

Module 3: Project 1 by Team 5

mothur vs. QIIME2 Microbiome Data Analysis

Karl Abuan

May Ho

Jonah Lin

Leilynaz Malekafzali

Tiffany Yang

Abdur Rahman M. A. Basher

March 22, 2018

Abstract

This is the abstract. It consists of two paragraphs.

Contents

1	Introduction	2
2	Problem Formulation	2
3	Materials and Experimental Configuration	2
3.1	Experimental Protocols	2
3.2	Dataset	2
3.3	Data Preporocessing	2
4	Results	2
4.1	Analysis of microbial community structure along with depth and oxygen concentration	2
4.2	Analysis of abundance information of Planctomyces along with depth and/or oxygen concentration	5
4.3	Estimate richness (number of OTUs/ASVs) for Planctomyces	8
4.4	Interpretation of abundance information of OTUs/ASVs of Planctomyces along with depth and/or oxygen concentration	9
5	Discussion	11
	References	12

1 Introduction

2 Problem Formulation

3 Materials and Experimental Configuration

3.1 Experimental Protocols

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

P1. Analysis of microbial community structure along with depth and oxygen concentration.

P2. Analysis of abundance information of Planctomyces along with depth and/or oxygen concentration.

P3. Estimate richness (number of OTUs/ASVs) for Planctomyces.

P4. Interpretation of abundance information of OTUs/ASVs of Planctomyces along with depth and/or oxygen concentration.

3.2 Dataset

Talk about mothur and QIIME2

3.3 Data Preprocessing

We used the Saanich Inlet dataset that was preprocessed using mothur and QIIME2. Afterward, samples were rarefied/normalized to 100,000 sequences per sample to facilitate comparisons between samples. The rarefied counts were then converted to relative abundance percentages. Next, we perform a series of filterings according to three rules: i)- exclude OTUs that are not observed for more than 4 samples; ii)- prune samples and OTUs with unknown values, such as `unclassified` value; and iii)- any phylum fail to have more than 5 OTUs should be trimmed. This has resulted in 371 and 190 taxa from mothur and QIIME2, respectively. No other preprocessing were applied.

4 Results

4.1 Analysis of microbial community structure along with depth and oxygen concentration

To understand the compositional complexity of a microbial community across Saanich Inlet samples, we use Shannon’s diversity index (SDI), which considers both the species abundances and the total number of distinct species in its diversity estimation. Figures 1(a) and 2(a) depicts Shannon’s diversity index for mothur and QIIME2 datasets. Immediately, we observe that SDI values peak at depth 10 and 100 before monotonically decreasing at 200. The SDI values are maximal when

the microbes are evenly distributed. Indeed, Figure 3 supports our claim and we see an uneven distribution of Phylum at 200 more than at 10 or 100 depths. However, the Shannon index values do not represent the number of different species or richness varying across depths. Instead, Figure 3 should be used together with Figures 1(a) and 2(a).

Now, if we analyze the association between oxygen concentration within a microbial community, we observe, as in Figures 1(b) and 2(b), that SDI values increase when oxygen become more abundant. This is not surprising since the abundance of oxygen indicates. . . . [FILL]. The boxplots in Figures 1(c) and 2(c) supports this evidence too. Another interesting observation that follows the observations in [papers] is the oxygen depletion along the water columns, as illustrated in Figures 1(d) and 2(d).

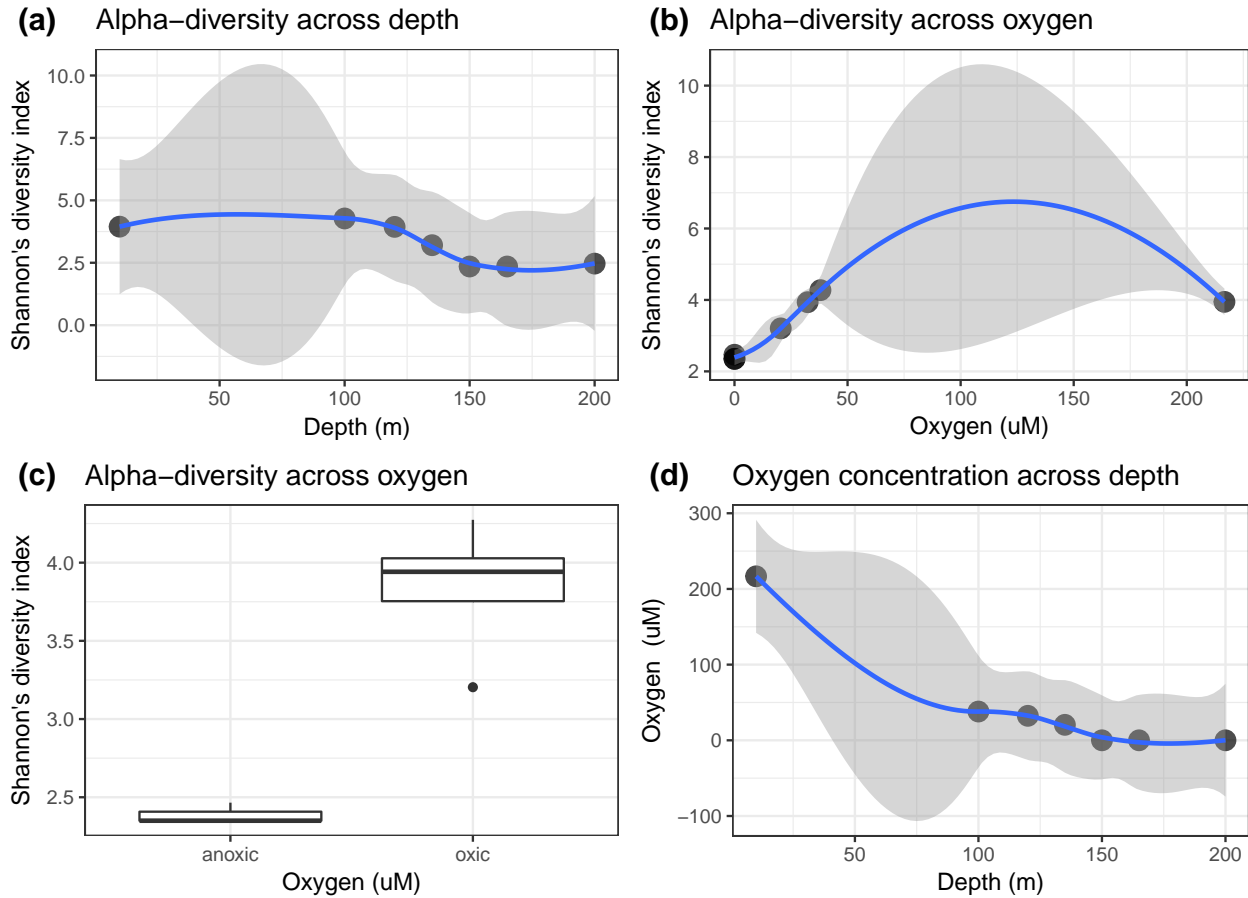
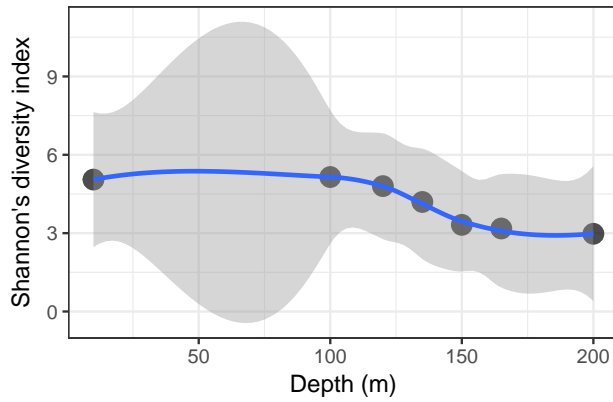
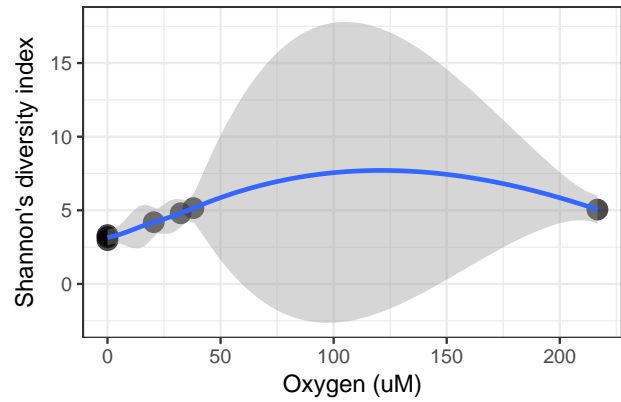


Figure 1: Mothur

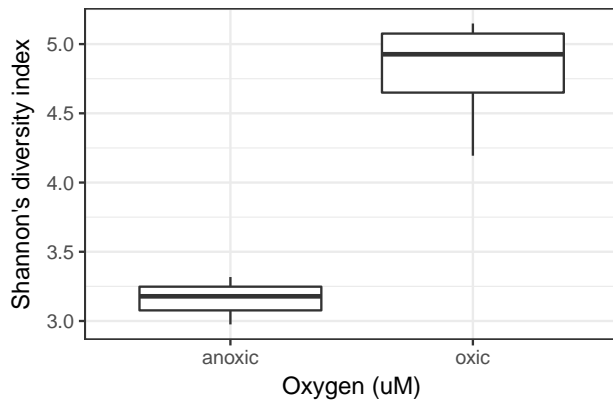
(a) Alpha-diversity across depth



(b) Alpha-diversity across oxygen



(c) Alpha-diversity across oxygen



(d) Oxygen concentration across depth

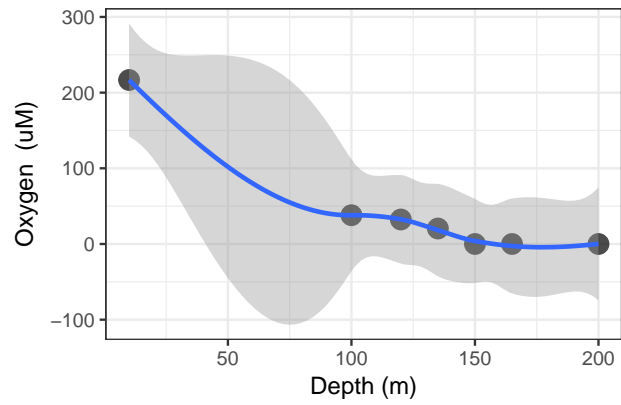


Figure 2: QIIME2

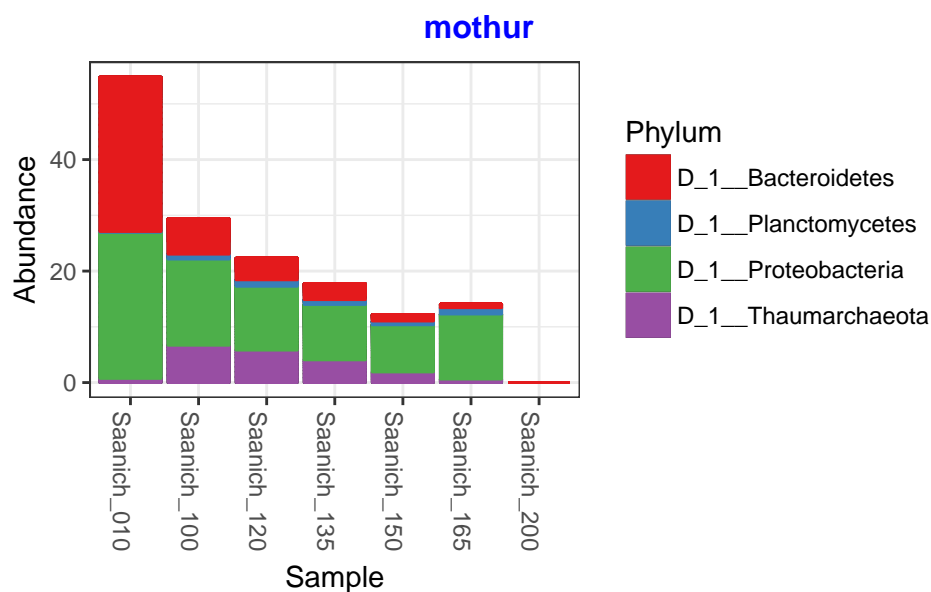
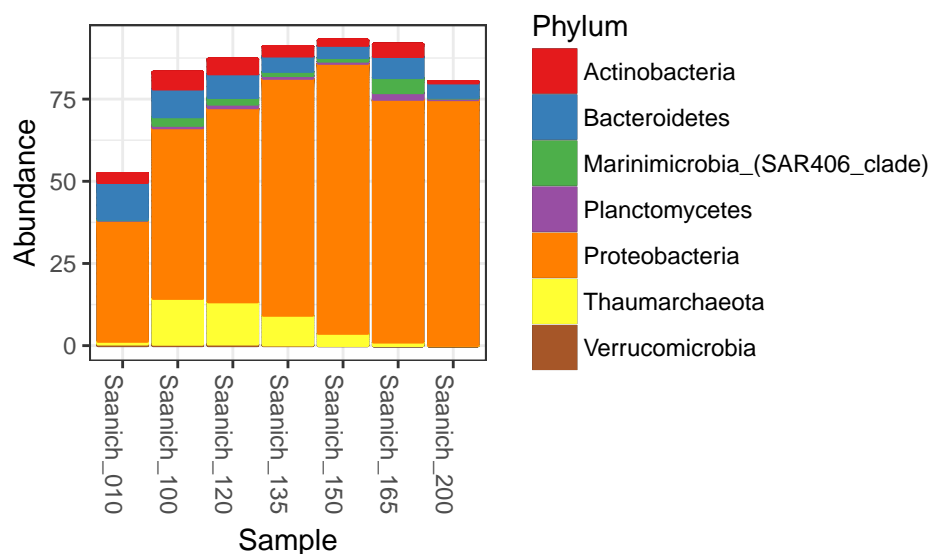


Figure 3: Phylum distribution across samples

4.2 Analysis of abundance information of Planctomyces along with depth and/or oxygen concentration

Table 1: Phylums from Mothur vs Phylums from QIIME2

Phylums from Mothur	Phylums from QIIME2
Proteobacteria	D_1__Proteobacteria
Bacteroidetes	D_1__Bacteroidetes
Thaumarchaeota	D_1__Planctomycetes
Actinobacteria	D_1__Thaumarchaeota
Marinimicrobia_(SAR406_clade)	

Phylums from Mothur	Phylums from QIIME2
Planctomycetes	
Verrucomicrobia	

Figure 3 provides an instinctive insight into the microbial species and their distributions in Saanich inlet samples based on the phylum rank. The number of hypothetical phyla present in mothur is estimated to be 7. Table 1 summarizes the phyla in both mothur and QIIME2. By in-depth investigation, it is clear that the *Planctomycetes* phylum is differentially presented across samples. And, after further exploring its genus distribution, as depicted in Figure 4, we exhibit that *Planctomyces* genus indeed are unevenly distributed across water columns. We hypothesize that *Planctomyces* would be represented differently across depths, and for this, we performed regression analysis.

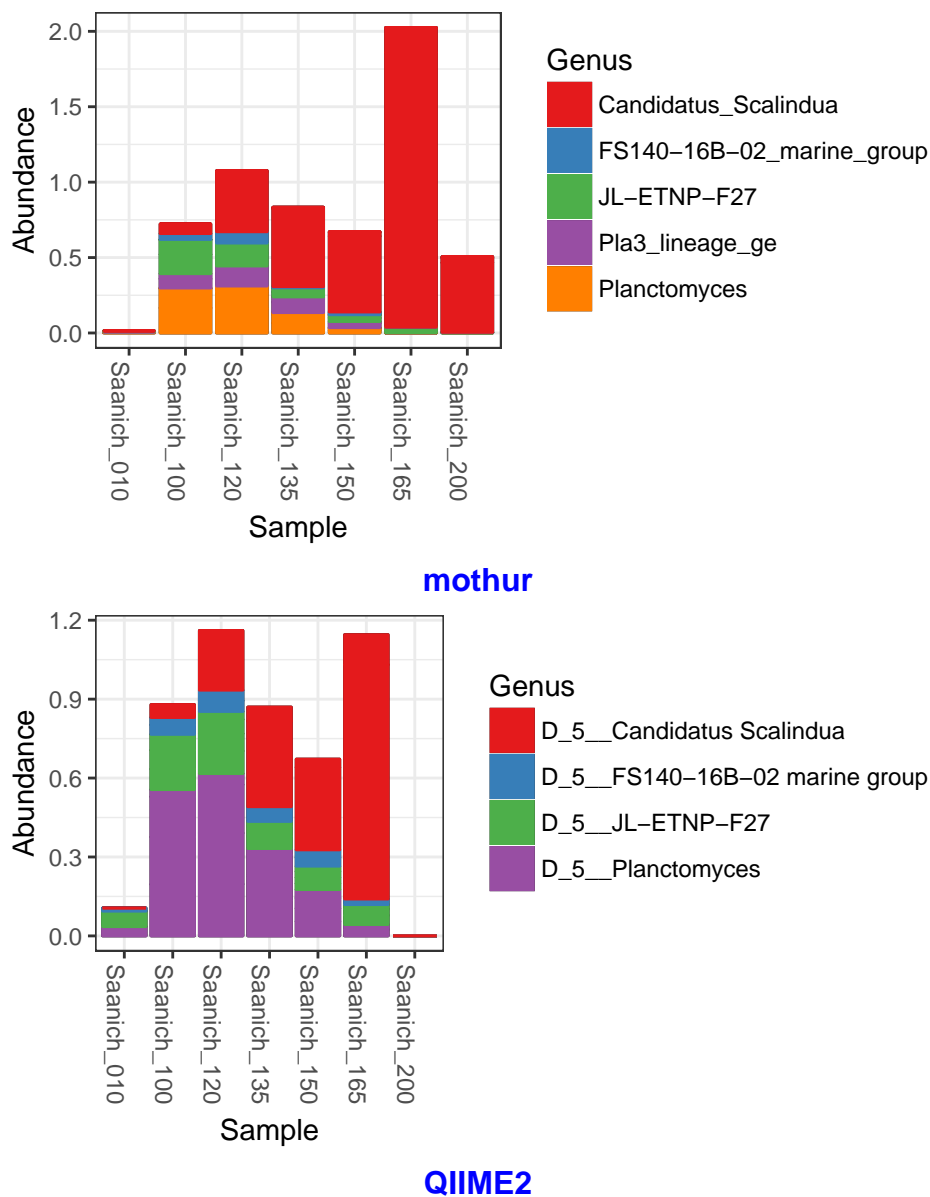


Figure 4: Genus distribution of Planctomycetes across samples

The well known general linear model is employed to recover relationships that might exist between explanatory variables and target variables. We investigate the correlation of *Planctomyces* abundance as a function of depth and as a function of oxygen concentration. The results of this analysis are represented in Table 2. As it can be seen, that no relationships exist with either the depth or the oxygen. Together the coefficients of depth and oxygen and their P-values were found to be (-0.0003609 , P-value = 0.7385598) and (-0.0002544 , P-value = 0.7616253), respectively. This is of course not statistically significant at 5% (an arbitrary cutoff) since their p-values are above > 0.05 . Hence, we might reason that there is a little statistical evidence to support our belief that *Planctomyces* indeed varies across depth and oxygen. This is due to the fitting problem associated with the general linear model that uses simple ordinary least squares to fit the data. The samples from depth 10, 150, 165 and 200 do not or are less *Planctomyces* abundant than at 100, 120, and 135 depths, which obviously imply an alternative proposal that *Planctomyces* is differentially abundant in Saanich Inlet dataset. But for now on, we stick with the statistical outputs.

Table 2: Correlation data of OTUs within *Planctomyces* genus across depth and oxygen concentration from mothur and QIIME2

Covariates	Estimate	Std. Error	t-value	Pr(> t)
Depth (mothur)	-0.0003609	0.0010227	-0.3528908	0.7385598
O2_uM (mothur)	-0.0002544	0.0007941	-0.3203956	0.7616253
Depth (QIIME2)	-0.0005878	0.0018774	-0.3130933	0.7668485
O2_uM (QIIME2)	-0.0005997	0.0014441	-0.4152436	0.6951812

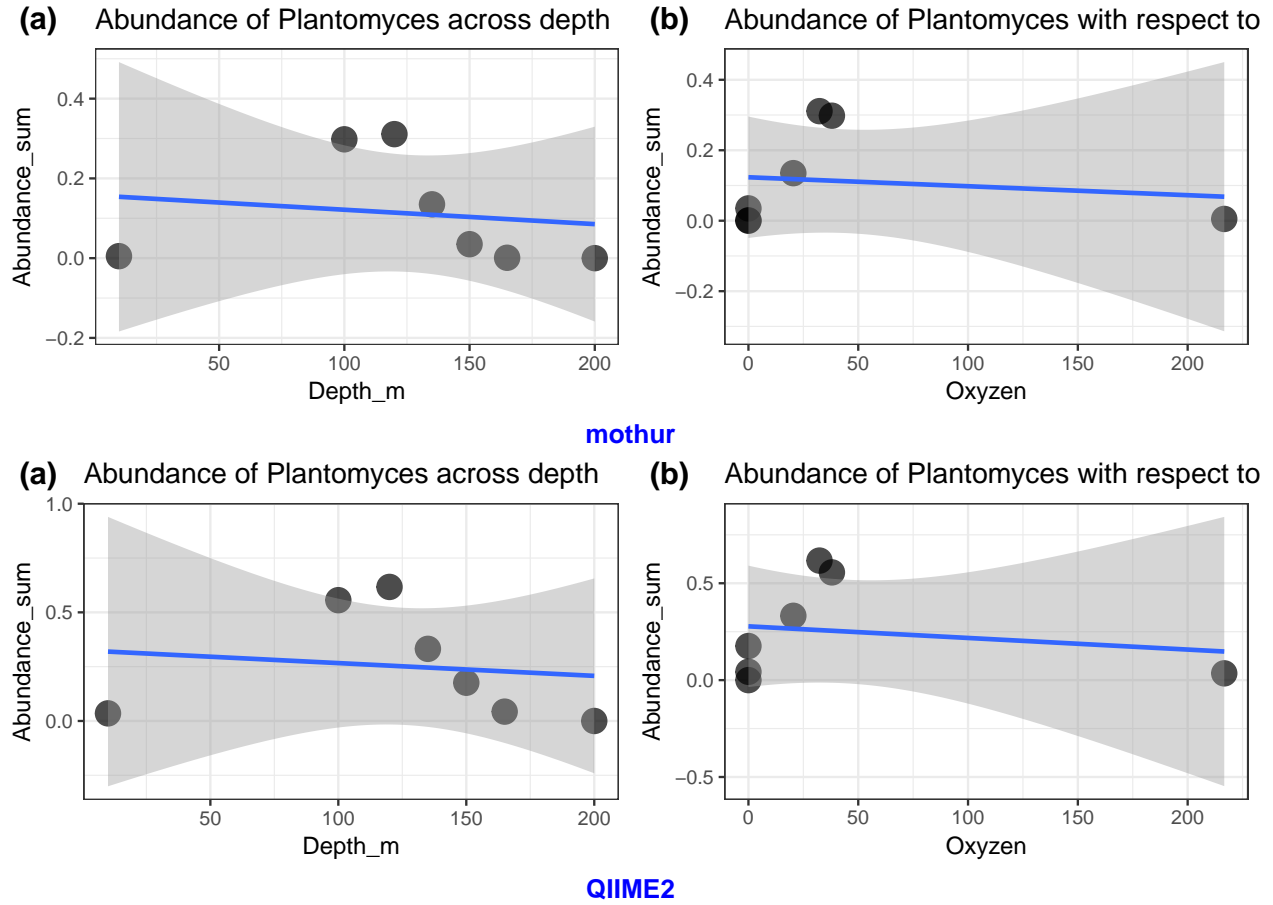


Figure 5: Regression analysis of Planctomyces across depth

4.3 Estimate richness (number of OTUs/ASVs) for Planctomyces

Table 3: OTUs from Mothur vs ASVs from QIIME2

OTUs from Mothur	ASVs from QIIME2
Otu0125	Asv232
Otu0144	Asv799
Otu0401	Asv1021
Otu0592	Asv1124

We explore the diversity of *Planctomyces* across depth.

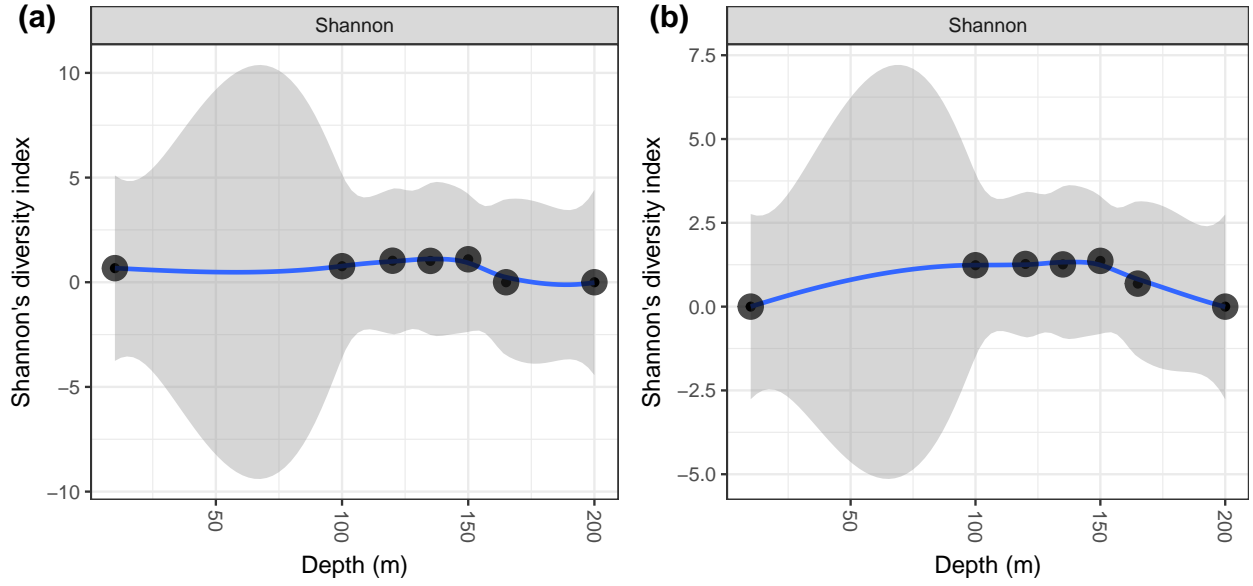


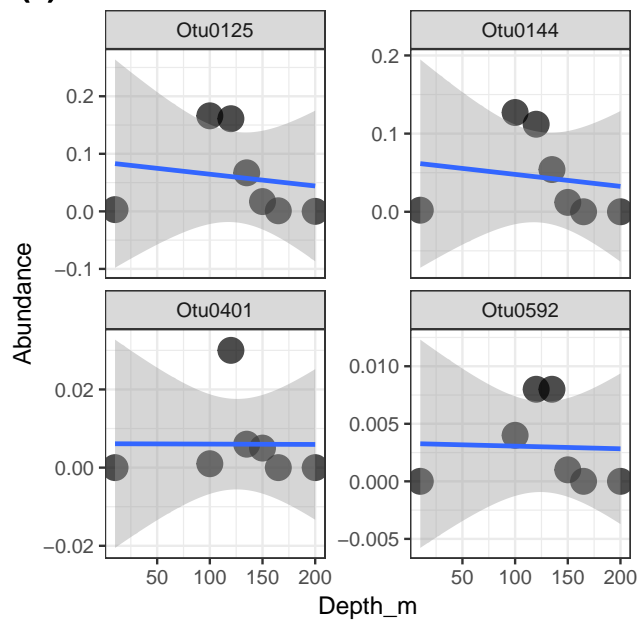
Figure 6: Alpha-diversity of *Planctomyces* across depth from (a) mothur and (b) QIIME2

4.4 Interpretation of abundance information of OTUs/ASVs of *Planctomyces* along with depth and/or oxygen concentration

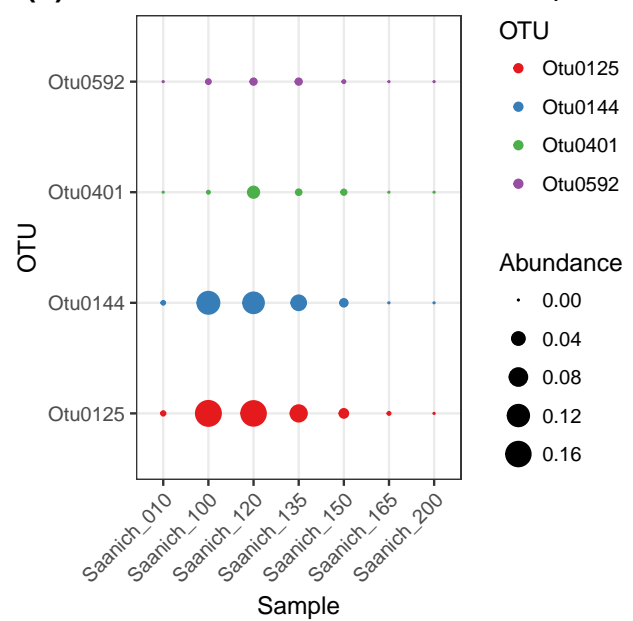
Table 4: Correlation data of OTUs within *Planctomyces* genus across depth

Covariates	Estimate	Std. Error	t-value	Pr(> t)
Otu0125 (mothur)	-0.0002045	0.0005479	-0.3731784	0.7243139
Otu0144 (mothur)	-0.0001533	0.0004035	-0.3798581	0.7196506
Otu0401 (mothur)	-0.0000009	0.0000807	-0.0113619	0.9913741
Otu0592 (mothur)	-0.0000023	0.0000274	-0.0836392	0.9365887
Asv232 (QIIME2)	-0.0001665	0.0006541	-0.2545201	0.8092302
Asv799 (QIIME2)	-0.0000953	0.0005682	-0.1676505	0.8734282
Asv1021 (QIIME2)	0.0000544	0.0001222	0.4454921	0.6745908
Asv1124 (QIIME2)	-0.0003805	0.0005859	-0.6493922	0.5447317

(a) Per OTU abundance information

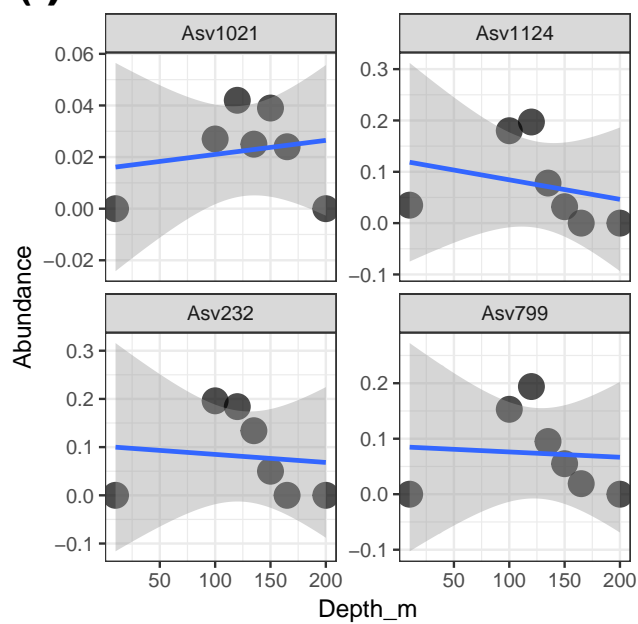


(b) Abundance of OTUs across samples

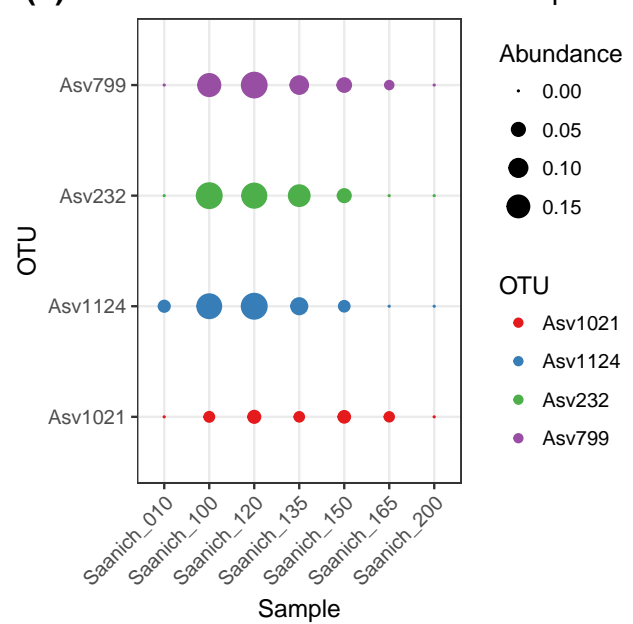


mothur

(a) Per ASV abundance information



(b) Abundance of ASVs across samples



QIIME2

Figure 7: Abundance of OTUs/ASVs within *Planctomyces* genus across depth

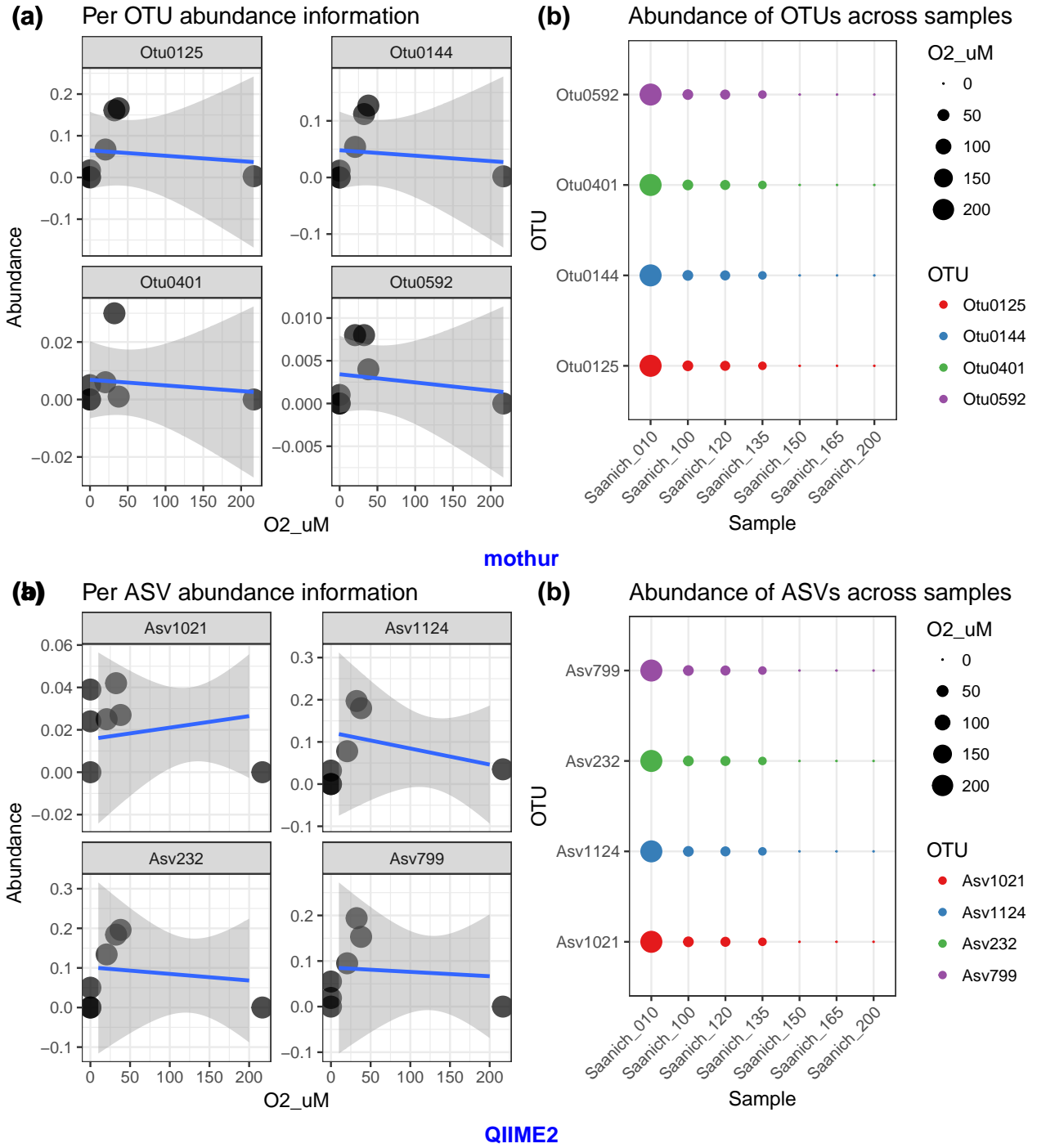


Figure 8: Abundance of OTUs/ASVs within *Planctomyces* genus across oxygen concentration

5 Discussion

(Hawley et al. 2017; Torres-Beltrán et al. 2017)

References

Hawley, Alyse K, Mónica Torres-Beltrán, Elena Zaikova, David A Walsh, Andreas Mueller, Melanie Scofield, Sam Kheirandish, et al. 2017. “A Compendium of Multi-Omic Sequence Information from the Saanich Inlet Water Column.” *Scientific Data* 4. Nature Publishing Group: 170160.

Torres-Beltrán, Mónica, Alyse K Hawley, David Capelle, Elena Zaikova, David A Walsh, Andreas Mueller, Melanie Scofield, et al. 2017. “A Compendium of Geochemical Information from the Saanich Inlet Water Column.” *Scientific Data* 4. Nature Publishing Group: 170159.