

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 3

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

load NMDS_new_regions, NMDS_old_regions and and NMDS_water_regions

### 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(986)

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

#####
### NMDS_new_regions  ###
#####
```

```

### Select new zostera growth from phyloseq object ####
phylo_merge_new <- subset_samples(phylo_merge_rare,
                                   sample_growth %in% c("zostera_new"))

### ordinate using weighted Unifrac, Bray-Curtis dissimilarity ####
NMDS_bray_new <- ordinate(phylo_merge_new, "NMDS", "bray")
NMDS_UNI_new <- ordinate(phylo_merge_new, "NMDS", "wunifrac")

### NMDS plot graph ####
NMDS_new_regions <- plot_ordination(phylo_merge_new, NMDS_UNI_new,
                                      type = "sample",
                                      color = "region",
                                      shape = "region",
                                      title = "New Zostera (Weighted Unifrac)") +
  geom_point(size=6) +
  scale_colour_brewer(name = "Region", palette = "Set1",
                      labels=c("Choked", "Goose", "McMullin", "Triquet")) +
  scale_shape_manual(name = "Region", values=c(18,15,16, 17),
                     labels=c("Choked", "Goose", "McMullin", "Triquet")) +
  stat_ellipse(type = "t", linetype = 3, size = 1) +
  theme_bw() +
  theme(text = element_text(size=18),
        axis.text.x=element_text(size=20),
        axis.text.y = element_text(size=20),
        plot.title = element_text(hjust = 0.5),
        panel.grid.major = element_blank(), #remove major grid
        panel.grid.minor = element_blank())

#####
### NMDS_old_regions #####
#####
### Select old zostera growth from phyloseq object #####
phylo_merge_old <- subset_samples(phylo_merge_rare,
                                   sample_growth %in% c("zostera_old"))

### ordinate using weighted Unifrac, Bray-Curtis dissimilarity ####
NMDS_bray_old <- ordinate(phylo_merge_old, "NMDS", "bray")
NMDS_UNI_old <- ordinate(phylo_merge_old, "NMDS", "wunifrac")

### NMDS plot graph ####
NMDS_old_regions <- plot_ordination(phylo_merge_old, NMDS_UNI_old,
                                      type = "sample",
                                      color = "region",
                                      shape = "region",
                                      title = "Old Zostera (Weighted Unifrac)") +
  geom_point(size=6) +
  scale_colour_brewer(name = "Region", palette = "Set1",
                      labels=c("Choked", "Goose", "McMullin", "Triquet")) +
  scale_shape_manual(name = "Region", values=c(18,15,16, 17),
                     labels=c("Choked", "Goose", "McMullin", "Triquet")) +
  stat_ellipse(type = "t", linetype = 3, size = 1) +
  theme_bw() +
  theme(text = element_text(size=18),

```

```

    axis.text.x=element_text(size=20),
    axis.text.y = element_text(size=20),
    plot.title = element_text(hjust = 0.5),
    panel.grid.major = element_blank(), #remove major grid
    panel.grid.minor = element_blank())

#####
### NMDS_water_regions #####
#####
### Select seawater from phyloseq object #####
phylo_merge_water <- subset_samples(phylo_merge_rare, sample_growth %in% c("seawater"))

### ordinate using weighted Unifrac, Bray-Curtis dissimilarity #####
NMDS_bray_water <- ordinate(phylo_merge_water, "NMDS", "bray")
NMDS_UNI_water <- ordinate(phylo_merge_water, "NMDS", "wunifrac")

### NMDS plot graph #####
NMDS_water_regions <- plot_ordination(phylo_merge_water, NMDS_UNI_water,
                                         type = "sample",
                                         color = "region",
                                         shape = "region",
                                         title = "Seawater (Weighted Unifrac)") +
  geom_point(size=6) +
  scale_colour_brewer(name = "Region", palette = "Set1",
                      labels=c("Choked", "Goose", "McMullin", "Triquet")) +
  scale_shape_manual(name = "Region", values=c(18,15,16, 17),
                     labels=c("Choked", "Goose", "McMullin", "Triquet")) +
  stat_ellipse(type = "t", linetype = 3, size = 1) +
  theme_bw() +
  theme(text = element_text(size=18),
        axis.text.x=element_text(size=20),
        axis.text.y = element_text(size=20),
        plot.title = element_text(hjust = 0.5),
        panel.grid.major = element_blank(), #remove major grid
        panel.grid.minor = element_blank())

```

load significance figures

```

new_signif <- ggdraw() +
  draw_image("new.png") +
  draw_label("New Zostera", x = 0.5, y = 0.65, size = 22, fontfamily = "Arial")
old_signif <- ggdraw() +
  draw_image("old.png") +
  draw_label("Old Zostera", x = 0.5, y = 0.65, size = 22, fontfamily = "Arial")
sea_signif <- ggdraw() +
  draw_image("sea.png") +
  draw_label("Seawater", x = 0.5, y = 0.65, size = 22, fontfamily = "Arial")

```

load venn_new, venn_old and venn_water

```

#### 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

##### Select zostera new growth data #####
project_data_zostera_new = subset_samples(project_data.shared,
                                            sample_growth == "zostera_new")
project_data_zostera_new

## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 1206 taxa and 54 samples ]
## sample_data() Sample Data: [ 54 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 based on groups you're interested in (regions) #####
zostera_new_groupC = subset_samples(project_data_zostera_new, region == "choked")
zostera_new_groupG = subset_samples(project_data_zostera_new, region == "goose")
zostera_new_groupM = subset_samples(project_data_zostera_new, region == "mcmillin")
zostera_new_groupT = subset_samples(project_data_zostera_new, region == "triquet")

##### remove all OTUs not found at threshold (N samples) #####
# do sums from presence absence OTU table
zostera_new_taxa_sums_grC <- as.data.frame(filter_taxa(zostera_new_groupC,
                                                       function(x) sum(x)))
# do sums from presence absence OTU table
zostera_new_taxa_sums_grG <- as.data.frame(filter_taxa(zostera_new_groupG,
                                                       function(x) sum(x)))
# do sums from presence absence OTU table
zostera_new_taxa_sums_grM <- as.data.frame(filter_taxa(zostera_new_groupM,
                                                       function(x) sum(x)))
# do sums from presence absence OTU table
zostera_new_taxa_sums_grT <- as.data.frame(filter_taxa(zostera_new_groupT,
                                                       function(x) sum(x)))

##### select OTUs present in at least 2 samples #####
#select OTUs with sample count over your threshold
keep_new_zostera_C <- row.names(
  zostera_new_taxa_sums_grC)[which(zostera_new_taxa_sums_grC[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_new_zostera_G <- row.names(
  zostera_new_taxa_sums_grG)[which(zostera_new_taxa_sums_grG[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_new_zostera_M <- row.names(

```

```

zostera_new_taxa_sums_grM [which(zostera_new_taxa_sums_grM[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_new_zostera_T <- row.names(
  zostera_new_taxa_sums_grT [which(zostera_new_taxa_sums_grT[,1] >= 2)]

### venn_new ####
new_regions <- list("Choked"=keep_new_zostera_C,
                      "Goose"=keep_new_zostera_G,
                      "McMullin"=keep_new_zostera_M,
                      "Triquet"=keep_new_zostera_T)
venn_new <- ggvenn(
  new_regions,
  fill_color = c("#e41a1c", "#377eb8", "#4daf4a", "#984ea3"),
  fill_alpha = 0.6,
  stroke_color = "black",
  stroke_alpha = 0.7,
  stroke_size = 1,
  stroke_linetype = "solid",
  set_name_color = "black",
  set_name_size = 5,
  text_color = "black",
  text_size = 5
)

venn_new <- venn_new +
  annotate("text", x = 0, y = 1.6, label = "New Zostera", size = 8)

### Select zostera old growth data ####
project_data_zostera_old = subset_samples(project_data.shared,
                                             sample_growth == "zostera_old")
project_data_zostera_old

## phyloseq-class experiment-level object
## otu_table()    OTU Table:      [ 1206 taxa and 51 samples ]
## sample_data() Sample Data:     [ 51 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 based on groups you're interested in (regions) ####
zostera_old_groupC = subset_samples(project_data_zostera_old, region == "choked")
zostera_old_groupG = subset_samples(project_data_zostera_old, region == "goose")
zostera_old_groupM = subset_samples(project_data_zostera_old, region == "mcmullin")
zostera_old_groupT = subset_samples(project_data_zostera_old, region == "triquet")

### remove all OTUs not found at threshold (N samples) ####
# do sums from presence absence OTU table
zostera_old_taxa_sums_grC <- as.data.frame(filter_taxa(zostera_old_groupC,
                                                       function(x) sum(x)))
# do sums from presence absence OTU table
zostera_old_taxa_sums_grG <- as.data.frame(filter_taxa(zostera_old_groupG,
                                                       function(x) sum(x)))
# do sums from presence absence OTU table
zostera_old_taxa_sums_grM <- as.data.frame(filter_taxa(zostera_old_groupM,
                                                       function(x) sum(x)))

```

```

# do sums from presence absence OTU table
zostera_old_taxa_sums_grT <- as.data.frame(filter_taxa(zostera_old_groupT,
                                                       function(x) sum(x)))

### select OTUs present in at least 2 samples ####
#select OTUs with sample count over your threshold
keep_old_zostera_C <- row.names(
  zostera_old_taxa_sums_grC)[which(zostera_old_taxa_sums_grC[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_old_zostera_G <- row.names(
  zostera_old_taxa_sums_grG)[which(zostera_old_taxa_sums_grG[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_old_zostera_M <- row.names(
  zostera_old_taxa_sums_grM)[which(zostera_old_taxa_sums_grM[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_old_zostera_T <- row.names(
  zostera_old_taxa_sums_grT)[which(zostera_old_taxa_sums_grT[,1] >= 2)]

### venn_old ####
old_regions <- list("Choked"=keep_old_zostera_C,
                     "Goose"=keep_old_zostera_G,
                     "McMullin"=keep_old_zostera_M,
                     "Triquet"=keep_old_zostera_T)
venn_old <- ggvenn(
  old_regions,
  fill_color = c("#e41a1c", "#377eb8", "#4daf4a", "#984ea3"),
  fill_alpha = 0.6,
  stroke_color = "black",
  stroke_alpha = 0.7,
  stroke_size = 1,
  stroke_linetype = "solid",
  set_name_color = "black",
  set_name_size = 5,
  text_color = "black",
  text_size = 5
)
venn_old <- venn_old +
  annotate("text", x = 0, y = 1.6, label = "Old Zostera", size = 8)

### Select seawater data ####
project_data_water = subset_samples(project_data.shared,
                                     sample_growth == "seawater")

### subset based on groups you're interested in (regions) ####
water_groupC = subset_samples(project_data_water, region == "choked")
water_groupG = subset_samples(project_data_water, region == "goose")
water_groupM = subset_samples(project_data_water, region == "mcmullin")
water_groupT = subset_samples(project_data_water, region == "triquet")

### remove all OTUs not found at threshold (N samples) ####
# do sums from presence absence OTU table
water_taxa_sums_grC <- as.data.frame(filter_taxa(water_groupC,

```

```

        function(x) sum(x)))
# do sums from presence absence OTU table
water_taxa_sums_grG <- as.data.frame(filter_taxa(water_groupG,
                                                 function(x) sum(x)))
# do sums from presence absence OTU table
water_taxa_sums_grM <- as.data.frame(filter_taxa(water_groupM,
                                                 function(x) sum(x)))
# do sums from presence absence OTU table
water_taxa_sums_grT <- as.data.frame(filter_taxa(water_groupT,
                                                 function(x) sum(x)))

### select OTUs present in at least 2 samples ###
#select OTUs with sample count over your threshold
keep_water_C <- row.names(
  water_taxa_sums_grC)[which(water_taxa_sums_grC[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_water_G <- row.names(
  water_taxa_sums_grG)[which(water_taxa_sums_grG[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_water_M <- row.names(
  water_taxa_sums_grM)[which(water_taxa_sums_grM[,1] >= 2)]
#select OTUs with sample count over your threshold
keep_water_T <- row.names(
  water_taxa_sums_grT)[which(water_taxa_sums_grT[,1] >= 2)]

### venn_water ####
sea_regions <- list("Choked"=keep_water_C,
                     "Goose"=keep_water_G,
                     "McMullin"=keep_water_M,
                     "Triquet"=keep_water_T)
venn_water <- ggvenn(
  sea_regions,
  fill_color = c("#e41a1c", "#377eb8", "#4daf4a", "#984ea3"),
  fill_alpha = 0.6,
  stroke_color = "black",
  stroke_alpha = 0.7,
  stroke_size = 1,
  stroke_linetype = "solid",
  set_name_color = "black",
  set_name_size = 5,
  text_color = "black",
  text_size = 5
)
venn_water <- venn_water +
  annotate("text", x = 0, y = 1.6, label = "Seawater", size = 8)

```

Arrange all graphs for Final figure 3

```

### arrange all NMDS plots ####
NMDS_new_regions <- NMDS_new_regions + theme(legend.position="none")
NMDS_old_regions <- NMDS_old_regions + theme(legend.position="none")
NMDS_water_regions <- NMDS_water_regions + theme(legend.position="none")

```

```

NMDS_sample_types_regions <- ggarrange(NMDS_new_regions +
                                         theme(plot.margin = unit(c(11,30,10,7), "pt")),
                                         NMDS_old_regions +
                                         theme(plot.margin = unit(c(11,30,10,7), "pt")),
                                         NMDS_water_regions +
                                         theme(plot.margin = unit(c(11,30,10,7), "pt")),
                                         labels = c("A", "B", "C"),
                                         ncol = 1, nrow = 3)

### get legend from one of the graphs ####
NMDS_new_regions_legend <- NMDS_new_regions + theme(legend.position = "bottom",
                                                       legend.spacing.x = unit(0.3, 'cm'),)
legend <- get_legend(NMDS_new_regions_legend)
nmds_samples_regions_legend <- ggarrange(NMDS_sample_types_regions,
                                           legend, ncol = 1, nrow = 2,
                                           heights=c(15, 1))

### add title to significance grids ####
signif_grids <- ggarrange(new_signif,
                           old_signif,
                           sea_signif,
                           labels = c("D", "E", "F"), ncol = 1, nrow=3,
                           heights=c(1, 1))

venn_diagrams <- ggarrange(venn_new,
                           venn_old,
                           venn_water,
                           labels = c("G", "H", "I"), ncol = 1, nrow=3)

Figure_3 <- ggarrange(nmds_samples_regions_legend,
                      signif_grids,
                      venn_diagrams,
                      ncol=3,
                      heights=c(1,0.6,0.9),
                      widths =c(1,0.6,0.9))

ggsave("final_figures/fig3_beta_venn.jpeg", plot = Figure_3 ,
       width=600, height=650, units="mm",dpi=600)

```

Figure 3 NMDS plots based on weighted UniFrac distances, pairwise PERMANOVA results table and Venn diagrams of shared taxa of microbial communities of *Z. marina* new growth leaves (A, D, G), old growth leaves (B, E, H) and seawater (C, F, I) samples among regions. Pairwise PERMANOVA grids coloured in grey are not significant, whereas grids coloured in yellow are significant, where * indicates $p>0.05$ and *** indicates $p>0.001$. Ellipses represent ordination confidence intervals (95%).

