Volume 7: 1-13

UJEMI+

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

Taxonomic diversity and abundance profiles of denitrification genes *nirK* and *nirS* in the Saanich Inlet water column

Catherine WY Wong, Cathy Yan, Josie Setiawan, Erin Tanaka, Daniela Yanez
Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

SUMMARY Oxygen minimum zone (OMZ) expansion throughout the global ocean prompts investigation into deoxygenation effects on marine microbial communities. Nitrogen cycling in OMZs is of interest because in anoxic conditions, nitrate is an alternative electron acceptor by various microorganisms resulting in biological nitrogen loss through denitrification. Previous research has highlighted the importance of two nitrite reductase genes, *nirK* and *nirS*, when assessing denitrification potential in marine ecosystems. However, more research on less prominent taxonomic groups is needed to better understand pathway distribution patterns. In this study, environmental samples are collected at 100, 120 and 200 m depths of the Saanich Inlet and metagenomic approaches were applied to analyze the differences in diversity and abundance of *nirK* and *nirS*. FastQC and Trimmomatic were utilized to trim and quality filter raw reads. MEGAHIT assembled the reads into contigs and TreeSAPP, ggplot, iTOL and Anvi'o were used to evaluate the abundance of *nirK* and *nirS* across taxonomies and the 3 depths. The results found *nirK* counts to be higher at 100 m and 120 m, but decreased at 200 m, whereas *nirS* steadily increased with depth. The difference is reflected in geochemical trends with oxygen levels highest near the surface and anoxic at 200 m. *nirS* was primarily associated with the Proteobacteria phylum and Thaumarcheota had the highest abundance for *nirK*. These results highlight differential abundance patterns of *nirK* and *nirS* genes in the water column relevant to the study of both ammonia oxidation and denitrification processes at different oxygen concentrations within OMZs.

INTRODUCTION

Importance of microbial metabolism in OMZs. Microorganisms utilize various energy sources for their survival, and the metabolic substrates and products they use depends on the geochemical environment and the surrounding microbial community. In marine ecosystems, oxygen minimum zones (OMZs) are widespread areas of low dissolved oxygen (1). Microbial communities use nitrogen and sulfur as alternate terminal electron acceptors and donors due to oxygen deficiency, contributing to the loss of fixed nitrogen in OMZs (1). These activities lead to changes in geochemical gradients of dissolved gasses and the production of greenhouse gasses such as nitrous oxide and methane (1). Due to anthropogenic activities, the ocean faces imminent changing conditions such as rising temperatures, increased stratification, acidification, and oxygen depletion, leading to an expansion of OMZs (2). Therefore, it is becoming increasingly important to study how microbial communities respond to these changes to better understand the impact of OMZ expansion on the future ocean.

Saanich Inlet as a model ecosystem for chemical and microbial interactions and measurements. The seasonally anoxic fjord, Saanich Inlet, provides a model ecosystem for studying microbial diversity and abundance in OMZs. Saanich Inlet is located on the

Published Online: September 2021

Citation: Catherine WY Wong, Cathy Yan, Josie Setiawan, Erin Tanaka, Daniela Yanez. 2021. **Taxonomic diversity and abundance profiles of denitrification genes *nirK* and *nirS* in the Saanich Inlet water column**. UJEMI+ 7:1-13

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

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

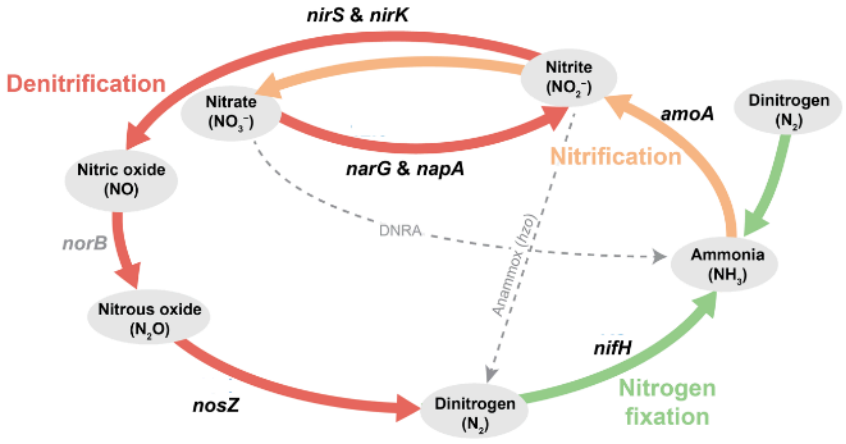


FIG. 1 Schematic of the nitrogen cycle. *nirS* and *nirK* catalyze the denitrification of nitrite to nitric oxide (6). The red arrows illustrate the four steps in the denitrification reaction, where each arrow represents a given gene or set of genes. The orange arrows depict the steps in the nitrification process and the green arrows represent the processes of nitrogen fixation. Figure was adapted from (6).

southeast coast of Vancouver, British Columbia, Canada, and is characterized by its distinctive water circulation properties, attributed to the shallow entrance sill situated at a depth of 75 m (3). This geological feature hinders deep water circulation beyond the extent of the basin, which creates a hypoxic environment at the bottom of the fjord. In the spring and early summer, coastal wind patterns cause partial mid-depth water renewal, which results in seasonal cycles of water oxygenation, with alternating phases of hypoxia and oxygenation (3). This annual pattern in oxygen availability creates a useful model for studying hypoxic marine microbial metabolism and its influence on the ocean's composition of metabolic substrates and biogeochemical cycles.

Metagenomic approaches. Metagenomics has been developed as a powerful, culture-independent technique that allows for the exhaustive identification of microorganisms present in complex environmental samples and the linkage of microbial taxonomy to gene function and outcomes. Utilizing data from next-generation sequencing platforms, metagenomics helps with the study of microbial diversity, and how microbial communities shape the physical and chemical properties of their environments (5). The process involves DNA extraction, library preparation using PCR, sequencing, assembly, functional annotation, and taxonomic profiling (5). One key step and major challenge in the metagenomics approach is binning, which identifies and merges contigs belonging to the same population. To perform this process, short reads are assembled into contigs and sorted into clusters representing the same source population.

Denitrification & *nirK/nirS*. Denitrification steps have been of particular interest in previous studies due to their contribution in releasing harmful greenhouse gases, such as nitrous oxide as well as resulting in a loss of fixed nitrogen (3). As shown in Figure 1, the reduction of nitrite (NO_2) to nitric oxide (NO) is catalyzed by nitrite reductase, which has two structurally different forms (6) (7). The *nirK* gene encodes a copper-containing nitrite reductase, whereas the *nirS* gene encodes a cytochrome cd_1 nitrite reductase (7). These genes play a role in allowing microorganisms to use nitrite as an alternative terminal electron acceptor during their metabolic processes when oxygen is not readily available (3). The denitrification produces nitric oxide, which also has a proposed role as a redox shuttle in aerobic ammonia oxidizers. The product of the denitrification pathway is dinitrogen, which in turn is the reactant for nitrogen fixation, resulting in ammonia (Figure 1) (6). The result of ammonia oxidation is nitrite which is an essential reactant in the first denitrification step (Figure 1) (6). There is competition for nitrite from ammonia oxidation between complete denitrification and the anammox pathway (8), which makes the step catalyzed by nitrite reductase a significant point of the cycle to investigate.

Phylogeny & *nirK/nirS* Abundance. Previous work has found that genes for the various steps of denitrification (*napA/narG, nirK/nirS, norB, nosZ*) were widely present throughout the Saanich Inlet water column (9). Although patterns in functional denitrifying gene

abundance between different depths of the Saanich Inlet have been investigated, few studies on the Saanich Inlet have specifically focused on comparing *nirK* and *nirS,* their distinct expression in different phylogenies of bacteria and the reason behind the disparity. Previous research has revealed a differential expression of these two genes in different environments, such as a greater abundance and diversity of bacteria encoding *nirS* genes compared to *nirK* in the South China Sea and Liaohe Estuary (10) (11). Furthermore, a study on soil denitrifier communities showed differences in *nirK* and *nirS* gene abundance with varying soil pH, with fewer *nirS* genes expressed at lower pH levels (12).

An extensive phylogenetical composition of the two denitrification genes is also under-explored in the literature. Previous works have investigated the interaction between denitrification rates to shifts in microbial community composition. Michiels et al. (8) found that SUP05 Gammaproteobacteria dominated during periods of intense denitrification and were heavily involved in reducing nitrate to nitrous oxide (8). Zaikova et al. (13) analyzed microbial community dynamics at different depths of the Saanich Inlet by quantifying SSU rRNA genes and found that classes within the Proteobacteria phylum dominated in abundance across all depths. Gammaproteobacteria were the most abundant, with the SUP05 clade of gammaproteobacteria notably present at greater depths of 125 to 215 m (13). Rhodobacterales, SAR11, Arctic96B-1, Bacteroides, Microthrix were also identified but mainly present at more shallow depths of 10 m (13).

Hypothesis and Objectives. Here a study of the Saanich Inlet water column was conducted using metagenomics methods to chart the abundance and distribution of *nirK* and *nirS* genes at depths of 100, 120, and 200 meters. Previous research has highlighted how microbial communities respond to changes in their environment, as well as the importance of considering both nitrite reductase genes when assessing denitrification (12). Studies have also underlined the significant role that Gammaproteobacteria have in Saanich Inlet denitrification. However, more research on less prominent taxonomic groups involved in specific denitrification steps, such as nitrite reduction, is needed to gain a more holistic understanding of the metabolic interactions involving microbial communities. Due to the geochemical gradient of oxygen and nitrogen species present across these depths, we predict that the abundance of, and phylogenies encoding nitrite reductases would differ.

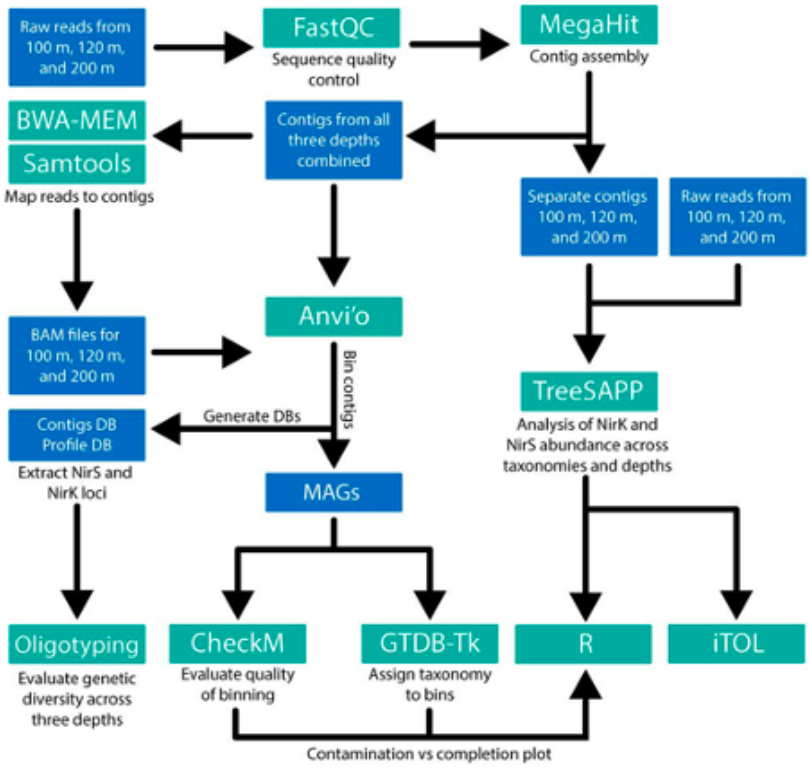


FIG. 2 Project workflow for DNA sequence data analysis. Programs used were FastQC, MegaHit, Bwa-Mem, Samtools, Anvi'o, Oligotyping, CheckM, GTDB-Tk, R, iTol and TreeSAPP.

METHODS AND MATERIALS

Environmental Sampling. Sampling methods used in this study are as described by Hawley et al. (14). Water samples were collected aboard the *MSV John Strickland* station S3 Cruise SI072 at the Saanich Inlet from depths 100 m, 120 m and 200 m (14). The water samples were collected in 2x12 Go-Flow bottles on a wire and after sampling for dissolved gases to decrease microbial gene expression changes, the water was immediately separated into Nalgene bottles with a sterile silicone tubing (14). In the Nalgene bottles, the water was filtered through a 0.22 µm Sterivex filter with a 2.7 µm GDF pre-filter for biomass collection (14). Residual seawater was removed by extrusion and 1.8 mL of sucrose lysis buffer was added to the Sterivex filters (14). The filters were then stored on dry ice until it was brought to the lab and stored at -80°C until DNA extraction (14).

Environmental DNA Extraction. Environmental DNA extraction methods used in this study are as described by Hawley et al. (14). The Sterivex filters were first thawed on ice and then lysozyme was added to the filters and rotationally incubated for 1 hr at 37°C (14). After 1 hr incubation, Proteinase K and 20% SDS was added and rotationally incubated for 2 hr at 55°C (14). The extrusion method was utilized to remove the lysate and after the removal, sucrose lysis buffer was used to rinse the filters (14). Extraction of combined lysate was performed with phenol: chloroform followed by chloroform (14). The collection of the aqueous layer was followed by an addition onto a 10 K Amicon filter cartridge and washed with TE buffer at pH 8.0 three times (14). Lastly, centrifugation was utilized for concentrating the aqueous layer to 150 - 400 µl (14).

Environmental DNA Sequencing. Environmental DNA sequencing methods used in this study are as described by Hawley et al. (14). PCR amplification of samples was completed with a 926F forward and 1392R reverse primers of the small subunit ribosomal RNA V6-V8 hypervariable region (14). A QIAquick PCR Purification Kit was used to purify the samples and paired end sequencing on an Illumina MiSeq platform was conducted using a 454-pyrosequencing at the McGill University Genome Quebec Innovation Center or DOE Joint Genome Institute (JGI) (14).

DNA Sequencing Data Workflow. After Illumina sequencing, FastQC (15) and Trimmomatic (16) were utilized to trim and quality filter the raw reads in the form of .fastq files from 100, 120 and 200 m depths (Figure 2). With the quality reads, MEGAHIT (17) was used to assemble them into contigs (Figure 2). MEGAHIT was used as the contig assembler as opposed to other assemblers such as SPAdes, IDBA-UD and metaSPAdes due to the significantly fewer computational resources required with an output of similar moderate and high-quality assemblies (18). SPAdes traditionally provided the best assembly statistics but MEGAHIT has a significantly lower memory and required the shortest time to assemble this complex dataset of 3 different depths. Thus, MEGAHIT was more computationally inexpensive compared to SPAdes, and was chosen as the program of choice (18).

The separate contigs for the three depths from MEGAHIT and trimmed quality reads from FastQC were placed through TreeSAPP (19) to evaluate the abundance of *nirK* and *nirS* genes expressed across taxonomies and the 3 depths (Figure 2). R Studio 3.6.0 (20) ggplot within the tidyverse was used with TreeSAPP marker output files for visualization of *nirK* and *nirS* relative abundance across depths and taxa at class level with a bubble plot and the distribution of *nirK* and *nirS* raw counts across depths and taxa at class level was showcased with a stacked bar chart (Figure 2). The Interactive Tree of Life (iTOL) (21) was utilized to determine taxonomic placements of *nirK* and *nirS* genes and the reads per kilobase of transcript per million mapped (RPKM) from TreeSAPP marker output files (Figure 2).

In a separate pipeline, the raw reads from the three depths were mapped to the combined contigs of the three depths from MEGAHIT with BWA-Mem (22) and BAM files were generated with Samtools (23) (Figure 2). The reasons for using BWA-Mem as a mapper is similar to the use of MEGAHIT. Another mapper, GMAP has the highest number of

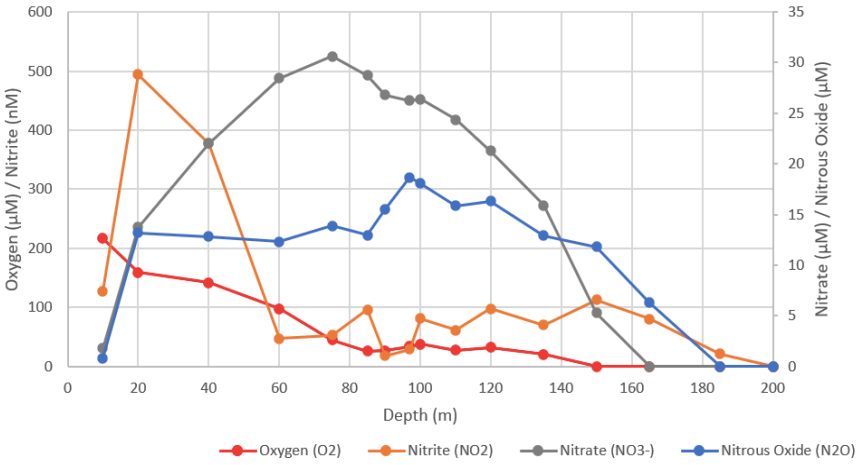


FIG. 3 Geochemical data from Saanich Inlet on August 1, 2012. Mean concentrations of oxygen (O_2), nitrite (NO_2^-), nitrate (NO_3^-), and nitrous oxide (N_2O) in water samples collected at various depths are shown in comparison with each other.

alignments but is computationally expensive with a significantly longer mapping time compared to BWA-Mem (24). BWA-Mem had less alignments compared to GMAP but aligned more than Bowtie and used less computational time and memory than GMAP and Bowtie (24). Anvi'o (25) was used to run BAM files from Samtools to predict open reading frames, identify marker genes and group information into bins using DasTools (Figure 2). DasTools with default settings was used to give coverage information about each bin. DasTools was chosen over CONCOCT, MaxBin2 and MetaBAT because it has been shown to recover substantially higher near-complete genomes compared to the other mentioned binning tools (26). The Anvi'o output produced metagenome assembled genomes (MAGs) and the quality of the MAGs binning was assessed with CheckM (27) (Figure 2). GTDB-Tk (28) assigned taxonomic classification and functional annotation of the MAGs (Figure 2). R ggplot tidyverse was used to produce a contamination vs completion plot with the MAGs, CheckM and GTDB-Tk (Figure 2).

Aside from predicting open reading frames, identifying marker genes and grouping information into bins, Anvi'o was also used to generate a contigs database (contigs DB) and a profile database (profile DB) (Figure 2). The sequences for *nirS* and *nirK* loci were extracted from the contigs DB and profile DB by using gene names as search terms. The resulting FASTA files were translated into amino acids, aligned, and translated back into DNA. Oligotyping (29) was used to evaluate the genetic diversity of the conserved amino acids within the active sites of these two genes across the three depths (Figure 2). The inputs for this tool were the aligned DNA FASTA file and numerical positional indices for the nucleotides of interest (i.e. the ones encoding for amino acids within the active sites).

RESULTS

Geochemical trends of the Saanich Inlet relating to the nitrogen cycle. Through a comparison of the oxygen, nitrate, nitrite, and nitrous oxide concentrations present at various depths of the Saanich Inlet from geochemical data collected (9), several trends can be observed (Figure 3). The water appears to be under oxic conditions between 0-60 m depths (>90 µM O_2), and gradually decreases in O_2 concentration beyond 60 m depth, finally reaching anoxic conditions at 150 m. The regions between 60-150 m can be characterized as dysoxic conditions (90-20 µM O_2) (9). As for the abundance of compounds involved in the nitrogen cycle at different depths, nitrite decreased drastically from a maximum of 500 nM to about 50 nM as conditions became dysoxic, and then maintained a fairly consistent concentration until decreasing again from 100 nM to 0 nM within anoxic conditions. In contrast, nitrate and nitrous oxide levels were at their highest within dysoxic conditions. Similar to nitrite, they also decreased in concentration as conditions became anoxic. Nitrous oxide was most abundant between depths of 100-120 m.

The *nirS* and *nirK* genes are differentially abundant across depths and taxa. As NirK and NirS catalyze nitrite reduction via different mechanisms, we sought to determine how

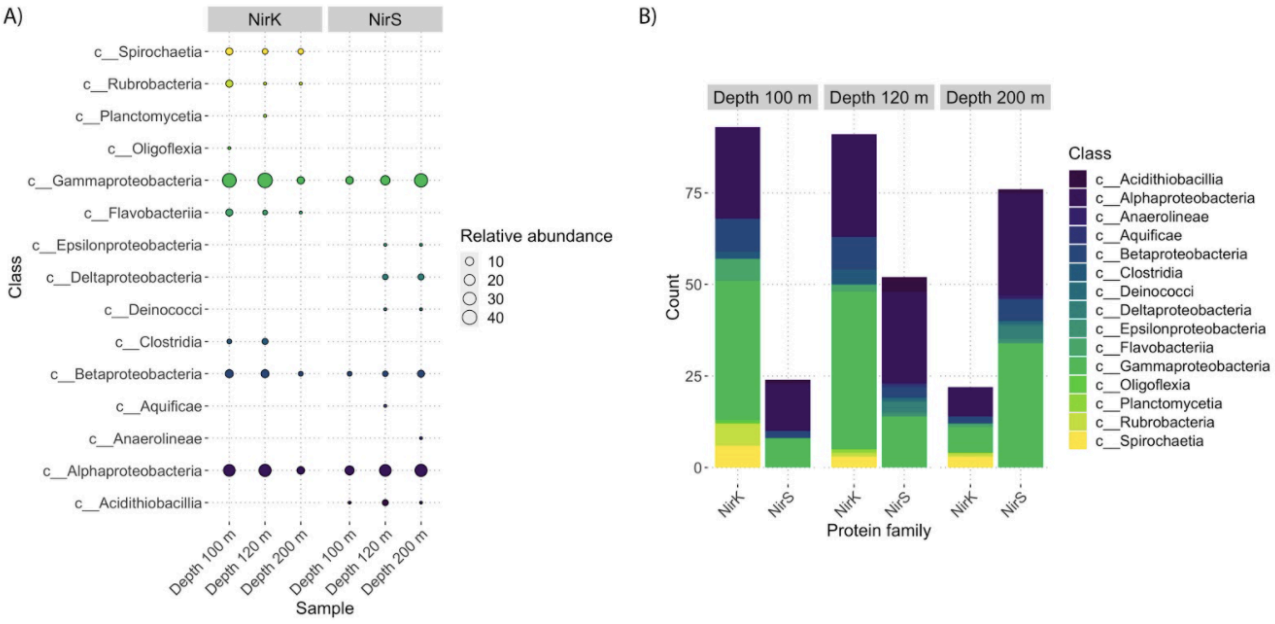


FIG. 4 *nirS* and *nirK* are differentially abundant across depths and taxa. (A) Bubble plot mapping *nirK* and *nirS* to different classes of organisms at 100 m, 120 m, and 200 m. The size of the circles represents relative abundance. (B) Stacked bar chart demonstrating how *nirK* and *nirS* are distributed across depths and taxa at the class level. Bar heights represent raw counts.

they are distributed across depths and whether they are encoded by the same taxa. Using TreeSAPP, ORFs for nir*K* and *nirS* were identified from contigs and assigned taxonomy. Broadly, while some classes of microbes like Gammaproteobacteria, Betaproteobacteria, and Alphaproteobacteria have both *nirK* and *nirS*, others only have one variant (Figure 4A). For both genes, however, the majority of reads are mapped to Gammaproteobacteria or Alphaproteobacteria. Furthermore, *nirK* and *nirS* were differentially abundant across depths. *nirK* counts are similar at depths of 100 m and 120 m but decrease sharply at 200 m (Figure 4B). *nirS* counts, on the other hand, increase steadily with depth.

The *nirK* gene maps to a more diverse range of taxa at the class level than *nirS*. Next, we sought to characterize how taxa encoding for these genes are phylogenetically related. Tree files and files with taxonomic and read abundance information for *nirK* and *nirS* were generated using TreeSAPP and visualized in iTOL (Figure 5). In general, the taxa dominant in *nirK* and *nirS* abundance are the same across all three depths. However, whereas *nirS* sequences primarily map to the classes Delta-, Alpha-, Beta-, and Gammaproteobacteria, *nirK* is more diversely distributed across classes including Bacilli, Alphaproteobacteria, and Deinococci, and the phylum Thaumarchaeota.

MAGs curated using Anvi'o have low contamination and moderate completion. Before oligotyping, we first assessed the quality of bins curated using Anvi'o. For all medium and high-quality MAGs, contamination was below ten percent (Figure 6). Completion rates spanned a wide range, but most were below fifty percent. At the class level, there were no trends between taxa and completeness or contamination. For original bins generated using MaxBin, MetaBat, CONCOCT, and DasTools, reads from the three depths generally mapped to the same ones (Figure 7A-D). The only exception is DasTools, where reads from 100 m and 120 m were mapped to different bins compared to 200 m (Figure 7D). Contamination and completion levels varied across the tools. MetaBat and CONCOCT had the highest completion and contamination rates (Figure 7B-C), while MaxBin had moderate levels of both (Figure 7A). DasTools had the lowest contamination rate while retaining high completion rates, though it also generated the fewest bins.

FIG. 5 *nirK* maps to a more diverse range of taxa at the class level than *nirS*. TreeSAPP reference tree mapping of *nirK* and *nirS* genes across different classes visualized in iTOL. Bar heights around the tree indicate metagenomic abundance levels (RPKM) of each gene, and bar colours indicate depth. Coloured ranges represent the different taxonomic classes, and red circles denote to which microbes genes are mapped. Taxa to which *nirS* or *nirK* are mapped are annotated at the class level.

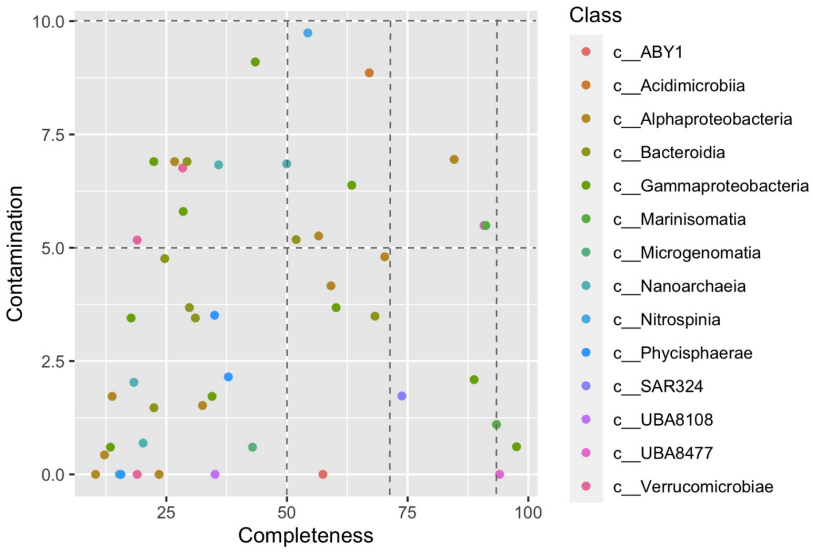


FIG. 6 MAGs curated using Anvi'o have low contamination and moderate completion. Contamination vs completeness plot for medium and high-quality bins curated from Anvi'o. Contamination and completion values were calculated using Check-M, and taxonomy was assigned using GTDB-Tk. Colours represent taxa at the class level.

The NirK active site is more conserved than the NirS active site across all depths. Based on the sequences encoding for two and three conserved amino acid residues in the active sites of NirK and NirS, respectively (Figure 8), *nirK* is more conserved than *nirS* across all depths. Looking at only the oligotypes where all nucleotides are present, the same two oligotypes dominate at all three depths for *nirK* (Figure 9). The number of oligotypes present does not change depending on the number of sequences extracted from the sample. Conversely, new oligotypes appear for *nirS* when there are more sequences in the sample that encode for the gene. The oligotypes also differ across depths, and more are shared between 120 m and 200 m than with 100 m.

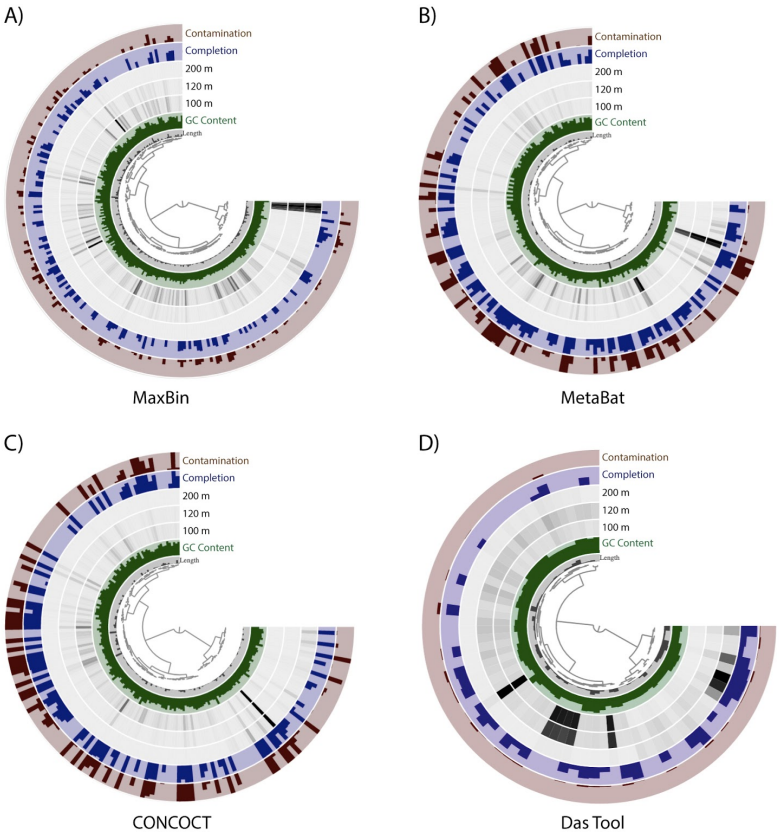


FIG. 7 Binning algorithms differ in contamination and completion rates. (A - D) Diagrams from Anvi'o interactive showing (from outermost to innermost layer) contamination, completion, where reads map to at 200 m, 120 m and 100 m, GC content, and Newick tree diagram for bins generated using MaxBin, MetaBat, CONCOCT, and DasTools.

DISCUSSION

How NirK and NirS, two functionally equivalent but structurally different nitrite reductases, are distributed across depths and phylogenies in Saanich Inlet, a model ecosystem for OMZs, has not yet been thoroughly investigated. Previously, studies have found that *nirS* and *nirK* are differentially abundant in shorelines along the Pacific northwest (30), cropland soils (31), and forested wetlands (32). In the Saanich Inlet, we found that *nirK* and *nirS* were differentially abundant at different depths, with *nirK* present at higher levels near the surface and most *nirS* found at 200 m (Figure 4). This disparity is reflected in the geochemical trends across depths (Figure 3). Firstly, oxygen levels are highest near the surface, but the waters are anoxic at 200 m. In an OMZ near northern Chile, it was found that higher O_2 levels suppressed transcription of *nirS*, but not *nirK* (33). This serves as a possible explanation for *nirS* being most abundant at the deepest part of the Inlet. Conversely, the genes' abundances also impact the concentration of N_2O, as *nirS* co-occurs with nitric oxide reductases and nitrous oxide reductases more frequently than *nirK* (34). Consequently, there was less N_2O due to it being completely denitrified into N_2. We see a similar trend in the Saanich Inlet where N_2O is relatively steady across the water column but decreases rapidly with anoxic conditions from 150 m onwards and is not present at 200 m where *nirS* is most abundant.

Regarding the distribution of *nirK* and *nirS* across taxa, both have been primarily found in the phylum Proteobacteria (30), which aligns with our data (Figure 4). When looked at in higher resolution using iTOL, we discovered that *nirK* is distributed more evenly across a broader range of organisms than *nirS*, which is dominated by Proteobacteria (Figure 5). Proteobacteria is one of the most successful microorganisms on earth. They are involved in the denitrification pathway in hypoxic and hypolimnetic areas (35). Furthermore, in a recent study, members of this phylum can perform chemolithotrophic and chemoorganotrophic metabolisms to thrive in extreme environments (36). It is not unexpected to see such a diverse spectrum of proteobacteria in our *nirS* and *nirK* data.

Surprisingly, the phylum with the highest RPKMs for the *nirK* gene was Thaumarcheota. Thaumarchaeota from Marine Group I (MGI) is one of the most abundant chemoautotrophs within the dark ocean and is found in very large numbers throughout the water column (37). Archaea from the phylum *Thaumarchaeota* are organisms that contribute to the

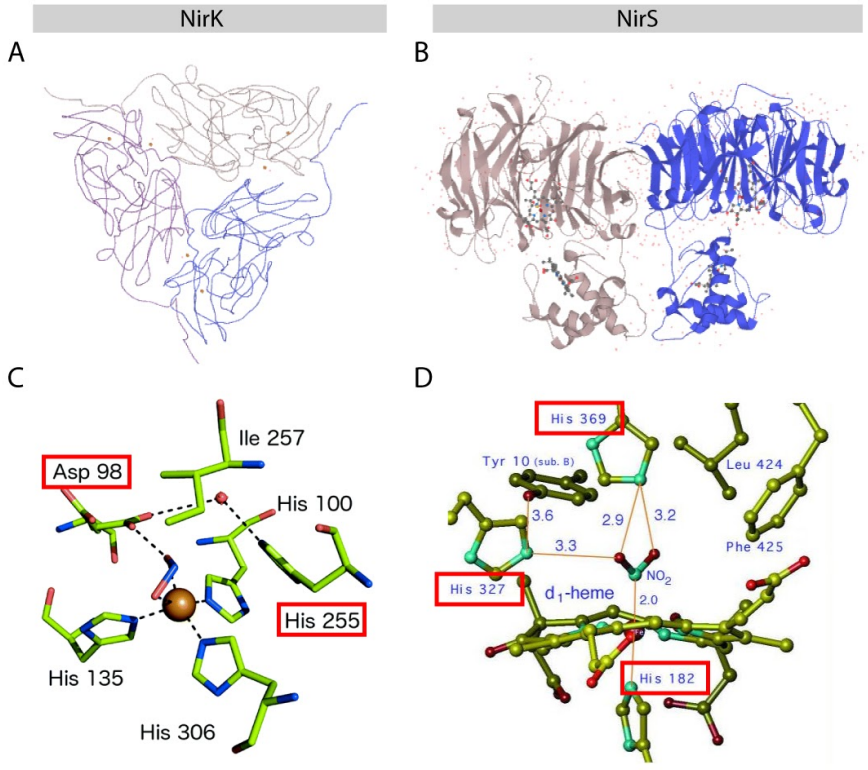


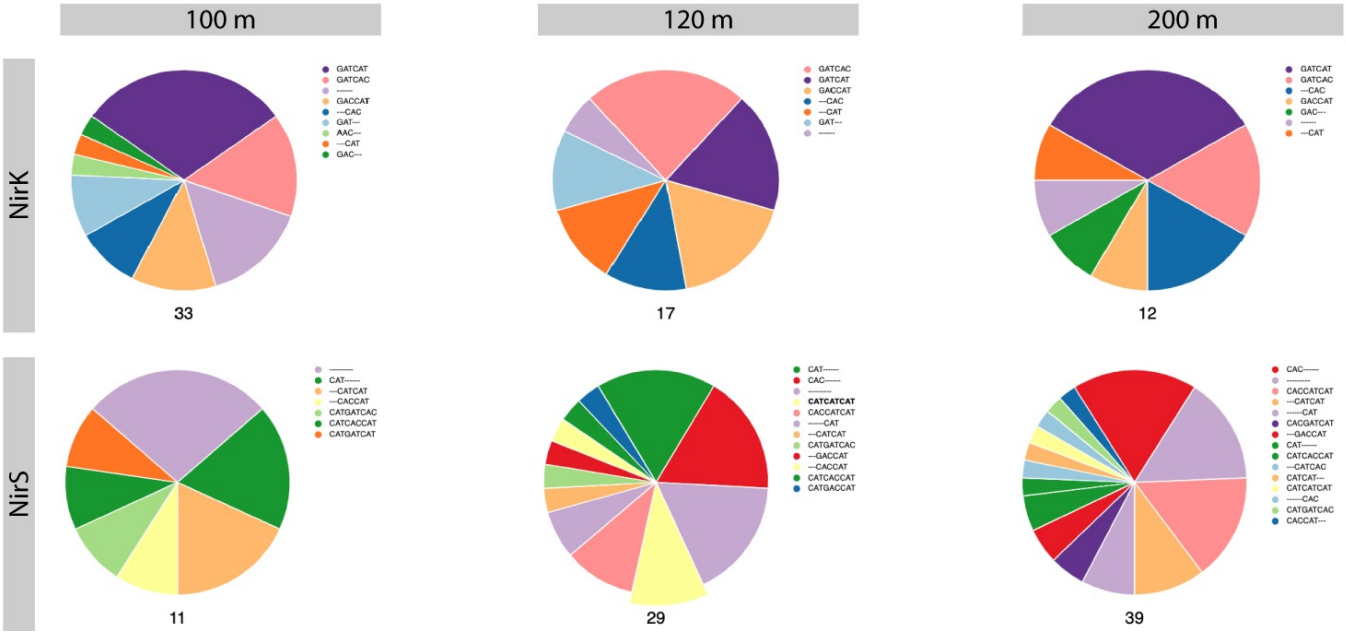
FIG. 8 NirK and NirS have structurally different active sites. (A, B) X-ray crystallography images for NirK and NirS from UniProt (accession IDs: P81445, P72181). (C, D) Diagrams of NirK and NirS active sites obtained from Horrell et al. and Cutruzzola et al, respectively. Conserved residues of interest are boxed in red.

biogeochemical cycling of both nitrogen and carbon (38). They fix CO_2 by using the energy from oxidizing ammonia (NH_3) to nitrite (NO_2^-) and have been shown to be one of the major ammonia oxidizers in the marine environment (37). Thaumarchaeota is one of the main players of the nutrient cycle in the water column, however, ecology and metabolism information remains unclear. As an ammonia oxidizer, *Thaumarchaoeta* has been clustered using the *amoA* gene, but recently, studies had assessed the *nirK* gene as an effective molecular marker for tracking the thaumarchael populations (37). Consistent with our findings in Figure 5, Thaumarchaeota highly expresses *nirK* at both surface and deeper levels (38). Interestingly, even within some bacterial species like *Nitrosomonas europaea*, *nirK* is implicated in ammonia oxidation (39). This functionality differentiates it from *nirS*. Based on our findings and results from literature, *nirK* has the potential to be used as a marker for determining the heterogeneity and community composition of thaumarchaeota and ammonia-oxidizing bacteria.

Delving deeper into the idea of heterogeneity, we used oligotyping to assess the diversity of *nirK* and *nirS* sequences. Since our MAGs have low completion rates (Figure 6A), we chose to profile using only the nucleotides encoding for conserved amino acids within the active sites. For NirK, these were aspartic acid and histidine in the T2CU site, which is where nitrite reduction takes place (40). These residues, Asp 98 and His 255, are responsible for the rate-determining step of providing two protons to the bound nitrite to yield nitric oxide as the product. For NirS, the conserved residues were three histidines that utilize heme as an electron donor for nitrite reduction (41). His 182, 327, and 369 directly interact with to stabilize the nitrite molecule, and are essential for maintaining the enzyme's conformation. Despite using only conserved amino acids, the low completion of our MAGs still led to gaps in our oligotypes (Figure 9). Looking at only the complete oligotypes, *nirK* was more conserved than *nirS*, which aligns with results from literature (32)(42). However, since the variations do not alter which amino acid is coded for, they are unlikely to impact functionality and instead indicators for phylogenetic diversity.

In summary, *nirK* and *nirS* both mediate nitrite reduction in the denitrification pathway, and the former has also been linked to ammonia oxidation. Through metagenomic analyses, we found that they are differentially abundant across depths, and are encoded by different phyla. Furthermore, we uncovered how they vary in their level of sequence conservation even

FIG. 9 The NirK active site is more conserved than the NirS active site across all depths. Pie charts generated by oligotyping showing the distribution of oligotypes across depths for each gene. Colours represent different oligotypes. The topmost segment corresponds to the first item listed in the legend and, moving clockwise, the next piece is referred to by the next item. The number of sequences extracted from each sample is shown under their respective pie chart.



at the active sites. The findings detailed in this paper allude to the diversity of microbial denitrifiers in OMZs. As climate change leads to decreased oceanic oxygenation, unveiling communities responsible for fixed nitrogen loss and greenhouse gas production is crucial.

Limitations Since we do not have transcriptomic data, we cannot draw any conclusions regarding the expression of *nirK* and *nirS*. In terms of technical limitations, the contigs used for the Anvi'o contigs database may vary from the ones used to generate TreeSAPP outputs. This is because, for TreeSAPP, in order to get separate counts for each depth, separate files had to be used. Anvi'o, on the other hand, was able to recognize that contigs came from different depths within one file. Additionally, for Oligotyping, FASTA files with *nirK* or *nirS* sequences had to be translated into amino acids before being aligned. The alignment file was then translated back into DNA. We did this because aligning the DNA sequences right away led to poorer outcomes that obscured the location of the amino acids in the active site.

Conclusions Despite being functionally-equivalent, *nirK* and *nirS* are distributed differently across depths and taxa in the Saanich Inlet water column, a model ecosystem for OMZs. Uncovering the location, taxonomic distribution, and diversity of these two genes enables a better understanding of microbial communities involved in denitrification and, in the case of *nirK*, ammonia oxidizers.

Future Directions Since the geochemical gradient in Saanich Inlet undergoes seasonal cycling, we aim to investigate patterns in the distribution of *nirK* and *nirS* throughout the year. It would be fascinating to see if *nirK* and *nirS* dominance is a seasonal or constant phenomenon. Additionally, *nirK* and *nirS* are responsible for only one step of denitrification. To gain a comprehensive insight into the nitrogen cycle, we will also conduct more analyses on genes for other pathways. Furthermore, although there was no clear correlation between the expression pattern and the chemical compositions, other factors such as temperature, salinity, and density could have contributed to the observed differences. All three physical properties tend to be steadier as depth increases (9), so it would be interesting to study how *nirK* and *nirS* distribution are impacted by these environmental variables.

ACKNOWLEDGEMENTS

We would like to express our gratitude to the teaching team of MICB 405, including Dr. Steven Hallam, Dr. Martin Hirst and Dr Stephan Koenig for providing us with knowledge for this course and Julia Anstett, Axel Hauduc and Marty Yue for their support in this project. We would like to extend a special recognition to Dr. Steven Hallam and Julia Anstett for their guidance in this manuscript. We are also grateful to the University of British Columbia and the Department of Microbiology and Immunology for providing the funding and resources for this study. We would also like to thank two anonymous reviewers for constructive feedback on this manuscript.

REFERENCES

1. **Ulloa O, Canfield DE, DeLong EF, Letelier RM, Stewart FJ.** 2012. Microbial oceanography of anoxic marine zones. *PNAS* **109:**15996-16003.
2. **Trucco-Pignata PN, Hernandez-Ayon JM, Santamaria-del-Angel E, Beier E, Sanchez-Velasco L, Godinez VM, Norzagaray O.** 2019. Ventilation of the upper oxygen minimum zone in the coastal region off mexico: implications of El Nino 2015-2016. *Front Mar Sci* **6:** 459.
3. **Hallam SJ, Torres-Beltrán M, Hawley.** 2017. Monitoring microbial responses to ocean deoxygenation in a model oxygen minimum zone. *Sci Data* **4:**170158.
4. **Wade W.** 2002. Unculturable bacteria--the uncharacterized organisms that cause oral infections. *J R Soc Med* **95:** 81–83.
5. **Delmont TO, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P, Hirsch PR, Vogel TM.** 2011. Accessing the soil metagenome for studies of microbial diversity. *Appl Environ Microbiol* **77**: 1315-24.
6. **Griffith JC.** 2016. Insights into the soil microbial communities in New Zealand's indigenous tussock grasslands. PhD Thesis. University of Otago.

7. **Murdock SA, Juniper SK.** 2017. Capturing compositional variation in denitrifying communities: a multiple-primer approach that includes Epsilonproteobacteria. *Appl Environ Microbiol* **83:** e02753-16.
8. **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:**27.
9. **Torres-Beltrán M, Hawley A, 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.
10. **Ming H, Fan J, Chen Q, Su J, Song J, Yuan J, Shi T, Li B.** (2020). Diversity and abundance of denitrifying bacteria in the sediment of a eutrophic estuary. *Geomicrobiol J.*
11. **Li M, Hong Y, Cao H, Klotz MG, Gu JD.** 2013. Diversity, abundance, and distribution of NO-forming nitrite reductase-encoding genes in deep-sea subsurface sediments of the South China Sea. *Geobiology* **11:** 170-9.
12. **Herold MB, Giles ME, Alexander CJ, Baggs EM, Daniell TJ.** 2018. Variable response of *nirK* and *nirS* containing denitrifier communities to long-term pH manipulation and cultivation. *FEMS Microbiology Letters* **365:** fny035.
13. **Zaikova E, Hawley A, Walsh DA, Hallam SJ.** 2009. Seawater sampling and collection. *J Vis Exp* **28:** e1159.
14. **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, Rio TGD, Suttle CA, Tringe S, Hallam SJ**. 2017. A compendium of multi-omic sequence information from the Saanich Inlet water column. *Sci Data* **4:**170160.
15. **Andrews S.** 2010. FastQC: A quality control tool for high throughput sequence data. [Online].
16. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* btu170.
17. **Dinghua Li, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, Tak-Wah Lam.** 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31:** 1674–1676.
18. **van der Walt AJ, van Goethem MW, Ramond J, Makhalanyane TP, Reva O, Cowan DA.** 2017. Assembling metagenomes, one community at a time. *BMC Genomics.* **18:** 521.
19. **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:**1–8.
20. **RStudio Team.** 2020. RStudio: Integrated development for R. RStudio, PBC, Boston, MA
21. **Letunic I, Bork P.** 2006. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23:** 127-8.
22. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25:**1754-60.
23. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup.** 2009. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25:** 2078-9.
24. **Clausen PTLC, Aarestrup FM, Lund O.** 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMA Bioinformatics.* **19:** 307.
25. **Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO.** 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**:e1319.
26. **Sieber CM, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF.** 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* **3:** 836-843.
27. **Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW.** 2014. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* **25:** 1043-1055.
28. **Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH.** 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database, *Bioinformatics* **36:** 1925–1927.
29. **Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML.** 2013. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol Evol* **4:** 1111-1119.
30. **Braker G, Zhou J, Wu L, Devol AH, Tiedje JM.** 2000. Nitrite reductase genes (nirK and nirS) as functional markers to investigate diversity of denitrifying bacteria in pacific northwest marine sediment communities. *Appl Environ Microbiol* **66:** 2096-2104.
31. **Yang Y, Nie J, Wang S, Shi L, Li Z, Zeng Z**. 2020. Differentiated responses of *nirS*- and *nirK*-type denitrifiers to 30 years of combined inorganic and organic fertilization in a paddy soil. *Arch. Agron. Soil Sci* ***67****: 79-92.*
32. **Prieme A, Braker, G, Tiedje JM.** 2002. Diversity of nitrite reductase (*nirK* and *nirS*) gene fragments in forested upland and wetland soils. *AEM* **68:** 1893-1900.
33. **Dalsgaard T, Stewart FJ, Thamdrup B, De Brabandere L, Revsbech NP, Ulloa O, Canfield DE, DeLong EF.** 2014. Oxygen at nanomolar levels reversibly suppresses process rates and gene expression in anammox and denitrification in the oxygen minimum zone off northern Chile. *mBio* **5:** e01966-14.

34. **Graf DRH, Jones CM, Hallin S.** 2014. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N_2O emissions. *PLoS One* **9:** e114118.
35. **Mori JF, Chen L-X, Jessen GL, Rudderham SB, McBeth JM, Lindsay MBJ, Slater GF, Banfield JF, Warren LA.** 2019. Putative mixotrophic nitrifying-denitrifying Gammaproteobacteria implicated in nitrogen cycling within the ammonia/oxygen transition zone of an oil sands pit lake. *Front Micro* **10:** 2435.
36. **Zhou Z, Tran PQ, Kieft K, Anantharaman K.** 2020. Genome diversification in globally distributed novel marine Proteobacteria is linked to environmental adaptation. *ISME J* **14**: 2060–2077
37. **Reji L, Tolar BB, Smith JM, Chavez FP, Francis CA.** 2019. Depth distributions of nitrite reductase (nirK) gene variants reveal spatial dynamics of thaumarchaeal ecotype populations in coastal Monterey Bay. *Env Microbiol*. **21:** 4032-4045.
38. **Lund M, Smith J, Francis C.** 2012. Diversity, abundance and expression of nitrite reductase (*nirK*)-like genes in marine thaumarchaea. *ISME J* **6:** 1966–1977.
39. **Kozlowski J, Price J, Stein LY.** 2014. Revision of N_2O-producing pathways in the ammonia-oxidizing bacterium *Nitrosomonas europaea* ATCC 19718. *Appl Environ Microbiol* **80:** 4930-4935
40. **Horrell S, Kekilli D, Strange RW, Hough MA.** 2017. Recent structural insights into the function of copper nitrite reductases. *Metallomics* **9:** 1470-1482.
41. **Cutruzzola F, Brown K, Wilson EK, Bellelli A, Arese M, Tegoni M, Cambillau C, Brunori M.** 2001. The nitrite reductase from *Pseudomonas aeruginosa:* Essential role of two active-ste histidines in the catalytic and structural properties. *PNAS* **98**: 2232-2237.
42. **Braker G, Zhou J, Wu L, Devol AH, Tiedje JM.** 2000. Nitrite reductase genes (nirK and nirS) as functional markers to investigate diversity of denitrifying bacteria in pacific northwest marine sediment communities. *Appl Environ Microbiol*. **66:** 2096-2104.