

Host-specificity and core taxa of seagrass leaf microbiome identified across tissue age and geographical regions | *Sanders-Smith, R. & Segovia, B.T.*(joint contribution), Forbes, C., Hessing-Lewis, M., Morien, E., Lemay, M.A., O'Connor, M. I., Parfrey, L.W.

Bianca Trevizan Segovia

20/11/2020

Final figure 2

```
### load packages ###
library(ggplot2)
library(ggpubr)
library(vegan)
library(dplyr)
library(reshape2)
library(phylloseq)
library(dplyr)
library(phylosmith)
library(QsRutils)
library(tidyverse)
library(ggvenn)
```

load alpha_all and alpha_choked

```
### importing master table ###
allsamples <- read.csv("data/Bact_3000_metadata_clean.csv", header = T)

### remove metadata to calculate alpha diversity ###
taxa_only <- allsamples %>% dplyr::select(-(1:4))

### disabling scientific notation ###
options(scipen = 999)

### set.seed for reproducibility ###
set.seed(1024)

### calculate alpha diversity metrics ###
Richness <- specnumber(taxa_only)
shannon <- diversity(taxa_only, index = "shannon")
pielou <- shannon/log(Richness)
chao1 <- estimateR(taxa_only)[2,]
```

```

#### create a data frame with alpha metrics values and metadata ####
alpha_div_df <- data.frame(chao1, pielou, allsamples$region, allsamples$sample_growth)

#### rename metadata columns ####
alpha_div_df <- alpha_div_df %>%
  dplyr::rename(region = allsamples.region, sample_type = allsamples.sample_growth)

#####
#### Boxplot alpha_all ####
#####
remove_artificial <- c("artificial")
alpha_metrics_no_artificial <- alpha_div_df %>%
  dplyr::filter(!sample_type %in% remove_artificial)

#### melting alpha metrics into :
#### one column called variable (containing alpha metrics labels)
#### and another column called value containing all values for those metrics ####
all_16s <- reshape2::melt(alpha_metrics_no_artificial,
                           id.var=c("sample_type", "region"))

all_16s$sample_type <- factor(all_16s$sample_type,
                               levels=c("zostera_new", "zostera_old", "seawater"))

#### rename sample types as I want them to appear in the graph ####
all_16s <- all_16s %>%
  dplyr::mutate(sample_type = dplyr::recode(sample_type,
                                              "zostera_new"="New Zostera",
                                              "zostera_old"="Old Zostera",
                                              "seawater"="Seawater"))

#### rename alpha metrics as I want them to appear in the graph ####
all_16s$variable <- factor(all_16s$variable, levels = c("chao1", "pielou"),
                            labels = c("Richness", "Evenness"))

# specify comparison brackets
my_comparisons <- list( c("New Zostera", "Old Zostera"),
                         c("Old Zostera", "Seawater"),
                         c("New Zostera", "Seawater"))
# remove labels and add later according to ANOVA's results
symnum.args <- list(
  cutpoints = c(0.0001, 0.001, 0.01, 0.05, 1),
  symbols = c("", "", "", ""))
)

#### boxplot graph ####
alpha_all <- ggplot2::ggplot(all_16s, aes(x = sample_type, y = value, fill = sample_type)) +
  facet_wrap(. ~ variable, scale="free") +
  geom_boxplot(notch = FALSE) +
  geom_jitter(width = 0.2, alpha = 0.2) +
  scale_fill_manual(values=c("yellow3", "#2a9958", "steelblue")) +
  stat_compare_means(comparisons = my_comparisons, symnum.args = symnum.args) +
  labs(y = "Alpha diversity measures") +
  theme_bw() +

```

```

theme (legend.position="none",
       axis.title.y = element_text(size = 16,
                                    margin = margin(t = 0, r = 10, b = 0, l = 0)),
       axis.title.x = element_blank(),
       axis.text.x = element_text(size = 14, angle = 90, hjust=1),
       axis.text.y = element_text(size = 18),
       panel.grid.major = element_blank(), #remove major grid
       panel.grid.minor = element_blank(), #remove minor grid
       axis.line = element_line(colour = "black"), #draw line in the axis
       strip.text.x = element_text(size = 14), # font size of face wrap
       strip.background =element_rect(fill="gray96", linetype="solid",color="black"),
       panel.border = element_blank()) #remove lines outside the graph

### add significance according to ANOVA's results #####
alpha_all <- alpha_all +
  geom_text(data = all_16s %>% filter(variable == "Richness"),
            x=1.5, y = 635, label = "***", size = 5, family="mono") +
  geom_text(data = all_16s %>% filter(variable == "Richness"),
            x=2.5, y = 700, label = "***", size = 5, family="mono") +
  geom_text(data = all_16s %>% filter(variable == "Richness"),
            x=2.0, y = 765, label = "**", size = 5, family="mono") +
  geom_text(data = all_16s %>% filter(variable == "Evenness"),
            x=1.5, y = 0.95, label = "***", size = 5, family="mono") +
  geom_text(data = all_16s %>% filter(variable == "Evenness"),
            x=2.5, y = 1.02, label = "***", size = 5, family="mono") +
  geom_text(data = all_16s %>% filter(variable == "Evenness"),
            x=2.0, y = 1.095, label = "***", size = 5, family="mono")

#####
### Boxplot alpha_choked #####
#####
remove_regions <- c("goose", "triquet", "mcmullin")
alpha_metrics_choked <- alpha_div_df %>%
  dplyr::filter (!region %in% remove_regions)

### melting alpha metrics into :
### one column called variable (containing alpha metrics labels)
### and another column called value containing all values for those metrics #####
choked_16s <- melt(alpha_metrics_choked , id.var=c("sample_type", "region"))
choked_16s$sample_type <- factor(choked_16s$sample_type,
                                   levels=c("zostera_new", "zostera_old", "seawater", "artificial"))

### rename sample types as I want them to appear in the graph #####
choked_16s <- choked_16s %>%
  dplyr::mutate(sample_type =dplyr::recode(sample_type,
                                              "zostera_new"="New Zostera",
                                              "zostera_old"="Old Zostera",
                                              "seawater"="Seawater",
                                              "artificial"="Artificial"))

choked_16s$variable <- factor(choked_16s$variable,
                               levels = c("chao1", "pielou"),
                               labels = c("Richness", "Evenness"))

```

```

# specify comparison brackets
my_comparisons_art <- list( c("Artificial", "New Zostera"),
                            c("Artificial", "Old Zostera"),
                            c("Artificial", "Seawater") )
# remove labels and add later according to ANOVA's results
symnum.args <- list(
  cutpoints = c(0.0001, 0.001, 0.01, 0.05, 1),
  symbols = c("", "", "", ""))
  
### boxplot graph ###
alpha_choked <- ggplot2::ggplot(choked_16s, aes(x = sample_type, y = value, fill = sample_type)) +
  facet_wrap(. ~ variable, scale="free") +
  geom_boxplot(notch = FALSE) +
  geom_jitter(width = 0.2, alpha = 0.2) +
  scale_fill_manual(values=c("yellow3", "#2a9958", "steelblue", "#D55E00")) +
  stat_compare_means(comparisons = my_comparisons_art, symnum.args = symnum.args) +
  labs(y = "Alpha diversity measures") +
  theme_bw() +
  theme (legend.position="none",
         axis.title.y = element_text(size = 16,
                                      margin = margin(t = 0, r = 10, b = 0, l = 0)),
         axis.title.x = element_blank(),
         axis.text.x = element_text(size = 14, angle = 90, hjust=1),
         axis.text.y = element_text(size = 18),
         panel.grid.major = element_blank(), #remove major grid
         panel.grid.minor = element_blank(), #remove minor grid
         axis.line = element_line(colour = "black"), #draw line in the axis
         strip.text.x = element_text(size = 14), # font size of face wrap
         strip.background =element_rect(fill="gray96", linetype="solid",color="black"),
         panel.border = element_blank()) #remove lines outside the graph

### add significance according to ANOVA's results ####
alpha_choked <- alpha_choked +
  geom_text(data = all_16s %>% filter(variable == "Richness"),
            x=2.5, y = 570, label = "***", size = 5, family="mono") +
  geom_text(data = all_16s %>% filter(variable == "Richness"),
            x=3.0, y = 630, label = "ns", size = 4, family="mono") +
  geom_text(data = all_16s %>% filter(variable == "Richness"),
            x=3.5, y = 680, label = "***", size = 5, family="mono") +
  geom_text(data = all_16s %>% filter(variable == "Evenness"),
            x=2.5, y = 0.93, label = "***", size = 5, family="mono") +
  geom_text(data = all_16s %>% filter(variable == "Evenness"),
            x=3.0, y = 1.0, label = "ns", size = 4, family="mono") +
  geom_text(data = all_16s %>% filter(variable == "Evenness"),
            x=3.5, y = 1.07, label = "ns", size = 4, family="mono")

```

load NMDS_all and NMDS_choked

```

### importing rarefied phyloseq object ####
phylo_merge_rare <- readRDS("data/phylo_merge_rarefied_16S_paper.rds")
phylo_merge_rare

```

```

### disabling scientific notation #####
options(scipen = 999)

### set.seed for reproducibility #####
set.seed(3245)

### root tree #####
phylo_merge_rare <- root_phyloseq_tree(phylo_merge_rare)
tree1 = phylo_tree(phylo_merge_rare)

#####
### NMDS_all #####
#####
### Remove artificial seagrass #####
phylo_NO_artificial_1 <- subset_samples(phylo_merge_rare, sample_growth!="artificial")

### IMPORTANT !!! #####
### Re-order phyloseq object to run PERMANOVA #####
phylo_NO_artificial <- set_sample_order(phylo_NO_artificial_1, c('sample_growth'))
# check if re-ordering worked, set same order for metadata later
SampleID_reordered <- as.data.frame(sample_data(phylo_NO_artificial)$SampleID)
#View(SampleID_reordered) # "seawater", "zostera_new", "zostera_old"

### ordinate using weighted Unifrac, Bray-Curtis and Jaccard dissimilarity
NMDS_bray_NO_artificial <- ordinate(phylo_NO_artificial, "NMDS", "bray")
NMDS_UNI_NO_artificial <- ordinate(phylo_NO_artificial, "NMDS", "wunifrac")
NMDS_jacc_NO_artificial <- ordinate(phylo_NO_artificial, "NMDS", "jaccard")

colours_no_artificial <- c(zostera_new="yellow3",
                             zostera_old="#2a9958",
                             seawater="steelblue")

NMDS_all <- plot_ordination(phylo_NO_artificial, NMDS_UNI_NO_artificial,
                           color = "sample_growth",
                           shape = "sample_growth",
                           title = "Weighted Unifrac") +
  geom_point(size=6) +
  scale_colour_manual(name = "Sample Type", values=colours_no_artificial,
                      labels=c("Seawater","New Zostera","Old Zostera")) +
  scale_shape_manual(name = "Sample Type", values=c(17,15,16),
                     labels=c("Seawater","New Zostera","Old Zostera")) +
  stat_ellipse(type = "t", linetype = 3, size = 1) +
  theme_bw() +
  theme (text = element_text(size=16),
        axis.text.x=element_text(size=12),
        axis.text.y = element_text(size=12),
        plot.title = element_text(hjust = 0.5),
        panel.grid.major = element_blank(), #remove major grid
        panel.grid.minor = element_blank())#remove minor grid

#####
### NMDS_choked #####
#####

```

```

### Select only Choked region ####
phylo_choked <- subset_samples(phylo_merge_rare,
                                region=="choked")

### IMPORTANT !!! ####
### Re-order phyloseq object to run PERMANOVA ####
phylo_choked <- set_sample_order(phylo_choked, c('sample_growth'))
# check if re-ordering worked, set same order for metadata later
SampleID_reordered_choked <- as.data.frame(sample_data(phylo_choked)$SampleID)
#View(SampleID_reordered_choked) # "artificial". "seawater", "zostera_new", "zostera_old"

### ordinate using weighted Unifrac, Bray-Curtis and Jaccard dissimilarity
NMDS_bray_CHOKED <- ordinate(phylo_choked, "NMDS", "bray")
NMDS_UNI_choked <- ordinate(phylo_choked, "NMDS", "wunifrac")
NMDS_jacc_choked <- ordinate(phylo_choked, "NMDS", "jaccard")

colours_choked <- c(zostera_new="yellow3",
                      zostera_old="#2a9958",
                      artificial="#D55E00",
                      seawater="steelblue")
NMDS_choked <- plot_ordination(phylo_choked, NMDS_UNI_choked,
                                 color = "sample_growth",
                                 shape = "sample_growth",
                                 title = "Weighted Unifrac") +
geom_point(size = 6) +
scale_colour_manual(name = "Sample Type", values=colours_choked,
                    labels=c("Artificial","Seawater", "New Zostera","Old Zostera")) +
scale_shape_manual(name = "Sample Type", values=c(18,17,15,16),
                   labels=c("Artificial","Seawater", "New Zostera","Old Zostera")) +
stat_ellipse(type = "t", linetype = 3, size = 1) +
theme_bw() +
theme (text = element_text(size=16),
       axis.text.x=element_text(size=12),
       axis.text.y = element_text(size=12),
       plot.title = element_text(hjust = 0.5),
       panel.grid.major = element_blank(), #remove major grid
       panel.grid.minor = element_blank())

```

load venn_all and venn_choked

```

##### importing files #####
phylo_merge_16S <-readRDS("data/phylo_merge_not_rarefied_16S_paper.rds")
phylo_merge_16S #1206 taxa

## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 1206 taxa and 149 samples ]
## sample_data() Sample Data: [ 149 samples by 4 sample variables ]
## tax_table() Taxonomy Table: [ 1206 taxa by 7 taxonomic ranks ]
## phy_tree() Phylogenetic Tree: [ 1206 tips and 1204 internal nodes ]

##### subset taxa based on pres/abs threshold per sample group #####
# make presence absence table
project_data.shared <- phylo_merge_16S # duplicate raw counts phyloseq object
otu <- as.data.frame(otu_table(project_data.shared)) #get OTU table
#set all positive values in OTU table of project_data.pres_abs to '1'

```

```

otu_table(project_data.shared)[otu >= 1] <- 1

##### subset based on groups you're interested in (sample types) #####
seawater = subset_samples(project_data.shared, sample_growth == "seawater")
zostera_new = subset_samples(project_data.shared, sample_growth == "zostera_new")
zostera_old = subset_samples(project_data.shared, sample_growth == "zostera_old")
artificial = subset_samples(project_data.shared, sample_growth == "artificial")

##### remove all OTUs not found at threshold (N samples) #####
# do sums from presence absence OTU table
taxa_sums_sea <- as.data.frame(filter_taxa(seawater,
                                              function(x) sum(x)))
# do sums from presence absence OTU table
taxa_sums_zos_new <- as.data.frame(filter_taxa(zostera_new,
                                                 function(x) sum(x)))
# do sums from presence absence OTU table
taxa_sums_zos_old <- as.data.frame(filter_taxa(zostera_old,
                                                function(x) sum(x)))
# do sums from presence absence OTU table
taxa_sums_artificial<- as.data.frame(filter_taxa(artificial,
                                                   function(x) sum(x)))

##### select OTUs present in at least 2 samples #####
#select OTUs with sample count over your threshold
keep_sea <- row.names(
  taxa_sums_sea )[which(taxa_sums_sea [,1] >= 2)]
#select OTUs with sample count over your threshold
keep_zos_new <- row.names(
  taxa_sums_zos_new)[which(taxa_sums_zos_new[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_zos_old <- row.names(
  taxa_sums_zos_old)[which(taxa_sums_zos_old[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_artificial <- row.names(
  taxa_sums_artificial)[which(taxa_sums_artificial[,1] >= 2)]

##### venn_all #####
count = dplyr::count
all <- list("New Zostera"=keep_zos_new,
            "Old Zostera"=keep_zos_old,
            "Seawater"=keep_sea)
venn_all <- ggvenn(
  all,
  fill_color = c("yellow3", "#2a9958","steelblue"),
  fill_alpha = 0.6,
  stroke_color = "black",
  stroke_alpha = 0.7,
  stroke_size = 1,
  stroke_linetype = "solid",
  set_name_color = "black",
  set_name_size = 6,
  text_color = "black",
  text_size = 5

```

```

)

### Select data from Choked region to compare to artificial seagrass ####
phylo_choked <- subset_samples(project_data.shared, region=="choked")
### Remove any OTUs that are absent from Choked samples ####
phylo_choked <- prune_taxa((taxa_sums(phylo_choked) > 0), phylo_choked)

##### subset based on groups you're interested in (sample types) #####
seawater_choked = subset_samples(phylo_choked, sample_growth == "seawater")
zostera_new_choked = subset_samples(phylo_choked, sample_growth == "zostera_new")
zostera_old_choked = subset_samples(phylo_choked, sample_growth == "zostera_old")
artificial_choked = subset_samples(phylo_choked, sample_growth == "artificial")

##### remove all OTUs not found at threshold (N samples) #####
# do sums from presence absence OTU table
taxa_sums_sea_choked <- as.data.frame(filter_taxa(seawater_choked,
                                                    function(x) sum(x)))
# do sums from presence absence OTU table
taxa_sums_zos_new_choked <- as.data.frame(filter_taxa(zostera_new_choked,
                                                       function(x) sum(x)))
# do sums from presence absence OTU table
taxa_sums_zos_old_choked <- as.data.frame(filter_taxa(zostera_old_choked,
                                                       function(x) sum(x)))
# do sums from presence absence OTU table
taxa_sums_artificial_choked <- as.data.frame(filter_taxa(artificial_choked,
                                                          function(x) sum(x)))

### select OTUs present in at least 2 samples ####
#select OTUs with sample count over your threshold
keep_sea_choked <- row.names(
  taxa_sums_sea_choked)[which(taxa_sums_sea_choked[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_zos_new_choked <- row.names(
  taxa_sums_zos_new_choked)[which(taxa_sums_zos_new_choked[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_zos_old_choked <- row.names(
  taxa_sums_zos_old_choked)[which(taxa_sums_zos_old_choked[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_artificial_choked <- row.names(
  taxa_sums_artificial_choked)[which(taxa_sums_artificial_choked[,1] >= 2)]

### venn_choked ####
choked <- list("Seawater"=keep_sea_choked,
               "New Zostera"=keep_zos_new_choked,
               "Old Zostera"=keep_zos_old_choked,
               "Artificial"=keep_artificial_choked)
venn_choked <- ggvenn(
  choked,
  fill_color = c("steelblue", "yellow3", "#2a9958", "#D55E00"),
  fill_alpha = 0.6,
  stroke_color = "black",
  stroke_alpha = 0.7,
  stroke_size = 1,
  stroke_linetype = "solid",

```

```

    set_name_color = "black",
    set_name_size = 6,
    text_color = "black",
    text_size = 5
)

```

Arrange all graphs for Final figure 2

```

alpha_beta <- ggarrange(alpha_all + theme(plot.margin = unit(c(0,30,15,7), "pt")),
                         alpha_choked + theme(plot.margin = unit(c(0,30,15,0), "pt")),
                         NMDS_all+ theme(plot.margin = unit(c(0,30,0,7), "pt")),
                         NMDS_choked+ theme(plot.margin = unit(c(0,30,0,0), "pt")),
                         labels = c("A", "B", "C", "D"), ncol = 2, nrow = 2)

alpha_beta_legend_title <- alpha_beta + ggtitle("All regions"
  plot.title = element_text(size = 30, face = "bold", margin = margin(10, 0, 20, 0)))

venn_diagrams <- ggarrange(venn_all, venn_choked, labels = c("E", "F"), ncol = 2)

Figure_2 <- ggarrange(alpha_beta_legend_title, venn_diagrams, nrow=2, heights=c(2, 1))
ggsave("final_figures/fig2_alpha_beta_venn.jpeg",
  plot = Figure_2 , width=455, height=650, units="mm",dpi=600)

```

Figure 2 Alpha diversity metrics, NMDS plots based on weighted UniFrac distances, and Venn diagrams of shared taxa of microbial communities on *Z. marina* new growth leaves, old growth leaves, and seawater for all regions (A, C, E), and for Choked region including artificial seagrass units (ASUs) (B, D, F). Ellipses represent ordination confidence intervals (95%).

