QIIME2

Module 3: Project 1 by Team 5

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

March 14, 2018

Abstract

This is the abstract. It consists of two paragraphs.

Contents

1 Introduction

2 Problem Formulation

3 Materials and Experimental Configuration

3.1 Experimental Protocols

Here . . . .:

P1. Analysis of microbial community structure along with depth and oxygen concentration (see Section 4.1).

P2. Analysis of abundance information of [OTU****] along with depth and/or oxygen concentration (see Section 4.2).

P3. Estimate richness (number of OTUs/ASVs) for [OTU****] (see Section 4.3).

P4. Interpretation of abundance information of OTUs/ASVs of [OTU****] along with depth and/or oxygen concentration (see Section 4.4).

3.2 Dataset

3.3 Parameters Configuration

```
## try http:// if https:// URLs are not supported
source("https://bioconductor.org/biocLite.R")
biocLite("phyloseq")
library("tidyverse")
library("gridExtra")
library("magrittr")
library("formatR")
```

3.4 Data Preporocessing

We use saanich inlet datasets that are propocssed using mothur and QIIME2

```
load("data/mothur_phyloseq.RData")
load("data/qiime2_phyloseq.RData")
```

Samples are then rarefied/normalized to $100,000$ sequences per sample to facilitate comparisons between samples. A random seed was set to ensure reproducibility.

```
set.seed(4832)
rarefied = rarefy_even_depth(qiime2, sample.size = 1e+05)
```

Rarefied counts were converted to relative abundance percentages.

```r
rarefiedPer = transform_sample_counts(rarefied, function(x) 100 * x/sum(x))
```

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. The codes used for applying the three rules are:

```r
# First rule
firstTaxa = filter_taxa(rarefiedPer, function(x) sum(x == 0) <= 4, TRUE)

# Second rule
basedOnGenus <- as.data.frame(tax_table(firstTaxa)) %>% filter(!str_detect(Genus,
    "uncultured"), !str_detect(Genus, "unclassified"))
secondTaxa = subset_taxa(firstTaxa, Genus %in% basedOnGenus$Genus)

# Third rule
basedOnphylums <- as.data.frame(tax_table(secondTaxa)) %>% group_by(Phylum) %>%
    count() %>% filter(n > 5)
## In contrary we can run the following: thirdTaxa <-
## prune_taxa(taxa_sums(secondTaxa) > 5, secondTaxa)
thirdTaxa <- subset_taxa(secondTaxa, Phylum %in% basedOnphylums$Phylum)
```

4   Results

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

We first estimate the overall taxa diversity using Shannon's diversity index.

```r
rarefiedRich <- estimate_richness(rarefied, measures = "Shannon")
rarefiedRichAlpha <- full_join(rownames_to_column(rarefiedRich), rownames_to_column(data.frame
    by = "rowname")

p1 <- rarefiedRichAlpha %>% ggplot() + geom_point(aes(x = Depth_m, y = Shannon),
    size = 4, alpha = 0.7) + geom_smooth(method = "loess", aes(x = as.numeric(Depth_m),
    y = Shannon)) + labs(title = "Alpha-diversity across depth", y = "Shannon's diversity inde:
    x = "Depth (m)")

p2 <- rarefiedRichAlpha %>% ggplot() + geom_point(aes(x = O2_uM, y = Shannon),
    size = 4, alpha = 0.7) + labs(title = "Alpha-diversity across oxygen", y = "Shannon's diver
    x = "Oxygen (uM)")

p3 <- rarefiedRichAlpha %>% mutate(O2_group = ifelse(O2_uM == 0, "anoxic", "oxic")) %>%
    ggplot() + geom_boxplot(aes(x = O2_group, y = Shannon)) + labs(title = "Alpha-diversity ac
    y = "Shannon's diversity index", x = "Oxygen (uM)")

p4 <- rarefiedRichAlpha %>% ggplot() + geom_point(aes(x = Depth_m, y = O2_uM),
```
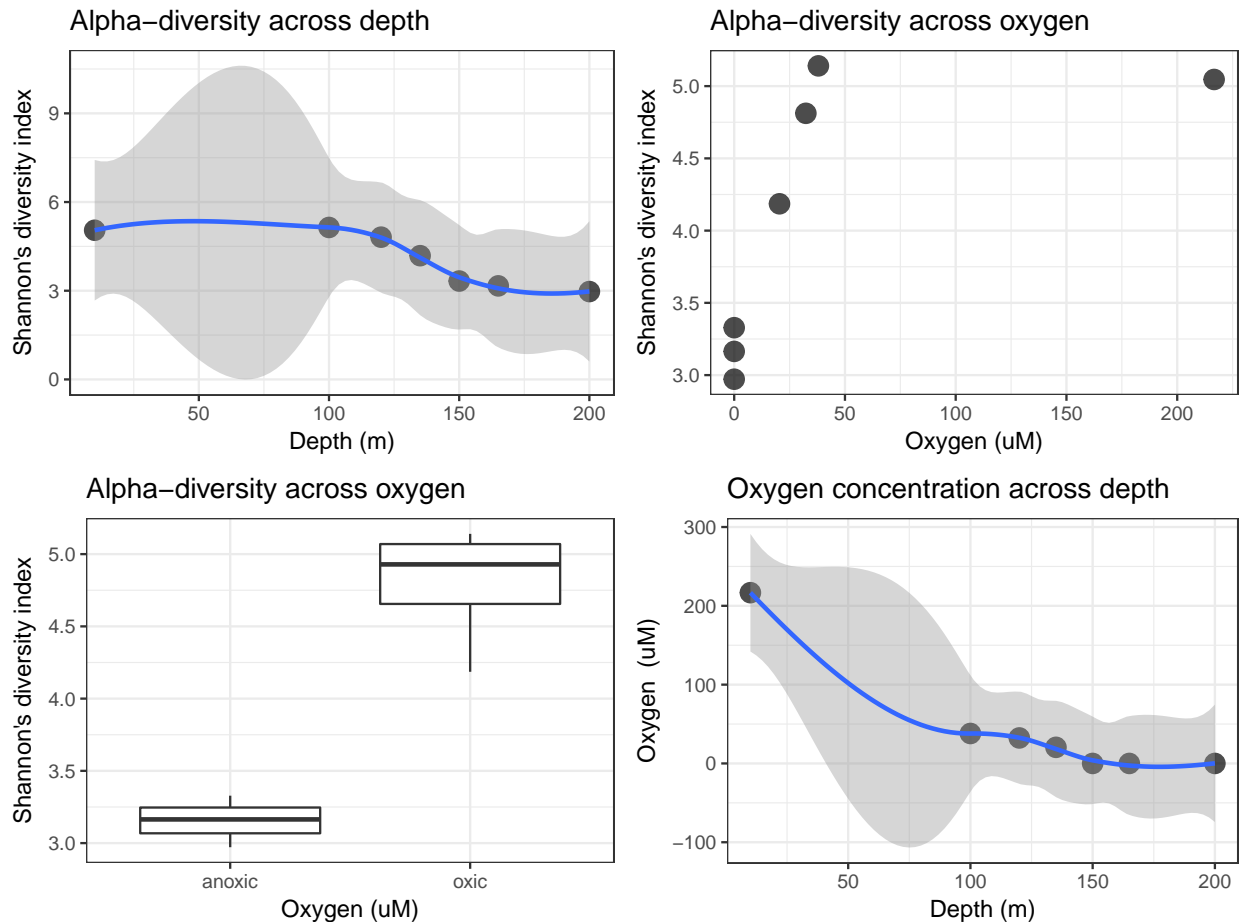
```
        size = 4, alpha = 0.7) + geom_smooth(method = "loess", aes(x = as.numeric(Depth_m),
        y = O2_uM)) + labs(title = "Oxygen concentration across depth", y = "Oxygen  (uM)",
        x = "Depth (m)")
```

```
grid.arrange(p1, p2, p3, p4, ncol = 2)
```


The following function will assist us to understand the unique Phylum rank:

```
get_taxa_unique(physeq = thirdTaxa, taxonomic.rank = "Phylum")
```

```
## [1] "D_1__Proteobacteria"  "D_1__Bacteroidetes"   "D_1__Planctomycetes"
## [4] "D_1__Thaumarchaeota"  "D_1__Verrucomicrobia" "D_1__Cyanobacteria"
```

We choose the *Planctomycetes* phylum, and explored the distribution of genera of this phylum.

```
subTaxa = subset_taxa(thirdTaxa, Phylum == "D_1__Planctomycetes")
plot_bar(subTaxa, fill = "Genus") + geom_bar(aes(color = Genus, fill = Genus),
    stat = "identity", position = "stack") + labs(title = "Genus distribution of Planctomycetes
```

Genus distribution of Planctomycetes across samples
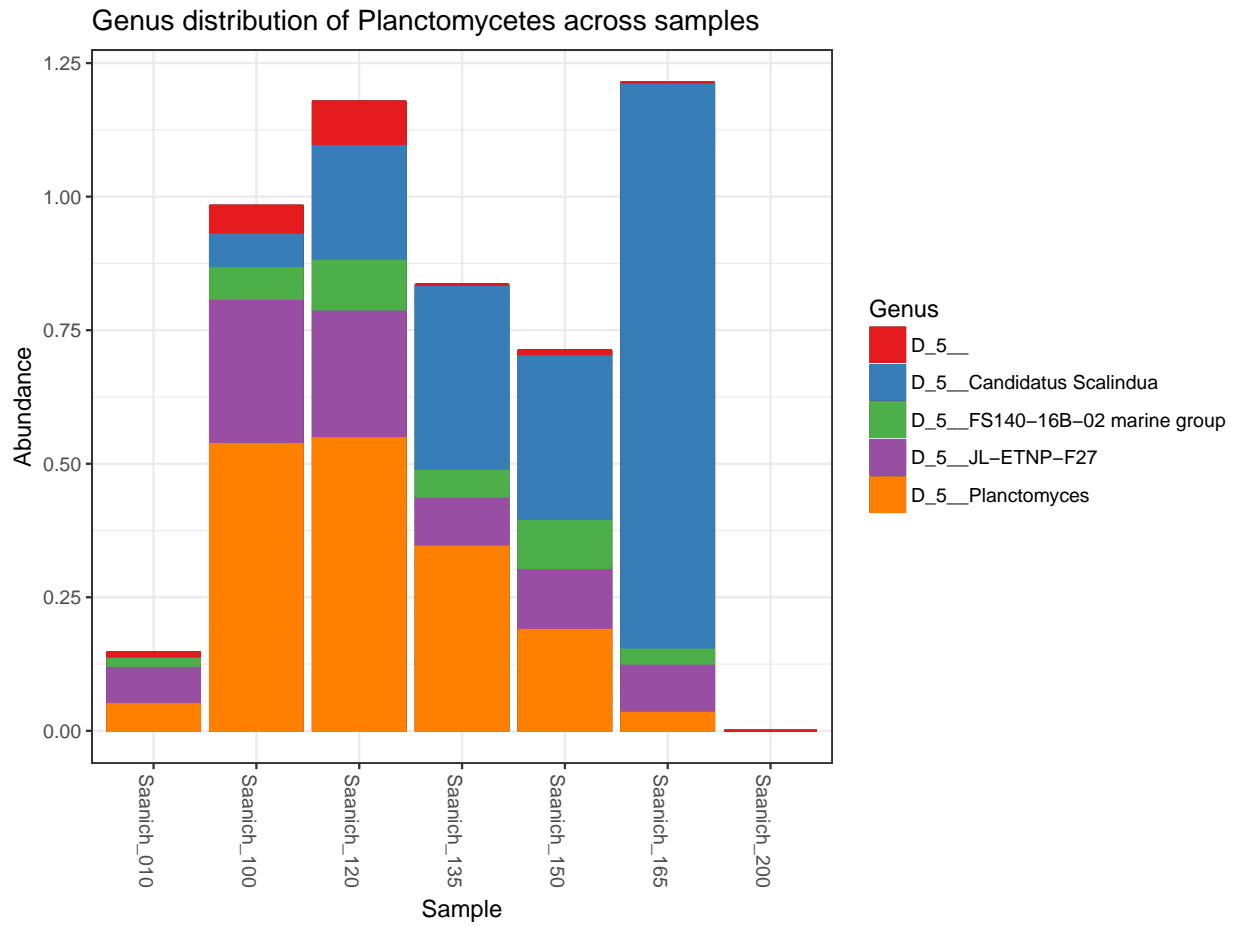
We further investigated the genus distribution of this phylum across samples grouped by family level.

```
plot_bar(subTaxa, fill = "Genus", facet_grid = ~Family) + geom_bar(aes(color = Genus,
    fill = Genus), stat = "identity", position = "stack") + labs(title = "Genus distribution o
```

Genus distribution of Planctomycetes across samples grouped by family level
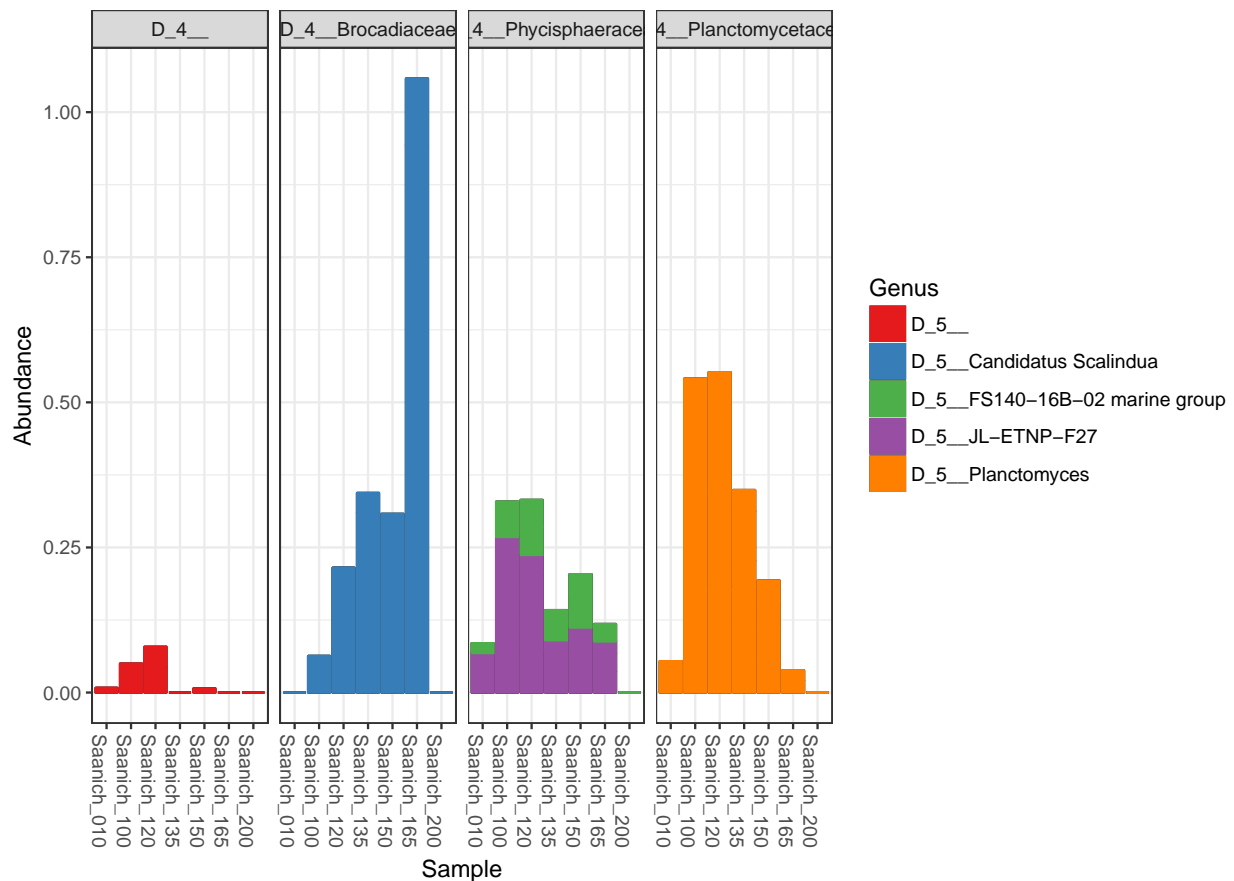
Finally, we settled on performing experimental analysis at *Planctomyces* genus level and it's associated OTUs.

```
workingTaxa = subset_taxa(thirdTaxa, Genus == "D_5__Planctomyces")
(suggestedOTUs <- colnames(otu_table(workingTaxa)))
```

```
## [1] "Saanich_010" "Saanich_100" "Saanich_120" "Saanich_135" "Saanich_150"
## [6] "Saanich_165" "Saanich_200"
```

4.2 Analysis of abundance information of [OTU****] along with depth and/or oxygen concentration

```
workingTaxa %>% tax_glom(taxrank = "Genus") %>% psmelt() %>% lm(Abundance ~
    Depth_m, .) %>% summary()
```

```
##
## Call:
## lm(formula = Abundance ~ Depth_m, data = .)
##
## Residuals:
##           3         2         4         5         1         6         7
```

6

```
##  0.30165   0.27791   0.10820 -0.03825 -0.26639 -0.18370 -0.19942
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.3267558  0.2377250   1.375    0.228
## Depth_m     -0.0006367  0.0017282  -0.368    0.728
##
## Residual standard error: 0.2553 on 5 degrees of freedom
## Multiple R-squared:  0.02643,    Adjusted R-squared:  -0.1683
## F-statistic: 0.1357 on 1 and 5 DF,  p-value: 0.7277
```

```
workingTaxa %>% psmelt() %>% group_by(Sample) %>% summarize(Abundance_sum = sum(Abundance),
    Depth_m = mean(Depth_m)) %>% ggplot() + geom_point(aes(x = Depth_m, y = Abundance_sum),
    size = 5, alpha = 0.7) + geom_smooth(method = "lm", aes(x = as.numeric(Depth_m),
    y = Abundance_sum)) + labs(title = "Abundance of Plantomyces across depth")
```


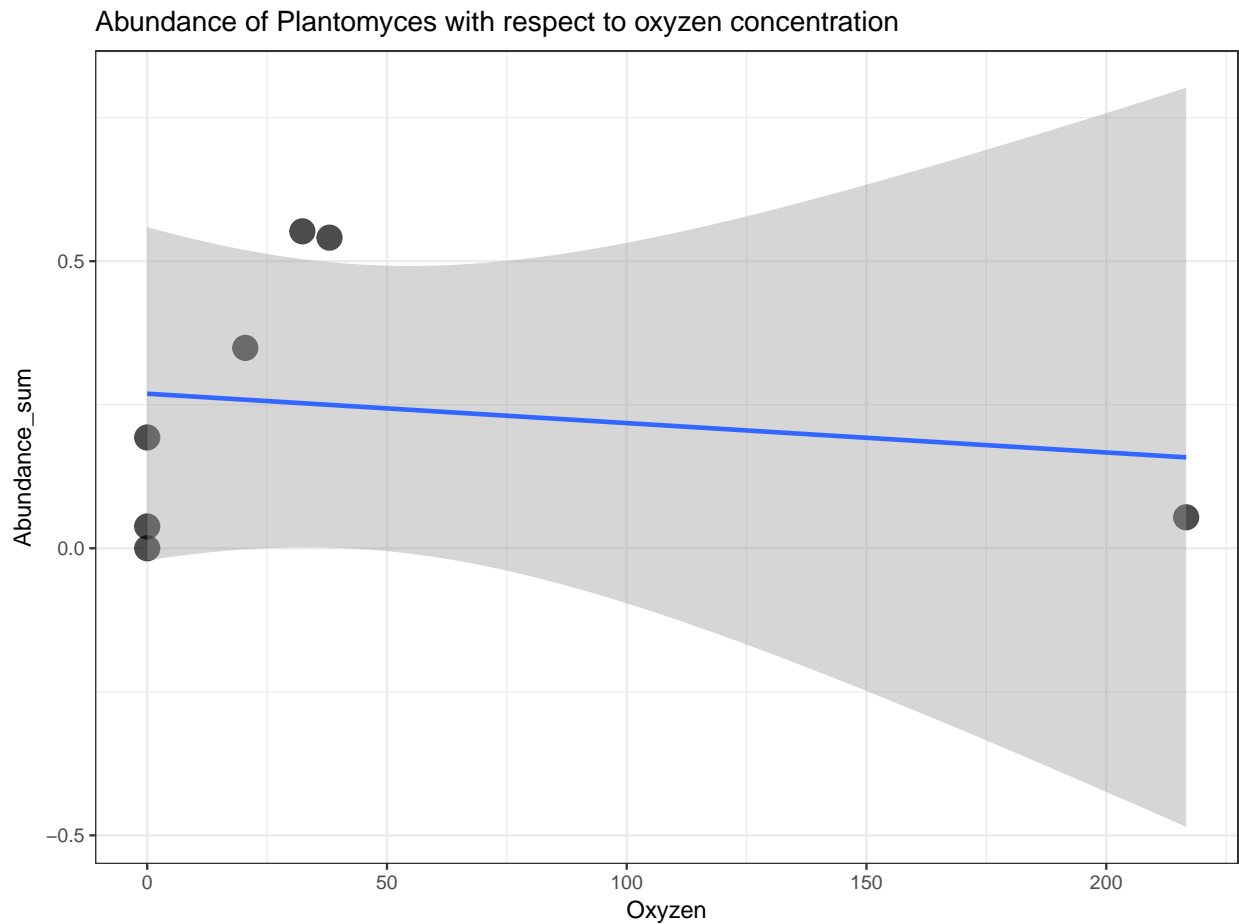Abundance of Plantomyces across depth
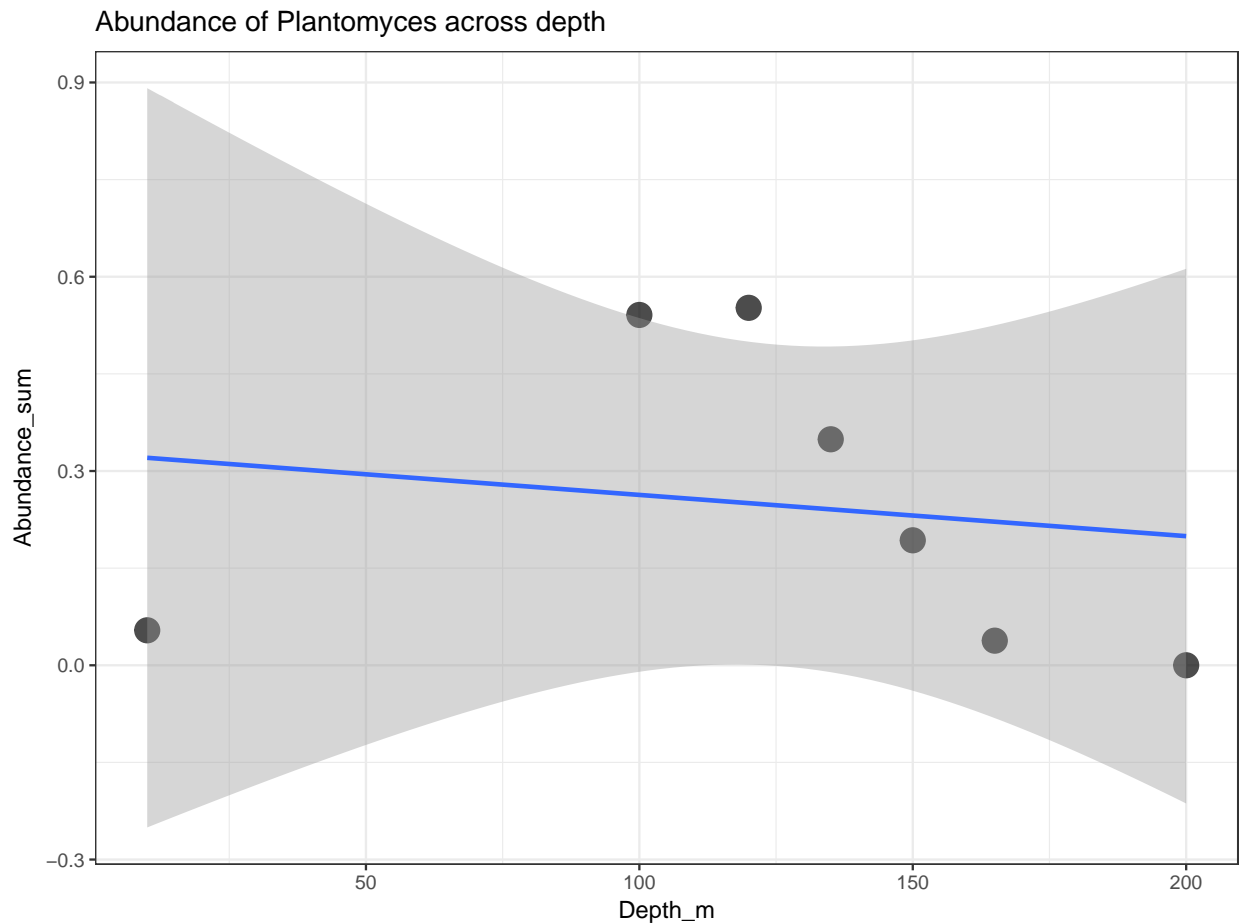
```
workingTaxa %>% tax_glom(taxrank = "Genus") %>% psmelt() %>% lm(Abundance ~
    O2_uM, .) %>% summary()
```

```
##
## Call:
## lm(formula = Abundance ~ O2_uM, data = .)
```

```
## 
## Residuals:
##        3         2         4         5         1         6         7
##   0.29936   0.29126   0.09027  -0.07619  -0.10432  -0.23119  -0.26919
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.2691917  0.1128893    2.385   0.0628 .
## O2_uM       -0.0005117  0.0013377   -0.383   0.7178
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.255 on 5 degrees of freedom
## Multiple R-squared:  0.02844,    Adjusted R-squared:  -0.1659
## F-statistic: 0.1463 on 1 and 5 DF,  p-value: 0.7178
```

```r
workingTaxa %>% psmelt() %>% group_by(Sample) %>% summarize(Abundance_sum = sum(Abundance),
    Oxyzen = mean(O2_uM)) %>% ggplot() + geom_point(aes(x = Oxyzen, y = Abundance_sum),
    size = 5, alpha = 0.7) + geom_smooth(method = "lm", aes(x = as.numeric(Oxyzen),
    y = Abundance_sum)) + labs(title = "Abundance of Plantomyces with respect to oxyzen concent
```


Abundance of Plantomyces with respect to oxyzen concentration

4.3 Estimate richness (number of OTUs/ASVs) for [OTU****]

We explore the diversity of *Planctomyces* across depth.

```
workingTaxa %>% plot_richness(x = "Depth_m", measures = "Shannon") + geom_smooth(method = "loes
    aes(x = as.numeric(Depth_m))) + geom_point(size = 5, alpha = 0.7) + labs(title = "Alpha-div
    y = "Shannon's diversity index", x = "Depth (m)")
```

Alpha−diversity of Plantomyces across depth
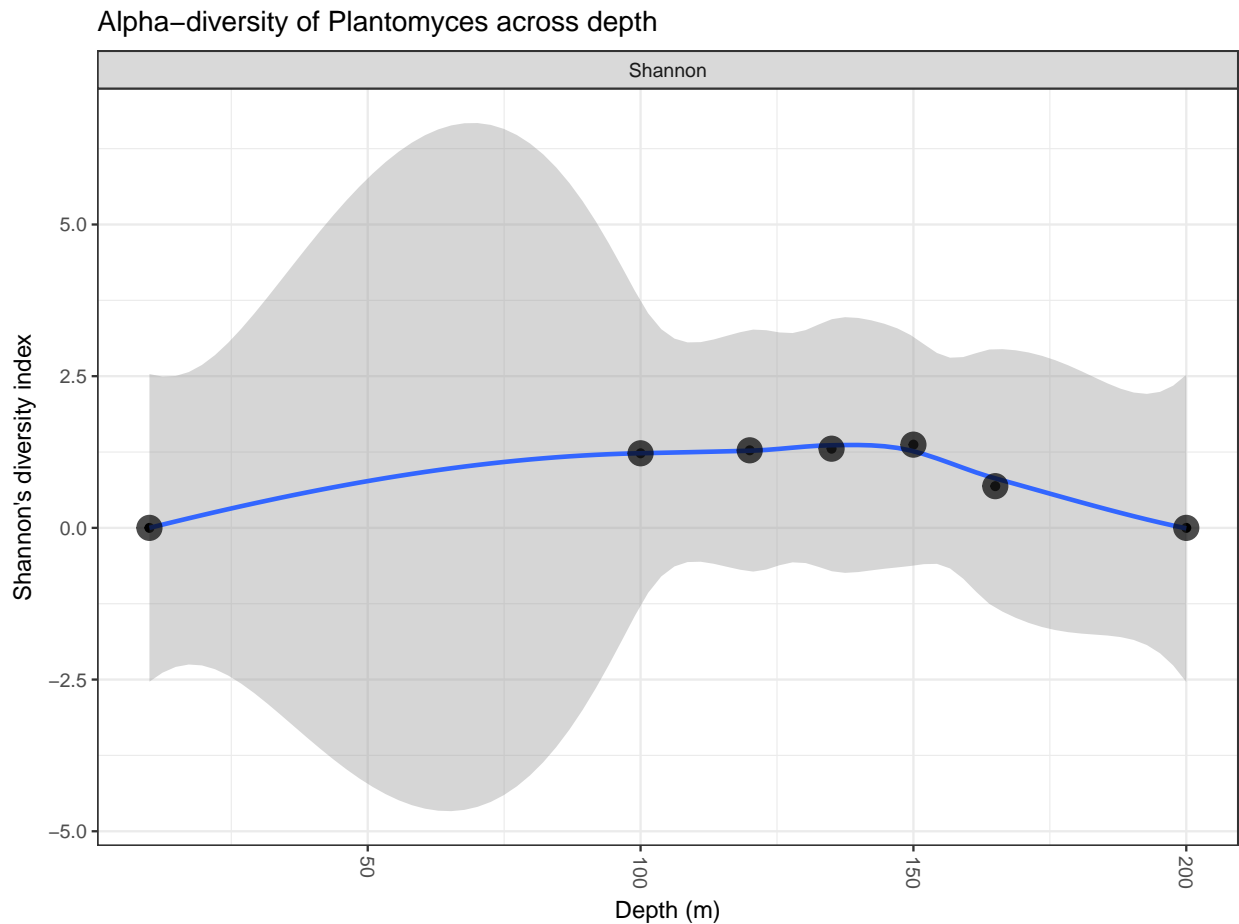
4.4 Interpretation of abundance information of OTUs/ASVs of [OTU****] along with depth and/or oxygen concentration

```
# Generalized linear model for each OTU for (otu in suggestedOTUs) {
# cat('### Generalized linear model for ', otu) workingTaxa %>% psmelt() %>%
# filter(OTU==otu) %>% lm(Abundance ~ Depth_m, .) %>% summary() %>% print()
# }

p.adjust(runif(length(suggestedOTUs), min = 0.005, max = 0.85), method = "fdr")
```

```
## [1] 0.5484101 0.5484101 0.5484101 0.5484101 0.2322418 0.5484101 0.5484101
```

```
# Abundance of OTUs within unclassified domain across depth
p5 <- workingTaxa %>% psmelt() %>% ggplot() + geom_point(aes(x = Depth_m, y = Abundance),
    size = 5, alpha = 0.7) + geom_smooth(method = "lm", aes(x = Depth_m, y = Abundance)) +
    facet_wrap(~OTU, scales = "free_y") + labs(title = "Abundance of OTUs within Planctomyces g

# Abundance of OTUs within unclassified depth by colour
p6 <- workingTaxa %>% psmelt() %>% ggplot() + geom_point(aes(x = Sample, y = OTU,
    size = Abundance, color = OTU)) + scale_size_continuous(range = c(0, 5)) +
    labs(title = "Abundance of OTUs within Planctomyces genus across depth")

grid.arrange(p5, p6, ncol = 2)
```
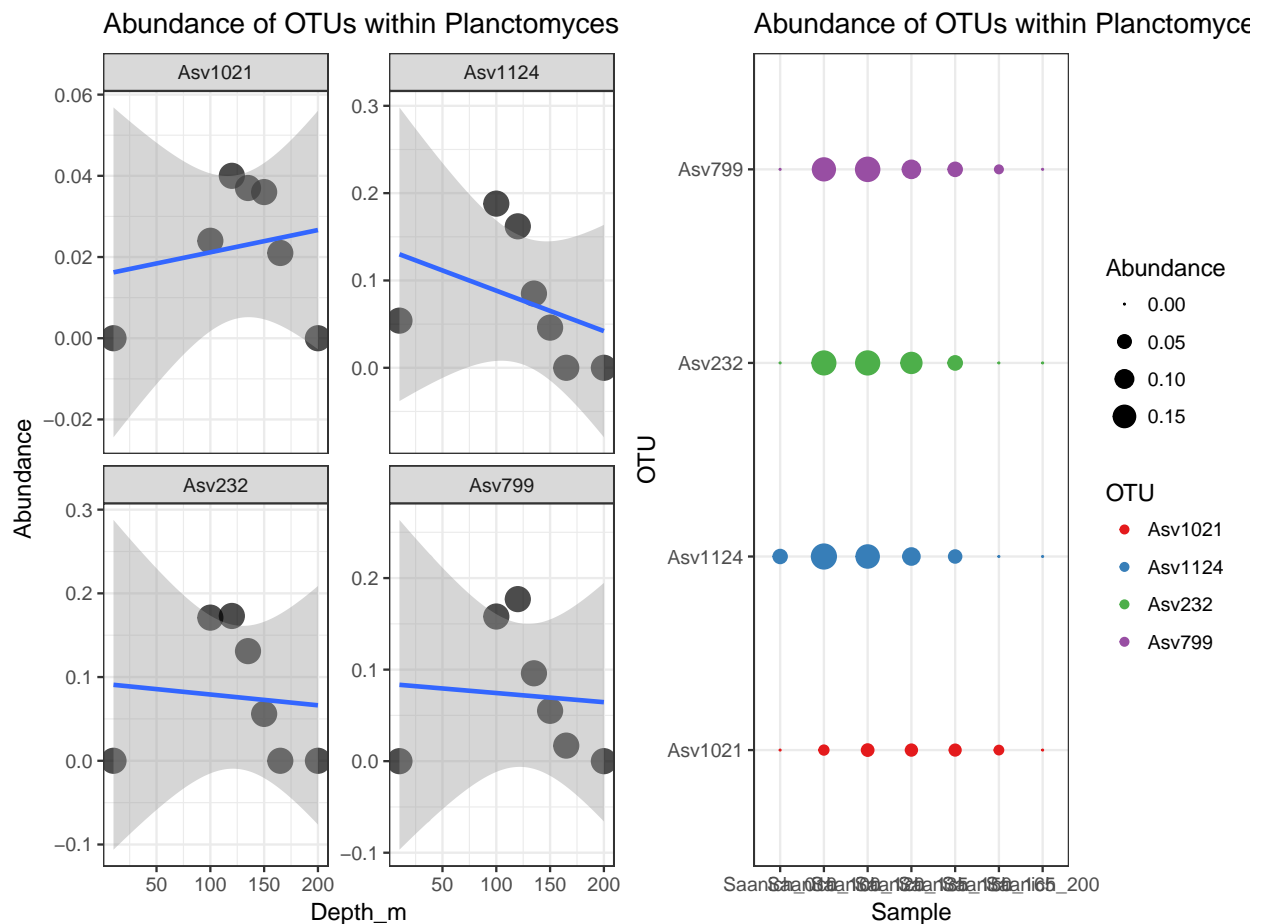


```
# Abundance of OTUs within Planctomyces genus across depth
p7 <- workingTaxa %>% psmelt() %>% ggplot() + geom_point(aes(x = O2_uM, y = Abundance),
    size = 5, alpha = 0.7) + geom_smooth(method = "lm", aes(x = O2_uM, y = Abundance)) +
    facet_wrap(~OTU, scales = "free_y") + labs(title = "Abundance of OTUs within Planctomyces g

# Abundance of OTUs within Planctomyces genus by colour
p8 <- workingTaxa %>% psmelt() %>% ggplot() + geom_point(aes(x = Sample, y = OTU,
    size = O2_uM, color = OTU)) + scale_size_continuous(range = c(0, 5)) + labs(title = "Abunda
```

```
grid.arrange(p7, p8, ncol = 2)
```
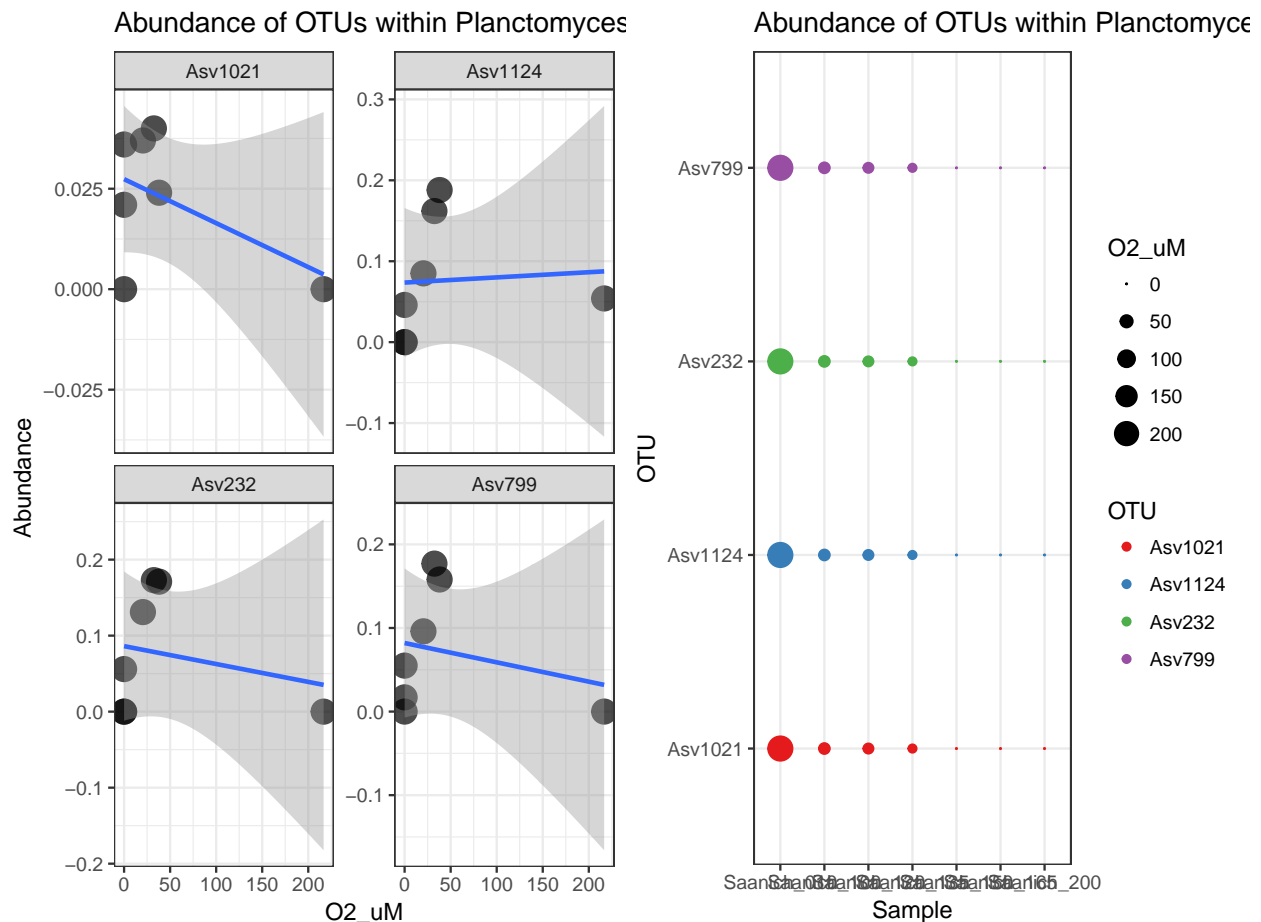

5 Discussion

(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.