

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

NMDS and PERMANOVAs comparison across regions for each sample type

```
#### load packages ####
library(QsRutils)
library(EcolUtils)
library(phyloseq)
library(vegan)
library(ggplot2)
library(dplyr)

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

## 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 ]

### 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 = phy_tree(phylo_merge_rare)
```

New growth leaves

```
#####
### NMDS plot ###
#####
```

```

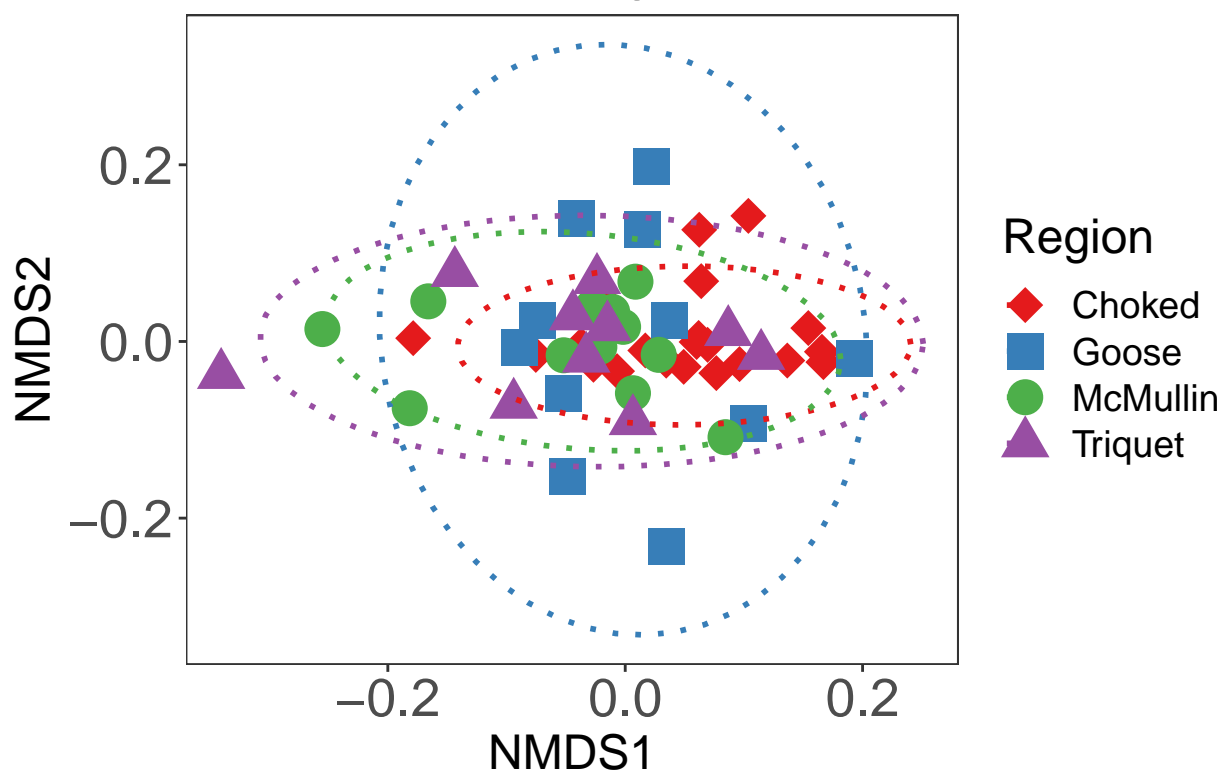
### Select new zostera growth from phyloseq object ###
phylo_merge_new <- subset_samples(phylo_merge_rare,
                                  sample_growth %in% c("zostera_new"))

### 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_new_regions

```

New Zostera (Weighted Unifrac)



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

### get metadata only ###
metadata <- allsamples[,1:4]

### Select new growth seagrass leaves ###
new <- c("zostera_new")
new_only <- allsamples %>%
  dplyr::filter(sample_growth %in% new)

### Get metadata with new growth seagrass leaves ###
metadata_new <- metadata %>%
  dplyr::filter(sample_growth %in% new)

### Calculate weighted Unifrac, Bray-Curtis and Jaccard dissimilarity ###
sampledist_new_bray <- phyloseq::distance(phylo_merge_new, method = "bray")
sampledist_new_UNI <- phyloseq::distance(phylo_merge_new, method = "wunifrac")

### Convert distances from phyloseq objects to matrices ###
bray_new <- as.matrix(sampledist_new_bray)
UNI_new <- as.matrix(sampledist_new_UNI)
dist_bray_new_mtx <- as.dist(as(bray_new, "matrix"))
dist_UNI_new_mtx <- as.dist(as(UNI_new, "matrix"))
```

```

### test for heteroscedasticity in the data ###
homogeneity <-betadisper(dist_bray_new_mtx ,metadata_new$region)
permutest(homogeneity, pairwise = TRUE)

##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 999
##
## Response: Distances
##      Df Sum Sq Mean Sq      F N.Perm Pr(>F)
## Groups   3 0.02088 0.0069600 0.8269   999  0.486
## Residuals 50 0.42086 0.0084172
##
## Pairwise comparisons:
## (Observed p-value below diagonal, permuted p-value above diagonal)
##      choked   goose mcmullin triquet
## choked           0.42700 0.50000 0.504
## goose    0.43390           0.12000 0.139
## mcmullin 0.51313 0.12558           0.949
## triquet  0.51442 0.14655 0.95264

#boxplot(homogeneity)

homogeneity <-betadisper(dist_UNI_new_mtx ,metadata_new$region)
permutest(homogeneity, pairwise = TRUE)

##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 999
##
## Response: Distances
##      Df Sum Sq Mean Sq      F N.Perm Pr(>F)
## Groups   3 0.015945 0.0053150 1.2776   999  0.309
## Residuals 50 0.208014 0.0041603
##
## Pairwise comparisons:
## (Observed p-value below diagonal, permuted p-value above diagonal)
##      choked   goose mcmullin triquet
## choked           0.024000 0.810000 0.410
## goose    0.031146           0.145000 0.422
## mcmullin 0.819625 0.150178           0.682
## triquet  0.421191 0.407157 0.656277

#boxplot(homogeneity)

### Data is homogeneous, so can continue with analyses ###

#####
#### PERMANOVA Bray Curtis ####
#####
adonis2(dist_bray_new_mtx ~ metadata_new$region, data= metadata_new,
        permutations=10000, by = "margin")

```

```

## Permutation test for adonis under NA model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 10000
##
## adonis2(formula = dist_bray_new_mtx ~ metadata_new$region, data = metadata_new, permutations = 10000)
##              Df SumOfSqs      R2      F    Pr(>F)
## metadata_new$region  3    2.5508 0.14296 2.78 0.00009999 ***
## Residual            50   15.2926 0.85704
## Total               53   17.8434 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# pairwise comparisons
set.seed(986)
adonis.pair(dist_bray_new_mtx, metadata_new$region,
            nper = 10000, corr.method = "BH")

##           combination SumsOfSqs  MeanSqs  F.Model      R2    P.value
## 1      choked <-> goose 0.8437890 0.8437890 2.651322 0.08120105 0.00099990
## 2      choked <-> mcmullin 1.3045788 1.3045788 4.386134 0.12395063 0.00009999
## 3      choked <-> triquet 0.7518639 0.7518639 2.511749 0.07970833 0.00299970
## 4      goose <-> mcmullin 0.8258839 0.8258839 2.623124 0.11104051 0.00029997
## 5      goose <-> triquet 0.4262319 0.4262319 1.333682 0.06558981 0.12598740
## 6      mcmullin <-> triquet 0.7471580 0.7471580 2.601047 0.11508525 0.00049995
##   P.value.corrected
## 1           0.00149985
## 2           0.00059994
## 3           0.00359964
## 4           0.00089991
## 5           0.12598740
## 6           0.00099990

#####
#### PERMANOVA weighted Unifrac ####
#####
adonis2(dist_UNI_new_mtx ~ metadata_new$region, data = metadata_new, permutations = 10000, by = "margin")

## Permutation test for adonis under NA model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 10000
##
## adonis2(formula = dist_UNI_new_mtx ~ metadata_new$region, data = metadata_new, permutations = 10000,
##              Df SumOfSqs      R2      F Pr(>F)
## metadata_new$region  3  0.24843 0.12376 2.3539 0.0018 **
## Residual            50   1.75900 0.87624
## Total               53   2.00743 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# pairwise comparisons
set.seed(986)
adonis.pair(dist_UNI_new_mtx, metadata_new$region,
            nper = 10000, corr.method = "BH")

```

```
##           combination SumsOfSqs   MeanSqs   F.Model      R2      P.value
## 1   choked <->  goose 0.09957521 0.09957521 2.8977982 0.08808487 0.00989901
## 2   choked <-> mcmullin 0.12270103 0.12270103 4.0769070 0.11622767 0.00349965
## 3   choked <-> triquet 0.08586388 0.08586388 2.6751737 0.08445648 0.02779722
## 4   goose <-> mcmullin 0.06875623 0.06875623 1.7434016 0.07665527 0.09209079
## 5   goose <-> triquet 0.06464573 0.06464573 1.4870023 0.07258272 0.16058394
## 6 mcmullin <-> triquet 0.02632438 0.02632438 0.7230714 0.03489210 0.61833817
##   P.value.corrected
## 1           0.02969703
## 2           0.02099790
## 3           0.05559444
## 4           0.13813619
## 5           0.19270073
## 6           0.61833817
```

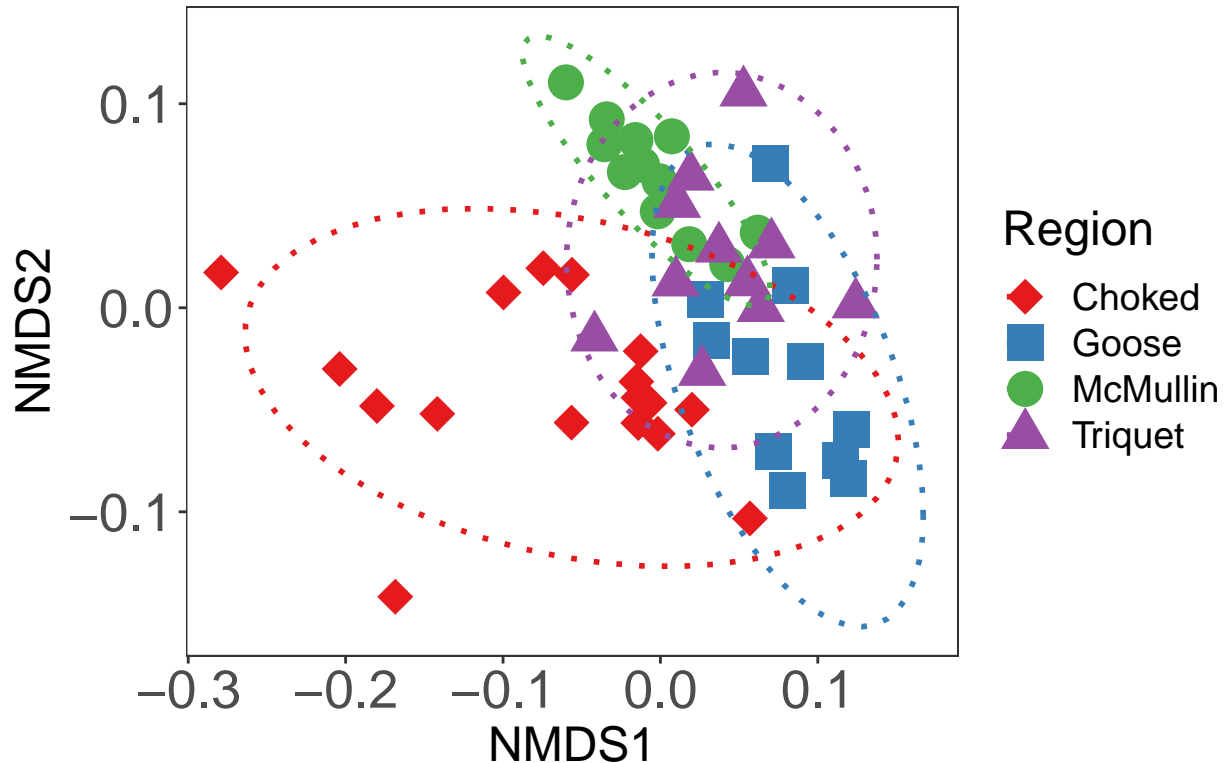
Old growth leaves

```
#####
### NMDS plot ###
#####
### 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_old_regions
```

Old Zostera (Weighted Unifrac)



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

### get metadata only ###
metadata <- allsamples[,1:4]

### Select new growth seagrass leaves ###
old <- c("zostera_old")
old_only <- allsamples %>%
  dplyr::filter(sample_growth %in% old)

### Get metadata with new growth seagrass leaves ###
metadata_old <- metadata %>%
  dplyr::filter(sample_growth %in% old)

### Calculate weighted Unifrac, Bray-Curtis and Jaccard dissimilarity ###
sampledist_old_bray <- phyloseq::distance(phylo_merge_old, method = "bray")
sampledist_old_UNI <- phyloseq::distance(phylo_merge_old, method = "wunifrac")

### Convert distances from phyloseq objects to matrices ###
bray_old <- as.matrix(sampledist_old_bray)
UNI_old <- as.matrix(sampledist_old_UNI)
dist_bray_old_mtx <- as.dist(as(bray_old, "matrix"))
dist_UNI_old_mtx <- as.dist(as(UNI_old, "matrix"))
```

```

### test for heteroscedasticity in the data ###
homogeneity <-betadisper(dist_bray_old_mtx ,metadata_old$region)
permutest(homogeneity, pairwise = TRUE)

##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 999
##
## Response: Distances
##      Df  Sum Sq  Mean Sq      F N.Perm Pr(>F)
## Groups   3 0.15569 0.051896 5.2825   999 0.007 **
## Residuals 47 0.46173 0.009824
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pairwise comparisons:
## (Observed p-value below diagonal, permuted p-value above diagonal)
##      choked      goose mcmullin triquet
## choked           0.0890000 0.0010000   0.174
## goose    0.0963699           0.0440000   0.595
## mcmullin 0.0016033 0.0555782           0.009
## triquet  0.1759778 0.6088588 0.0102831

#boxplot(homogeneity)

homogeneity <-betadisper(dist_UNI_old_mtx ,metadata_old$region)
permutest(homogeneity, pairwise = TRUE)

##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 999
##
## Response: Distances
##      Df  Sum Sq  Mean Sq      F N.Perm Pr(>F)
## Groups   3 0.027322 0.0091074 5.8226   999 0.001 ***
## Residuals 47 0.073515 0.0015642
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pairwise comparisons:
## (Observed p-value below diagonal, permuted p-value above diagonal)
##      choked      goose mcmullin triquet
## choked           0.0350000 0.0010000   0.110
## goose    0.0340915           0.0780000   0.424
## mcmullin 0.0019648 0.0861416           0.017
## triquet  0.1104201 0.4149748 0.0233036

#boxplot(homogeneity)

### Because we have heteroscedasticity, we should have the same n so PERMANOVA is robust ###
# GET LIST OF RANDOMLY SUBSET OF DATA FOR PHYLOSEQ FILTERING / PERMANOVA #
sum(old_only$region == "choked") # 17samples

```



```

## [1] 17
sum(old_only$region == "triquet") # 11samples

## [1] 11
sum(old_only$region == "mcmullin") # 12samples

## [1] 12
sum(old_only$region == "goose") # 11samples

## [1] 11
### Randomly get 11 samples from each region they have same n: ###
bact_random_old <- old_only %>% group_by(region) %>% sample_n(size = 11)

## See if it worked ###
sum(bact_random_old$region == "choked")

## [1] 11
sum(bact_random_old$region == "triquet")

## [1] 11
sum(bact_random_old$region == "mcmullin")

## [1] 11
sum(bact_random_old$region == "goose")

## [1] 11
### list of randomly subset samples ###
bact_random_list_old <- bact_random_old$SampleID
length(bact_random_list_old)

## [1] 44
### Now use the list above to subsample data in phyloseq object ###
phylo_random_old <- subset_samples(phylo_merge_old,
                                   sample_names(phylo_merge_old) %in%
                                   bact_random_list_old)

### Calculate weighted Unifrac, Bray-Curtis dissimilarity again, but for random samples ###
sampledist_random_old_bray <- phyloseq::distance(phylo_random_old, method = "bray")
sampledist_random_old_UNI <- phyloseq::distance(phylo_random_old, method = "wunifrac")

### Convert distances from phyloseq objects to matrices ###
bray_random_old <- as.matrix(sampledist_random_old_bray)
UNI_random_old <- as.matrix(sampledist_random_old_UNI)
dist_bray_random_old_mtx <- as.dist(as(bray_random_old , "matrix"))
dist_UNI_random_old_mtx <- as.dist(as(UNI_random_old, "matrix"))

#### get list of subsamples from metadata ###
old_subsample <- metadata_old %>%
  filter(SampleID %in% bact_random_list_old)
old_subsample$region <- factor(old_subsample$region)

```

```
#####
### RANDOM samples (11 samples for each region) ###
#####

#####
#### PERMANOVA Bray Curtis ####
#####
adonis2(dist_bray_random_old_mtx ~ region,
        data= old_subsample, permutations=10000, by = "margin")

## Permutation test for adonis under NA model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 10000
##
## adonis2(formula = dist_bray_random_old_mtx ~ region, data = old_subsample, permutations = 10000, by =
##           Df SumOfSqs      R2      F      Pr(>F)
## region      3   3.6410 0.30594 5.8774 0.00009999 ***
## Residual    40   8.2599 0.69406
## Total       43  11.9008 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# pairwise comparisons
set.seed(986)
adonis.pair(dist_bray_random_old_mtx, old_subsample$region,
            nper = 10000, corr.method = "BH")

##           combination SumsOfSqs  MeanSqs F.Model      R2    P.value
## 1      choked <-> goose 1.7600977 1.7600977 7.465745 0.2718202 0.00009999
## 2      choked <-> mcmullin 1.5703717 1.5703717 7.608675 0.2755900 0.00009999
## 3      choked <-> triquet 1.3751176 1.3751176 5.656919 0.2204832 0.00009999
## 4      goose <-> mcmullin 1.2727736 1.2727736 7.491020 0.2724897 0.00009999
## 5      goose <-> triquet 0.7360495 0.7360495 3.562674 0.1511999 0.00009999
## 6      mcmullin <-> triquet 0.5675311 0.5675311 3.202120 0.1380098 0.00019998
##      P.value.corrected
## 1              0.000119988
## 2              0.000119988
## 3              0.000119988
## 4              0.000119988
## 5              0.000119988
## 6              0.000199980

#####
#### PERMANOVA Weighted Unifrac ####
#####
adonis2(dist_UNI_random_old_mtx ~ region,
        data= old_subsample, permutations=10000, by = "margin")

## Permutation test for adonis under NA model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 10000
##
## adonis2(formula = dist_UNI_random_old_mtx ~ region, data = old_subsample, permutations = 10000, by =
```

```
##           Df SumOfSqs      R2      F      Pr(>F)
## region    3  0.40662 0.3868 8.4107 0.00009999 ***
## Residual 40  0.64461 0.6132
## Total    43  1.05123 1.0000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# pairwise comparisons
set.seed(986)
adonis.pair(dist_UNI_random_old_mtx, old_subsample$region,
            nper = 10000, corr.method = "BH")

##           combination SumsOfSqs      MeanSqs      F.Model      R2      P.value
## 1      choked <-> goose 0.18247777 0.18247777  9.422415 0.3202461 0.00009999
## 2      choked <-> mcmullin 0.14186377 0.14186377  8.156100 0.2896744 0.00009999
## 3      choked <-> triquet 0.12882842 0.12882842  6.215170 0.2370830 0.00009999
## 4      goose <-> mcmullin 0.20519425 0.20519425 17.839426 0.4714507 0.00009999
## 5      goose <-> triquet 0.08087609 0.08087609  5.451056 0.2141780 0.00009999
## 6      mcmullin <-> triquet 0.07399767 0.07399767  5.752305 0.2233705 0.00039996
##      P.value.corrected
## 1              0.000119988
## 2              0.000119988
## 3              0.000119988
## 4              0.000119988
## 5              0.000119988
## 6              0.000399960
```

Seawater

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

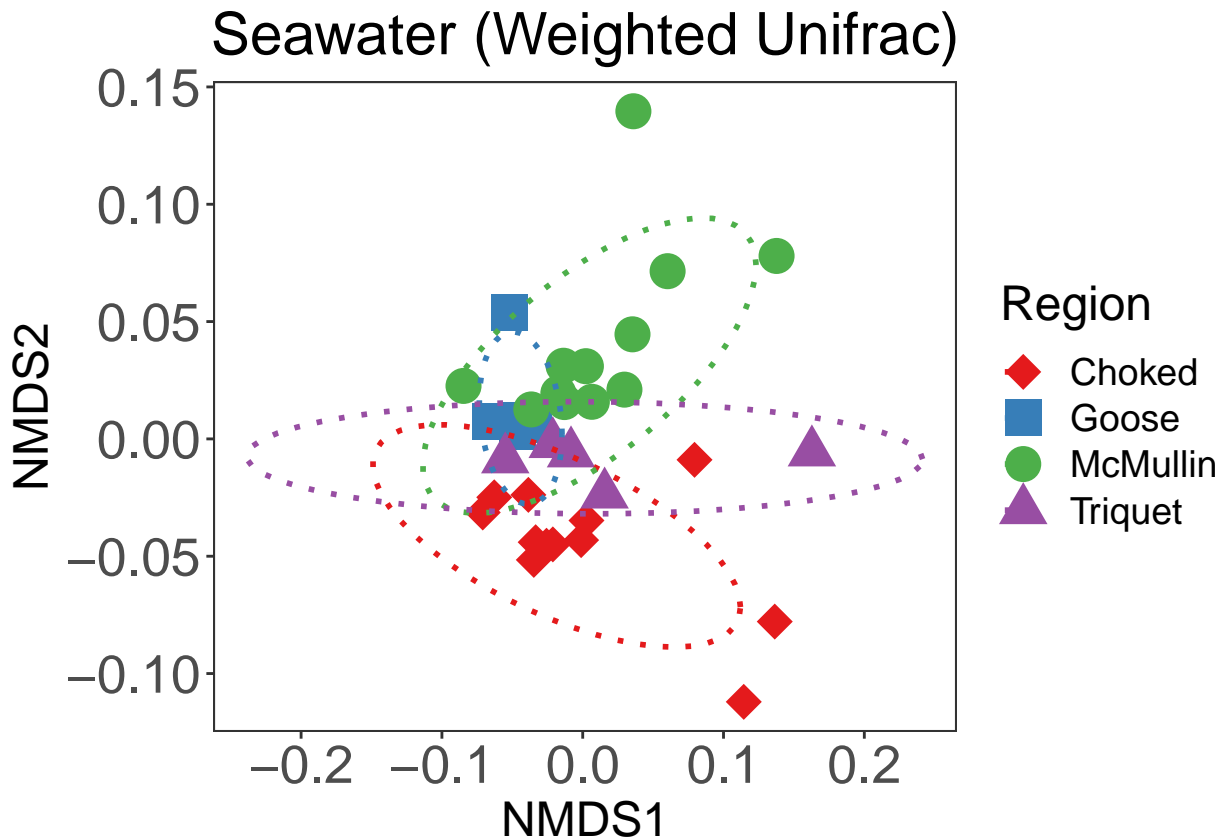
### ordinate using weighted Unifrac, Bray-Curtis dissimilarity ###
NMDS_bray_water <- ordinate(phylo_merge_water, "NMDS", "bray")
NMDS_UNI_water <- ordinate(phylo_merge_water, "NMDS", "wunifrac")

### 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()
NMDS_water_regions

```



```

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

### get metadata only ###
metadata <- allsamples[,1:4]

### Select new growth seagrass leaves ###
water <- c("seawater")
water_only <- allsamples %>%
  dplyr::filter(sample_growth %in% water)

### Get metadata with new growth seagrass leaves ###
metadata_water <- metadata %>%
  dplyr::filter(sample_growth %in% water)

### Calculate weighted Unifrac, Bray-Curtis and Jaccard dissimilarity ###
sampledist_water_bray <- phyloseq::distance(phylo_merge_water, method = "bray")
sampledist_water_UNI <- phyloseq::distance(phylo_merge_water, method = "wunifrac")

```

```

### Convert distances from phyloseq objects to matrices ###
bray_water <- as.matrix(sampled_dist_water_bray)
UNI_water <- as.matrix(sampled_dist_water_UNI)
dist_bray_water_mtx <- as.dist(as(bray_water, "matrix"))
dist_UNI_water_mtx <- as.dist(as(UNI_water, "matrix"))

### test for heteroscedasticity in the data ###
homogeneity <- betadisper(dist_bray_water_mtx, metadata_water$region)
permutest(homogeneity, pairwise = TRUE)

##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 999
##
## Response: Distances
##           Df Sum Sq Mean Sq      F N.Perm Pr(>F)
## Groups      3 0.09729 0.032429 2.529   999 0.078 .
## Residuals 31 0.39750 0.012823
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pairwise comparisons:
## (Observed p-value below diagonal, permuted p-value above diagonal)
##           choked      goose mcmullin triquet
## choked              0.045000 0.387000   0.705
## goose      0.042702              0.096000   0.038
## mcmullin 0.363810 0.088893              0.200
## triquet  0.693658 0.036533 0.208263

#boxboxplot(homogeneity)

homogeneity <- betadisper(dist_UNI_water_mtx, metadata_water$region)
permutest(homogeneity, pairwise = TRUE)

##
## Permutation test for homogeneity of multivariate dispersions
## Permutation: free
## Number of permutations: 999
##
## Response: Distances
##           Df Sum Sq Mean Sq      F N.Perm Pr(>F)
## Groups      3 0.009578 0.0031926 1.6253   999 0.201
## Residuals 31 0.060894 0.0019643
##
## Pairwise comparisons:
## (Observed p-value below diagonal, permuted p-value above diagonal)
##           choked      goose mcmullin triquet
## choked              0.052000 0.797000   0.851
## goose      0.046143              0.042000   0.105
## mcmullin 0.792294 0.047038              0.970
## triquet  0.881706 0.108158 0.967393

#boxboxplot(homogeneity)

```

```

### Because we have heteroscedasticity, we should have the same n so PERMANOVA is robust ###
# GET LIST OF RANDOMLY SUBSET OF DATA FOR PHYLOSEQ FILTERING / PERMANOVA #
sum(water_only$region == "choked")

## [1] 12

sum(water_only$region == "triquet")

## [1] 5

sum(water_only$region == "mcmullin")

## [1] 12

sum(water_only$region == "goose")

## [1] 6

### Randomly get 5 samples from each region they have same n: ###
bact_random_water <- water_only %>% group_by(region) %>% sample_n(size = 5)

## See if it worked ###
sum(bact_random_water$region == "choked")

## [1] 5

sum(bact_random_water$region == "triquet")

## [1] 5

sum(bact_random_water$region == "mcmullin")

## [1] 5

sum(bact_random_water$region == "goose")

## [1] 5

### list of randomly subset samples ###
bact_random_list_water <- bact_random_water$SampleID
length(bact_random_list_water)

## [1] 20

### Now use the list above to subsample data in phyloseq object ###
phylo_random_water <- subset_samples(phylo_merge_water , sample_names(phylo_merge_water) %in% bact_random_list_water)

### Calculate weighted Unifrac, Bray-Curtis dissimilarity again, but for random samples ###
sampledist_random_water_bray <- phyloseq::distance(phylo_random_water, method = "bray")
sampledist_random_water_UNI <- phyloseq::distance(phylo_random_water, method = "wunifrac")

### Convert distances from phyloseq objects to matrices ###
bray_random_water <- as.matrix(sampledist_random_water_bray)
UNI_random_water <- as.matrix(sampledist_random_water_UNI)
dist_bray_random_water_mtx <- as.dist(as(bray_random_water , "matrix"))
dist_UNI_random_water_mtx <- as.dist(as(UNI_random_water, "matrix"))

#### get list of subsamples from metadata ###
water_subsample <- metadata_water %>%
  filter(SampleID %in% bact_random_list_water)

```

```

water_subsample$region <- factor(water_subsample$region)

#####
### RANDOM samples (5 samples for each region) ###
#####

#####
#### PERMANOVA Bray Curtis ####
#####
adonis2(dist_bray_random_water_mtx ~ water_subsample$region, data= water_subsample, permutations=10000)

## Permutation test for adonis under NA model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 10000
##
## adonis2(formula = dist_bray_random_water_mtx ~ water_subsample$region, data = water_subsample, permut
##
##          Df SumOfSqs      R2      F    Pr(>F)
## water_subsample$region 3  0.97583 0.32564 2.5755 0.007699 **
## Residual              16  2.02077 0.67436
## Total                  19  2.99660 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# pairwise comparisons
set.seed(986)
adonis.pair(dist_bray_random_water_mtx, water_subsample$region,
            nper = 10000, corr.method = "BH")

##          combination SumsOfSqs  MeanSqs  F.Model      R2    P.value
## 1      choked <-> goose 0.4111866 0.4111866 4.0481615 0.3359983 0.00759924
## 2      choked <-> mcmullin 0.4026122 0.4026122 2.7024678 0.2525088 0.07859214
## 3      choked <-> triquet 0.1597679 0.1597679 0.9240996 0.1035510 0.35466453
## 4      goose <-> mcmullin 0.4370177 0.4370177 5.4828387 0.4066531 0.02249775
## 5      goose <-> triquet 0.3072318 0.3072318 2.9650606 0.2704099 0.01589841
## 6      mcmullin <-> triquet 0.2338358 0.2338358 1.5483441 0.1621584 0.15628437
##      P.value.corrected
## 1              0.0449955
## 2              0.1178882
## 3              0.3546645
## 4              0.0449955
## 5              0.0449955
## 6              0.1875412

#####
#### PERMANOVA weighted Unifrac ####
#####
adonis2(dist_UNI_random_water_mtx ~ water_subsample$region, data= water_subsample, permutations=10000,

## Permutation test for adonis under NA model
## Marginal effects of terms
## Permutation: free
## Number of permutations: 10000
##
## adonis2(formula = dist_UNI_random_water_mtx ~ water_subsample$region, data = water_subsample, permut

```

```
##              Df SumOfSqs    R2      F Pr(>F)
## water_subsample$region  3 0.057489 0.342 2.772 0.0023 **
## Residual              16 0.110611 0.658
## Total                  19 0.168100 1.000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# pairwise comparisons
set.seed(986)
adonis.pair(dist_UNI_random_water_mtx, water_subsample$region,
            nper = 10000, corr.method = "BH")

##      combination SumsOfSqs    MeanSqs F.Model      R2    P.value
## 1      choked <-> goose 0.02694189 0.02694189 4.221451 0.3454132 0.00759924
## 2      choked <-> mcmullin 0.02496175 0.02496175 2.925743 0.2677844 0.01629837
## 3      choked <-> triquet 0.01099018 0.01099018 1.102659 0.1211360 0.33656634
## 4      goose <-> mcmullin 0.02095646 0.02095646 5.430004 0.4043189 0.00729927
## 5      goose <-> triquet 0.01722530 0.01722530 3.253376 0.2891022 0.02519748
## 6      mcmullin <-> triquet 0.01390337 0.01390337 1.867675 0.1892720 0.08389161
##      P.value.corrected
## 1      0.02279772
## 2      0.03259674
## 3      0.33656634
## 4      0.02279772
## 5      0.03779622
## 6      0.10066993
```