Module 4: Project 2 by Team 5

Charting the distributed nitrogen cycle

Karl Abuan
May Ho
Jonah Lin
Leilynaz Malekafzali
Tiffany Yang
Abdur Rahman M. A. Basher

April 23, 2018

Abstract

Recent developments of high-throughput sequencing and various bioinformatics platforms have provided us with further understand of microbial functions on an ecological scale, including microbes' roles in global biogeochemical cycles. In particular, the analyses have provided us great insights and answers to important questions: which microbes are present, what are they doing, and how do they respond to changes in a given environment? Here, we use DNA and RNA sequence data to examine the nitrogen cycle in Saanich Inlet – which is an intermittently anoxic fjord and also a suitable model to study microbial community responses to changes in oxygen level. Although there are several well-defined gene encoding key steps in the nitrogen cycle, the primary investigation of this report focuses on the nitric oxide reductase, *norB*, and its abundance at multiple depths in Saanich Inlet. Through a new software pipeline, TreeSAPP, we found that DNA abundance of *norB* followed an upward trend, increasing from 2.5% at depth 100m to 10.0% at depth 200m. Similarly, RNA abundance was observed to increase from 5% at depth 150m to 15% at depth 200m. The resulting trees from TreeSAPP identified the different genera responsible for the presence of *norB* belonging to the Phyla Bacteriodetes, Chlorobi and Proteobacteria. While our results unveiled the abundance of *norB* at various depths in Saanich, the methods we used can be extended to other genes of interest in an effort to further examine the biogeochemical processes and predict microbial responses to changes or disturbances in an environment.

Contents

---

1 Introduction

Saanich Inlet, located in southeastern Vancouver Island, is a seasonally anoxic fjord and therefore an ideal model to study microbial community responses to ocean deoxygenation (1, 2). Established by the shallow sill opening to the Strait of Georgia, circulation of basin water in Saanich Inlet is reduced. Together with poor ventilation and intense downward flux of organic matter by aerobically respiring organisms living on surface waters (3), oxygen levels down the inlet water column decreases followed by changes in nutrient availability. With the lack of oxygen, microbes turn to use alternative terminal electron acceptors such as nitrate, followed by hydrogen sulfide which is developed at the bottom of the water column (4). Thus, these oxygen-deprived regions, termed as oxygen minimum zones (OMZs), post a diverse microbial community that mediate important biogeochemical processes such the nitrogen cycle (4). Since OMZs provide appropriate conditions to enable substantial nitrogen loss, we aim to reassemble the nitrogen cycle in Saanich Inlet to study the microbial community responses to varying levels of oxygen deficiency with specific emphasis on the denitrification pathway.

The Nitrogen Cycle

Nitrogen is the fourth most abundant element in the cellular biomass (5). Microbial activities entirely control the interchange between the inert N_2 gas in the atmosphere to usable nitrogen that can support cellular metabolism and growth (5). The nitrogen cycle consists of four significant nitrogen-transformation flows: ammonification, nitrification, denitrification, anammox (5). Ammonification, includes nitrogen fixation, where N_2 is reduced to ammonium by bacteria and archaea that encode nitrogenase enzyme, and also the assimilatory and dissimilatory reduction of nitrate to ammonium (5). It should be noted that nitrogen fixation is an oxygen sensitive process as nitrogenase enzyme is inhibited by oxygen (6). Nitrification is the process of oxidation of ammonia to nitrite, and further oxidation of nitrite to nitrate (5). Denitrification is the process of dissimilatory reduction of nitrate (NO_3^-) and nitrite (NO_2^-) to the gaseous oxides, nitric oxide (NO), and nitrous oxide (N_2O), which may further get reduced to dinitrogen gas (N_2) (5). In anaerobic ammonium oxidation commonly abbreviated as anammox, the pools of nitrite (N_2O^-) and ammonium (NH_4^+) are utilized to form N_2 (5). Anammox is used for wastewater treatment, as it removes both ammonium and nitrite simultaneously without producing N_2O (5).

The denitrification process plays an integral role in the global nitrogen balance. It returns the nitrogen into the atmosphere and maintains homeostasis between the nitrogen content of soil and air. (8). The four enzymes of denitrification are: nitrate, nitrite, nitric oxide and nitrous oxide reductase which are encoded by *nar*, *nir*, *nor*, and *nos*, respectively (9). Denitrification takes place in aquatic, marine, and terrestrial ecosystems, and the rate of this process is affected by various factors such as pH, temperature, and oxygen concentration (8). In OMZs, nitrate (NO_3^-) and nitrite (NO_2^-) act as the main terminal electron acceptors in respiration (10). Nitrogen oxide reductases are sensitive to oxygen and it is known that oxygen represses denitrification (8). Among the four N-oxide reductases, nitrous oxide reductase is the most sensitive enzyme to oxygen, leading to the increase

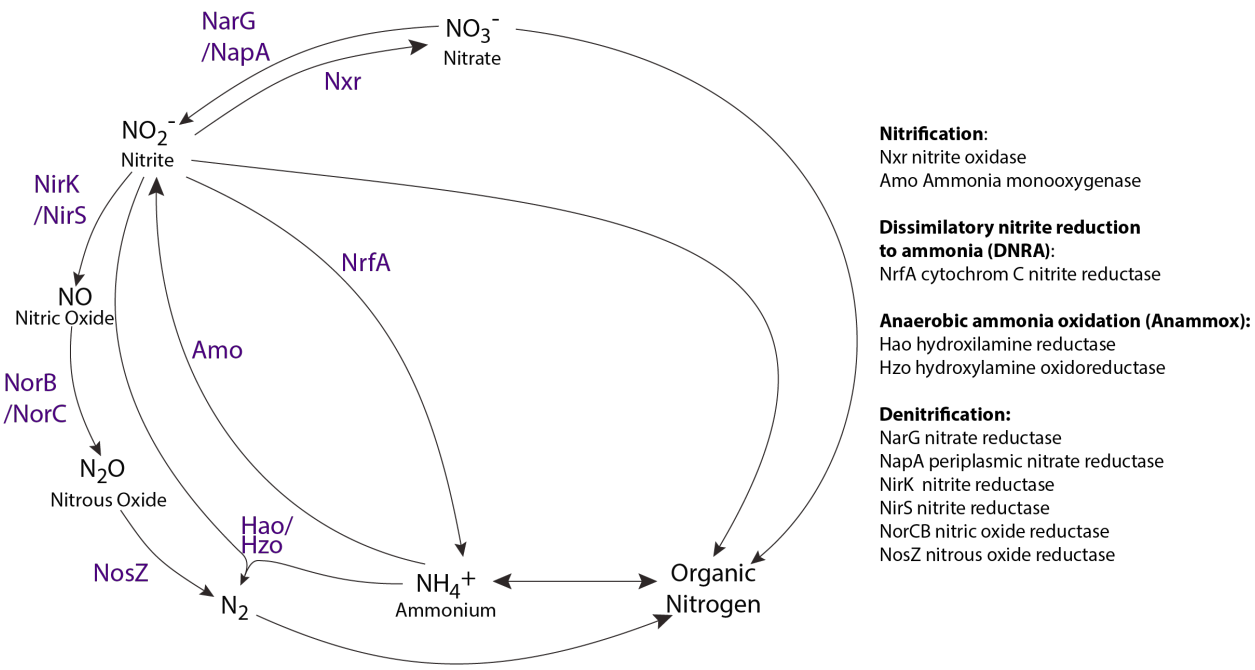

Figure 1: Nitrogen Cycle

in N_2O production in the suboxic and anoxic conditions (10). The interest in denitrification process is for several reasons. Denitrification is the process contributing to the loss of fertilizer nitrogen, therefore reducing the agriculture productivity (7). Additionally, denitrification can be utilized to remove nitrogen from high-nitrate wastewaters to decrease the contamination and eutrophication (7). However, denitrification can have negative effects on the climate system, as the emission of N_2O leads to stratospheric ozone layer depletion, contributing to global warming (7). Some of the major denitrifying bacteria are from genera: *Alcaligenes*, *Agrobacterium*, *Bacillus*, *Hyphomicrobium*, *Moraxella*, *Pseudomonas*, and *Paracoccus* (7). Due to the high frequency of isolation, the genera *Pseudomonas* and *Alcaligenes* are of the greatest significance (4).

The objective of this project is to reassemble the nitrogen cycle as a distributed metabolic process which may provide new insight into the roles that microbes play on nutrient and energy fluxes in the ocean. There are several well-defined gene encoding key steps in the nitrogen cycle and they serve as basis for functional anchor screening to determine their distribution across the Tree of Life. Here, we focused on the nitric oxide reductase, *norB*, and we examined its DNA and RNA abundance across the different depths in Saanich Inlet. We also investigated how the nitrogen species in Saanich relate to the abundances of *norB*.

2 Materials and Experimental Configuration

2.1 Experimental Protocols

To understand the correlation of microbial diversity and oxygen concentration across samples, we report four experimentally designed test protocols:

P1. Analysis of the DNA abundance of *norB* with depth.

P2. Analysis of similarities between RNA and DNA abundance information of *norB* with depths.

P3. Reconstructing the associated taxa with *norB* and analyze variances of DNA and RNA based on depths.

P4. Analysis of the abundance of *norB* in relation to nitrogen species in Saanich.

2.2 Dataset

2.3 Data Preporocessing

Manipulate the data into a single data frame

2.4 Methods

Sequences for *norB* were obtained from various depths at Saanich Inlet. Listed below are the method of processing the tree-based sensitive and accurate protein profiler (TreeSAPP):

1) The query sequences were first mapped to the reference sequences in BLASTp.
2) Coding sequences were then extracted and queried with putative functions
3) Profile-alignment (HmmAlign) with multiple alignment of references and query
4) Classification of sequence insertions were then done in reference trees (RAxML)
5) Reference trees were then updated with novel sequences
6) Construction of trees at various depths with Interactive Tree of Life (iTOL)

Majority of TreeSAPP pipeline was run on Google Cloud for metagenomic (MetaG, DNA) and metatranscriptomic (MetaT, RNA) assemblies of norB gene at the depths of 10m, 100m, 120m, 135m, 150m, 165m, and 200m. Reads per kilo base per million mapped reads (RPKM) was used for normalization when comparing the gene converage values, which corrected differences in sampling sequencing depth and gene length (16). The formula used in RPKM is indicated below:

RPKM = numReads / (geneLength/1000 * totalNumReads/1,000,000)

Where numReads was the number of reads mapped to a gene sequence, geneLength was the length of the gene sequence, and totalNumReads was the total number of mapped reads of a sample.

The following codes below were used in the pipeline (15). More information can be obtained from the TreeSAPP Demo on Github:

time treesapp.py -T 8 –verbose –delete –pairing pe
Various functions were stated with the following line above for: Tracking the runtime, specifying the maximum number of threads (8) to be used for efficiency, printing the intermediate steps, deleting the intermediate files to reduce the storage space occupied, and lastly specifying the usage of paired end files for comparing our data.

-t D0501
Target gene was specified here (D0501, which corresponded to *norB*).

-i bucket/MetaG_assemblies/depth#_assembly.fasta –rpkm
Input files and path to store files was specified with the following line above.

-r bucket/MetaG_reads/depth#.fastq.gz
The input files for MetaG and MetaT reads were also specified above.

-o treesapp_out_dir_depth#
Output file name was specified above.

rm treesapp_out_dir_depth1/RPKM_outputs/*.sam
Last instruction here was used to remove gigabyte size intermediate SAM files.

Visualization of output files at various depths were done on iTOL v4 for *norB* gene.

Using the processed results, DNA abundance and RNA abundance analyses were done for *norB* at multiple depths (Fig. 2 and Fig. 3). In addition, various genera were also compared in regards to their DNA vs. RNA abundances to determine the expression of *norB* for each genus at different depths (Fig. 4a and Fib 4b). Lastly, the relationship of abundance of *norB* and concentration of several nitrogen species – such as nitrate, nitrogen dioxide, and nitrous oxide – were also observed (Fig. 5).

3 Results

3.1 Analysis of the DNA abundance of *norB* with depth

Table 1: Normalized abundance of the norB gene (DNA) at different depths by RPKM

Depth_m	Abundance_DNA
10	1.581131
100	101.397580
120	197.248612
135	345.156101
150	541.756037
165	137.576279
200	854.085568

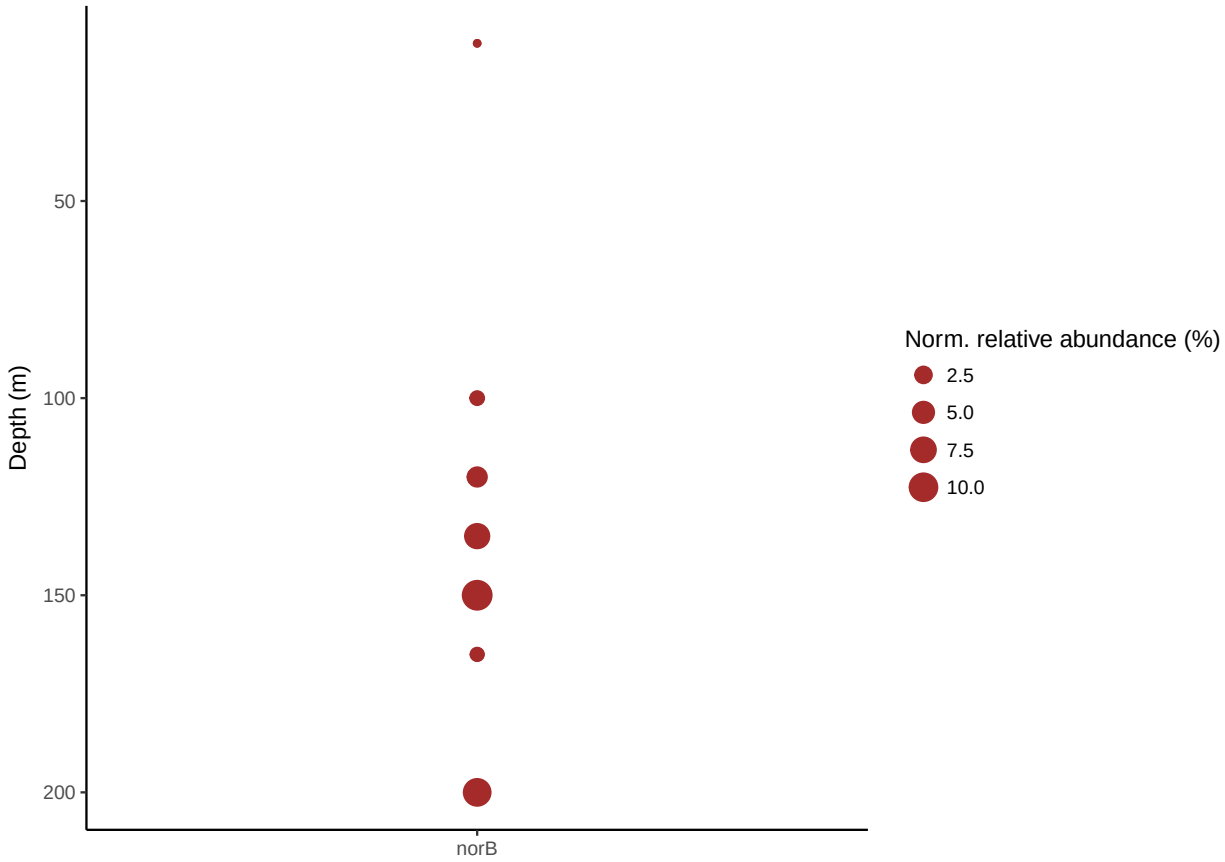

Figure 2: Normalized relative abundance (0−1) of norB gene (DNA) at different depths

The DNA abundance of the *norB* followed an upwards trend where it increased from a normalized relative abundance of 2.5% at depth 100m to 10.0% at depth 200m. This was consistent with what had been shown in previous studies, where genes associated with partial denitrification to nitrous oxide were most abundant at a depth of 150m (11) and were not present at depth 50m. An exception of this trend occurs at depth 165 meters, where the abundance declined sharply. Since the gene encodes the nitric oxide reductase enzyme—responsible for the reduction of nitric oxide to nitrous oxide—the microbes possessing these genes would be considered partial denitrifiers. It is likely that conditions at lower depths are more ideal for partial denitrifiers, which would explain why the gene's presence is more abundance at lower depths. Depth 165m could be less ideal for microbial growth because it sits between the sulfide nitrate transition zone (SNTZ) and the sulfide zone.

3.2 Analysis of similarities between RNA and DNA abundance information of *norB* with depths

Table 2: Normalized abundance of norB gene (DNA vs. RNA) at different depths by RPKM

Depth_m	Abundance_DNA	Abundance_RNA
10	1.581131	0.000000
100	101.397580	4.256517

Depth_m	Abundance_DNA	Abundance_RNA
120	197.248612	8.066146
135	345.156101	55.932741
150	541.756037	788.040985
165	137.576279	912.117306
200	854.085568	944.915908

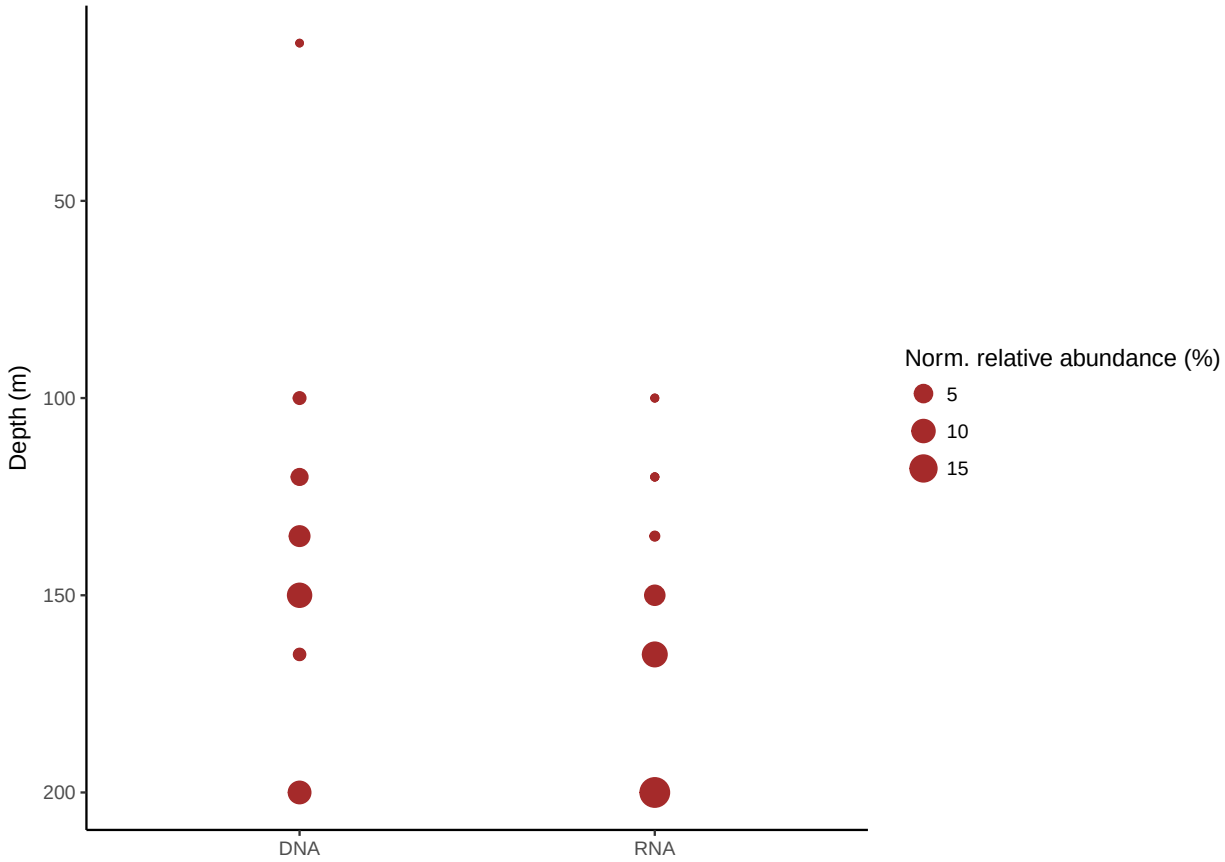

Figure 3: Normalized relative abundance (0−1) of norB gene (DNA vs. RNA) at different depths

The RNA abundance of the norB gene also seemed to follow a similar linear trend like its DNA abundance. It has a normalized relative abundance of 5% at depth 150m and it increased to 15% at depth 200m. This was expected since Hawley *et al.* (12) observed that expressed pathways progressed from ammonia oxidation and nitrification at higher depths to denitrification at lower depths.

3.3 Reconstructing the associated taxa with norB and analyze variances of DNA and RNA based on depths

Table 3: Taxa associated (including NA values) with normalized abundance of norB gene (DNA vs. RNA) at all depths by RPKM

Class	Genus	Abundance_DNA	Abundance_RNA
Alphaproteobacteria	unclassified Rhizobiales	1.037770	0.000000
Bacteria candidate phyla	NA	9.950229	3.264051
Betaproteobacteria	Gallionella	227.856175	24.256090
Betaproteobacteria	NA	266.265765	542.734521
Cytophagia	NA	0.582960	0.197474
Flavobacteriia	Formosa	1.603260	0.000000
Flavobacteriia	Muricauda	23.368146	37.471377
Flavobacteriia	Zobellia	3.161290	0.000000
Gammaproteobacteria	Achromatium	32.188300	120.183600
Gammaproteobacteria	Beggiatoa	0.961576	0.000000
Gammaproteobacteria	Candidatus Competibacter	238.590000	112.869170
Gammaproteobacteria	Endozoicomonas	0.398583	0.000000
Gammaproteobacteria	Thiocapsa	57.156660	60.359952
Gammaproteobacteria	NA	314.796488	1470.055217
unclassified Bacteroidetes	NA	4.218460	0.000000
unclassified Chlorobi	NA	7.638090	13.089740
NA	NA	989.027556	328.848411

The total normalized abundances computed from DNA and RNA data is 2178.801308 and 2713.3296034, respictavely; however, only 26.9102904% and 13.0887227% have known genera for both DNA and RNA. We impute `NA` with `unclassified` for phylum and genus.

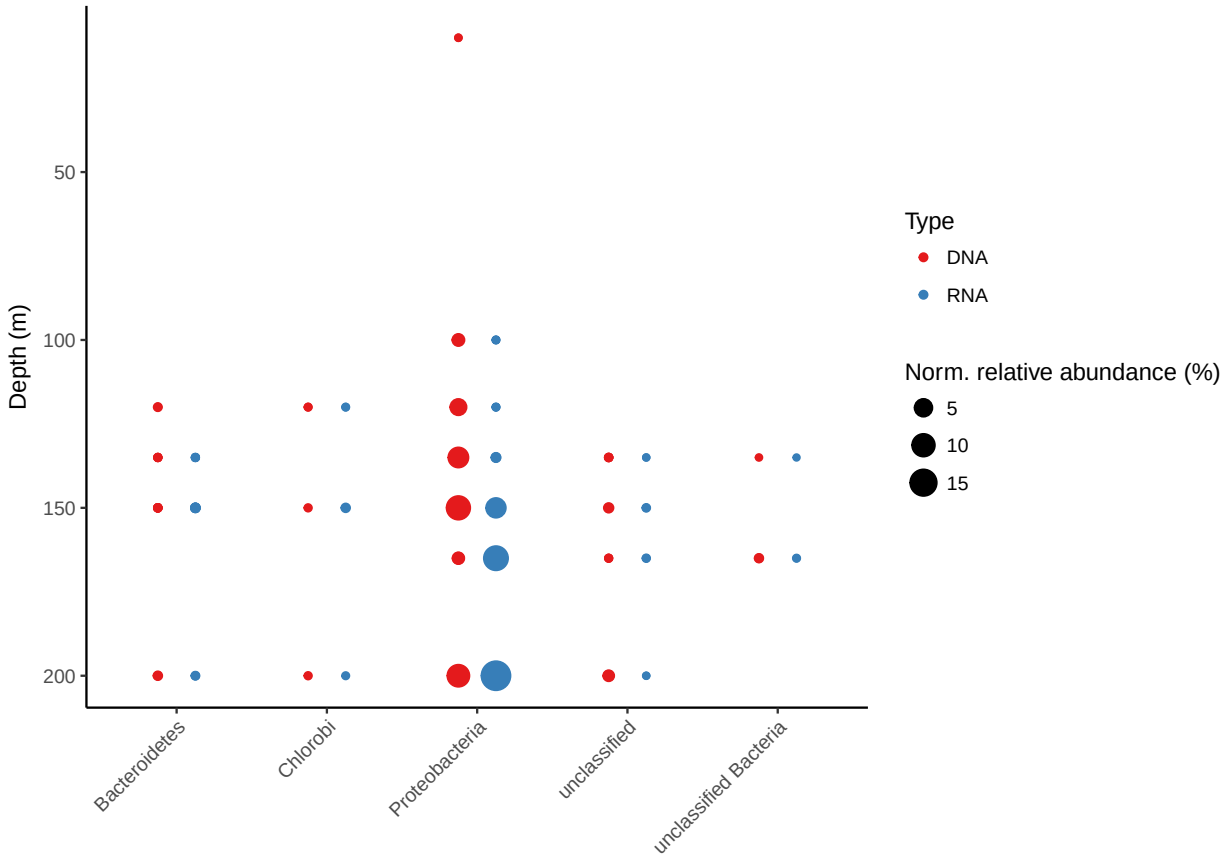

Figure 4: Abundance of Phyla with norB (DNA vs. RNA) at different depths

According to Table 3, the genus associated with abundance of *norB* includes *Formosa*, *Muricauda*, *Zobellia*, *Achromatium*, *Beggiatoa*, *Candidatus Competibacter*, *Endozoicomonas*, *Gallionella*, *Thiocapsa*, and *Rhizobiales*. These genera belong to the three Phyla Bacteriodetes, Chlorobi and Proteobacteria. The abundance of phylum, as shown in Fig. 4, differed with respect to depth and with DNA versus RNA. For Proteobacteria, the abundance of *norB* was greater at 200m in comparison with 100m. At depth of 200m, greater abundance of RNA was observed in comparison with DNA.

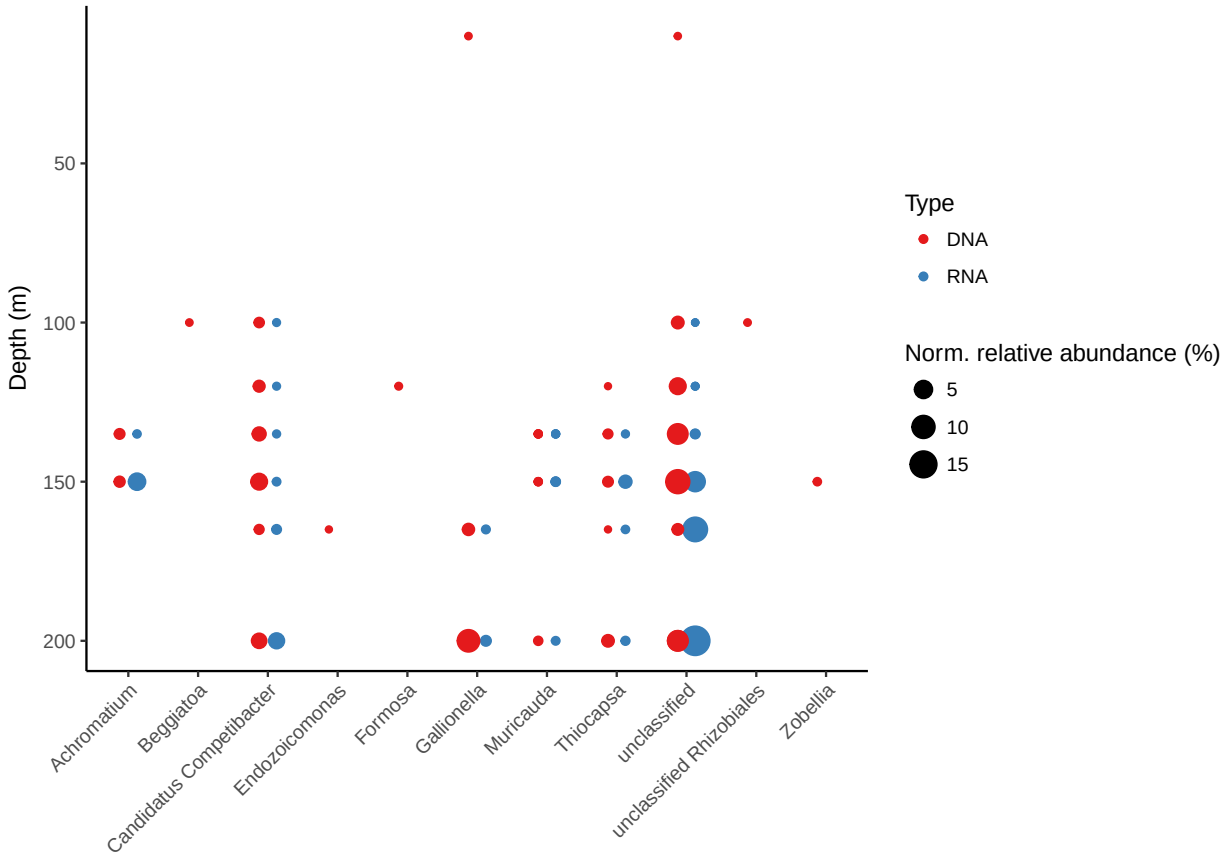

Figure 5: Abundance of genera with norB (DNA vs. RNA) at different depths

3.4 Analysis of the abundance of *norB* in relation to nitrogen species in Saanich

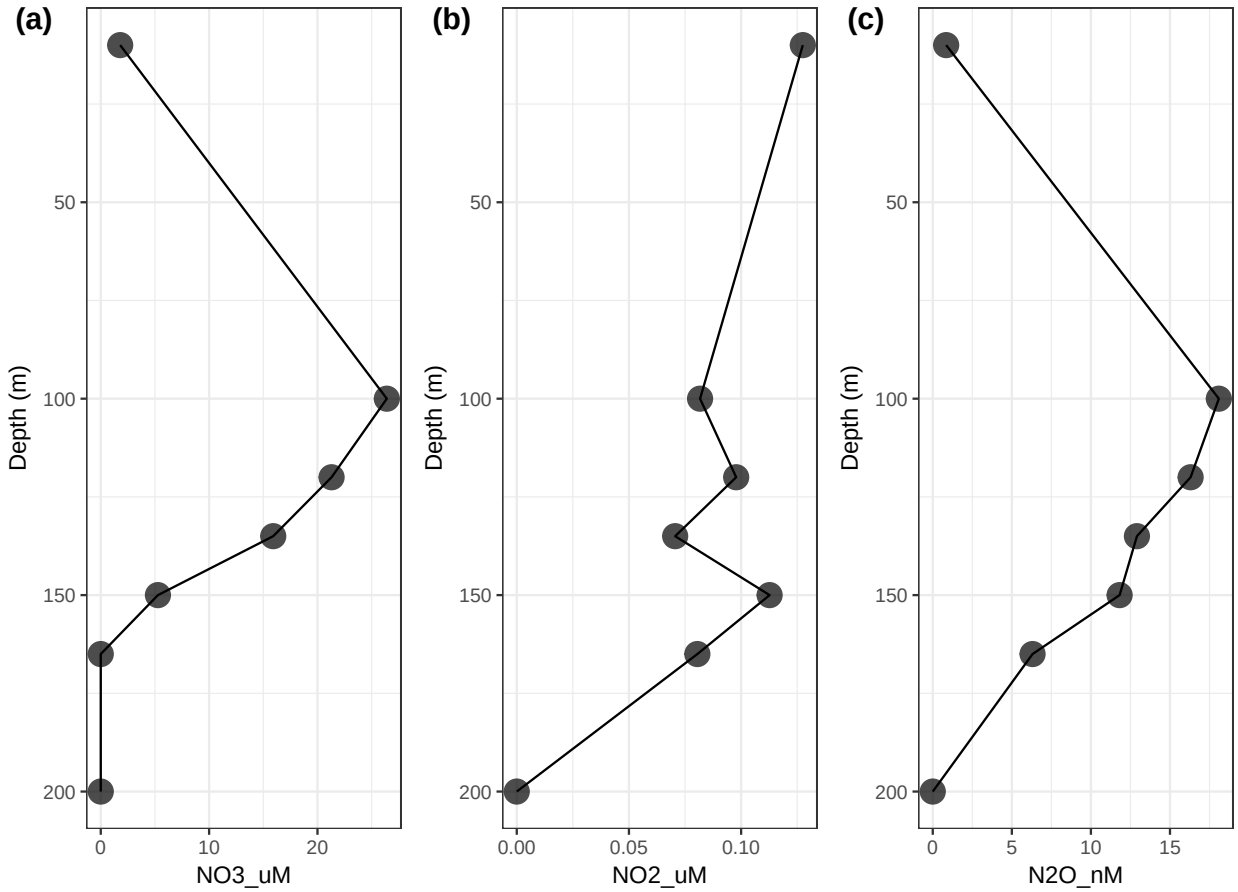

Figure 6: Abundance of norB in relation to nitrogen species across depths

According to Fig. 6, related to the abundance of NO_3 and N_2O, the abundance of *norB* increased from depth 10m to 100m, and declined from 100m to 200m. The abundance of *norB* related to NO_3 has more of a sharp declined in comparison with that of N_2O. Regarding to the abundance of *norB* related to NO_2, a sharp declined is observed from 10m to 100m, the abundance level fluctuated from 100m to 150m, and declined from 150m to 200m.

4 Discussion

At the depth of 100m, it appeared that most genes involved in the denitrification pathway like *norB* were present, but they were neither abundant nor highly expressed. However, the *nirK*/*nirS* genes encoding nitrite reductase had high abundance and high expression, suggesting that nitrite reductase was actively being reduced to nitric oxide at this depth. The rest of the pathway had low activity by comparison, including the gene encoding the nitrate reductase subunit, *napA*, which had low abundance and low expression. At the depth of 200m, most genes involved in the denitrification pathway were also present and were expressed in high levels. This was observed especially for *norB*, which was shown to be the most abundantly expressed at depth 200m. Interestingly, *norC* genes

was found to have similar expression level compared with *norB* across all depth levels; this finding suggested that *norB* and *norC* could potentially be involved in similar denitrification pathways.

The environmental and evolutionary standpoints are interconnected in regards to the distributed metabolism seen in the nitrogen cycle. At various depths, the temperatures and presence of other chemical species will differ, with some being more significant than others. The differences in temperature and presence of other chemical species may in turn affect the entropy production (13). According to one study (13), biological systems acquire information through evolution and store it within its metagenome for maximum entropy production (MEP). Biological systems may then use this information to achieve MEP and further refine it through evolution. In other words, biological systems that can achieve MEP efficiently are able to allocate resources in expressing the optimal genes for survival in terms of their functional efficiency at the optimal temperatures and resource availability in the environment (13).

In terms of the nitrogen cycle, the differences of biogeochemical gradients at low and high depths – such as the high nitrogen species and low oxygen species at depths 100m and beyond – can be used to explain the increases in abundance of DNA and RNA for *norB* gene (Fig. 3). This observation of oxygen control on denitrifiers was also made in another study (14) that described the decrease and inhibition of denitrification processes at low oxygen concentrations. Other studies (11 and 14) have created models to explain the distributed metabolism through gene abundance, biogeochemical gradients, or both. One model suggested looking at the models separately in terms of MEP and metabolic perspectives (13). In the MEP perspective, the form of the molecular machines is unimportant for biogeochemical MEP. Their model stated that living systems can be viewed as molecular machines that catalyzes reactions that both synthesize and degrade molecular machines and dissipate energy via redox reactions in the process, which are undifferentiated from abiotic systems (13). In the metabolic perspective, their model only predicts the functional characteristics in terms of allocation of molecular machinery (13). In other words, the metabolic model looks at the functional groups and community level gene expression from experiments to explain the distributed metabolism. On the other hand, the model proposed by the other study suggested looking at both perspectives in an integrated manner (11). Their model reasoned that both the gene and transcript abundances and biogeochemical concentrations are linked such that gene expression patterns are determined by gene abundance and biogeochemical conditions with biogeochemical fluxes being excellent indicators for microbial community structures (11). By looking at the DNA, mRNA, and protein profiles with the biogeochemical concentrations, different perspectives on the community activity can be obtained in terms of the immediate catalytic potential of a community, the level of expression of control environmental mRNA and protein distributions, and the prediction of pathway expression and growth in relation to biogeochemical conditions (11). The logics applied to these models and studies can then be used to explain the figure results, which demonstrated how the distributed metabolism in the nitrogen cycle are connected to the gene abundances (Fig. 3) and biogeochemical gradients (Fig. 5) at various depths.

For future experiments, it would be interesting to examine the relationship between the genera associated with the *norB* gene. It is possible that the abundance of one genus can have an effect on the other at certain depths. Looking into which factors favour the growth of certain genera over the other could also be a viable project. It is worth noting that although our study was primarily concerned with the presence of one particular gene in the denitrification pathway, this method can be extended to examine other metabolic pathways in biogeochemical cycles such as decomposition.

5 References

1. 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, Tijana Glavina Del Rio, Suttle, CA, Tringe, S, Hallam, SJ. 2017. A compendium of multi-omic sequence information from the Saanich Inlet water column. Scientific Data. 4:170160. doi: 10.1038/sdata.2017.160.

2. Hallam, SJ, Torres-beltrán, M, Hawley, AK. 2017. Monitoring microbial responses to ocean deoxygenation in a model oxygen minimum zone. Scientific Data. 4:170158. doi: 10.1038/sdata.2017.158.

3. Ulloa, O, Canfield, DE, DeLong, EF, Letelier, RM, Stewart, FJ. 2012. Microbial oceanography of anoxic oxygen minimum zones. Proc. Natl. Acad. Sci. U. S. A. 109:15996-16003. doi: 10.1073/pnas.1205009109.

4. Wright, JJ, Konwar, KM, Hallam, SJ. 2012. Microbial ecology of expanding oxygen minimum zones. Nature Reviews.Microbiology. 10:381. doi: 10.1038/nrmicro2778.

5. Stein LY, Klotz MG. 2016. The nitrogen cycle . Cell Press 26:R94–R98.

6. Falkowski PG, Fenchel T, Delong EF. 2008. The Microbial Engines That Drive Earth's Biogeochemical Cycles. Science 320:1034–1039.

7. Knowles R. 1982. Denitrification . Microbiological Reviews 46: 43-70.

8. Bergaust L, Spanning RJMV, Frostegard A, Bakken LR. 2011. Expression of nitrous oxide reductase in Paracoccus denitrificans is regulated by oxygen and nitric oxide through FnrP and NNR. Microbiology 158:826–834.

9. Dalsgaard T, Stewart FJ, Thamdrup B, Brabandere LD, 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.

10. Qu Zhi, Bakken LR, Molstad L, Frostegard A, Bergaust LL. 2016. Transcriptional and metabolic regulation of denitrification in Paracoccus denitrificans allow low but significant activity of nitrous oxide reductase under oxic conditions. Environmental Microbiology 18: 2951-2963.

11. Louca S, Hawley AK, Katsev S, Torres-Beltran M, Bhatia MP, Kheirandish S, Michiels CC, Capelle D, Lavik G, Doebeli M, Crowe SA, Hallam SJ. 2016. Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone. Proc Natl Acad Sci 113:E5925–E5933.

12. Hawley AK, Brewer HM, Norbeck AD, Pa a-Toli L, Hallam SJ. 2014. Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. Proc Natl Acad Sci 111:11395–11400.

13. Vallino JJ. 2010. Ecosystem biogeochemistry considered as a distributed metabolic network ordered by maximum entropy production. Philosophical Transactions: Biological Sciences. 365:1417-1427.

14. Christensen S. 1988. Oxygen control prevents denitrifiers and barley plant roots from directly competing for nitrate. FEMS Microbiology Letters. 53:217-221.

15. TreeSAPP Demo by Connor Morgan-Lang & Kim Dill-McFarland.

16. Metagenomics. Date of Access: Apr.15.2018. Link: http://www.metagenomics.wiki/pdf/definition/rpkm-calculation.