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

Alpha diversity (Chao1 richness and Pielou's eveness) comparison across samples types

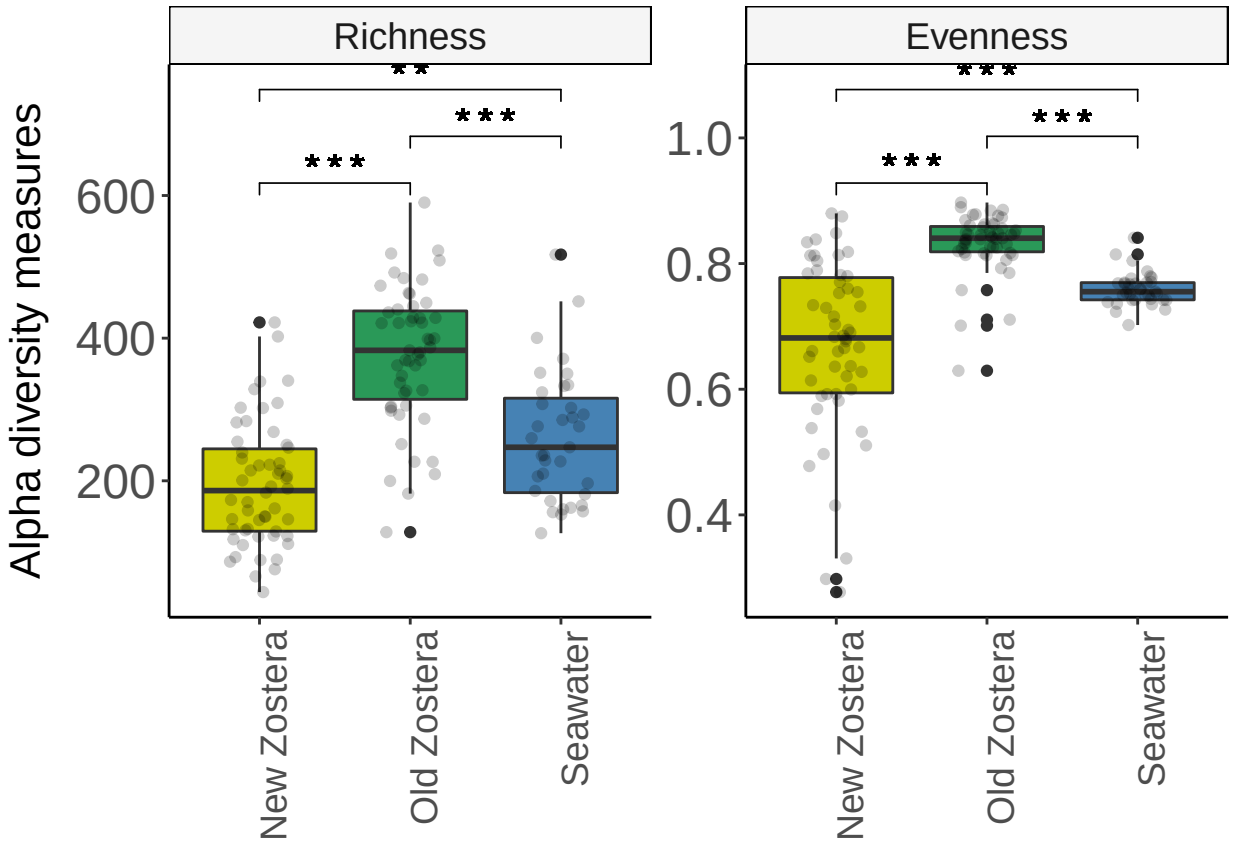
```
### load packages ###
library(vegan)
library(dplyr)
library(reshape2)
library(ggplot2)
library(car)
library(ggpubr)

### 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)
```

whole dataset *excluding artificial*

```
################
### Boxplot ###
################
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")
alpha_all
```




```
###############
### ANOVAs ###
```

```
###############
### test for heteroscedasticity in the data ###
car::leveneTest(chao1 ~ sample_type, data=alpha_metrics_no_artificial)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   2  0.1798 0.8356
##       137
```

```
car::leveneTest(pielou ~ sample_type, data=alpha_metrics_no_artificial)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value           Pr(>F)
## group   2  27.956 0.00000000006583 ***
##       137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
### Because we have heteroscedasticity, we should have the same n so ANOVA is robust ###
sum(alpha_metrics_no_artificial$sample_type == "seawater") #35 samples
```

```
## [1] 35
```

```
sum(alpha_metrics_no_artificial$sample_type == "zostera_new") #54 samples
```

```
## [1] 54
```

```
sum(alpha_metrics_no_artificial$sample_type == "zostera_old") #51 samples
```

```
## [1] 51
```

```
### Randomly get 35 samples from seawater and zostera so they have same n: ###
bact_random_no_artificial <- alpha_metrics_no_artificial %>%
  group_by(sample_type) %>%
  sample_n(size = 35)

##See if it worked
sum(bact_random_no_artificial$sample_type == "seawater")
```

```
## [1] 35
```

```
sum(bact_random_no_artificial$sample_type== "zostera_new")
```

```
## [1] 35
```

```
sum(bact_random_no_artificial$sample_type == "zostera_old")
```

```
## [1] 35
```

```
#############################################################
### RANDOM samples (35 samples for each sample type) ###
#############################################################

##############
### Chao1 ###
##############
### with geographical regions as a random effect ###
anova(lm(chao1 ~ sample_type + region, data=bact_random_no_artificial))
```

```
## Analysis of Variance Table
```

```
##
## Response: chao1
##             Df Sum Sq Mean Sq F value           Pr(>F)
## sample_type  2 540886  270443 31.5173 0.00000000002561 ***
## region       3 111733   37244  4.3404          0.00644 **
## Residuals   99 849496    8581
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(aov(chao1 ~ sample_type + region, data=bact_random_no_artificial))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = chao1 ~ sample_type + region, data = bact_random_no_artificial)
##
## $sample_type
##                               diff        lwr        upr     p adj
## zostera_new-seawater    -61.95158 -114.64120  -9.261953 0.0168766
## zostera_old-seawater    111.51048   58.82086 164.200107 0.0000064
## zostera_old-zostera_new 173.46206  120.77244 226.151684 0.0000000
##
## $region
##                       diff         lwr       upr     p adj
## goose-choked     25.98205 -40.1543741  92.11847 0.7342879
## mcmullin-choked  63.10389   0.4335293 125.77425 0.0477700
## triquet-choked   80.40888  15.2383898 145.57938 0.0091349
## mcmullin-goose   37.12184 -34.5315033 108.77519 0.5311892
## triquet-goose    54.42684 -19.4231544 128.27683 0.2239294
## triquet-mcmullin 17.30499 -53.4577689  88.06776 0.9190995
```

```
#########################
### Pielou's Evenness ###
#########################
### with geographical regions as a random effect ###
anova(lm(pielou ~ sample_type + region, data=bact_random_no_artificial))
```

```
## Analysis of Variance Table
##
## Response: pielou
##             Df  Sum Sq  Mean Sq F value           Pr(>F)
## sample_type  2 0.35517 0.177583 30.3405 0.00000000005283 ***
## region       3 0.03772 0.012573  2.1482          0.09898 .
## Residuals   99 0.57945 0.005853
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(aov(pielou ~ sample_type + region, data=bact_random_no_artificial))
```

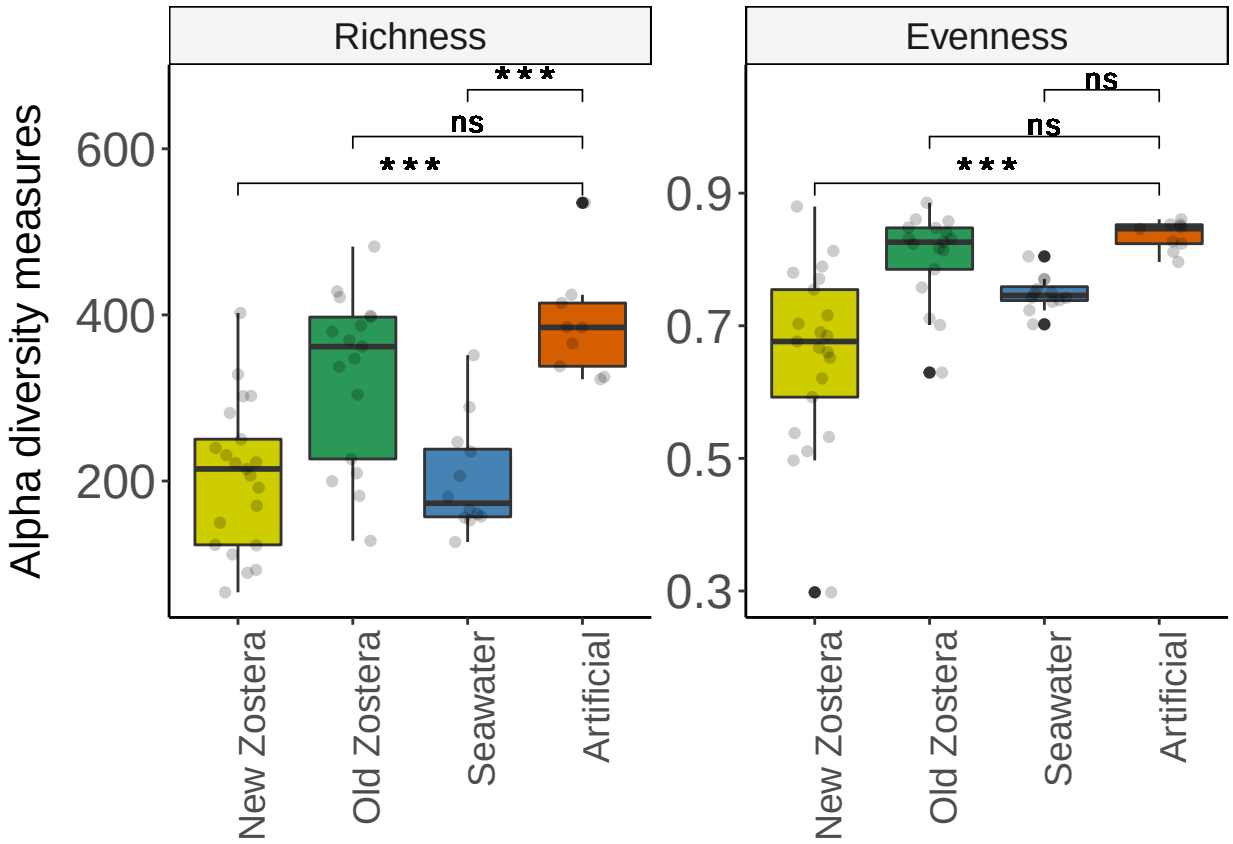
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = pielou ~ sample_type + region, data = bact_random_no_artificial)
##
## $sample_type
##                                diff         lwr         upr     p adj
```

```
## zostera_new-seawater    -0.07584979 -0.11936601 -0.03233358 0.0002086
## zostera_old-seawater     0.06650925  0.02299304  0.11002547 0.0012708
## zostera_old-zostera_new  0.14235904  0.09884283  0.18587526 0.0000000
##
## $region
##                           diff         lwr        upr     p adj
## goose-choked     -0.004014132 -0.05863602 0.05060776 0.9974731
## mcmullin-choked   0.038754552 -0.01300473 0.09051383 0.2115888
## triquet-choked    0.033475114 -0.02034902 0.08729925 0.3692870
## mcmullin-goose    0.042768683 -0.01640962 0.10194698 0.2395692
## triquet-goose     0.037489246 -0.02350326 0.09848175 0.3798695
## triquet-mcmullin -0.005279437 -0.06372221 0.05316333 0.9953405
```

choked dataset *including artificial*

```
################
### Boxplot ###
################
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")
alpha_choked
```




```
###############
### ANOVAs ###
###############
### test for heteroscedasticity in the data ###
car::leveneTest(chao1 ~ sample_type, data=alpha_metrics_choked)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  3  0.9816 0.4082
##       55
```

```
car::leveneTest(pielou ~ sample_type, data=alpha_metrics_choked)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value   Pr(>F)
## group  3   5.642 0.001916 **
##       55
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
### Because we have heteroscedasticity, we should have the same n so ANOVA is robust ###
sum(alpha_metrics_choked$sample_type == "seawater") #12 samples
```

```
## [1] 12
```

```
sum(alpha_metrics_choked$sample_type == "zostera_new") #21 samples
```

```
## [1] 21
```

```r
sum(alpha_metrics_choked$sample_type == "zostera_old") #17 samples
```

```
## [1] 17
```

```r
sum(alpha_metrics_choked$sample_type == "artificial")  #9 samples
```

```
## [1] 9
```

```r
### Randomly get 9 samples from seawater and zostera so they have same n: ###
bact_random_choked <- alpha_metrics_choked %>%
  group_by(sample_type) %>%
  sample_n(size = 9)

##See if it worked
sum(bact_random_choked$sample_type == "seawater")
```

```
## [1] 9
```

```r
sum(bact_random_choked$sample_type== "zostera_new")
```

```
## [1] 9
```

```r
sum(bact_random_choked$sample_type == "zostera_old")
```

```
## [1] 9
```

```r
sum(bact_random_choked$sample_type == "artificial")
```

```
## [1] 9
```

```r
############################################################
### RANDOM samples (9 samples for each sample type) ###
############################################################

##############
### Chao1 ###
##############
anova(lm(chao1 ~ sample_type, data=bact_random_choked))
```

```
## Analysis of Variance Table
##
## Response: chao1
##             Df Sum Sq Mean Sq F value       Pr(>F)
## sample_type  3 294275   98092  20.098 0.0000001653 ***
## Residuals   32 156180    4881
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
TukeyHSD(aov(chao1 ~ sample_type, data=bact_random_choked))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = chao1 ~ sample_type, data = bact_random_choked)
##
## $sample_type
##                               diff        lwr        upr     p adj
## seawater-artificial     -198.96089 -288.18836 -109.73342 0.0000056
## zostera_new-artificial  -219.08345 -308.31092 -129.85597 0.0000010
```

```
## zostera_old-artificial   -72.33666 -161.56414   16.89081 0.1459319
## zostera_new-seawater     -20.12256 -109.35003   69.10492 0.9278303
## zostera_old-seawater     126.62423   37.39675  215.85170 0.0028919
## zostera_old-zostera_new  146.74678   57.51931  235.97426 0.0005314
```

```
#########################
### Pielou's Evenness ###
#########################
anova(lm(pielou ~ sample_type, data=bact_random_choked))
```

```
## Analysis of Variance Table
##
## Response: pielou
##             Df  Sum Sq  Mean Sq F value    Pr(>F)
## sample_type  3 0.23242 0.077474  7.8901 0.0004442 ***
## Residuals   32 0.31421 0.009819
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(aov(pielou ~ sample_type, data=bact_random_choked))
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = pielou ~ sample_type, data = bact_random_choked)
##
## $sample_type
##                               diff         lwr          upr     p adj
## seawater-artificial     -0.07896979 -0.20552988  0.047590306 0.3451224
## zostera_new-artificial  -0.21360252 -0.34016261 -0.087042426 0.0003817
## zostera_old-artificial  -0.03984568 -0.16640577  0.086714418 0.8286580
## zostera_new-seawater    -0.13463273 -0.26119283 -0.008072637 0.0336064
## zostera_old-seawater     0.03912411 -0.08743598  0.165684206 0.8361925
## zostera_old-zostera_new  0.17375684  0.04719675  0.300316938 0.0040505
```

new and old growth seagrass leaves

to account for sampling from the same plant

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

### rename metadata columns ###
alpha_new_old <- alpha.div_new_old %>%
  rename(region = allsamples.region,
         sample_type = allsamples.sample_growth,
         individuals = allsamples.individuals)

### remove other sample types ###
remove_artificial_water <- c("artificial", "seawater")
alpha_metrics_new_old <- alpha_new_old %>%
  dplyr::filter(!sample_type %in% remove_artificial_water)
```

```
###############
### ANOVAs ###
###############
### test for heteroscedasticity in the data ###
car::leveneTest(chao1 ~ sample_type, data=alpha_metrics_new_old)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value Pr(>F)
## group   1   0.354 0.5532
##       103
```

```
car::leveneTest(pielou ~ sample_type, data=alpha_metrics_new_old)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##        Df F value       Pr(>F)
## group   1  29.938 0.0000003149 ***
##       103
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
### Because we have heteroscedasticity, we should have the same n so ANOVA is robust ###
sum(alpha_metrics_new_old$sample_type == "zostera_new") #54 samples
```

```
## [1] 54
```

```
sum(alpha_metrics_new_old$sample_type == "zostera_old") #51 samples
```

```
## [1] 51
```

```
### Randomly get 51 samples from zostera so they have same n: ###
bact_random_new_old <- alpha_metrics_new_old %>%
  group_by(sample_type) %>%
  sample_n (size = 51)

##See if it worked
sum(bact_random_new_old$sample_type == "zostera_new") #51 samples
```

```
## [1] 51
```

```
sum(bact_random_new_old$sample_type == "zostera_old") #51 samples
```

```
## [1] 51
```

```
#############################################################
### RANDOM samples (51 samples for each sample type) ###
#############################################################

##############
### Chao1 ###
##############
### with individuals as a random effect ###
anova(lm(chao1 ~ sample_type + individuals, data=bact_random_new_old))
```

```
## Analysis of Variance Table
##
## Response: chao1
##             Df Sum Sq Mean Sq  F value             Pr(>F)
## sample_type  1 878773  878773 106.7761 0.0000000000003162 ***
```

```
## individuals 57 461920    8104   0.9847             0.5268
## Residuals   43 353893    8230
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#########################
### Pielou's Evenness ###
#########################
### with individuals as a random effect ###
anova(lm(pielou ~ sample_type + individuals, data=bact_random_new_old))
```

```
## Analysis of Variance Table
##
## Response: pielou
##             Df  Sum Sq Mean Sq F value           Pr(>F)
## sample_type  1 0.67005 0.67005 75.1636 0.00000000005454 ***
## individuals 57 0.61971 0.01087  1.2196             0.25
## Residuals   43 0.38333 0.00891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```