

group, while anoxic peat was inhabited mostly by Zavarzinella- and Pirellula-like *Planctomyces* [20]. Therefore, the fact that there was no significance between the *Planctomyces* population versus the oxygen concentration and depth was justifiable, since different strains of Planctomycetes colonized different parts of the ocean. The specific strain of anammox Planctomycetes was found to be more abundant in anoxic part of the ocean, but different strains of Planctomycetes were found to be distributed across parts of the ocean.

In terms of the bioinformatic pipelines commonly used in metagenomics, some differences were observed between the Mothur and QIIME2 data:

- 1) Higher Shannon’s diversity indexes were observed in the QIIME2 data when compared to the Mothur data in terms of oxygen concentrations and oxic versus anoxic levels. One possible explanation for the higher Shannon’s diversity indexes in the QIIME2 data could be due to the difference in treatment of OTU and ASV. QIIME2 treats each ASV as an individual species, whereas Mothur uses the representative sequence of each OTU to determine the taxonomy [21]. This could potentially lead to higher Shannon’s diversity indexes in the QIIME2 when compared to the Mothur data.
- 2) The absence of *Pla3_lineage_ge* at the genus level in the QIIME2 data were also noted when compared to the Mothur data. One possibility for this observation is that QIIME2 discards more data through stricter filtering [22]. Another possibility for this observation is that Mothur keeps more of these types of data even if they might not represent the “real” taxa in the community [21].
- 3) The relative abundance of *Planctomyces* at the genus level were indicated to be higher in the QIIME2 data in comparison to the Mothur data. The removal of *Pla3_lineage_ge* by QIIME2 described above would affect the relative abundance calculation, which is the percent composition of an organism of a particular kind relative to the total number of organisms in the area. By not considering *Pla3_lineage_ge* in the total number of organisms, this would lead to a higher relative abundance of *Planctomyces* across all depths in the QIIME2 data.
- 4) The relative abundance in the genus distribution of planctomycetes across samples from Mothur data were also noted to be higher than the relative abundance of QIIME2 data. This could be due to Mothur keeping more data than QIIME2, which would affect the scaling of relative abundance. Similar to the previous point, these extra data kept by Mothur might not represent the “real” taxa in the community [21].

In addition to the comparisons of Mothur and QIIME2 in our analyses, another study demonstrated that between QIIME, Mothur, and MG-RAST, differences were mostly observed at the genus level due to Mothur’s tendency to have unclassified reads [2]. These inconsistencies highlight the limitations of bioinformatics pipelines and their ability to distinguish between some 16S rRNA sequences at a genus and species level because of their near identical 16s rRNA sequences.

Additional research questions can be addressed with this dataset for more in-depth exploration of future directions. In terms of the unclassified OTUs/ASVs, future research can be done to determine how significant the unclassified taxonomies are in impacting the geochemical gradients in Saanich Inlet in order to better understand their roles in the nutrient cycles. Another alternative question that can be addressed would be whether there are any significant differences with the classification of the other genera (including the ones that were unclassified) by both Mothur and QIIME2.