Mothur

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")
```

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(mothur, sample.size = 1e+05)
```

Rarefied counts were converted to relative abundance percentages.

```
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:

```
# 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)
```

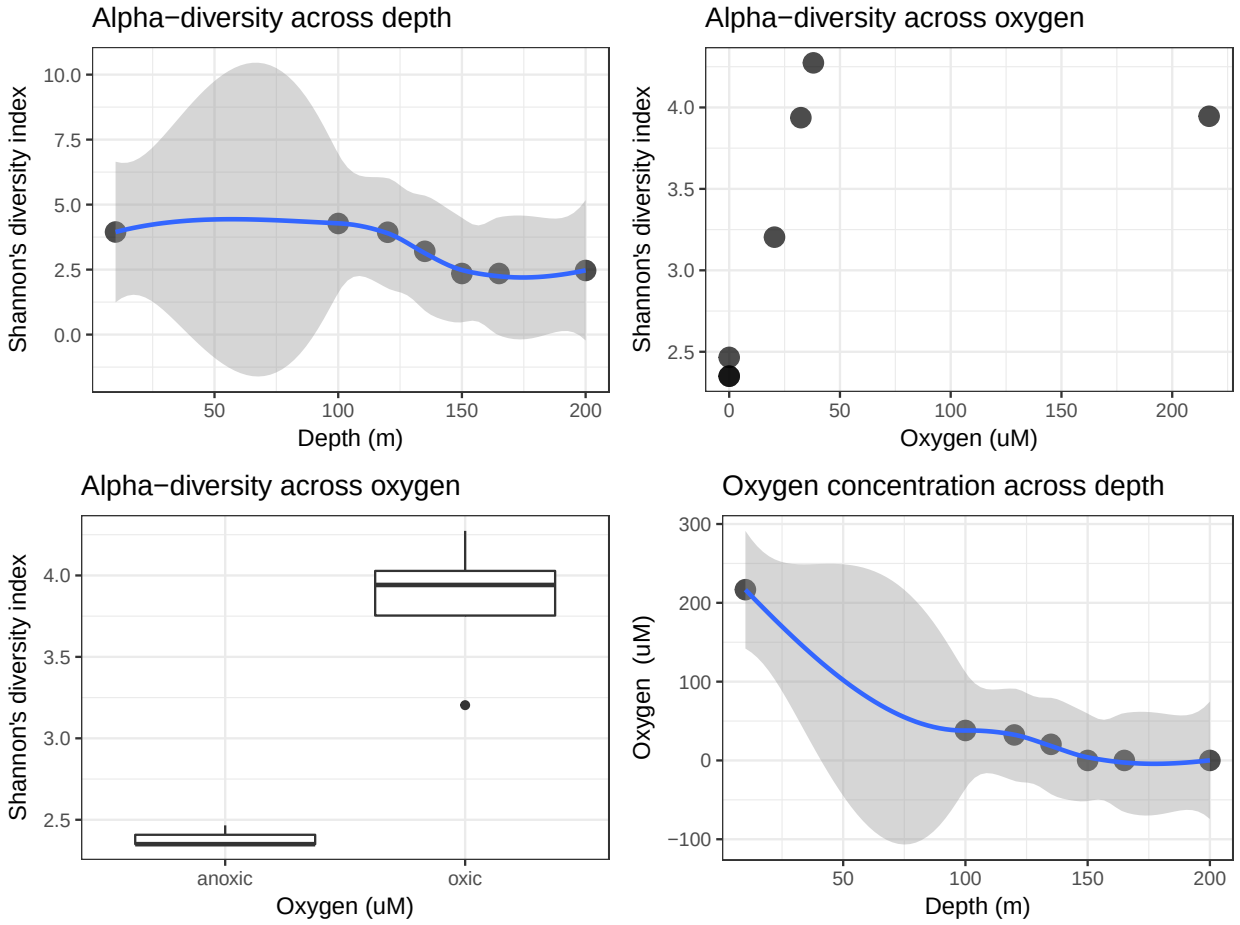
4 Results

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.

```
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 dive
    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)
```



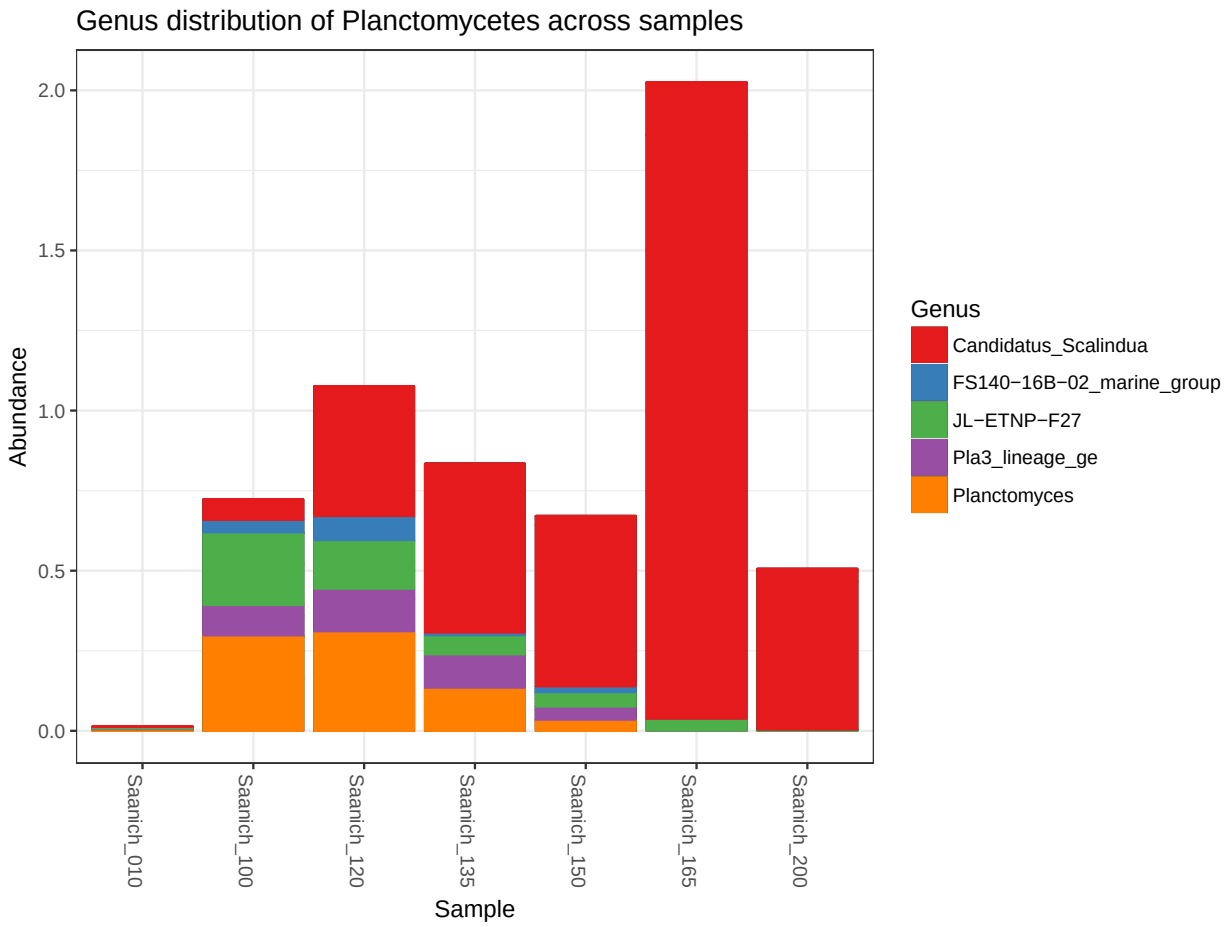
The following function will assist us to understand the unique Phylum rank:

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

```
## [1] "Proteobacteria"                "Bacteroidetes"
## [3] "Thaumarchaeota"                "Actinobacteria"
## [5] "Marinimicrobia_(SAR406_clade)" "Planctomycetes"
## [7] "Verrucomicrobia"
```

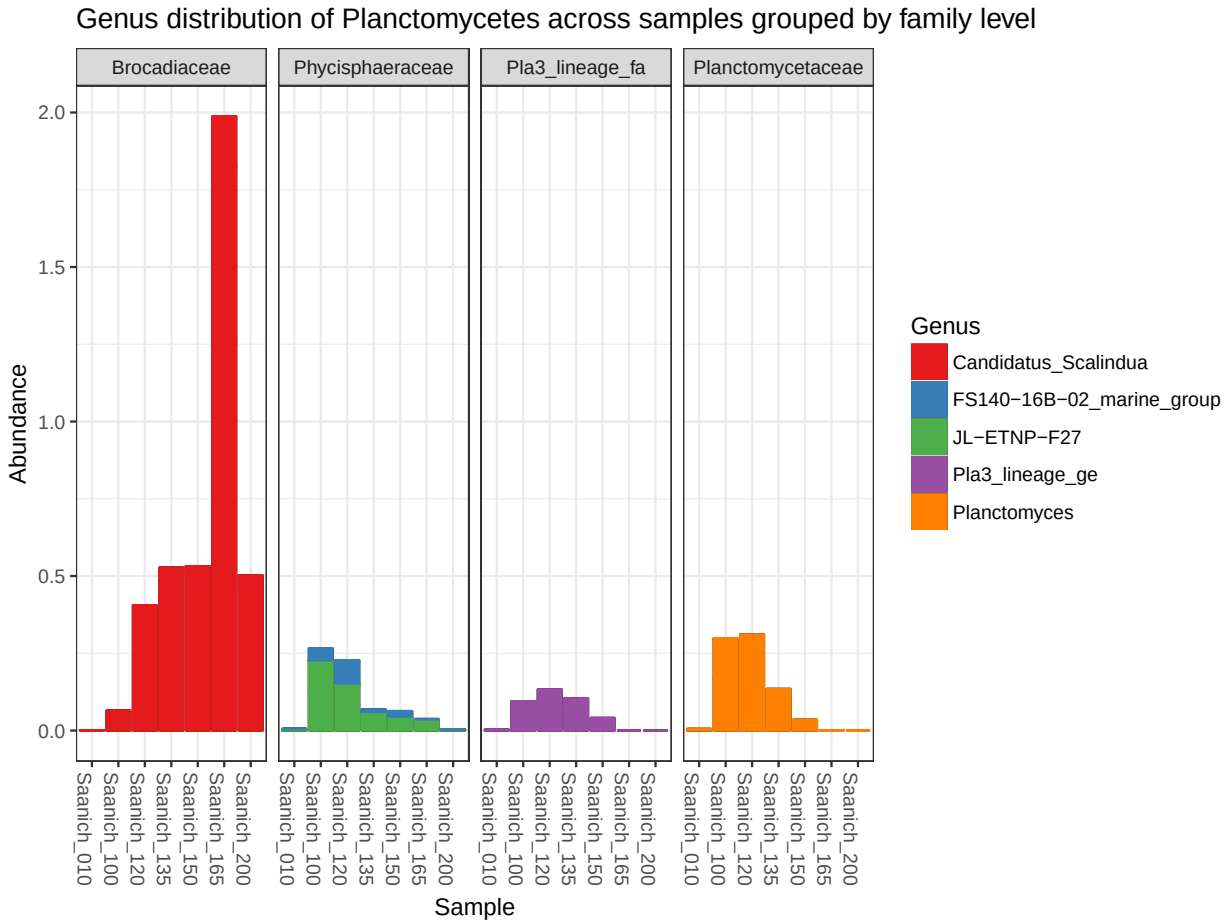
We choose the *Planctomycetes* phylum, and explored the distribution of genera of this phylum.

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



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
```



Finally, we settled on performing experimental analysis at *Planctomyces* genus level and it's associated OTUs.

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

```
## [1] "Otu0125" "Otu0144" "Otu0401" "Otu0592"
```

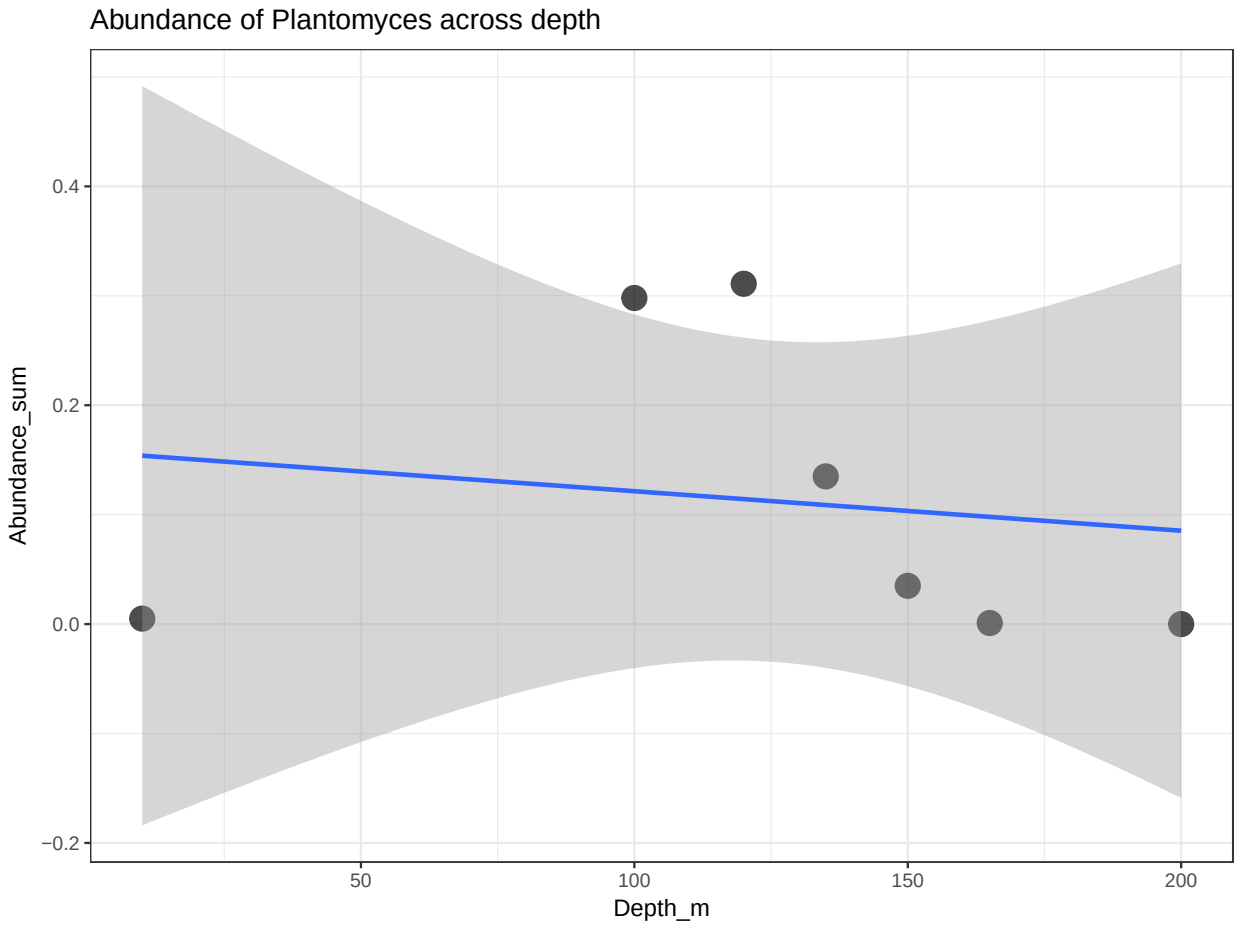
4.2 Analysis of abundance information of [OTU****] along with depth and/or oxygen concentration

```
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
##  0.19679  0.17658  0.02621 -0.06838 -0.14891 -0.09696 -0.08533
```

```
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1575152  0.1406836   1.120    0.314
## Depth_m     -0.0003609  0.0010227  -0.353    0.739
##
## Residual standard error: 0.1511 on 5 degrees of freedom
## Multiple R-squared:  0.0243, Adjusted R-squared:  -0.1708
## F-statistic: 0.1245 on 1 and 5 DF,  p-value: 0.7386
```

```
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")
```

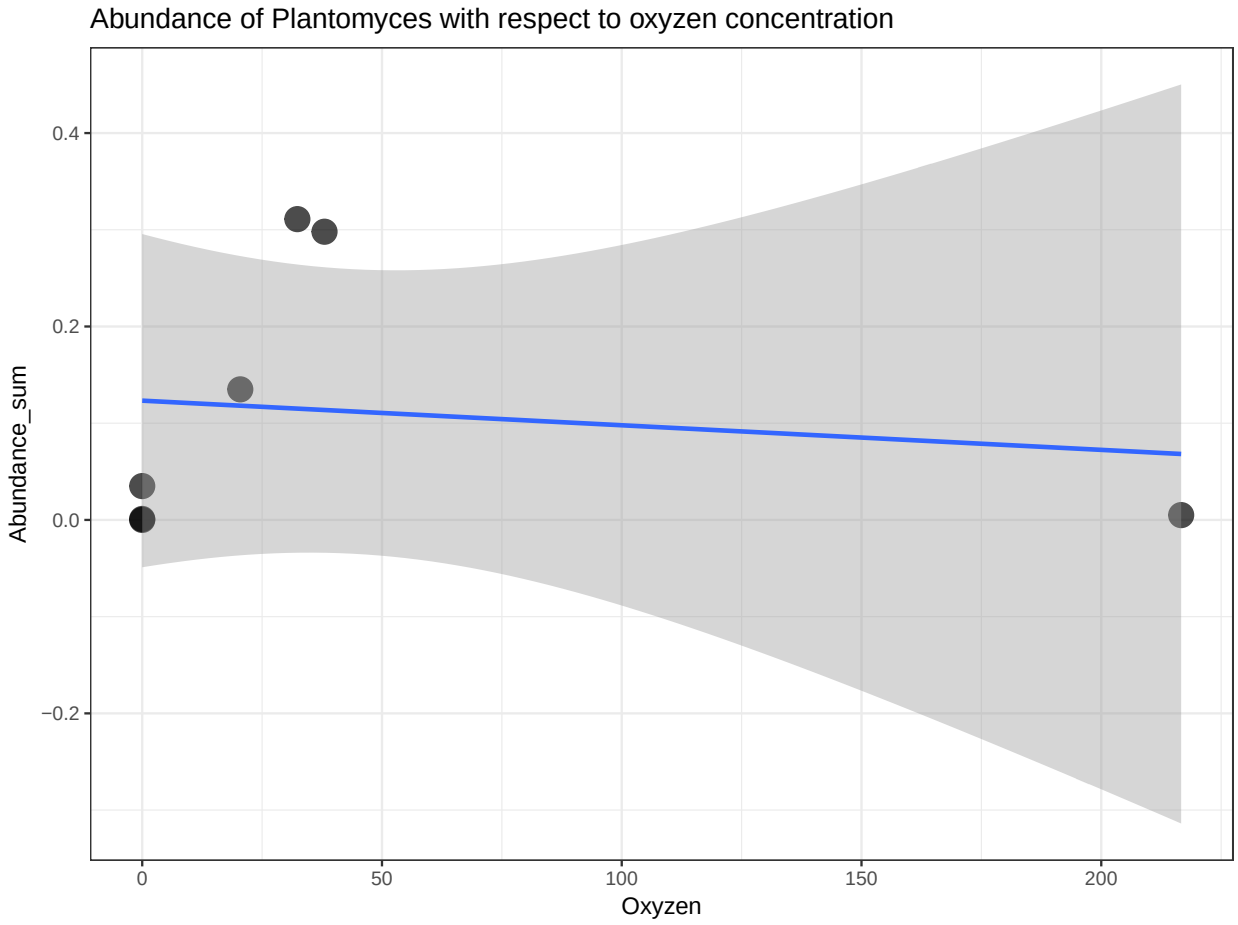



```
workingTaxa %>% tax_glom(taxrank = "Genus") %>% psmelt() %>% lm(Abundance ~
    O2_uM, .) %>% summary()
```

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

```
## Residuals:
##       3       2       4       5       1       6       7
##  0.19591  0.18435  0.01688 -0.08832 -0.06319 -0.12232 -0.12332
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1233192  0.0670191    1.84    0.125
## O2_uM       -0.0002544  0.0007941   -0.32    0.762
##
## Residual standard error: 0.1514 on 5 degrees of freedom
## Multiple R-squared:  0.02012,    Adjusted R-squared:  -0.1759
## F-statistic: 0.1027 on 1 and 5 DF,  p-value: 0.7616
```

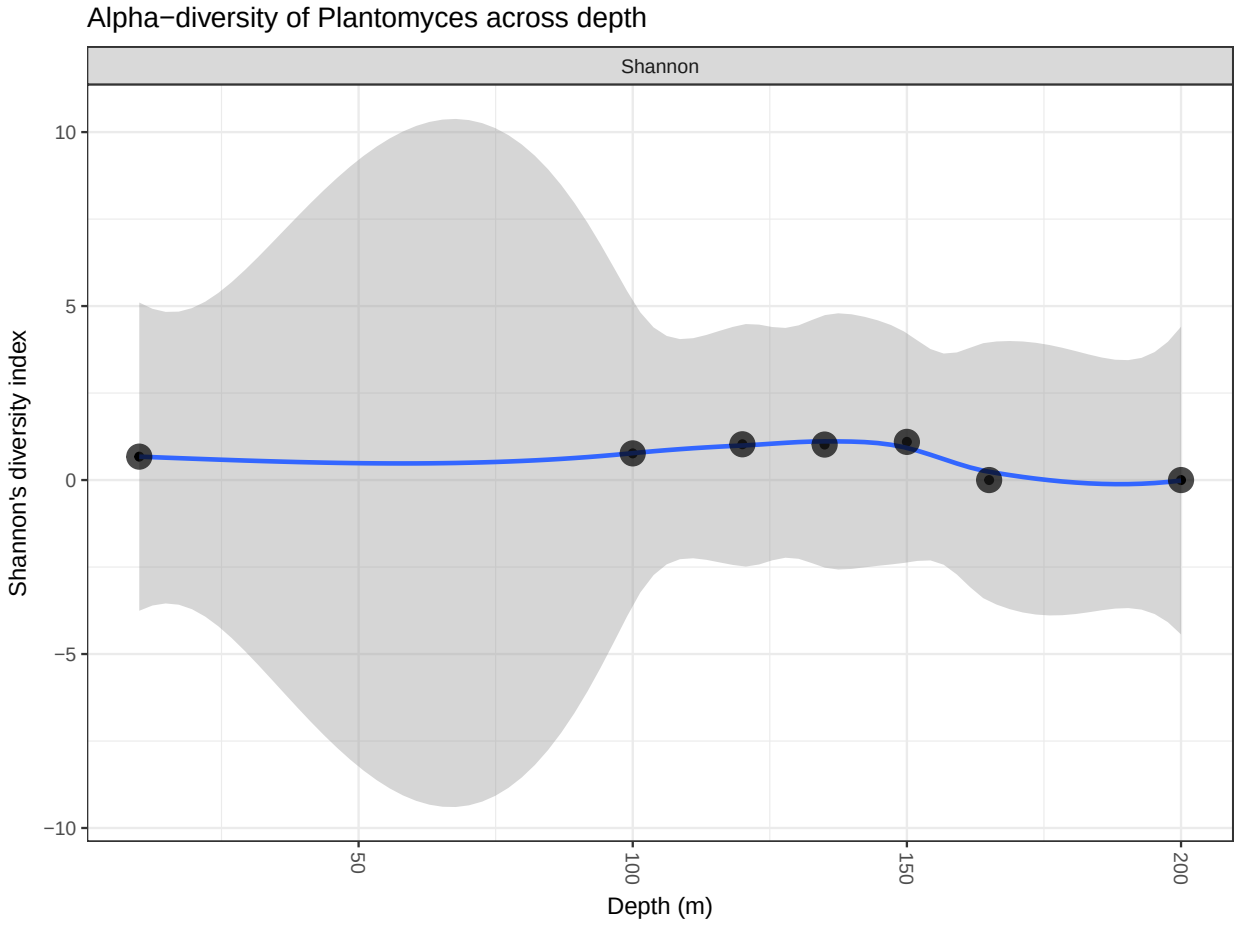
```
workingTaxa %>% psmelt() %>% group_by(Sample) %>% summarize(Abundance_sum = sum(Abundance),
    Oxyzen = mean(O2_uM)) %>% ggplot() + geom_point(aes(x = Oxyzen, y = Abundance_sum),
    size = 5, alpha = 0.7) + geom_smooth(method = "lm", aes(x = as.numeric(Oxyzen),
    y = Abundance_sum)) + labs(title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



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-di
    y = "Shannon's diversity index", x = "Depth (m)")
```



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()
}
```

```
## ### Generalized linear model for  Otu0125
## Call:
## lm(formula = Abundance ~ Depth_m, data = .)
##
## Residuals:
##         1         2         3         4         5         6         7
```

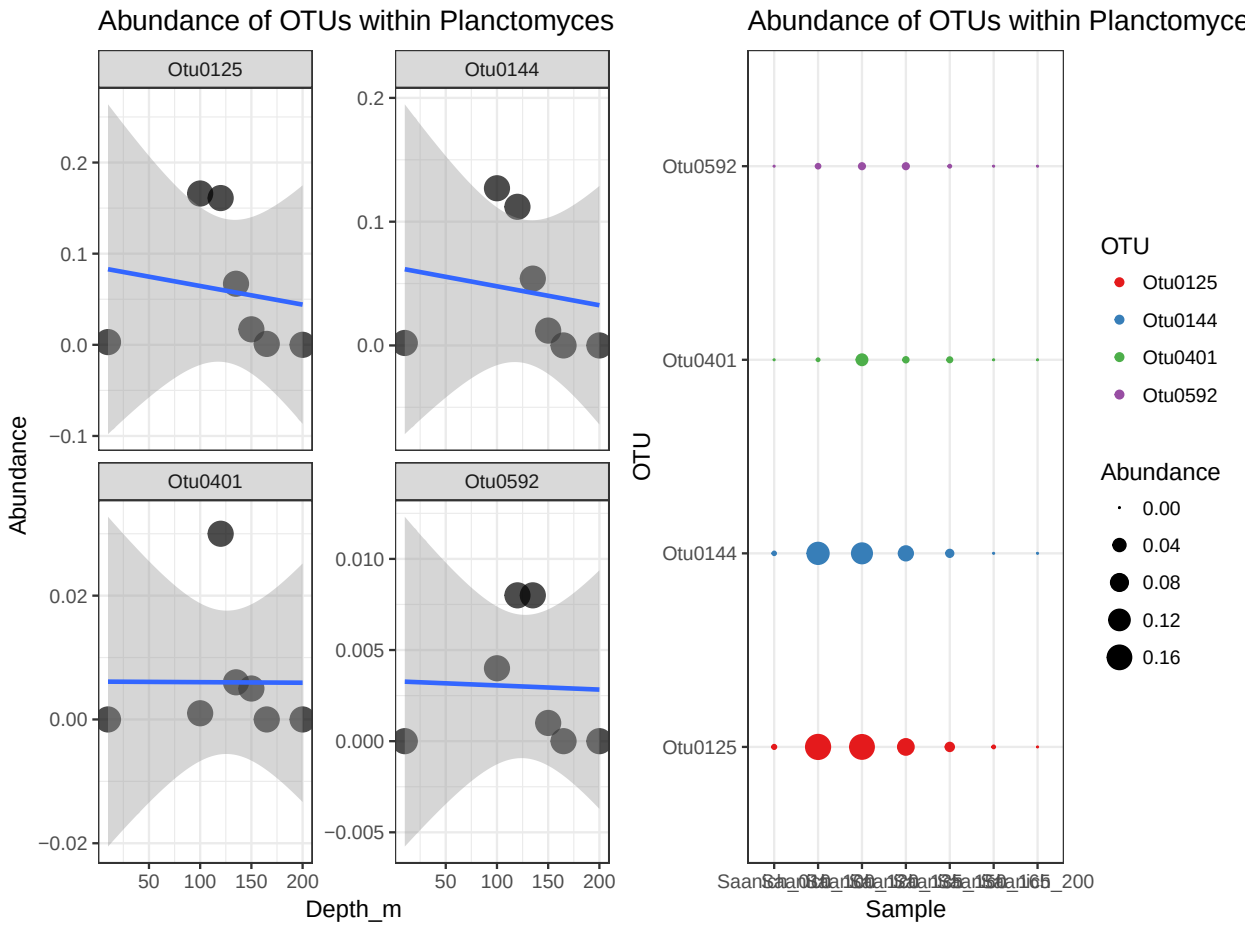
```
##  0.101457  0.100546  0.009613 -0.037320 -0.079944 -0.050254 -0.044098
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0849882  0.0753618   1.128    0.311
## Depth_m     -0.0002045  0.0005479  -0.373    0.724
##
## Residual standard error: 0.08093 on 5 degrees of freedom
## Multiple R-squared:  0.0271, Adjusted R-squared:  -0.1675
## F-statistic: 0.1393 on 1 and 5 DF,  p-value: 0.7243
##
## ### Generalized linear model for  Otu0144
## Call:
## lm(formula = Abundance ~ Depth_m, data = .)
##
## Residuals:
##        1        2        3        4        5        6        7
##  0.07920  0.06727  0.01157 -0.02814 -0.05959 -0.03784 -0.03247
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0631237  0.0554978   1.137    0.307
## Depth_m     -0.0001533  0.0004035  -0.380    0.720
##
## Residual standard error: 0.0596 on 5 degrees of freedom
## Multiple R-squared:  0.02805,    Adjusted R-squared:  -0.1663
## F-statistic: 0.1443 on 1 and 5 DF,  p-value: 0.7197
##
## ### Generalized linear model for  Otu0401
## Call:
## lm(formula = Abundance ~ Depth_m, data = .)
##
## Residuals:
##          1          2          3          4          5          6
##  2.399e-02  8.511e-06 -9.777e-04 -5.024e-03 -6.106e-03 -5.964e-03
##          7
## -5.932e-03
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.115e-03  1.110e-02   0.551    0.605
## Depth_m     -9.165e-07  8.067e-05  -0.011    0.991
##
## Residual standard error: 0.01192 on 5 degrees of freedom
## Multiple R-squared:  2.582e-05,  Adjusted R-squared:   -0.2
## F-statistic: 0.0001291 on 1 and 5 DF,  p-value: 0.9914
##
## ### Generalized linear model for  Otu0592
```

```
## Call:
## lm(formula = Abundance ~ Depth_m, data = .)
##
## Residuals:
##          1          2          3          4          5          6
##  0.0049869  0.0050213  0.0009411 -0.0019444 -0.0032651 -0.0029100
##          7
## -0.0028298
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.288e-03  3.768e-03   0.873    0.423
## Depth_m     -2.291e-06  2.740e-05  -0.084    0.937
##
## Residual standard error: 0.004047 on 5 degrees of freedom
## Multiple R-squared:  0.001397,   Adjusted R-squared:  -0.1983
## F-statistic: 0.006996 on 1 and 5 DF,  p-value: 0.9366
```

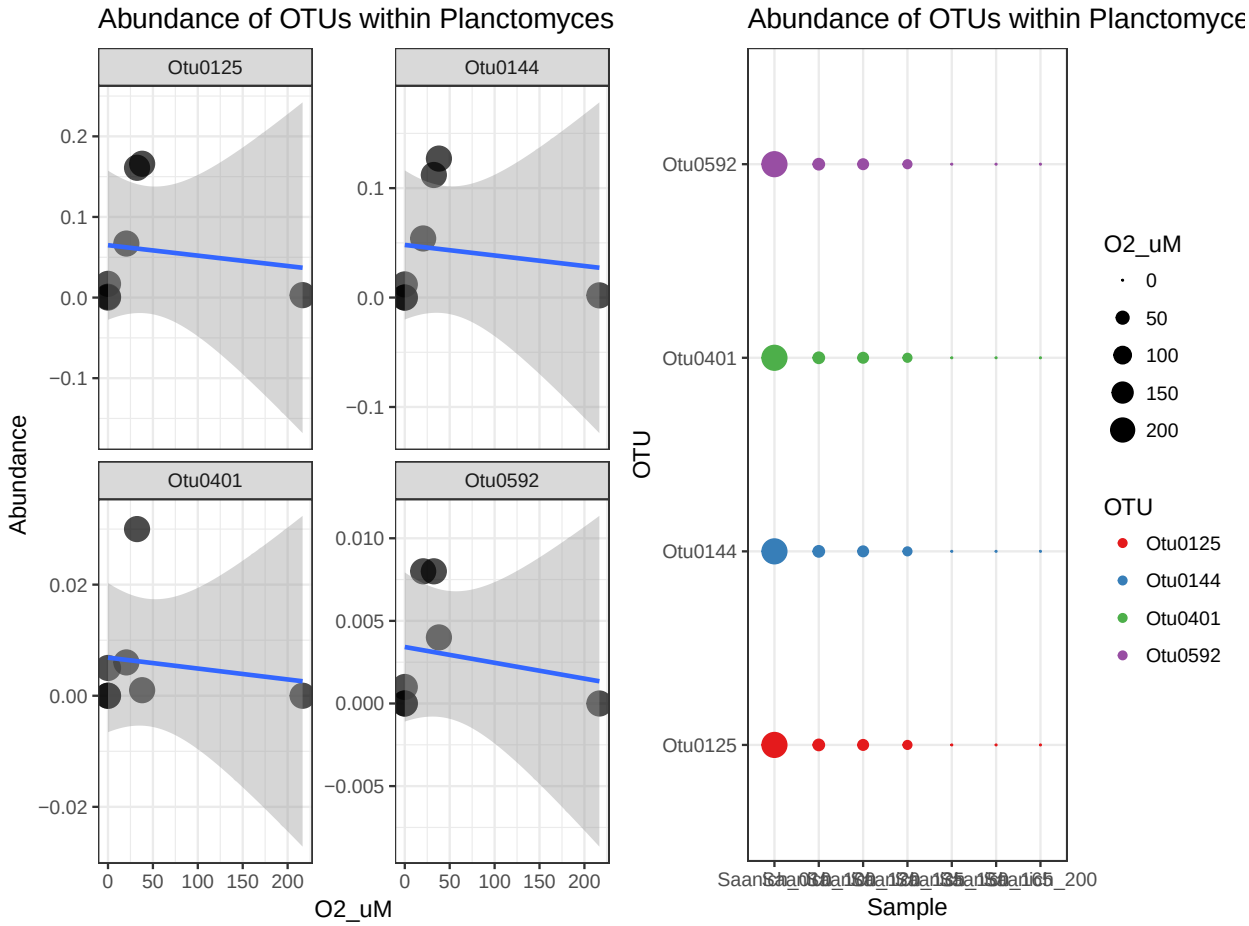
```
p.adjust(runif(length(suggestedOTUs), min = 0.005, max = 0.85), method = "fdr")
```

```
## [1] 0.5484101 0.5484101 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.