Project 1: an analysis of Planctomycetes using mothur vs. QIIME2 Microbiome Data Analysis

Abstract

The Saanich Inlet is host to a diverse community of microbes with varying biogeochemical processes. In this report, two amplicon sequencing processors, mothur and qiime2 were simultaneously used to process 16S sequencing data for the phylum Planctomycetes from seven different depths of the Saanich Inlet. The data was then analyzed to compare the two processing methods based on their unique criteria of clustering and denoising. Using clustering techniques of mothur we generated 110 OTUs in contrast to the 122 ASVs in QIIME2. Our findings demonstrated a similar trend in both OTU and ASV abundance data across depth, oxygen and ammonium concentrations. In particular, an increase in OTUs was identified at higher ammonium concentrations which are indicative of the anammox process of planctomycetes. However, the only statistically significant relationships found were OTU abundance with depth and ammonium concentration. No trends identified for ASV abundance were statistically significant likely due to the differences in stringency between the two pipelines.

Introduction

The Saanich Inlet has been studied for four decades as a model system for various oceanographic and biogeochemical processes. It is a seasonally anoxic fjord located off the coast of British Columbia, Vancouver Island specifically (1). During the spring and summer, the deep waters of the Saanich Inlet basin become increasingly anoxic, and CH_4, NH_4, and H_2S accumulate. However, in the late summer and fall, nutrient rich and oxygenated water enter the Inlet due to oceanic upwelling. This seasonal phenomenon is known as deep water renewal, and alters the composition of the water column (2,3). As a result, the deepest 80 m of the inlet undergoes fluctuation between oxygenated and reduced states every year.

At the surface of the water column, primary production provides a fixed carbon source for aerobic respiration, which progressively depletes the water column of dissolved O_2. This phenomenon generates a redoxcline, as microbes begin to use alternative respiratory substrates such as NO_2^-, NO_3^-, SO_4^-, Fe. or CO_2 (4). This in turn has been previously shown to elicit “taxonomic partitioning” as microbes with different metabolic capabilities are shown to persist at different depths(5).

In this report, we analyzed how the microbial community composition differs across metagenomic data sets collected from 7 depths of the Saanich Inlet. We chose to focus on the phylum Planctomycetes in particular, due to its unique metabolic capabilities. Specifically, members of Planctomycetes are the only clades of bacteria that are capable of the anaerobic oxidation of ammonium, also known as anammox. This phylum is morphologically unique, displaying a ribosome free region termed as the paryphoplasm, that surrounds the perimeter of

the cell, a pirellulosome and a double membrane nucleoid zone (6). Planctomycetes also have a specialized organelle called an anammoxosome. They first exist in a sessile state, later budding into a flagellated stage as part of their lifecycle, before settling to begin reproduction (7,8). Anammox bacteria are responsible for converting fixed nitrogen back into N_2 gas, and are thus gatekeepers of the biological nitrogen cycle. Indeed, it has been estimated that anammox bacteria produce approximately half of the nitrogen gas in the atmosphere(9).

Our reason for the selection of planctomycetes as our subject phylum is that previously planctomycetes have been reported as being present in largely oxic environments. They have been shown to exhibit a linear trend of presence vs pH, with a sharp decrease in abundance at greater depths, returning to a second maximum at an anoxic depth (10). Because of the suggested correlation of abundance at various oxygen and pH profiles, we felt that it would be interesting to study the profile of abundance for this phylum with respect to various depths and elemental concentrations within the Saanich Inlet to glimpse more information towards their preferred environmental niches.

Additionally, we wanted to compare two traditional metagenomic analysis programs used for the analysis of amplicon sequencing data: mothur and qiime2. Essentially, the main difference between these two programs is how they classify species and other taxonomic groups. Mothur clusters amplicon reads into operational taxonomic units (OTUs) whereas qiime2 quantifies the amplicon sequences using amplicon sequence variants (ASVs) (11,12). During amplicon analysis using OTUs, amplicon reads are clustered based upon similarity, typically 16S sequences that have 97% similarity can be considered a “species”(7). In ASV-based amplicon analysis, one de-noises amplicon sequences with 99% similarity to end up with singular unique amplicon sequences that are used for further downstream analysis. As data is discarded during the de-noising, this is thought to compensate somewhat for errors that occur during amplification of the target gene(8).

Methods

Water samples from the Saanich inlet were sampled at depths of 10, 100, 120, 135, 150 and 200m for large volume SSU rRNA gene tagging. Sixteen depth locations where sharp oxygen gradients were present were sampled for high resolution SSU rRNA gene tagging at monthly intervals. Samples were then transported in pooled batches for filtration using 0.22µm Sterivex Millipore filters to isolate biomass. Cell lysis was executed by addition of lysozyme, Proteinase K and 20% SDS. Isolation of genetic material was done via phenol chloroform fractionation. Samples were then washed 3 times on 10K Amicon Millipore cartridge with Tris EDTA prior to centrifugation and resuspension in 150-400µl volumes. Extracted DNA samples were subject to 16s rRNA sequencing of the V4-5 ribosomal region using 515F and 808R primers on the Illumina MiSeq platform with 2 X 300bp technology. A metagenomic library was generated and was subjected to JGI library quality control standards to check for DNA degradation (13). Samples collected were processed through the mothur and QIIME 2 pipelines to determine the number of OTUs and ASVs present at various depths.

In the mothur pipeline, contiguous regions of 16s rRNA reads were joined to create fewer sequences for analysis and to discard low quality sequences. Sequences were then aligned to the SILVA database and cut to derive consistent start and end sites. Sequences were clustered according to similarity and single occurrence and chimeric sequences were removed. Clustered sequences were classified as an OTU at 97% similarity. A phyloseq object was then created from this data containing an OTU table, taxonomy, and sample metadata. Data was imported into R studio and extra columns were removed, and taxonomy level was arranged into single columns (14).

In the QIIME2 pipeline, data clean up was done alongside ASV determination. Sequence quality was visualized using heads and tails of reads. Sequences were trimmed to approximate the mothur sequence trims. ASV determination was done using the Dada2 protocol, and non-qualifying sequences were removed. ASVs were then classified using Silva version 119, at 99% sequence similarity. Following this, results were formatted into a table and converted into a phyloseq object containing the ASVs, sample meta data and taxonomies.

The phyloseq object was imported into R (15). Both QIIME 2 and Mothur pipeline phyloseq objects were fed into R studio version 1.4.3. The tidyverse and phyloseq packages were first loaded into R. Sample size was set by random selection to 100,000 samples. Alpha-diversity was visualized across depth and oxygen concentration by plotting Shannon's diversity index against oxygen concentration. The percent abundance of Planctomycetes was visualized across depth and oxygen concentration with other phyla in a faceted plot. The abundance of planctomycetes across depth, oxygen, and ammonium concentration was visualized by plotting abundance sum against depth. The p-value of planctomycetes across depths and oxygen concentration was calculated in order to determine whether there is significant correlation with OTU abundance. The abundances of OTUs within planctomycetes was plotted against depth and oxygen concentration. QIIME 2 data was loaded onto R and the sample size normalized to 100,000. Alpha diversity across depth and oxygen was plotted. The observed data counts in the Saanich Inlet were compiled by depth. Next, the abundance of Planctomycetes across depth, oxygen, and ammonium concentration was plotted. The abundance of ASVs of Planctomycetes across depth and oxygen concentration was plotted. To generate the statistical data, including p values, for all of the OTUs and ASVs identified, Miguel Desmarais' code for looping linear models was used (16).

Results

We characterized the taxonomic presence of the phylum Planctomycetes through our depth profile to examine their abundance through the water column as well as comparing the two methods of sequence clusters for OTUs and ASVs. This aims to elucidate and characterize how a type of organism may fulfill an ecological niche which may be enhanced or suppressed according to the environmental selection pressures and nutrient availability.

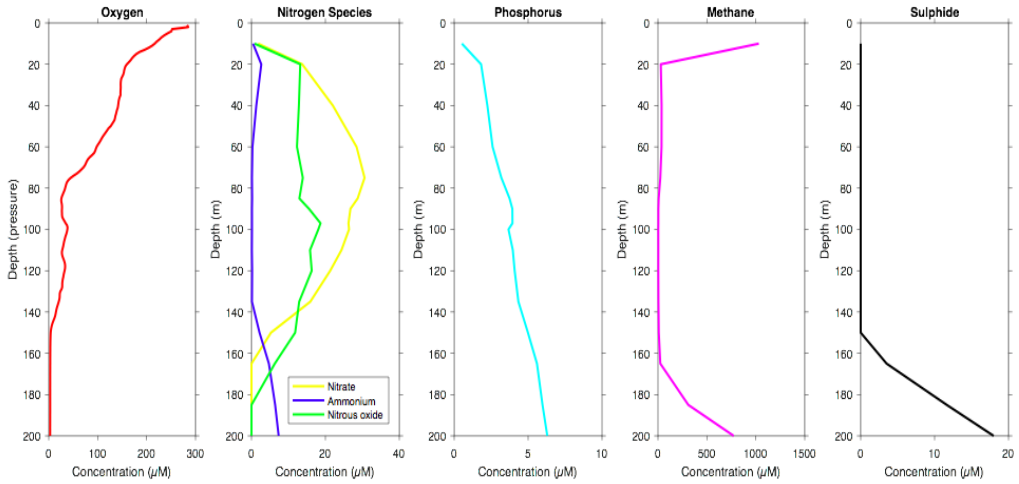

Fig 1. Nutrient depth profile for Saanich Inlet (adapted from MICB 405, group 7)

Beginning with the nutrient depth profile for Saanich Inlet (Figure 1), it is observed that oxygen concentration can be approximately described with a logarithmic decline across depths down to roughly zero micromolars past 160 meters. With reference to the commonly accepted operational definition of anoxia, being any body of water containing less than 20μM oxygen (17), all depths below 140 meters encompasses the anoxic zone in Saanich Inlet. This corresponds strongly to the availability of nitrogen species, as reduced nitrogen species (ammonium), C1 molecules (methane), and sulphides can be observed to inversely correlate with the presence of oxygen. These biochemical species show notable presence in the water column beginning from 140-160 meters down to the sampled 200 meters.

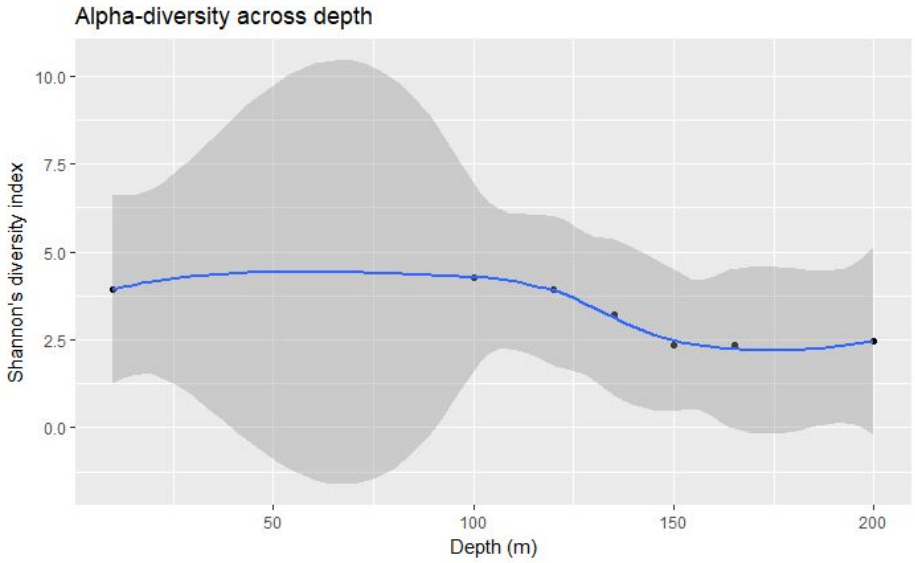

Fig 2a. Shannon's diversity across depth for OTUs

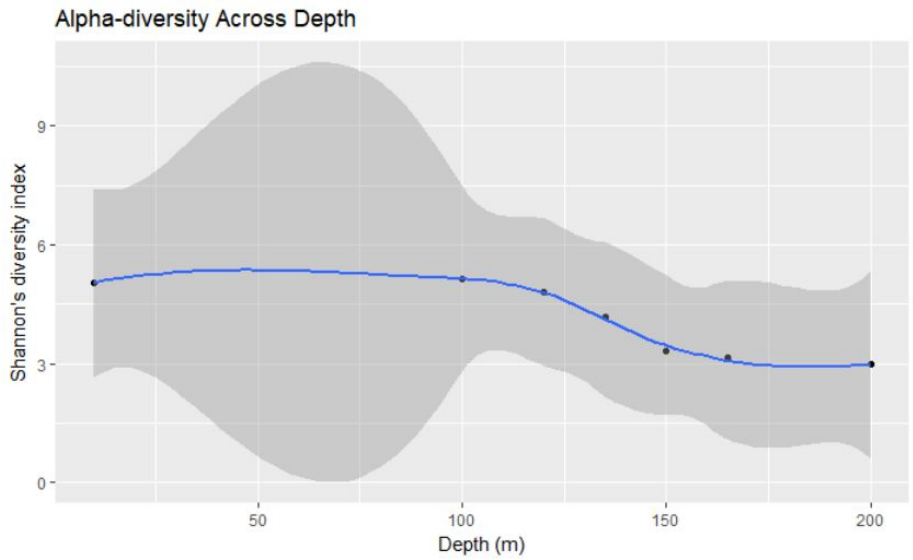

Fig. 2b Shannon's diversity across depth for ASVs

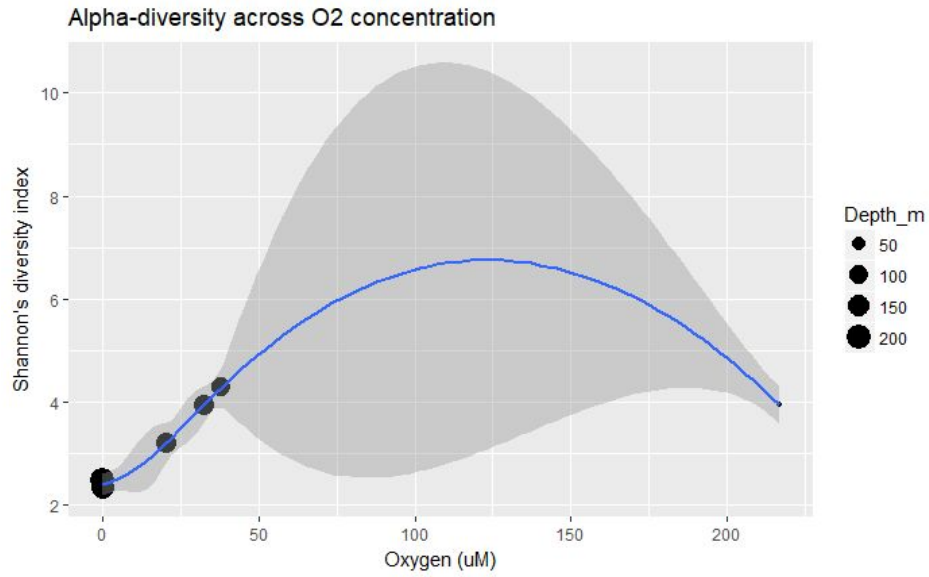

Figure 2c. Alpha-diversity across O2 concentration (OTU)

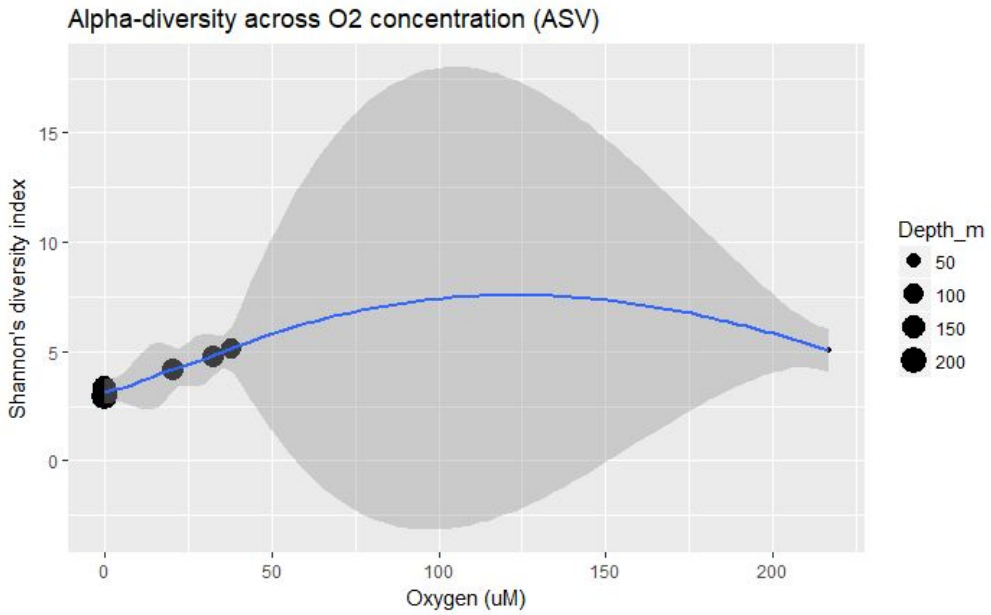

Figure 3d. Alpha-diversity across O2 concentration (ASV)

Following sequence cleanup and clustering, the clustered 16s sequences were first examined for alpha diversity through calculating Shannon's diversity index for both the OTU and ASV group sets (Figure 2a). Alpha diversity is shown to exist in two steady states. From the water surface to approximately depth 125 meters, the alpha diversity is highest for both the OTU-based and ASV-based data. Roughly below 125 meters a transition zone is observed where the alpha diversity decreases to a second steady state to 150 meters where it is consistent down the rest of the water column. OTU and ASVs both demonstrate this overall trend, with the notable observation that ASVs has higher values for the initial steady state (above 125 meters) while showing lower values for the second steady state (below 150 meters) (Fig 2b).

Correlating Shannon's diversity index values for OTUs and ASVs with the oxygen gradient at the respective depths, it is observed that alpha diversity sharply increases with respect to the presence of oxygen up to the point of around 25-50 micromolars, where it then remains plateaued with greater oxygen concentration (Fig 2c). Both OTU- and ASV-based methods exhibit this trend. ASV-based data showed an increase of diversity index values across the chart with a 0.5-1 point increase at each index value. A clear breakdown of oxic versus anoxic for both OTUs and ASVs (with our previous operational definitions of these two terms) shows oxic depths having definitively higher diversity compared to anoxic zones (Fig 2d).

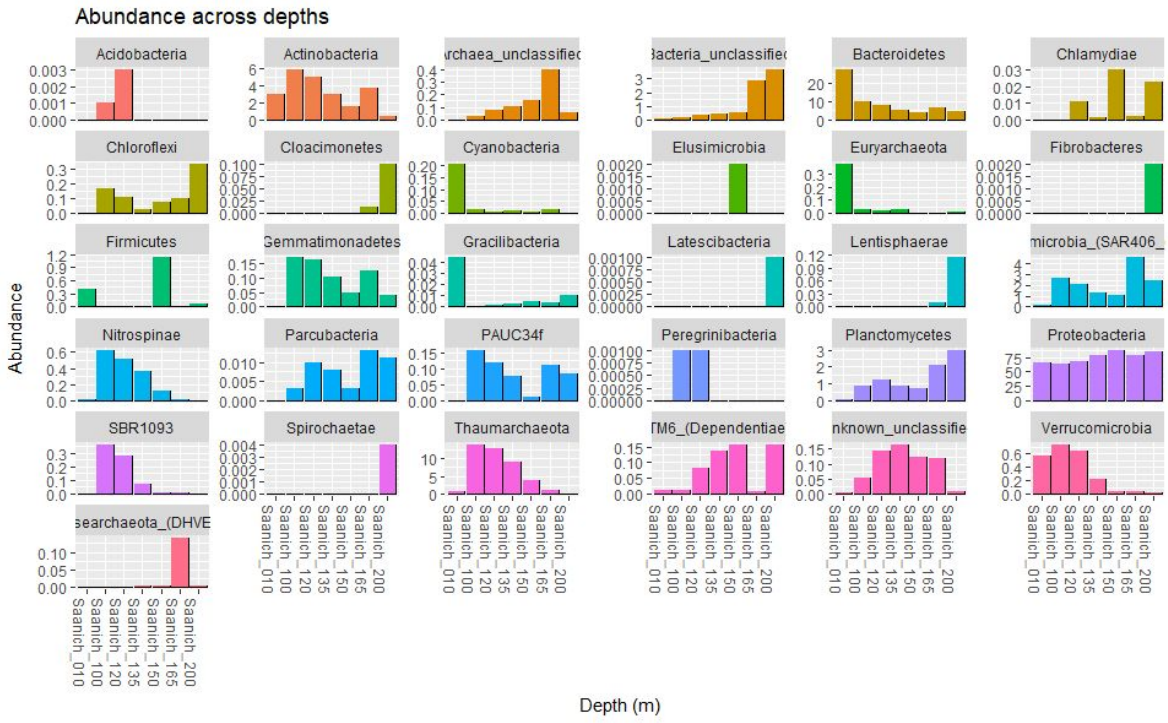

Fig 3a. Abundance of Planctomycetes across depth (OTUs)

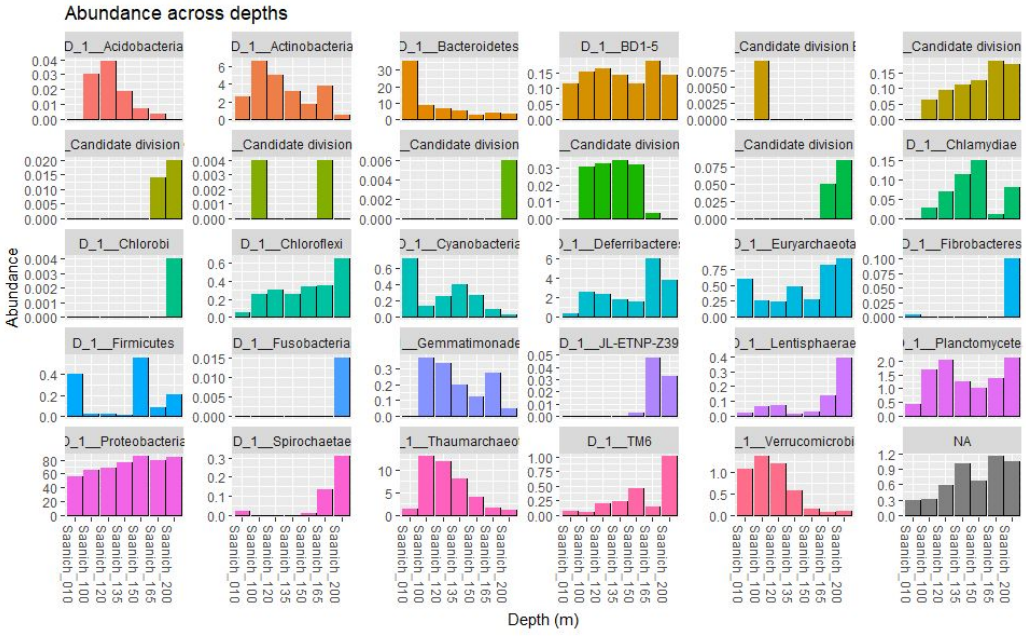

Fig 3b. Abundance of Planctomycetes across depth (ASVs)

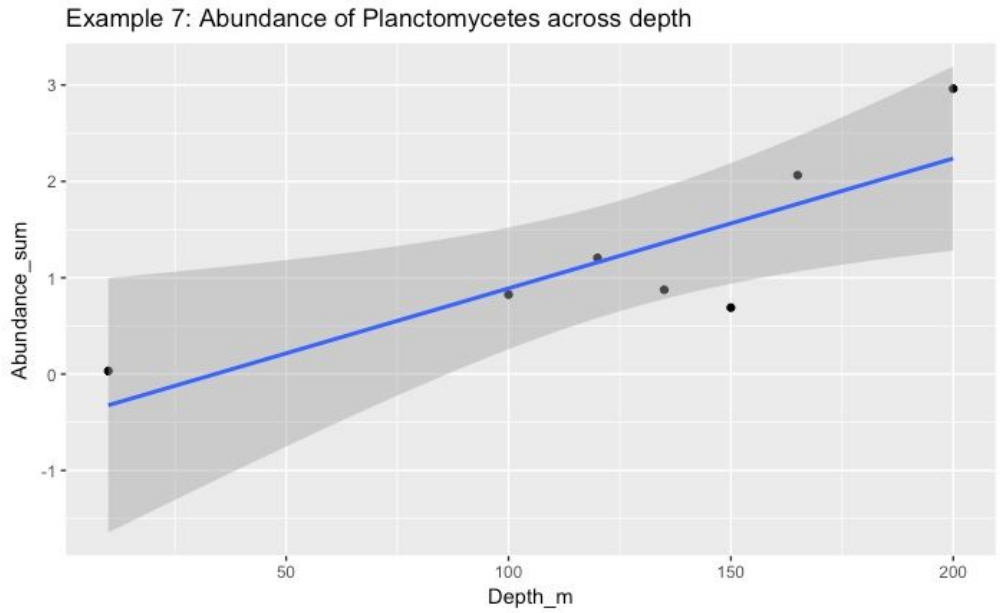

Figure 3c. Abundance of Planctomycetes across depth (OTUs)

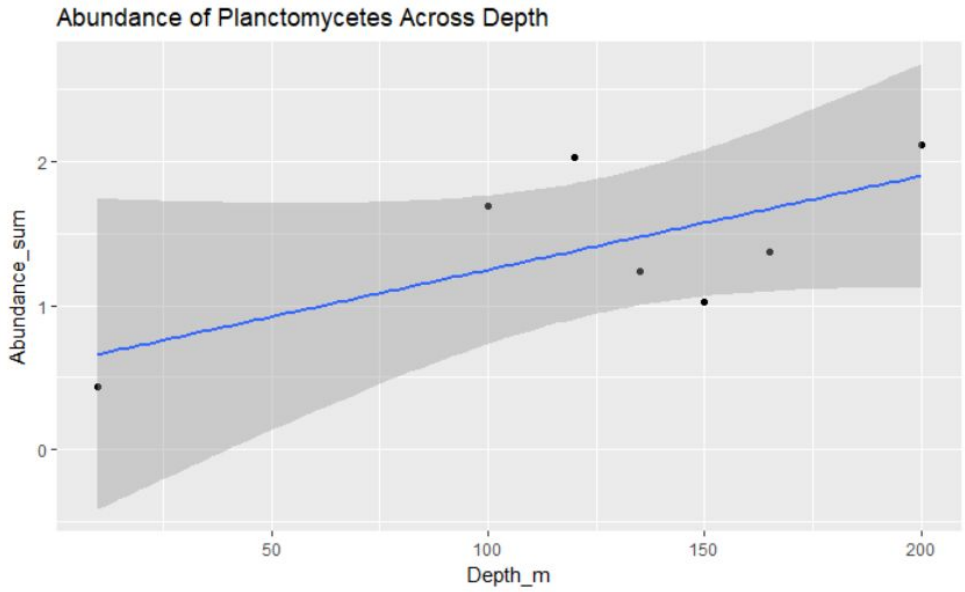

Figure 3d. Abundance of Planctomycetes across depth (ASVs)

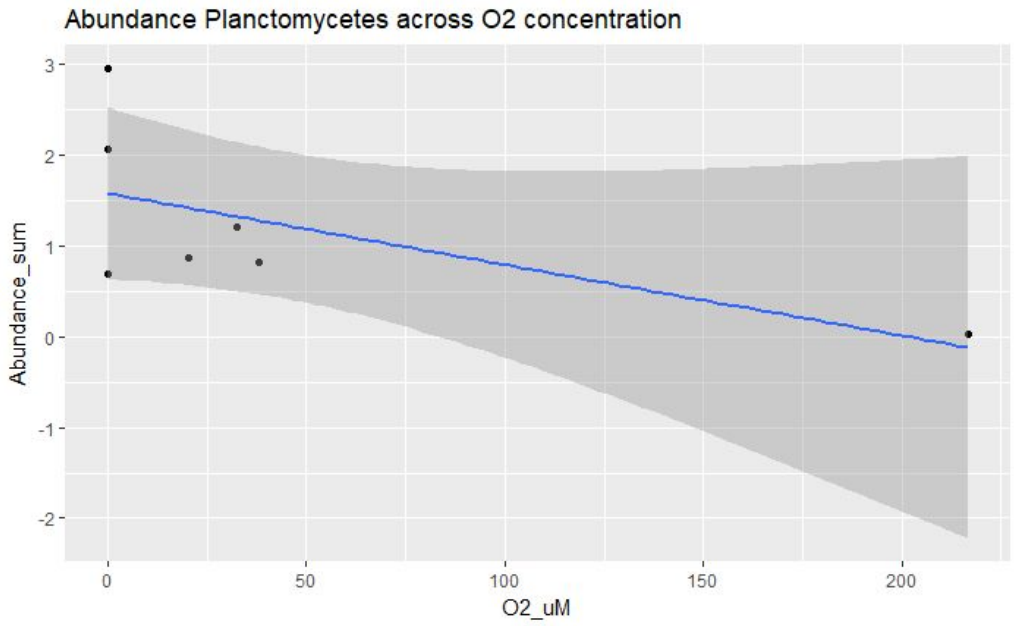

Figure 3e. Abundance of Planctomycetes across O2 concentration (OTUs)

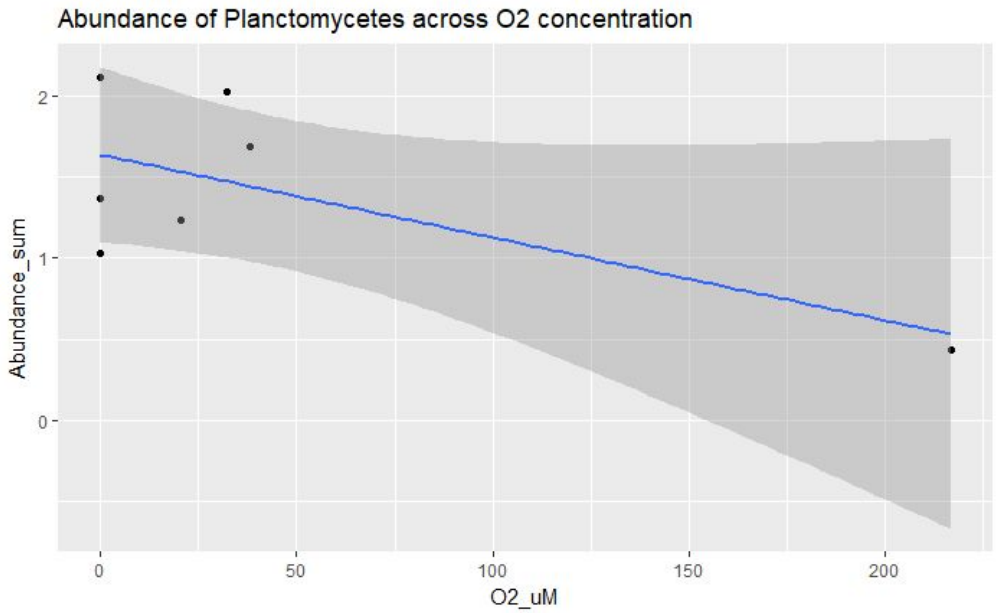

Figure 3f. Abundance of Planctomycetes across O2 concentration (ASVs)

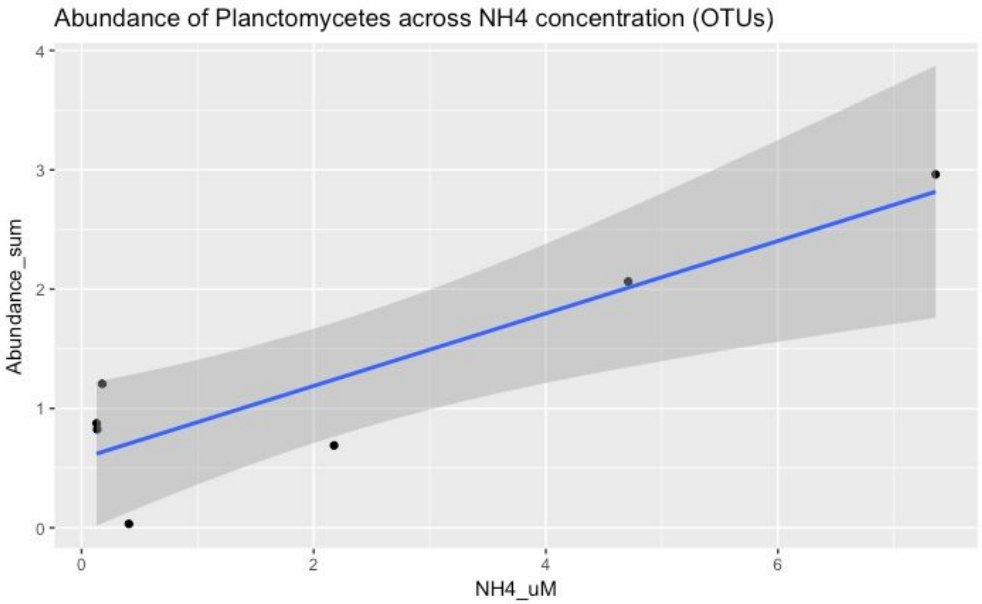

Figure 3g. Abundance of Planctomycetes across NH4 concentration (OTUs)

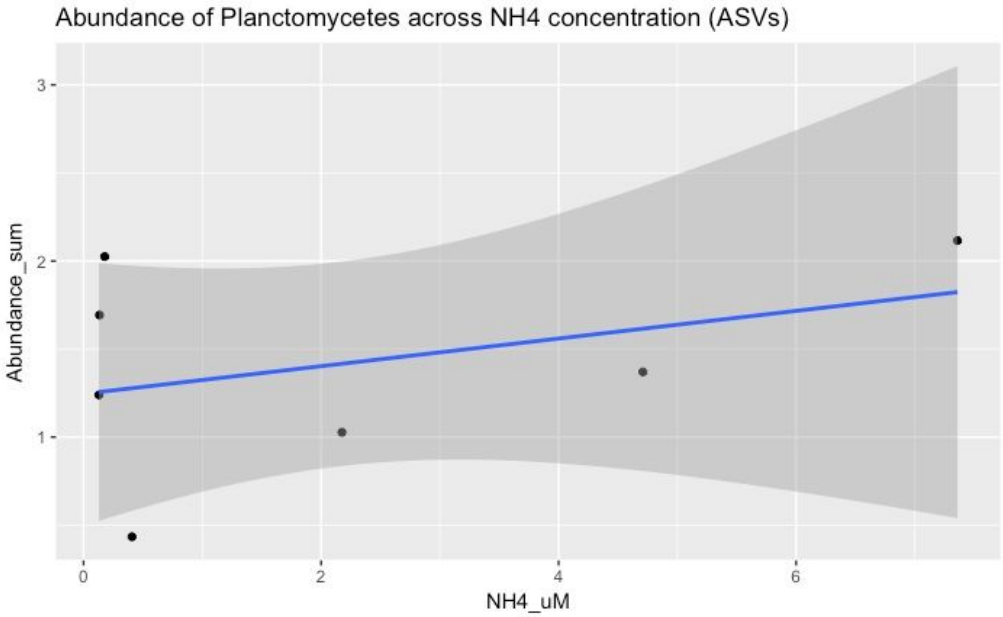

Figure 3h. Abundance of Planctomycetes across NH4 concentration (ASVs)

Table 1: p-values for Planctomycetes abundance with respect to depth and oxygen and ammonium concentration

	Depth	Oxygen concentration	Ammonium Concentration
OTU	0.01989	0.1332	0.007673

ASV	0.0961	0.1024	0.4011

Proceeding to a breakdown of OTU and ASV taxonomy, OTU clustering using mothur generated 110 OTUs in comparison to the 122 ASVs generated by QIIME2. A simple breakdown of taxonomy at the Phylum level with respect to Planctomycetes abundance reveals a sharp increase beneath the 165 meter depth profiles when clustered using mothur (Figure 3a). ASV data show Planctomycetes having two peak abundances at 120 meter and 200 meters respectively (Figure 3b). An isolated plot of Planctomycete abundance across depth (Figure 3c) shows a positive trend where abundance increases with depth, consistently observed between OTUs and ASVs. For all OTU and ASV clusters, it is observed that both clustering methods showed an overall Planctomycetes abundance decrease past depth 100 meters (Figure 3c, 3d). Although the trend for the abundance of Planctomycetes across depth is similar for both OTU and ASV data, the p-values of 0.01989 for OTUs and 0.0961 for ASVs, respectively, show that the abundance of Planctomycetes at various depths differs significantly under OTU clustering methods, but conversely, is insignificant under ASV methods (Table 1).

The Planctomycete phylum is not known to require substantial oxygen for survival, and this is observed with both OTU- and ASV-based methods, as a majority of the abundance of Planctomycetes fall under 40uM of oxygen and only a small minority found in 250 uM of oxygen (Figure 3e, 3f). Taking that into account, the calculated p-value for OTUs is 0.1332 and ASVs, 0.1204 rendering Planctomycete abundance across oxygen concentrations insignificant (Table 1). Nonetheless, when referencing the observed positive correlation between Planctomycete abundance and depth, the inverse trend with decreasing oxygen concentration appears to be a convenient reinforcement as well as an explanation for the survival of the phylum in environments lacking oxygen.

Lastly, as the Planctomycetes phylum is known to perform anammox using ammonium as a substrate, we created a plot of the abundance of Planctomycetes across ammonium concentration (Figures 3g, 3h). Interestingly, while the trendlines for this relationship for both OTU and ASV derived data are both positive, the plots look significantly different. The p-values of 0.007673 for OTUs and 0.4011 for ASVs, respectively, show that the abundance of Planctomycetes differs significantly under OTU clustering methods, but is insignificant under ASV methods (Table 1). The significance of abundance in depth and ammonia for ASV-derived methods but not OTU derived data can be attributed to the clustering algorithm by QIIME2 and MOTHUR and the denoising protocols of these two pipelines. QIIME2 uses a more stringent 99% similarity threshold for clustering while MOTHUR uses 97% for ASV clustering. The fact that the OTU derived data is found to exhibit a statistically significant relationship between planctomycetes abundance is indicative of the presence of anammox capable bacteria in the Saanich Inlet. However, without further characterization of the metagenomic data set, one cannot conclusively prove this theory.

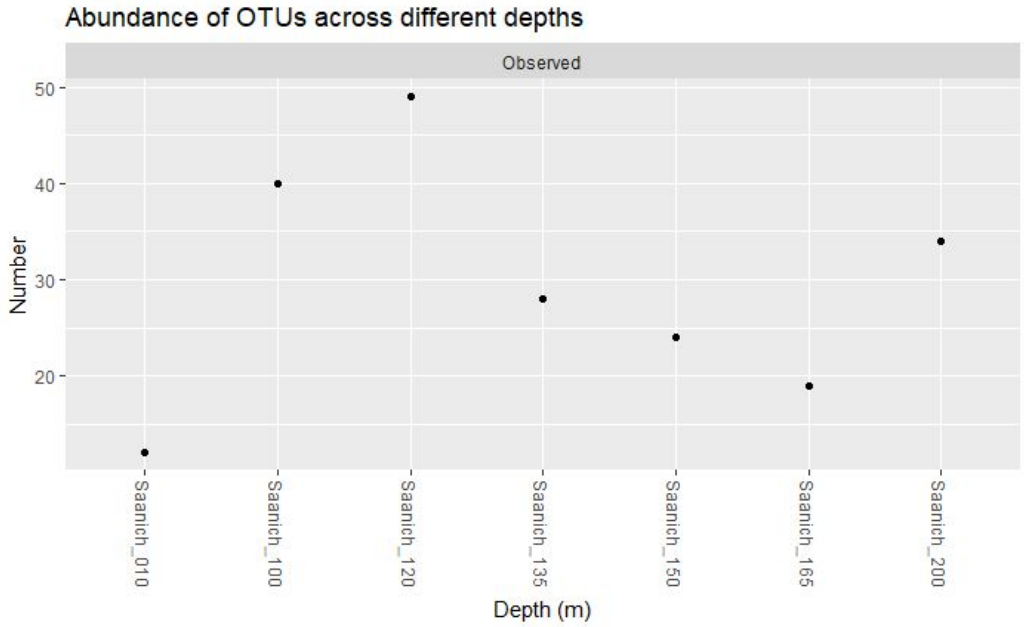

Figure 5a. Abundance of OTUs across different depths

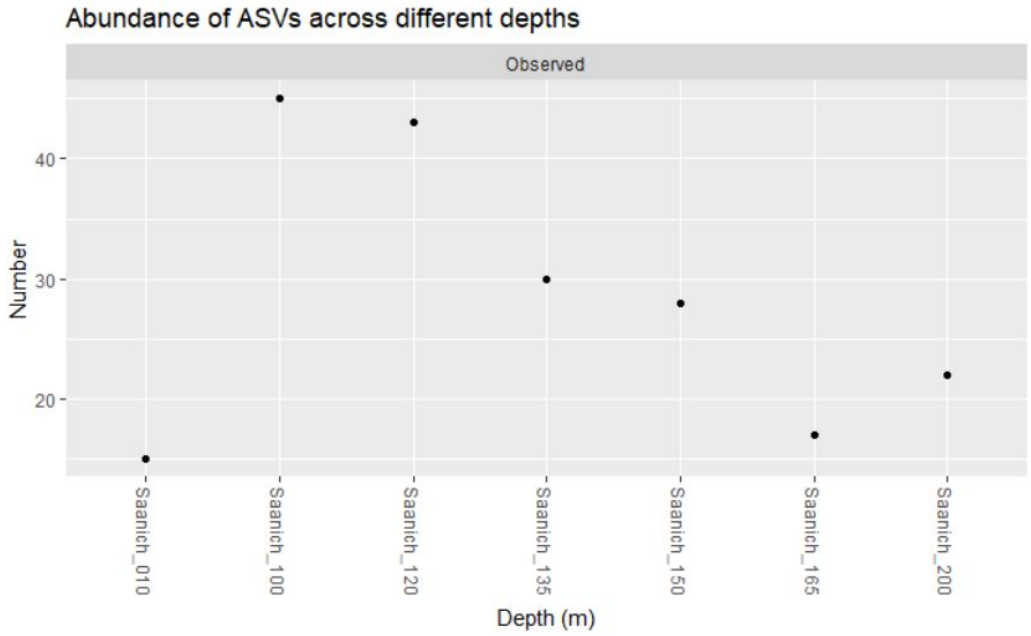

Figure 5b. Abundance of ASVs across different depths

While there are a greater number of total ASVs in comparison to OTUs, the general trend of abundance of OTU/ASVs are similar (Fig 4a, 4b). The abundance of OTUs increase as the depth increases, reaches peak abundance, and decreases starting at 120 meters (Fig 4a). ASV

abundance peaks at 100 meters before decreasing (Fig 4b). Normalizing Planctomycete abundance across depth against all OTU/ASV abundance across depth, suggests that Planctomycete abundance increases despite overall OTU/ASV abundance decreasing (Fig 4a, 4b). This suggests at the expansion of an ecological niche which can be fulfilled by a single taxonomic group.

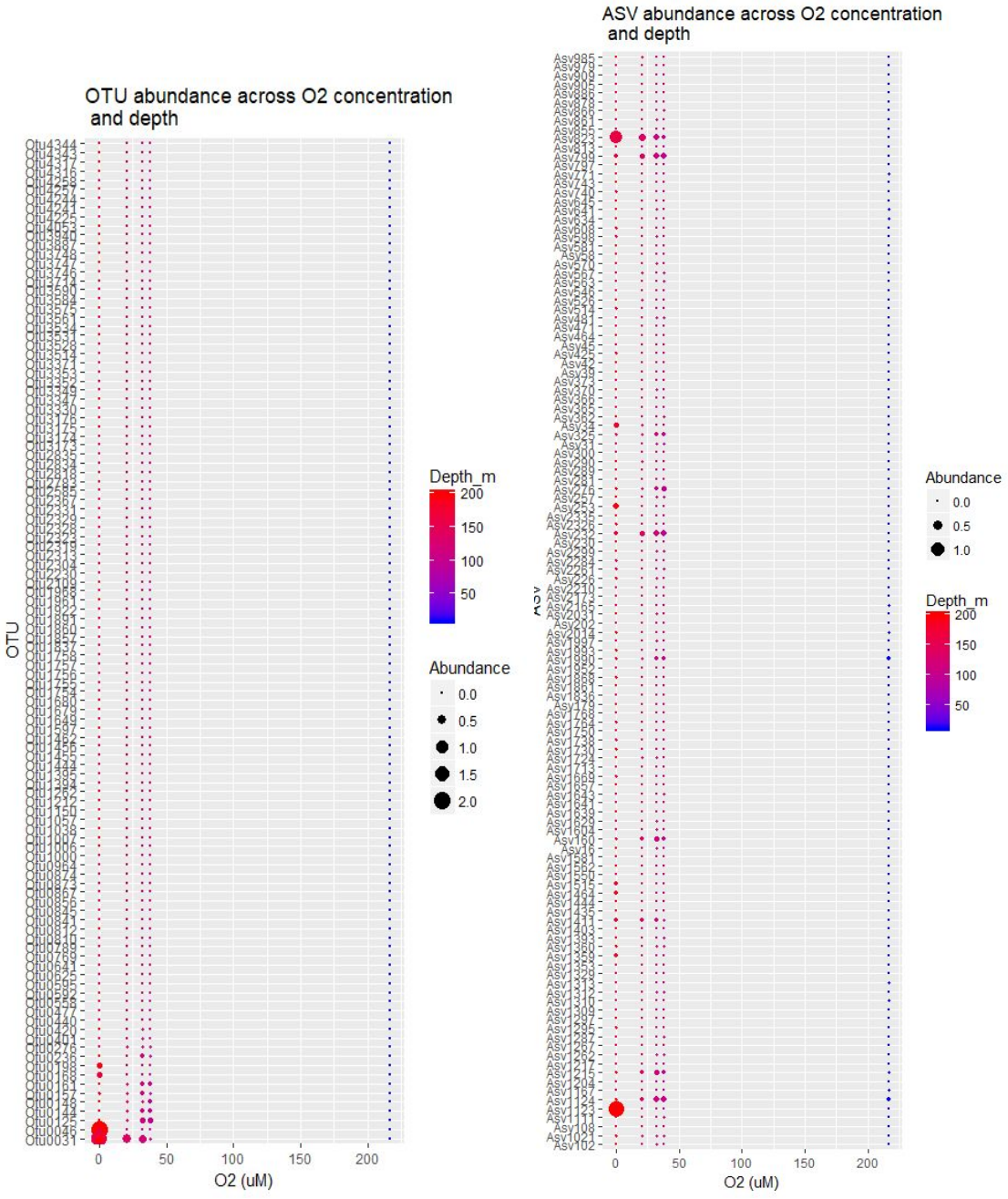

Figure 5. (a) Left: OTU abundance across O2 concentration and depth **(b)** Right: ASV abundance across O2 concentration and depth

Despite the observed trends of abundance increase across depth, a multivariable ANOVA was performed on the Planctomycete clusters with respect to their depth group and oxygen group. Taking a closer look at abundances of OTU/ASVs assigned to the Planctomycetes phyla on an individual cluster basis, it is observed that the OTU/ASVs are primarily below a depth of 100m and the oxygen gradient is present within a select few clusters where these clusters make up most of the abundance of the phylum (Figure 5a, 5b). There are only 4 out of 110 OTUs that are significant but there are 8 out of 122 ASVs that have a p-value of less than 0.05 The abundance is skewed heavily towards a small number of clusters, which potentially explains the large amount of variance between groups that was observed and the absence of statistical significance within the observed trends.

Discussion

Should we assume a statistically significant relationship between increasing depth and planctomycetes abundance, this trend can be explained by analyzing the chemical composition of the water column based upon depth (Figure 1). One of the most famous metabolic processes performed by planctomycetes is anammox, or the anaerobic oxidation of ammonium. Ammonium concentrations remain fairly steady throughout the first 140 m of the water column, whereupon they increase with further depth. Thus, it would make sense if the abundance of anammox capable planctomycetes species would increase in conjunction with increasing levels of their metabolite.

The linear relationship between oxygen concentration and planctomycetes abundance in both OTU and ASV derived data is not significant. As we know that planctomycetes are facultative anaerobes, their metabolic activity, and thus their abundance should not necessarily be correlated with specific oxygen concentration levels(18,19). Thus, this theory is in line with our findings. Although some members of the planctomycetes phylum are obligate anaerobes (18), indeed, the canonical anammox pathway is an anaerobic process, the fact that we have chosen to investigate an entire phylum belies an large amount of possible metabolic diversity. It is thus likely that individual OTUs and ASVs found in the data set are composed of bacteria that use different terminal electron acceptors, electron donors, and sources of organic carbon. To put this taxonomic classification (phylum) in terms of multicellular eukaryotes, members of a eukaryotic phylum, Chordata, contains all organisms with a spinal cord, from humans to lampreys.

The anammox pathway utilizes 3 key enzymes for metabolic conversion of ammonium into N_2. These are nitrate reductase (NIR), hydrazine synthase (HZS) and hydrazine dehydrogenase (HDH). Ammonium to nitrate oxide conversion is mediated by HZR and NIR working in tandem, followed by immediate conversion to hydrazine, requiring input of nitrite which is converted into nitrate. From here, HDH catalyzes the conversion of hydrazine to dinitrogen gas (20). This illustrates the need for large reduced nitrogen inputs in order to supply substrate sufficient for considerable biomass production in anammox bacteria, and is in line with our findings of higher abundance of planctomycetes at depths where ammonium substrate increases using the mothur pipeline. However, using the QIIME 2 pipeline, a clear trend is not visible. Nonetheless,

the general trends in abundance of Planctomycetes across depth, oxygen concentration, and ammonium concentration were similar in both OTU clustering methods and ASV de-noising methods. What differed between the two methods were whether or not the abundance of Planctomycetes were statistically different across depth and ammonium concentration (Table 1). It appears that the higher stringency of ASV denoising produced fewer hits relative to the OTU generation pipeline. As a result, it is possible that some of the OTUs generated by mothur may not be representative of Planctomycetes, which is reflected by the significant differences found in OTUs but not ASVs.

As one can see from table 1: p-values for Planctomycetes abundance with respect to depth and oxygen and ammonium concentration, p values from the same statistical analysis conducted using data from the two pipelines vary significantly depending on the pipeline used. Specifically, OTU based data from the mothur pipeline resulted in much lower p values than the corresponding analysis of the ASV based data from the qiime2 pipeline. This is most likely due to the different cutoffs required to create OTUs versus ASVs. As we discussed in the methods section, while both mothur and qiime determines some level of taxonomy by comparing the amplicon 16S rRNA reads to the SILVA database, the cutoffs for what defines an OTU and an ASV are different. Clustered sequences were classified as an OTU when sequences had 97% and above similarity, while ASVs are classified 99% sequence similarity. This more stringent requirement for ASVs means that the data set will be broken up into a greater number of ASVs than corresponding OTUs.

Future directions for this project may involve more specific characterizations of taxonomy down the tree of life followed by recruitment of non-PCR amplified assembled genomes for the analysis of single assembled genomes (SAGs). This way we can recreate the metabolic potential on a cellular level from a diverse metagenomic sequence set. Through the characterization of sequence clusters, we are only able to infer the metabolic relationship of anammox with the ecological niches of Planctomycetes within Saanich Inlet. With further analysis into assembled genomes, a direct metabolic link can be established that will shed greater insight on the metabolic causes that are attributed to the ecological presence and abundance that was observed in this study.

References

1. Torres-Beltrán, M., et al., Methanotrophic community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. Frontiers in Marine Science, 2016. 3(268).
2. Saanich Inlet. 2009 [cited 2017 11/25/17]; Available from: http://www.pac.dfo-mpo.gc.ca/science/oceans/BCinlets/saanich-eng.htm.
3. 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.
4. Orsi W, Song Y, Hallam S, Edgcomb V. 2012. Effect of oxygen minimum zone formation on communities of marine protists. *ISME JOURNAL*. 2012;6:1586-1601 https://www.ncbi.nlm.nih.gov/pubmed/22402396

5. 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. 2010;12:172–191. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1462-2920.2009.02058.x
6. Lindsay, M., Webb, R., Strous, M., Jetten, M., Butler, M., Forde, R. and Fuerst, J. (2001). Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. *Archives of Microbiology*, 175(6), pp.413-429.
7. Lee, K., Webb, R. and Fuerst, J. (2009). The cell cycle of the planctomycete Gemmata obscuriglobus with respect to cell compartmentalization. *BMC Cell Biology*, 10(1), p.4.
8. Jogler, C., Glöckner, F. and Kolter, R. (2011). Characterization of Planctomyces limnophilus and Development of Genetic Tools for Its Manipulation Establish It as a Model Species for the Phylum Planctomycetes. *Applied and Environmental Microbiology*, 77(16), pp.5826-5829.
9. Teeseling MCV, Mesman RJ, Kuru E, Espaillat A, Cava F, Brun YV, Vannieuwenhze MS, Kartal B, Niftrik LV. 2015. Anammox Planctomycetes have a peptidoglycan cell wall. Nature Communications 6. https://www.nature.com/articles/ncomms7878
10. Ivanova, A. and Dedysh, S. (2012). Abundance, Diversity, and Depth Distribution of Planctomycetes in Acidic Northern Wetlands. *Frontiers in Microbiology*, 3.
11. Callahan BJ, McMurdie PJ, Holmes SP. 2017. "Exact sequence variants should replace operational taxonomic units in marker-gene data analysis". *The ISME Journal* volume 11, pages 2639–2643. doi:10.1038/ismej.2017.119. https://www.nature.com/articles/ismej2017119
12. Fierer N, Brewer T, Choudoir M. 2017. "Lumping versus splitting – is it time for microbial ecologists to abandon OTUs?"http://fiererlab.org/2017/05/02/lumping-versus-splitting -is-it-time-for-microbial-ecologists-to-abandon-otus/
13. 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.
14. Dill-McFarland, K. (2018). mothur pipeline. [HTML Document] University of British Columbia, MICB 425 Project 1 Data Folder. Vancouver.
15. Beni, J. (2018). QIIME2 pipeline. [HTML Document] University of British Columbia, MICB 425 Project 1 data. Vancouver.
16. Desmarais, M. (2018). Linear OTU loop code. University of British Columbia, MICB 425 Project 1 Data Folder. Vancouver.
17. Hawley, A. K., Brewerb, H.M., Norbeck, A.D, Pasa-Tolic, L., Hallam, S. J. (2014), Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. PNAS, 111(31): 11395-11400.
18. Slobodkina, G. B., Panteleeva, A. N., Beskorovaynaya, D. A., Bonch-Osmolovskaya, E. A., & Slobodkin, A. I. (2016). Thermostilla marina gen. nov., sp. nov., a thermophilic, facultatively anaerobic planctomycete isolated from a shallow submarine hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology,66*(2), 633-638. doi:10.1099/ijsem.0.000767
19. Grismer, M. and Collison, R. (2017). The Zeolite-Anammox Treatment Process for Nitrogen Removal from Wastewater—A Review. *Water*, 9(11), p.901.

20. Peng, X., Wu, L., Yu, H., Ai, L. and Fu, L. (2013). CANON process for nitrogen removal from effluents of municipal sewage treatment plants. *Transactions of Tianjin University*, 19(4), pp.255-259.