Project 2: Assessing Distributed Metabolic Potential of the Nitrogen Cycle in Saanich Lake Metagenome

Abstract

The nitrogen cycle is responsible for the fluxes of fixed nitrogen in and out of an aquatic environment like the Saanich Inlet. The presence of *nirK,* an enzyme which converts nitrite to nitric oxide, and serves as a functional marker for nitrite-reducing microbes, was investigated in metagenomic and metatranscriptomic data from the Saanich Inlet. This data was then analyzed in relation to the depth and nitrite (NO_2) concentrations of the sampling locations, and a phylogenetic analysis of the Phylum and Class of the microbes was performed. Our findings illuminated several trends in the abundance of the *nirK* gene in the metagenome and its subsequent transcription profile throughout the water column. Based upon the depth profile of transcribed *nirK*, it appears that maximum transcription levels are reached between 100 and 150 meters. Additionally, it appears that Proteobacteria are the most abundant of the *nirK*-containing bacteria in both the metagenome and the metatranscriptome. However, when analyzed in terms of Class, Gammaproteobacteria are the most abundant in the metagenome, while unclassified Proteobacteria are the most abundant in the metatranscriptome. This indicates that these unclassified Proteobacteria are more metabolically active than their relative abundance in the metagenome would suggest, and that they are probably the dominant denitrifiers in this environment.

Introduction

For four decades, the Saanich Inlet has been studied as a model ecosystem for diverse phenomenon such as the formation of marine oxygen minimum zones, seasonal nutrient cycles, and other biogeochemical processes (1). Located off the southeastern coast of Vancouver Island, British Columbia, the majority of Saanich Inlet undergoes cycles of seasonal anoxia and oxygen depletion annually. The deeper portion of the water column becomes progressively oxygen-limited in the spring and summer, and is re-oxygenated in the fall due to oceanic upwelling (2). In addition, warming occurs at surface depths during spring, while deeper portions of the inlet remain at a relatively stable temperature. This variance in biochemical conditions contributes to the metabolic diversity of microbial life in the region. Different species thrive at different times of the year, at varying depths, and even at varying sampling locales within the fjord. This variance in predominant microbial species impacts larger-scale phenomena in and around the inlet. For example, seasonal phytoplankton blooms help to fulfill the energy requirements of spawning salmon and herring, thus illustrating the impact of biogeochemical processes on the wildlife ecology of the region (3). Previous studies have characterized the geochemical and microbial genomic makeup of the Saanich Inlet, making the study of microbial distribution correlated with biogeochemical trends possible(4,5).

The global nitrogen cycle is the biogeochemical cycling of nitrogen throughout the earth's terrestrial, aquatic, and atmospheric environments. Atmospheric nitrogen is reduced into biomass through biological nitrogen fixation, and ammonium is released into the environment through decomposition. Humans also use the Haber process to generate ammonium from atmospheric nitrogen. Ammonium is converted into nitric oxides through nitrification, and this process is mediated primarily by microbes(6). Biological denitrification is a process in which oxidized nitrogen is used as a terminal electron acceptor for energy production using a carbon source, and is reduced stepwise to gaseous end products like NO, N_2O, and N_2. This completes the nitrogen cycle, converting nitric oxides back to nitrogen gas. Denitrification occurs primarily in anoxic environments of <10% O_2 concentration, although it occasionally occurs in oxic environments as some microbes that are responsible for denitrification are facultative anaerobes(7). Denitrifying microbes consist of both heterotrophs and autotrophs, and are represented in a wide variety of phylogenetic groups(8). This is a significant biological process for a variety of reasons. Nitric oxide (NO) and nitrous oxide (N_2O) are potent greenhouse gases, and contribute to climate change. Increased nitrogen influx into marine environments from man-made fertilizer has also increased rates of nitrification in these environments, and denitrification allows for the flux of nitrogen from these environments back into the atmosphere. Thus, denitrification could potentially help to eliminate pools of nitrogen in marine environments which can lead to algal blooms and subsequent eutrophication(9).

In this project, we used metagenomic and metatranscriptomic data derived from seven depths of the Saanich Inlet to study the distribution of the nitrate reductase gene *nirK*. NirK serves as a functional marker for nitrate-reducing microbes, which we analyze with relation to depth, phylum, class, and nitrite (NO_2) concentrations.

Methods

Water was collected from Saanich Inlet at depths between 10 m to 200m and filtered for its biomass to extract RNA and DNA. RNA was then reverse transcribed to cDNA to generate a shotgun Illumina library. The data was sequenced using Illumina Hiseq and a metagenome was assembled for further analysis. Additional information on the sequencing and processing methods can be found in Hawley et al. 2017 (10). The metagenome and metatranscriptome was input into a Tree-based Sensitive and Accurate Protein Profiler (TreeSAPP) pipeline with the use of Google Cloud to reconstruct the nitrogen cycle in terms of nirK.

Google Cloud was used as it is a consistent, scalable, and on-demand resource that is available publicly. It provided enough RAM and CPU capacity to run the TreeSAPP protocol without the need for a dedicated computing system (11). TreeSAPP was used to map our metagenomic and metatranscriptomic reads to a reference genome, perform profile-alignments, classifying sequence insertions in reference trees, and finally generating abundance information, and phylogenetic classifications that were viewed as a phylogenetic tree in iTOL (11, 12).

In TreeSAPP, we set up a screen so that the commands will continue to run despite logging-out.

```
screen -S depth
```

Then, we created a bash script to map metagenomic and metatranscriptomic reads to a reference metagenome. The full bash script used will be found in Appendix 1 and 2 for DNA and RNA, respectively. An example command for the depth of 10m is as follows:

```
time treesapp.py -T 8 --verbose --delete \
-t D0301 -i bucket/MetaG_assemblies/SI072_LV_10m_DNA.scaffolds.fasta
--rpkm -r bucket/MetaG_reads/SI072_LV_10m.anqdp.fastq.gz -o
treesapp_out_dir_depth10m
```

The command `time` was used to track how long TreeSAPP needs to run the data. -T 8 was specified so that 8 threads were running in parallel to increase the speed. --verbose was used to print out information as TreeSAPP was running. --delete was indicated to delete any intermediate files. --pairing pe indicated that the paired end files were to be used and -t D0301 was specified as it encodes our gene of interest, the nitrate reductase gene, *nirK.*

Finally, using the following command, we removed all the .sam files as they were unneeded, and very large in size.

```
rm
treesapp_out_dir_depth120m/RPKM_outputs/SI072_LV_10m_DNA.scaffolds_ge
nes.sam
```

The following is an example command of mapping a metatranscriptomic read for the depth of 10m:

```
time treesapp.py -T 8 --verbose --delete --pairing pe -t D0301 -i
bucket/MetaG_assemblies/SI072_LV_10m_DNA.scaffolds.fasta --rpkm -r
bucket/MetaT_reads/SI072_LV_10m.qtrim.3ptrim.artifact.rRNA.clean.fast
q.gz -o treesapp_out_dir_depth10mRNA

rm treesapp_out_dir_depth10RNA/RPKM_outputs/*.sam
```

TreeSAPP produced .jplace files that were input into iTOL version 4.2 to generate abundance phylogenetic trees for the metagenomic and metatranscriptomic data. Labels and abundance bars were generated by uploading the other text files in the iTOL output file. (13). Additionally, the classification table in a .tsv output file that was generated was used for further analysis when input into R version 1.4.3. We used ggplot2, cowplot, and tidyverse packages to perform the analysis.

.tsv data were loaded into R studio using the read_tsv command, and named according to depth and DNA/RNA compositions, repeated for each depth. Here is an example for the depth of 10m but the full code for all the Figures can be found in the supplemental .rmd file:

```
nirK.DNA.10m = read_tsv("nirk_DNA_10m_contig_map.tsv") %>%
  select(Tax.DNA.10 = Confident_Taxonomy, Abund.DNA.10 = Abundance,
Query)
```

DNA and RNA data were combined into a single table using the full_join command and the code below was repeated for each depth.

```
nirK.all = nirK.DNA.10m %>%

      full_join(nirK.RNA.10m, by = "Query")
```

A taxonomy variable aggregating all taxonomy columns to fill in any non-applicable data points was created using the mutate function, repeated for each depth.

```
mutate(Taxonomy = ifelse(!is.na(Tax.RNA.10), Tax.RNA.10,
                         "unclassified")))))))))))))) %>%
```

Old taxonomic variables were removed. All abundance data was then gathered into a single column and the variables under key were seperated. The Key variable was transformed into Depth and DNA/RNA variables, and the Key variable was removed. Depth was then mutated into a numeric entity for R to recognise, and taxonomy was separated into columns for plotting against depth.

```
select(-starts_with("Tax.")) %>%
gather("Key", "Abundance", starts_with("Abund")) %>%
separate(Key, c("Key","Type","Depth_m"), by = ".") %>%
select(Depth_m, Type, Abundance, Taxonomy, Query) %>%
mutate(Depth_m = as.numeric(Depth_m)) %>%
separate(Taxonomy, into = c("Domain", "Phylum", "Class", "Order",
"Family", "Genus", "Species"), sep="; ")
```

DNA and RNA abundance was plotted according to depth using the mutate and ggplot functions. NA categories were changed to reflect unclassified phyla.

```
nirK.all %>%
  mutate(Phylum = ifelse(is.na(Phylum), "unclassified", Phylum)) %>%

ggplot(aes(x = Type, y = Depth_m)) +
  geom_point(aes(size = Abundance)) +
  scale_y_reverse(lim=c(200,10)) +
  labs(title = "Abundance of the nirK gene (DNA vs. RNA) at different
depths",
       x = "") +
  theme_classic()
```

A nitrite concentration plot was created using the arrange function to group by depth, and the ggplot function. A separate colour coded plot reflecting DNA and RNA data types was also created in the same manner, and combined with the nitrite plot.

```
plot1 = nirK.all %>%
  mutate(Phylum = ifelse(is.na(Phylum), "unclassified", Phylum)) %>%

ggplot(aes(x = Phylum, y = Depth_m)) +
  geom_point(aes(size = ifelse(Abundance == 0, NA, Abundance), color
= Type), position = position_dodge(0.5)) +
  scale_y_reverse(lim=c(200,10)) +
  labs(y = "") +
  theme_classic() +
  scale_size_continuous(name = "Abundance")

plot2 = metadata %>%
  arrange(Depth_m) %>%

ggplot(aes(x = NO2_uM, y = Depth_m)) +
    geom_point() +
    geom_path(aes(group = 1)) +
    scale_y_reverse(lim=c(200,10)) +
  theme_classic() +
  labs(y = "Depth (m)",
       x = "Nitrite NO2- (uM)")

plot_grid(plot2, plot1, labels=c("A", "B"), rel_widths=c(1/4, 3/4))
```

Abundance of Phyla containing nirK DNA and RNA by depth was then plotted using filter, mutate and ggplot functions and plot_grid was used to combine the Figures together.

```
plot1 = nirK.all %>%
  filter(Type == "DNA") %>%
  mutate(Phylum = ifelse(is.na(Phylum), "unclassified", Phylum)) %>%

ggplot(aes(x = Depth_m,  y = Abundance)) +
  geom_point(aes(colour = Phylum, size = Abundance)) +
  theme(legend.position="none")

plot2 = nirK.all %>%
# Filter the RNA data
  filter(Type == "RNA") %>%
```

```
  mutate(Phylum = ifelse(is.na(Phylum), "unclassified", Phylum)) %>%

ggplot(aes(x = Depth_m,  y = Abundance)) +
  # Use the size aesthetic to show abundance
  geom_point(aes(colour = Phylum, size = Abundance)) +
  guides(size=FALSE) +
  theme_classic()

plot_grid(plot1, plot2, labels=c("A", "B"), rel_widths = c(2/5, 3/5))
```

The abundance of Classes containing nirK DNA and RNA was plotted using ggplot and combined together using code similar to above.

```
plot1 = nirK.all %>%
  filter(Type == "DNA") %>%
  mutate(Class = ifelse(is.na(Class), "unclassified", Class)) %>%

ggplot(aes(x = Depth_m,  y = Abundance)) +
  geom_point(aes(colour = Class, size = Abundance)) +
  theme(legend.position="none")
plot2 = nirK.all %>%
  filter(Type == "RNA") %>%
  mutate(Class = ifelse(is.na(Class), "unclassified", Class)) %>%

ggplot(aes(x = Depth_m,  y = Abundance)) +
    geom_point(aes(colour = Class, size = Abundance)) +
  guides(size=FALSE) +
  theme_classic()
```

Results

It appears that the *nirK* gene is the most abundant in the metagenome at 150 and 200 m, whereas in the metatranscriptome it is most abundant from 100 to 150m, after which it drops off abruptly. But overall, the abundance of RNA is greater at almost all the studied depths, except at 165 and 200 m. At these two depths, the expression of nirK is much lower than there are genes. The abundance profile by depth for DNA is relatively constant compared to the abundance profile by depth for RNA. This leads one to believe that the majority of denitrification via nirK takes place from 100-150 m.

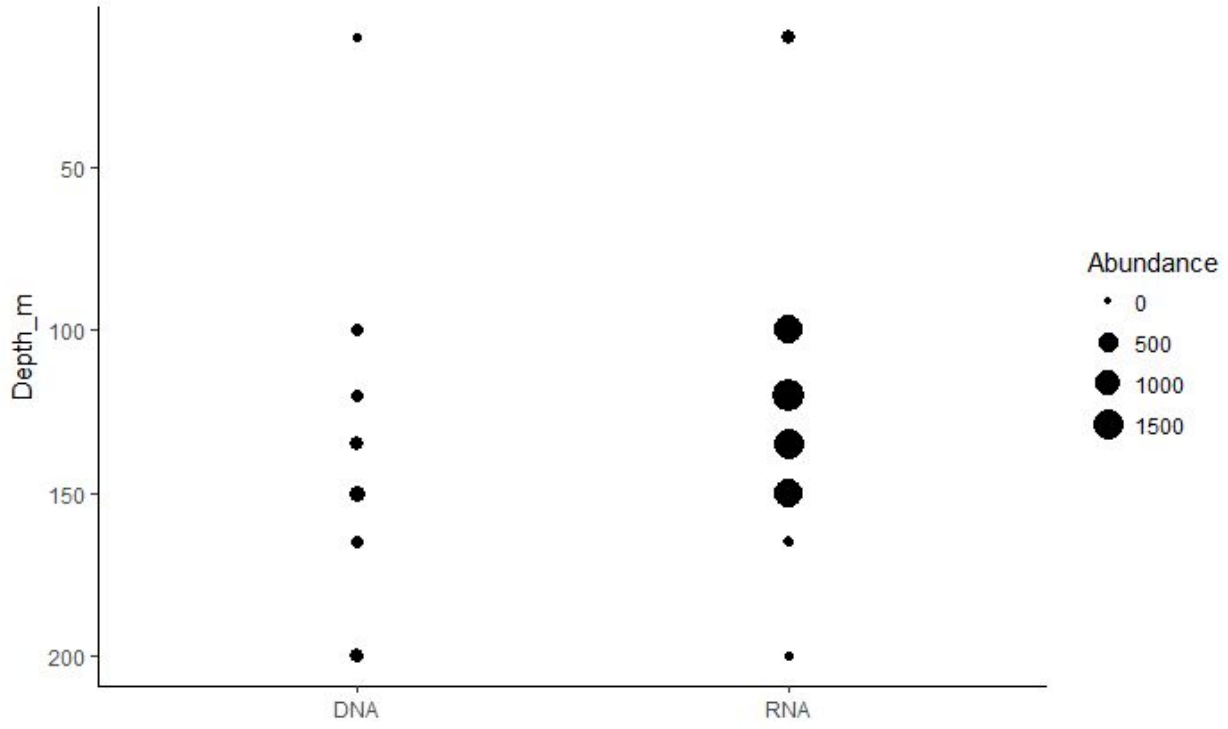


Figure 1. Abundance of the nirK gene DNA vs RNA at different depths. The size of the point represents the abundance.

Figure 2 illustrates that the majority of the nirK RNA present in the metatranscriptome is produced by bacteria in the Phylum Proteobacteria. The nirK present in these bacteria appears to be upregulated to significant levels given the difference between DNA and RNA abundance. It can be observed that as the nitrite levels start to decrease in the water column at about 150m, the level of expression of nirK also begins to drastically decrease by Proteobacteria and all Phyla (Fig. 2) Additionally, we can observe the same trend of increased transcription of the nirK gene at 100-150m and decreased transcription as at depths of lower than 150m in Figure 1.

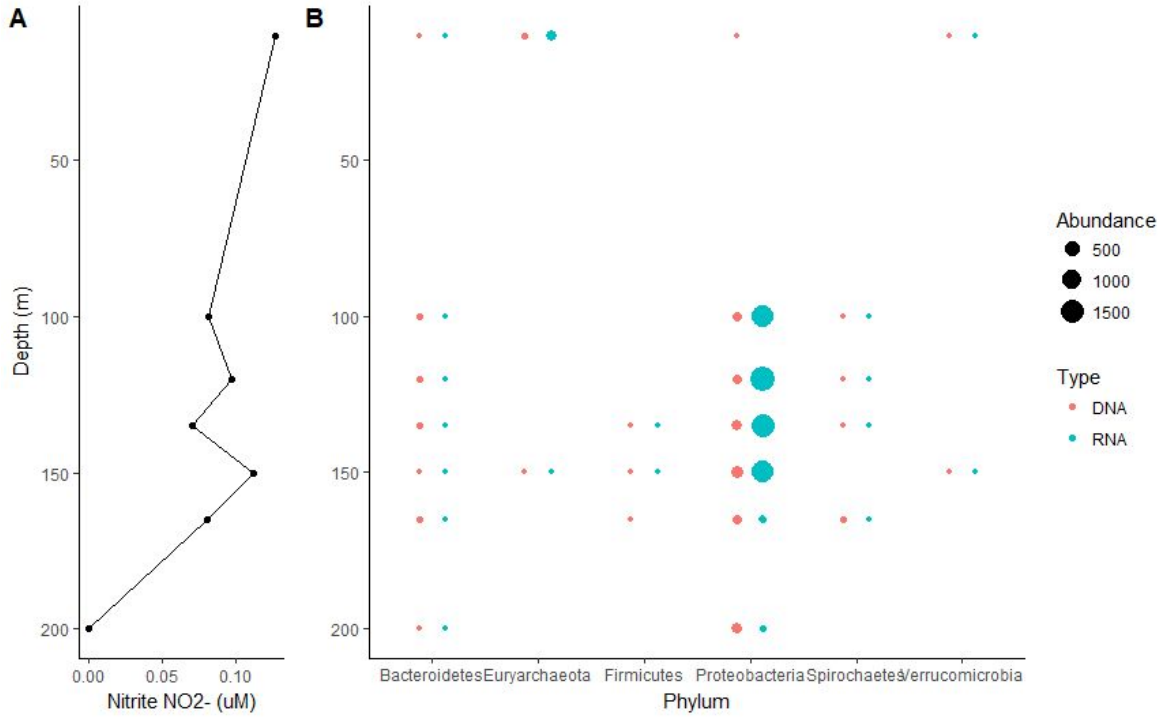


Figure 2. (**A**) Nitrite concentration throughout the Saanich Inlet water column. (**B**) Abundance of Phyla with nirK DNA (pink) and RNA (blue) at different depths.

According to Figure 3, we can tell that not only does Proteobacteria dominates the species with *nirK* at most depths, but there are outliers within the phylum of Proteobacteria (Figure 3a), specifically Gammaproteobacteria (Figure 3c) which are present at much higher abundances. However, when we look at the RNA profile or metatranscriptome (Figure 3b and 3d), we can see that while Proteobacteria still overwhelmingly dominates the species with nirK at most depths, the outliers within the phylum which are present at much higher abundances are actually unclassified Proteobacteria rather than Gammaproteobacteria.

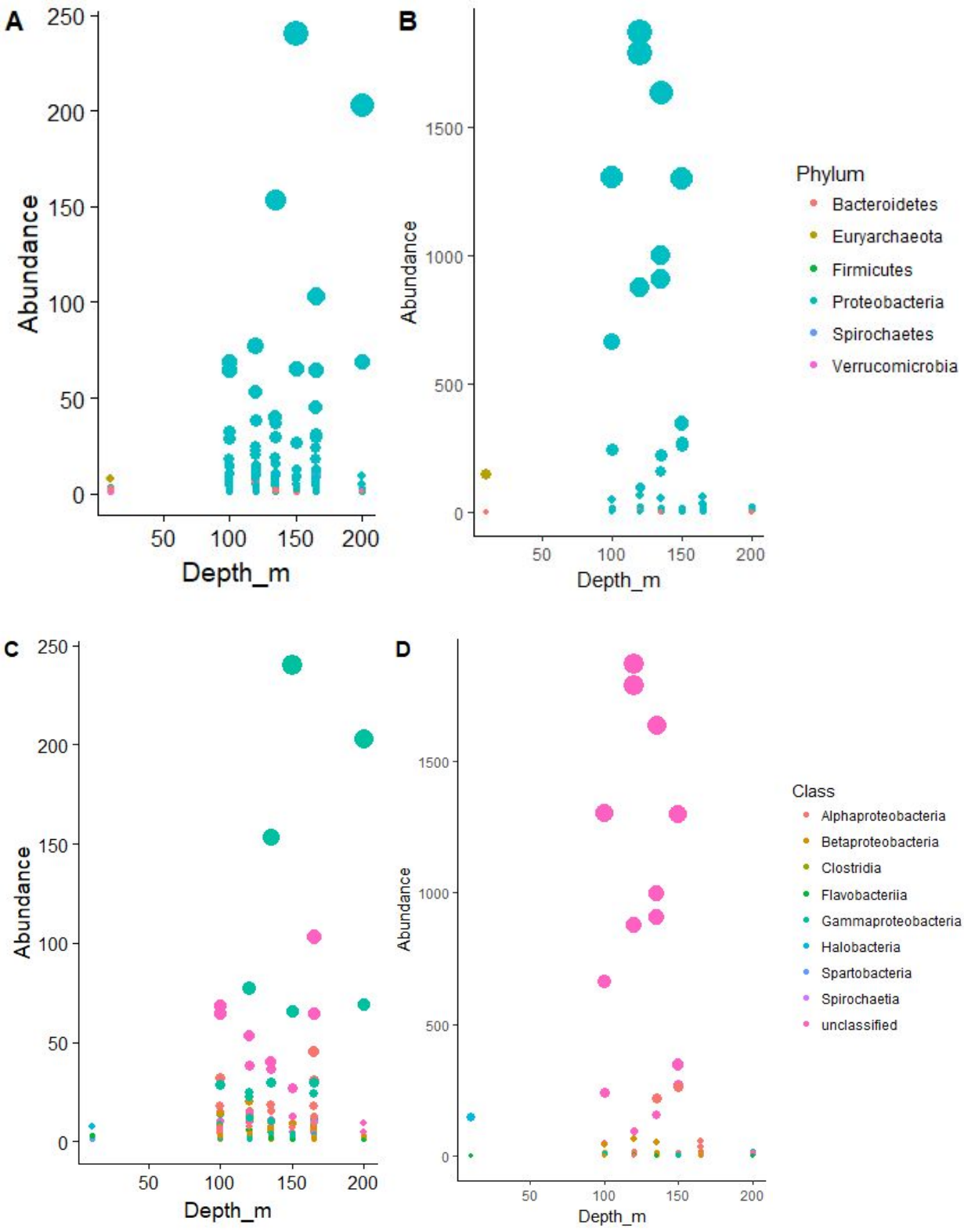
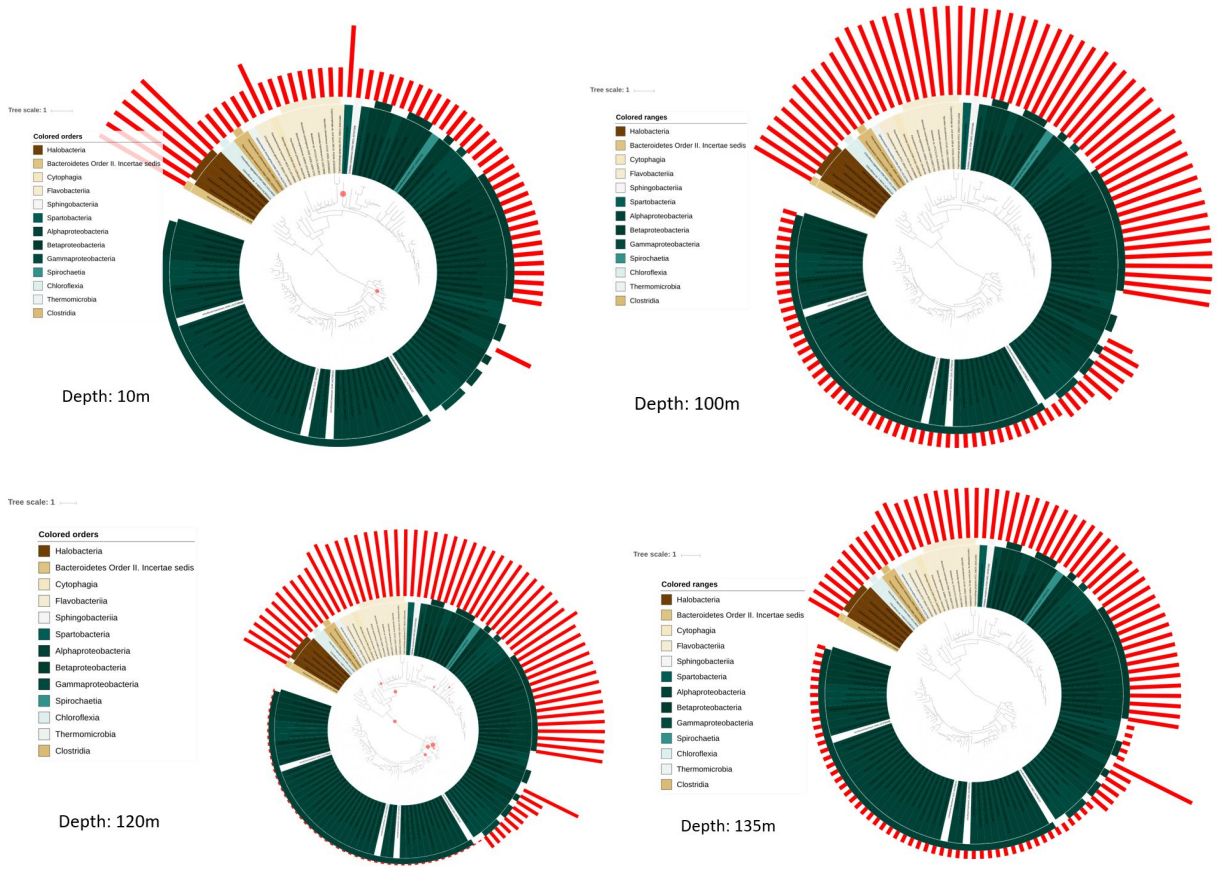


Figure 3. Abundance of taxa containing nirK DNA and RNA by depth. (**A**) Abundance of Phyla containing nirK DNA by depth with size representing abundance and colour representing

the different Phyla. The largest abundance is within the Phylum Proteobacteria. (**B**) Abundance of Phyla containing nirk RNA. The largest abundance is within the Phylum Proteobacteria. (**C**) Abundance of Classes containing nirK DNA showing that the largest abundance is within the Class Gammaproteobacteria (**D**) Abundance of Classes containing nirK RNA with the largest abundance being unclassified.

Figure 4 and Figure 5 show that there is a high abundance of organisms containing the *nirK* gene among Flavobacteria, Halobacteria and Betaproteobacteria at almost all depths. However, at 10m, *nirK* is abundant in a variety of different Classes of organisms in the metagenome and not only those 3 Classes. The transcriptome differs in that the majority of its expression is derived from the archaeal Class Halobacteria. At 200m, the abundance of the *nirK* is low levels amongst all the Classes with the exception of Gammaproteobacteria. However, the levels of RNA are quite high amongst all Classes, indicating that there is greater expression than there are genes amongst many organisms at a depth of 200m.



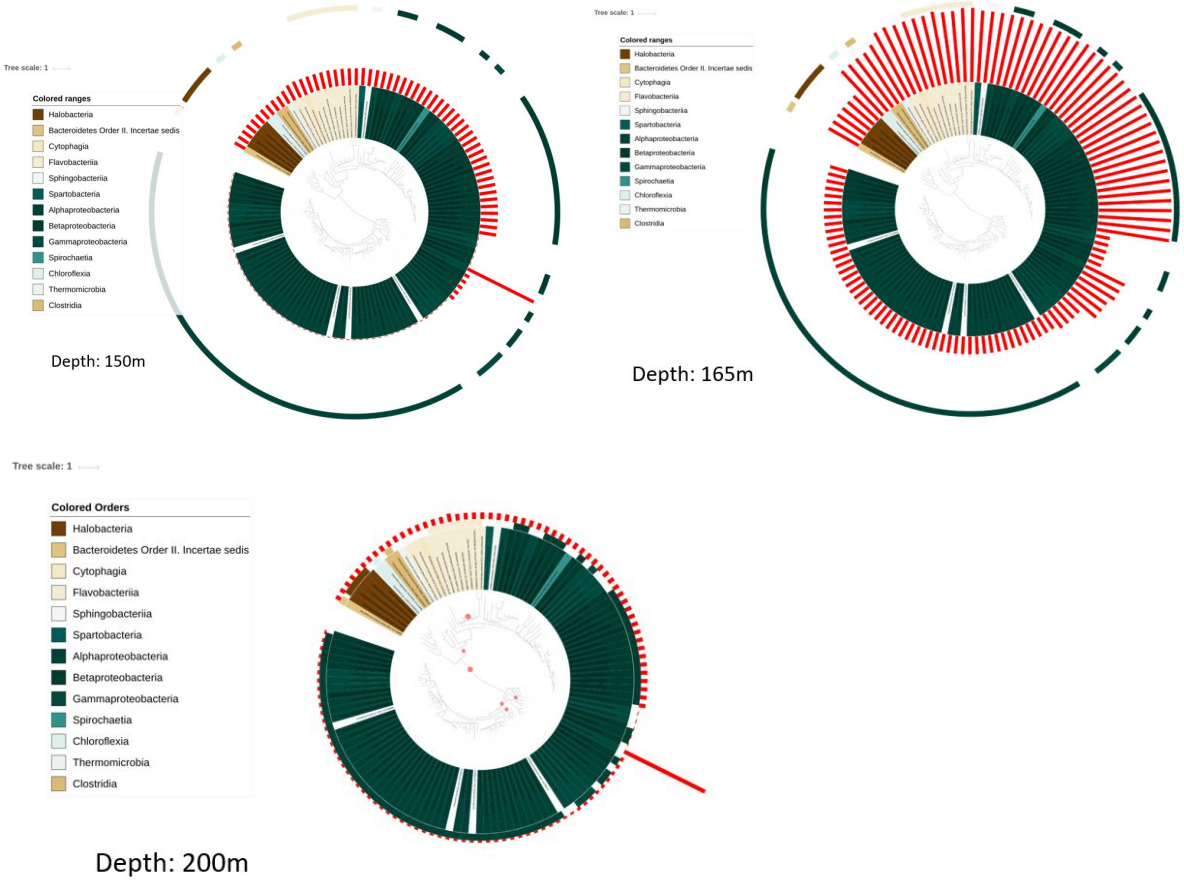
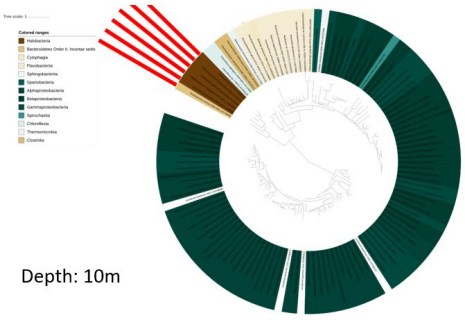
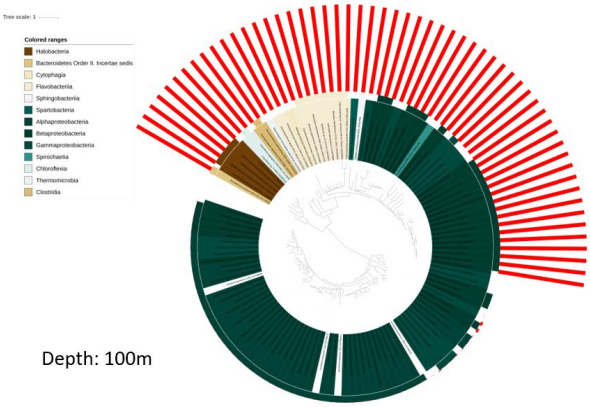


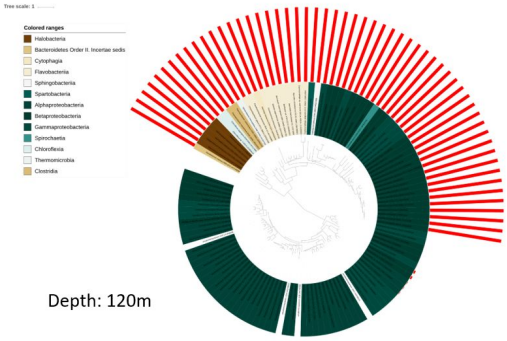
Figure 4. nirK DNA iTOL at depths of 10m to 200m, from left to right, top to bottom, respectively. The red bars indicate the abundance of DNA of a certain Class of organisms at that depth whereas the colours around the tree indicate the different Classes of organisms.



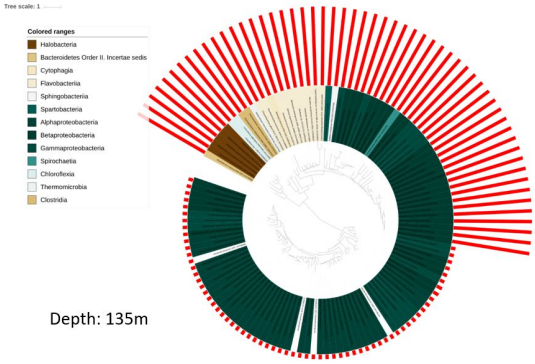
Depth: 10m



Depth: 100m



Depth: 120m



Depth: 135m

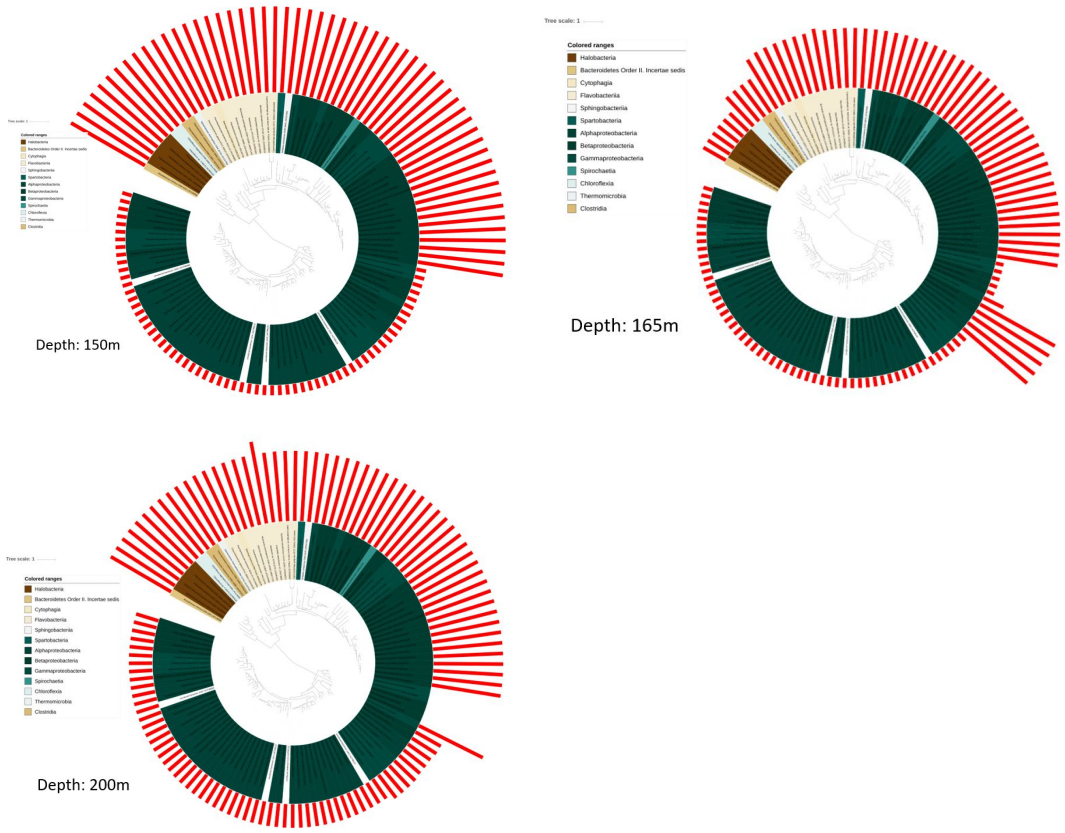


Figure 5: nirK RNA iTOL trees at seven depths from 10m to 200m from left to right, top to bottom. The red bars indicate the abundance of the metatranscriptome and the different colours indicate the Classes of organisms expressing the *nirK* gene.

Discussion

NirK is a copper-dependent nitrate reductase, which similar to NirS, is responsible for the conversion of Nitrite to Nitric Oxide (14, 15). This is one of the early steps in the process of denitrification. While *nirK* and *nirS* both perform the same function, they are structurally different single-copy genes (15). Previous phylogenetic analysis of 297 NirK sequences downloaded from genbank resulted in the classification of NirK into two clades: Clade I composed primarily of Alphaproteobacteria, and Clade II which contained members of more diverse taxonomic groups such as Archaea, Chloroflexi, and Gamma and Betaproteobacteria(15). However, without further analysis of the *nirK* gene sequences found in these samples, we cannot be sure which Clade of *nirK* predominates in the Saanich Inlet.

From the analysis of the metatranscriptome and metagenome, (Figure 1), we can find that there appears to be a peak of *nirK* transcription abundance between 100 and 150 meters deep. This could be an indication that this is where the majority of this denitrification activity occurs. This is somewhat correlated with the abundance of nitrite in the water column (Figure 2a), which begins

to decrease after 165 meters. It suggests that as the concentration of nitrite is less available for use, the expression of nirK decreases. Additionally, from Figures 3a-d and Figures 4 and 5, we can conclude that while the majority of nirK transcripts are produced by unclassified Proteobacteria at all depths except for 10 m, and at 10 meters the majority of *nirK* transcripts are actually produced by Archaea. Specifically, an archaeal species in the phylum of Euryarchaeota (class Halobacteria). Even though overall the *nirK* gene is not very abundant in the metagenome or metatranscriptome at this depth, the fact that this gene is present in both Archaea and Bacteria is still an interesting finding. This indicates that *nirK* is either so ancient that it was present in the last common ancestor of both Archaea and Bacteria, or that its presence in Archaea is indicative of a horizontal gene transfer sometime after the bifurcation of the prokaryotic evolutionary tree.

The low abundance of all nirK-containing transcripts at a depth of 10 m could be indicative of the fact that not much nitrite reduction occurs at this depth. Even though nitrite levels are the highest at this depth (Figure 2a), this is likely due to the fact that nitrogenase, the enzyme responsible for nitrogen fixation via the conversion of nitrogen gas to ammonium is inhibited by oxygen. Thus, no ammonium is available for the conversion into nitrite, the substrate of nirK. Therefore, it makes sense that bacteria capable of denitrification reside deeper in the water column, where fixed nitrogen is available due to the activities of nitrogen-fixing bacteria.

At depths of 100 to 165m, the fact that unclassified Proteobacteria dominate the transcriptome (Figures 3c and 3d) indicates that these unclassified Proteobacteria perform much more denitrification than their abundance in the metagenome would suggest. Additionally, the abundant *nirK*-possessing Gammaproteobacteria do not actually perform as much denitrification as one would assume based upon their abundance in the metagenomic data set. However, one could also interpret this data as evidence that more annotation is needed in the reference database for Gammaproteobacteria. Perhaps the unclassified Proteobacteria that are so prevalent in Figure 3d are actually also Gammaproteobacteria.

Future analysis of this data set could be focus on analyzing the *nirK* sequences that are present to determine the level of sequence variation for this enzyme in the Saanich Inlet. This could potentially be used to hypothesize about horizontal gene transfer, and to determine which *nirK* clade predominates in the Inlet.

It is well understood that microbial life does not exist in isolation. Microbes are capable of sharing of environmental resources, and varying degrees of syntrophy provides evidence for productive cooperation in metabolic cycles. This in turn reduces the genomic and energetic burden of executing the transcriptional and translational processes for all the enzymes necessary for the facilitation of steps in the pathway/cycle for each individual microbial taxa. By splitting the genomic and energetic burden of the enzymes required in distributed metabolism, microbial taxonomies are able to specialize over evolutionary time. Therefore, individual microbes lose the ability to conduct all metabolic reactions, but become more adapted to the specific reactions required to fill its ecological niche. This phenomenon is directly observable in the nitrogen cycle of the Saanich Inlet.

The conversion of nitrite to nitric oxide facilitated by *nirK* is a crucial step in the overall denitrification pathway, without which denitrification could not proceed. However, numerous additional enzymes are also involved in this pathway, and are responsible for other enzymatic reactions.The respective enzymes involved in conversion between nitrogen species is shown to be distributed amongst several taxonomic groups, with differing abundance across the depth gradient. The distribution of these denitrification pathway steps in marine environments over various depths and taxonomic groups can be attributed to both environmental and evolutionary factors.It is observed that as the energetic balance sustaining metabolic activity becomes less favorable and nutrient rich, greater degrees of microbial specialization occurs with respect to *nirK* (Figure 4). For example, the distribution of nutrients and light impact the depth at which certain enzymes can be produced and utilized. Additionally, horizontal gene transfer has contributed to the distribution of certain denitrification genes amongst different taxonomic groups. Given the ubiquitous nature of these processes, it is likely that they evolved early on in Earth's history(16).

References

1. Zaikova E, Walsh DA, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ. 2010. Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. Environ Microbiol.
2. Herlinveaux, R.H., Oceanography of Saanich Inlet in Vancouver Island, British Columbia. Journal of the Fisheries Research Board of Canada, 1962. 19(1): p. 1-37.
3. Ocean Networks Canada. (n.d.). Retrieved April 05, 2018, from http://www.oceannetworks.ca/introduction-saanich-inlet
4. Torres-Beltrán, M., Hawley, A. K., Capelle, D., Zaikova, E., Walsh, D. A., Mueller, A., . . . Hallam, S. J. (2017). A compendium of geochemical information from the Saanich Inlet water column. *Scientific Data,4*, 170159. doi:10.1038/sdata.2017.159
5. Hawley AK, Torres-Beltrán M, Zaikova E, et al. A compendium of multi-omic sequence information from the Saanich Inlet water column. *Scientific Data*. 2017;4:170160. doi:10.1038/sdata.2017.160.
6. Mosier, A., Syers, J. K., & Freney, J. R. (2013). *Agriculture and the Nitrogen Cycle Assessing the Impacts of Fertilizer Use on Food Production and the Environment*. Retrieved April 5, 2018.
7. Carlson CA, Ingraham JL. Comparison of denitrification by Pseudomonas stutzeri, Pseudomonas aeruginosa, and Paracoccus denitrificans. *Applied and Environmental Microbiology*. 1983;45(4):1247-1253.
8. Zumft WG. Cell biology and molecular basis of denitrification. *Microbiology and Molecular Biology Reviews*. 1997;61(4):533-616.
9. Mulvaney, R., Khan, S. & Mulvaney, C. Biol Fertil Soils (1997) 24: 211. https://doi.org/10.1007/s003740050233
10. Hawley, A., Torres-Beltrán, M., Zaikova, E., Walsh, D., Mueller, A., Scofield, M., Kheirandish, S., Payne, C., Pakhomova, L., Bhatia, M., Shevchuk, O., Gies, E., Fairley, D., Malfatti, S., Norbeck, A., Brewer, H., Pasa-Tolic, L., del Rio, T., Suttle, C., Tringe, S. and Hallam, S. (2017). A compendium

of multi-omic sequence information from the Saanich Inlet water column. Scientific Data, 4, p.170160.
11. TreeSAPP. (2018). *GitHub,* from https://github.com/hallamlab/treesapp
12. Morgan-Lang, C, Dill-McFarland, K (2018). TreeSAPP demo. [HTML Document] University of British Columbia, MICB 425 Project 2 Data Folder. Vancouver.
13. iTOL: Interactive Tree of Life. *iTOL,* from http://itol.embl.de/
14. Helen D, Kim H, Tytgat B, Anne W. Highly diverse *nirK* genes comprise two major clades that harbour ammonium-producing denitrifiers. *BMC Genomics*. 2016;17:155. doi:10.1186/s12864-016-2465-0.
15. Braker, G., et al. “Nitrite Reductase Genes (NirK and NirS) as Functional Markers To Investigate Diversity of Denitrifying Bacteria in Pacific Northwest Marine Sediment Communities.” *Applied and Environmental Microbiology*, vol. 66, no. 5, Jan. 2000, pp. 2096–2104., doi:10.1128/aem.66.5.2096-2104.2000.
16. Falkowski, P., Fenchel, T., & Delong, E. (2008). The Microbial Engines That Drive Earth’s Biogeochemical Cycles. Science Special Reviews,320.

Appendix

Appendix 1.

```
#create screen
$ screen -S depth
#return to screen
$ screen -r depth


#!/bin/bash

export
WISECONFIGDIR=/home/connor/TreeSAPP//data//genewise_support_files//wi
secfg

time treesapp.py -T 8 --verbose --delete \
-t D0301 -i bucket/MetaG_assemblies/SI072_LV_10m_DNA.scaffolds.fasta
--rpkm -r bucket/MetaG_reads/SI072_LV_10m.anqdp.fastq.gz -o
treesapp_out_dir_depth10m

rm
treesapp_out_dir_depth120m/RPKM_outputs/SI072_LV_10m_DNA.scaffolds_ge
nes.sam

time treesapp.py -T 8 --verbose --delete \
-t D0301 -i bucket/MetaG_assemblies/SI072_LV_100m_DNA.scaffolds.fasta
```

```
--rpkm
-r bucket/MetaG_reads/SI072_LV_100m.anqdp.fastq.gz -o
treesapp_out_dir_depth100m

rm
treesapp_out_dir_depth120m/RPKM_outputs/SI072_LV_100m_DNA.scaffolds_g
enes.sam

time treesapp.py -T 8 --verbose --delete \
-t D0301 -i bucket/MetaG_assemblies/SI072_LV_120m_DNA.scaffolds.fasta
--rpkm -r bucket/MetaG_reads/SI072_LV_120m.anqdp.fastq.gz -o
treesapp_out_dir_depth120m

rm
treesapp_out_dir_depth120m/RPKM_outputs/SI072_LV_120m_DNA.scaffolds_g
enes.sam

time treesapp.py -T 8 --verbose --delete \
-t D0301 -i bucket/MetaG_assemblies/SI072_LV_135m_DNA.scaffolds.fasta
--rpkm -r bucket/MetaG_reads/SI072_LV_135m.anqdp.fastq.gz -o
treesapp_out_dir_depth135m

rm
treesapp_out_dir_depth135m/RPKM_outputs/SI072_LV_135m_DNA.scaffolds_g
enes.sam

time treesapp.py -T 8 --verbose --delete \
-t D0301 -i bucket/MetaG_assemblies/SI072_LV_150m_DNA.scaffolds.fasta
--rpkm -r bucket/MetaG_reads/SI072_LV_150m.anqdp.fastq.gz -o
treesapp_out_dir_depth150m

rm
treesapp_out_dir_depth150m/RPKM_outputs/SI072_LV_150m_DNA.scaffolds_g
enes.sam

time treesapp.py -T 8 --verbose --delete \
-t D0301 -i bucket/MetaG_assemblies/SI072_LV_165m_DNA.scaffolds.fasta
--rpkm -r bucket/MetaG_reads/SI072_LV_165m.anqdp.fastq.gz -o
treesapp_out_dir_depth165m

rm
treesapp_out_dir_depth165m/RPKM_outputs/SI072_LV_165m_DNA.scaffolds_g
enes.sam
```

```
time treesapp.py -T 8 --verbose --delete \
-t D0301 -i bucket/MetaG_assemblies/SI072_LV_200m_DNA.scaffolds.fasta
--rpkm -r bucket/MetaG_reads/SI072_LV_200m.anqdp.fastq.gz -o
treesapp_out_dir_depth200m

rm
treesapp_out_dir_depth200m/RPKM_outputs/SI072_LV_200m_DNA.scaffolds_g
enes.sam
```

Appendix 2.

```
#create screen
$ screen -S depth
#return to screen
$ screen -r depth


#!/bin/bash
export
WISECONFIGDIR=/home/connor/TreeSAPP//data//genewise_support_files//wi
secfg

time treesapp.py -T 8 --verbose --delete --pairing pe -t D0301 -i
bucket/MetaG_assemblies/SI072_LV_10m_DNA.scaffolds.fasta --rpkm -r
bucket/MetaT_reads/SI072_LV_10m.qtrim.3ptrim.artifact.rRNA.clean.fast
q.gz -o treesapp_out_dir_depth10mRNA

rm treesapp_out_dir_depth10RNA/RPKM_outputs/*.sam


time treesapp.py -T 8 --verbose --delete --pairing pe -t D0301 -i
bucket/MetaG_assemblies/SI072_LV_100m_DNA.scaffolds.fasta --rpkm -r
bucket/MetaT_reads/SI072_LV_100m.qtrim.3ptrim.artifact.rRNA.clean.fas
tq.gz -o treesapp_out_dir_depth100mRNA

rm treesapp_out_dir_depth100RNA/RPKM_outputs/*.sam


time treesapp.py -T 8 --verbose --delete --pairing pe -t D0301 -i
bucket/MetaG_assemblies/SI072_LV_120m_DNA.scaffolds.fasta --rpkm -r
bucket/MetaT_reads/SI072_LV_120m.qtrim.3ptrim.artifact.rRNA.clean.fas
tq.gz -o treesapp_out_dir_depth120mRNA

rm treesapp_out_dir_depth120RNA/RPKM_outputs/*.sam
```

```
time treesapp.py -T 8 --verbose --delete --pairing pe -t D0301 -i
bucket/MetaG_assemblies/SI072_LV_135m_DNA.scaffolds.fasta --rpkm -r
bucket/MetaT_reads/SI072_LV_135m.qtrim.3ptrim.artifact.rRNA.clean.fas
tq.gz -o treesapp_out_dir_depth135mRNA

rm treesapp_out_dir_depth135RNA/RPKM_outputs/*.sam


time treesapp.py -T 8 --verbose --delete --pairing pe -t D0301 -i
bucket/MetaG_assemblies/SI072_LV_150m_DNA.scaffolds.fasta --rpkm -r
bucket/MetaT_reads/SI072_LV_150m.qtrim.3ptrim.artifact.rRNA.clean.fas
tq.gz -o treesapp_out_dir_depth150mRNA

rm treesapp_out_dir_depth150RNA/RPKM_outputs/*.sam

time treesapp.py -T 8 --verbose --delete --pairing pe -t D0301 -i
bucket/MetaG_assemblies/SI072_LV_165m_DNA.scaffolds.fasta --rpkm -r
bucket/MetaT_reads/SI072_LV_165m.qtrim.3ptrim.artifact.rRNA.clean.fas
tq.gz -o treesapp_out_dir_depth165mRNA

rm treesapp_out_dir_depth165RNA/RPKM_outputs/*.sam


time treesapp.py -T 8 --verbose --delete --pairing pe -t D0301 -i
bucket/MetaG_assemblies/SI072_LV_200m_DNA.scaffolds.fasta --rpkm -r
bucket/MetaT_reads/SI072_LV_200m.qtrim.3ptrim.artifact.rRNA.clean.fas
tq.gz -o treesapp_out_dir_depth200mRNA

rm treesapp_out_dir_depth200RNA/RPKM_outputs/*.sam
```