023. https://doi.org/10.1016/j.jinorgbio.2005.11.024
13. **Hawley, A. K., Torres-Beltrán, M., Zaikova, E., Walsh, D. A., Mueller, A., Scofield, M., Kheirandish, S., Payne, C., Pakhomova, L., Bhatia, M., Shevchuk, O., Gies, E. A., Fairley, D., Malfatti, S. A., Norbeck, A. D., Brewer, H. M., Pasa-Tolic, L., Del Rio, T. G., Suttle, C. A., … Hallam, S. J.** (2017). A compendium of multi-omic sequence information from the Saanich Inlet water column. *Scientific Data*, *4*, 170160. https://doi.org/10.1038/sdata.2017.160
14. **Torres-Beltrán, M., Hawley, A. K., Capelle, D., Zaikova, E., Walsh, D. A., Mueller, A., Scofield, M., Payne, C., Pakhomova, L., Kheirandish, S., Finke, J., Bhatia, M., Shevchuk, O., Gies, E. A., Fairley, D., Michiels, C., Suttle, C. A., Whitney, F., Crowe, S. A., … Hallam, S. J.** (2017). A compendium of geochemical information from the Saanich Inlet water column. *Scientific Data*, *4*, 170159. https://doi.org/10.1038/sdata.2017.159
15. **Zaikova, E., Walsh, D. A., Stilwell, C. P., Mohn, W. W., Tortell, P. D., & Hallam, S. J.** (2009). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environmental Microbiology*, ***12***(1), 172–191. https://doi.org/10.1111/j.1462-2920.2009.02058.x

16. **Bolger, A. M., Lohse, M., & Usadel, B.** (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, ***30***(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170
17. **Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O.** (2015). Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ*, *3*, e1319. https://doi.org/10.7717/peerj.1319
18. **Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A., & Singer, S. W.** (2014). MaxBin: An automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome*, ***2***(1), 26. https://doi.org/10.1186/2049-2618-2-26
19. **Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z.** (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, *7*, e7359. https://doi.org/10.7717/peerj.7359
20. **Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W.** (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, ***25***(7), 1043–1055. https://doi.org/10.1101/gr.186072.114
21. **Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., & Banfield, J. F.** (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, ***3***(7), 836–843. https://doi.org/10.1038/s41564-018-0171-1
22. **Seemann, T.** (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, ***30***(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153
23. **Morgan-Lang, C., McLaughlin, R., Armstrong, Z., Zhang, G., Chan, K., & Hallam, S. J.** (2020). TreeSAPP: The Tree-based Sensitive and Accurate Phylogenetic Profiler. *Bioinformatics*, ***36***(18), 4706–4713. https://doi.org/10.1093/bioinformatics/btaa588
24. **Parte, A. C.** (2014). LPSN—list of prokaryotic names with standing in nomenclature. *Nucleic Acids Research*, ***42***(D1), D613–D616. https://doi.org/10.1093/nar/gkt1111
25. **Parte, A. C.** (2018). LPSN – List of Prokaryotic names with Standing in Nomenclature (bacterio.net), 20 years on. *International Journal of Systematic and Evolutionary Microbiology*, ***68***(6), 1825–1829. https://doi.org/10.1099/ijsem.0.002786
26. **Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., & Göker, M.** (2020). List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology*, ***70***(11), 5607–5612. https://doi.org/10.1099/ijsem.0.004332
27. **Euzéby, J.P.** (1997). List of Bacterial Names with Standing in Nomenclature: A Folder Available on the Internet. *International Journal of Systematic and Evolutionary Microbiology*, ***47***(2), 590–592. https://doi.org/10.1099/00207713-47-2-590
28. **Schoch CL, et al.** (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford). Retrieved December **10**, 2020, from https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi
29. **Chen, J.-S.** (2005). Nitrogen Fixation in the Clostridia. In W. Klipp, B. Masepohl, J. R. Gallon, & W. E. Newton (Eds.), *Genetics and Regulation of Nitrogen Fixation in Free-Living Bacteria* (53–64). Springer Netherlands. https://doi.org/10.1007/1-4020-2179-8_3
30. **Kadnikov, V. V., Mardanov, A. V., Podosokorskaya, O. A., Gavrilov, S. N., Kublanov, I. V., Beletsky, A. V., Bonch-Osmolovskaya, E. A., & Ravin, N. V.** (2013). Genomic Analysis of Melioribacter roseus, Facultatively Anaerobic Organotrophic Bacterium Representing a Novel Deep Lineage within Bacteriodetes/Chlorobi Group. *PLOS ONE*, ***8***(1), e53047. https://doi.org/10.1371/journal.pone.0053047
31. **Garrity, G. M., Holt, J. G., Reysenbach, A.-L., Huber, H., Stetter, K. O., Zillig, W., Itoh, T., Suzuki, K.-I., Sanchez, P. C., Nakase, T., Huber, R., Sako, Y., & Nomura, N.** (2001). Phylum Al. Crenarchaeota phy. Nov. In D. R. Boone, R. W. Castenholz, & G. M. Garrity (Eds.), *Bergey's Manual® of Systematic Bacteriology: Volume One: The Archaea and the Deeply Branching and Phototrophic Bacteria* (169–210). Springer. https://doi.org/10.1007/978-0-387-21609-6_16
32. **Könneke, M., Schubert, D. M., Brown, P. C., Hügler, M., Standfest, S., Schwander, T., Schada von Borzyskowski, L., Erb, T. J., Stahl, D. A., & Berg, I. A.** (2014). Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proceedings of the National Academy of Sciences of the United States of America*, ***111***(22), 8239–8244. https://doi.org/10.1073/pnas.1402028111
33. **Pollet, T., Berdjeb, L., Garnier, C., Durrieu, G., Le Poupon, C., Misson, B., & Briand, J.-F.** (2018). Prokaryotic community successions and interactions in marine biofilms: The key role of Flavobacteriia. *FEMS Microbiology Ecology*, ***94***(fiy083). https://doi.org/10.1093/femsec/fiy083
34. **Slade, D., & Radman, M.** (2011). Oxidative Stress Resistance in Deinococcus radiodurans. *Microbiology and Molecular Biology Reviews*, ***75***(1), 133–191. https://doi.org/10.1128/MMBR.00015-10
35. **Gupta, R. S.** (2000). The phylogeny of proteobacteria: Relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiology Reviews*, ***24***(4), 367–402. https://doi.org/10.1111/j.1574-6976.2000.tb00547.x
36. **Costello, M. J., & Chaudhary, C.** (2017). Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation. *Current Biology*, ***27***(11), R511–R527. https://doi.org/10.1016/j.cub.2017.04.060