Bird Park Environmental Microbiomes

Dustin Foote, Chris Balakrishnan, Ariane Peralta

Last updated on 02 April, 2023

Project Description: Fill out

Initial Setup

#Import Files ## Environmental Data

Microbial Data

Diversity Metrics - Hypothesis Testing

```
# Rarefy Abundances (min abundance is 8106. We are sampling to 8000)
min(rowSums(otus))
```

```
## [1] 8106
```

```
max(rowSums(otus))
```

```
## [1] 37883
```

```
mean(rowSums(otus))
```

```
## [1] 19232.32
```

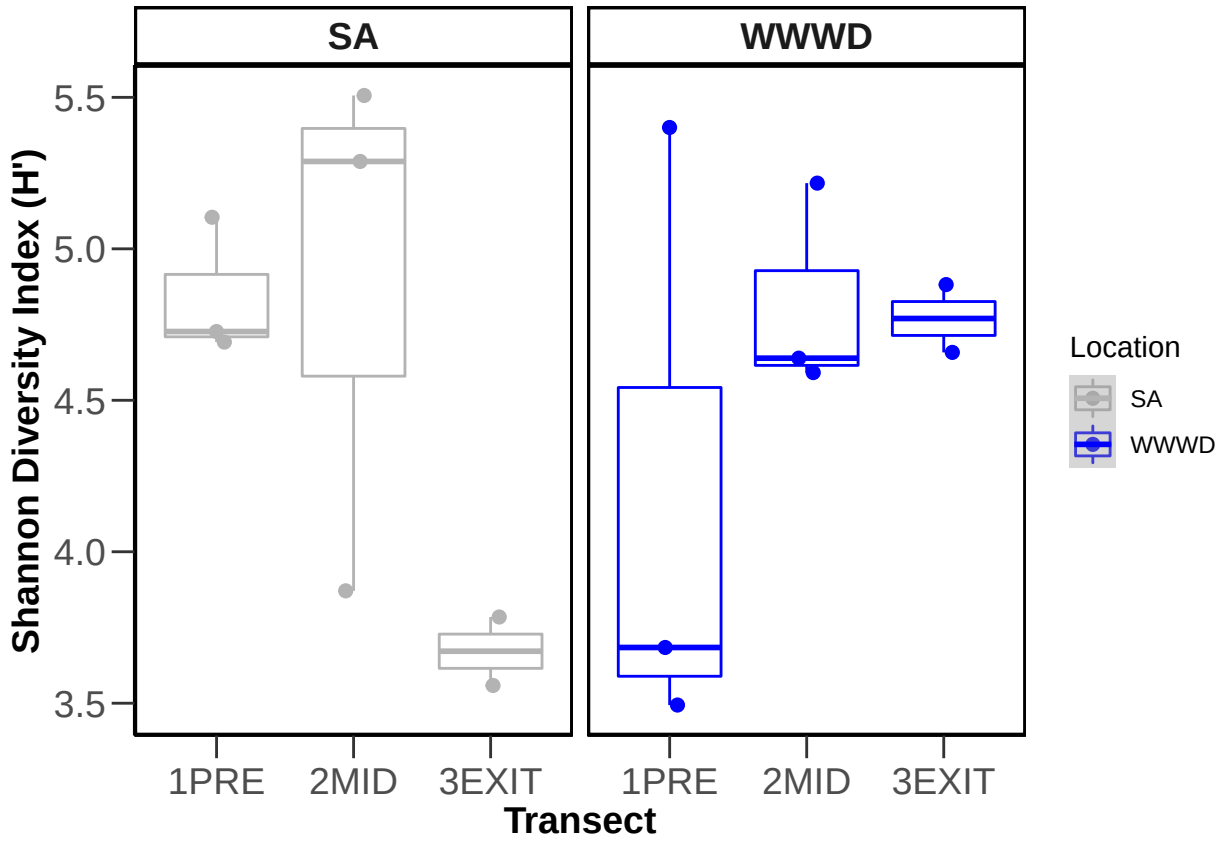
```
SH.r <- rrarefy(otus, 8000)

# Fisher's Alpha
fisher <- fisher.alpha(SH.r)

# Species Richness
richness <- rowSums((SH.r >= 1))

# Shannon Diversity
shannon <- diversity(SH.r, "shannon")

# Simpson's Evenness
```

```r
simp.even <- apply(SH.r, 1, simp_even)

#Pielou's evenness
J <- shannon/log(specnumber(SH.r[,-c(1:1)]))

#combined richness, diversity, evenness
diversity <- cbind(design,richness,shannon,simp.even,J)

diversity$Transect <- as.factor(diversity$Transect)
diversity$Location <- as.factor(diversity$Location)
diversity$Location_ID <- as.factor(diversity$Location_ID)
diversity$Location_ID_order <- as.factor(diversity$Location_ID_order)
str(diversity)
```

```
## 'data.frame':    19 obs. of  10 variables:
##  $ Location         : Factor w/ 3 levels "SA","SOURCE",..: 3 3 3 3 3 3 3 3 1 1 ...
##  $ Transect         : Factor w/ 4 levels "1PRE","2MID",..: 1 1 1 2 2 2 3 3 1 1 ...
##  $ Location_ID      : Factor w/ 7 levels "SA","SA_EXIT",..: 7 7 7 5 5 5 6 6 3 3 ...
##  $ Location_ID_order: Factor w/ 7 levels "1SA_PRE","1WWWD_PRE",..: 2 2 2 4 4 4 6 6 1 1 ...
##  $ Description      : chr  "EMU" "EMU" "EMU" "INFLOW" ...
##  $ Rep              : int  1 2 3 1 2 3 1 2 1 2 ...
##  $ richness         : num  1035 649 561 738 905 ...
##  $ shannon          : num  5.4 3.68 3.49 4.59 5.22 ...
##  $ simp.even        : num  0.0679 0.0134 0.0122 0.0535 0.0704 ...
##  $ J                : num  0.778 0.569 0.552 0.695 0.766 ...
```

```r
diversity.nosource <- diversity[c(1:16),]

# Graphing Shannon Diversity
p <- ggplot(diversity.nosource, aes(x=Transect, y=shannon, color=Location))+ geom_boxplot() +
      geom_point(aes(color=Location), size=2, position = position_jitterdodge())+ scale_color_manual(na
p1=p+geom_smooth(method="lm")+facet_wrap(~Location)
shannon<-p1 + theme_bw() +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(), axis.line
          =element_line(colour = "black")) +
    theme(axis.title=element_text(vjust=1,size=14,face="bold"),
          axis.text=element_text(size=14), axis.text.x = element_text(vjust=0.65, hjust=0.5,
          size=14), panel.border = element_rect(colour = "black",size=1)) +
    theme(axis.ticks.length=unit(0.3,"cm")) + labs(x = "Transect", y = "Shannon Diversity Index (H')") +
    theme(strip.text.x = element_text(size=14, face="bold"), strip.text.y =
          element_text(size=14, face="bold"), strip.background = element_rect(colour="black",
          fill="white", size=1))
shannon
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggsave("../figures/SH.bacteria.shannon.png", plot=last_plot(), device=NULL, path=NULL, scale=1, width=7
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
# shannon anova
shannon.lm <- lm(shannon ~ Location*Transect, data = diversity.nosource)
summary(shannon.lm) #NS
```

```
##
## Call:
## lm(formula = shannon ~ Location * Transect, data = diversity.nosource)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0176 -0.1887 -0.1125  0.2973  1.2077
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  4.84110    0.37329  12.969  1.4e-07 ***
## LocationWWWD                -0.64827    0.52791  -1.228   0.2476
## Transect2MID                 0.04744    0.52791   0.090   0.9302
## Transect3EXIT               -1.16947    0.59023  -1.981   0.0757 .
## LocationWWWD:Transect2MID    0.57548    0.74658   0.771   0.4586
## LocationWWWD:Transect3EXIT   1.74666    0.83471   2.093   0.0629 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6466 on 10 degrees of freedom
## Multiple R-squared:  0.4034, Adjusted R-squared:  0.1051
## F-statistic: 1.352 on 5 and 10 DF,  p-value: 0.3193
```

```
require("emmeans")
```

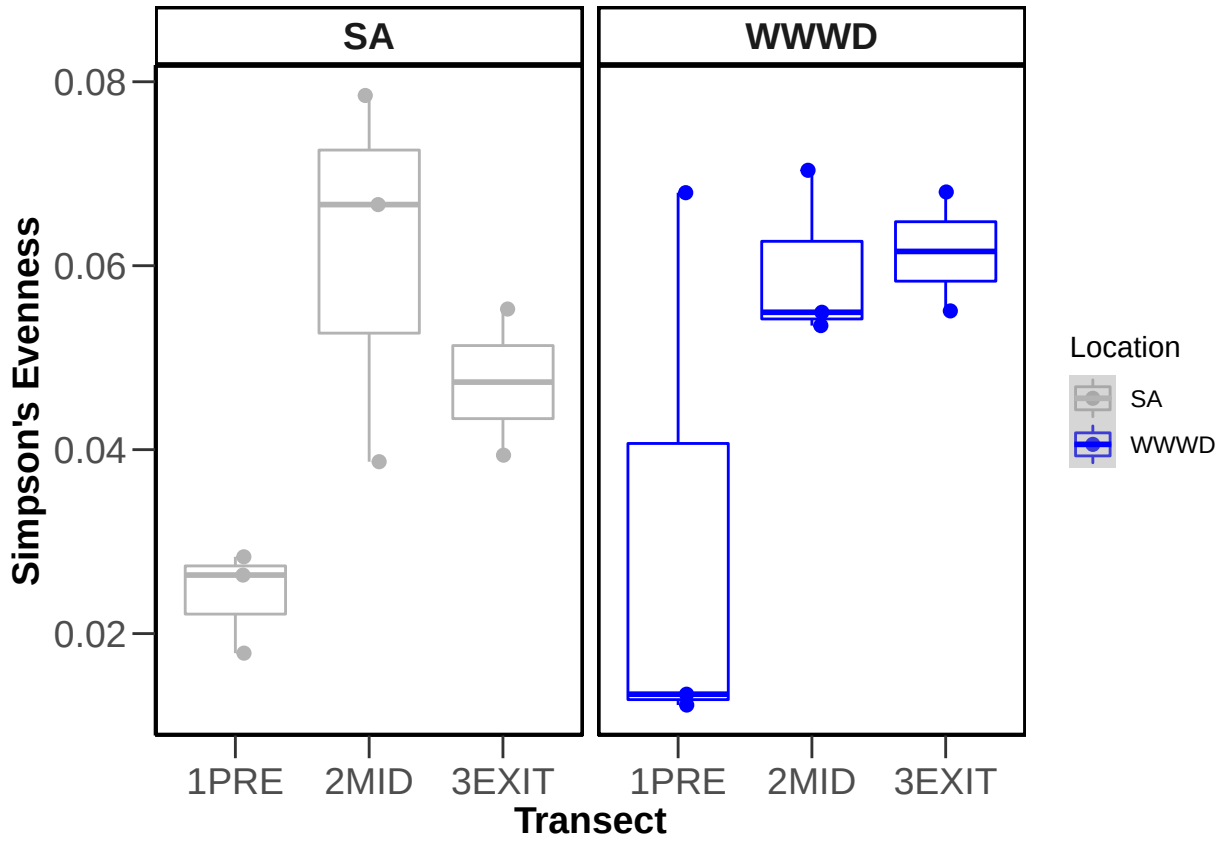
```
## Loading required package: emmeans
```

```
anova(shannon.lm)
```

```
## Analysis of Variance Table
##
## Response: shannon
##                   Df Sum Sq Mean Sq F value Pr(>F)
## Location           1 0.0001 0.00007  0.0002 0.9899
## Transect           2 0.9822 0.49110  1.1748 0.3481
## Location:Transect  2 1.8447 0.92236  2.2064 0.1608
## Residuals         10 4.1804 0.41804
```

```
# Graphing Evenness
p <- ggplot(diversity.nosource, aes(x=Transect, y=simp.even, color=Location))+ geom_boxplot() +
      geom_point(aes(color=Location), size=2, position = position_jitterdodge())+ scale_color_manual(na
p1=p+geom_smooth(method="lm")+facet_wrap(~Location)
even<-p1 + theme_bw() +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(), axis.line
          =element_line(colour = "black")) +
    theme(axis.title=element_text(vjust=1,size=14,face="bold"),
          axis.text=element_text(size=14), axis.text.x = element_text(vjust=0.65, hjust=0.5,
          size=14), panel.border = element_rect(colour = "black",size=1)) +
    theme(axis.ticks.length=unit(0.3,"cm")) + labs(x = "Transect", y = "Simpson's Evenness") +
    theme(strip.text.x = element_text(size=14, face="bold"), strip.text.y =
          element_text(size=14, face="bold"), strip.background = element_rect(colour="black",
          fill="white", size=1))
even
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggsave("../figures/SH.bacteria.evenness.png", plot=last_plot(), device=NULL, path=NULL, scale=1, width="
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
# evenness anova
even.lm <- lm(simp.even ~ Location*Transect, data = diversity.nosource)
summary(even.lm) #NS
```

```
##
## Call:
## lm(formula = simp.even ~ Location * Transect, data = diversity.nosource)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.022591 -0.006835 -0.001248  0.006835  0.036745
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                  0.024198   0.010502   2.304   0.0439 *
## LocationWWWD                 0.006997   0.014852   0.471   0.6477
## Transect2MID                 0.037083   0.014852   2.497   0.0316 *
## Transect3EXIT                0.023147   0.016605   1.394   0.1935
## LocationWWWD:Transect2MID   -0.008678   0.021004  -0.413   0.6882
## LocationWWWD:Transect3EXIT   0.007210   0.023483   0.307   0.7651
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01819 on 10 degrees of freedom
## Multiple R-squared:  0.5358, Adjusted R-squared:  0.3037
## F-statistic: 2.309 on 5 and 10 DF,  p-value: 0.1219
```

```
anova(even.lm)
```

```
## Analysis of Variance Table
##
## Response: simp.even
##                   Df    Sum Sq    Mean Sq F value  Pr(>F)
## Location           1 0.0001230 0.00012298  0.3717 0.55569
## Transect           2 0.0035398 0.00176988  5.3490 0.02632 *
## Location:Transect  2 0.0001565 0.00007826  0.2365 0.79367
## Residuals         10 0.0033088 0.00033088
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
emmeans(even.lm, pairwise ~ Transect)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
## $emmeans
##  Transect emmean      SE df lower.CL upper.CL
##  1PRE     0.0277 0.00743 10   0.0111   0.0442
##  2MID     0.0604 0.00743 10   0.0439   0.0770
##  3EXIT    0.0544 0.00910 10   0.0342   0.0747
##
## Results are averaged over the levels of: Location
## Confidence level used: 0.95
##
## $contrasts
##  contrast     estimate     SE df t.ratio p.value
##  1PRE - 2MID  -0.03274 0.0105 10  -3.118  0.0269
##  1PRE - 3EXIT -0.02675 0.0117 10  -2.278  0.1056
##  2MID - 3EXIT  0.00599 0.0117 10   0.510  0.8681
##
## Results are averaged over the levels of: Location
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Simple Hypothesis Testing - Microbes

```
#PERMANOVA
new.data <-cbind(design,dataREL)
adonis = adonis2(new.data[,-c(1:7)]~Transect*Location, method = "bray", data = new.data, perm=1000)
adonis
```

```
## Permutation test for adonis under reduced model
```

```
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 1000
##
## adonis2(formula = new.data[, -c(1:7)] ~ Transect * Location, data = new.data, permutations = 1000, m
##                   Df SumOfSqs      R2      F   Pr(>F)
## Transect           3   1.2841 0.25434 3.0966 0.001998 **
## Location           1   0.9296 0.18411 6.7249 0.000999 ***
## Transect:Location  2   1.1765 0.23302 4.2556 0.000999 ***
## Residual          12   1.6587 0.32853
## Total             18   5.0488 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Microbial Ordinations

Principal Coordinates Ordination

```
# Principal Coordinates Analysis
dataREL.dist <- vegdist(dataREL, method="bray")

pcoa <- cmdscale(dataREL.dist, k=3, eig=TRUE, add=FALSE)
  # Classical (Metric) Multidimensional Scaling; returns PCoA coordinates
  # eig=TRUE returns eigenvalues; k = # of dimensions to calculate

explainvar1b <- round(pcoa$eig[1] / sum(pcoa$eig), 3) * 100
explainvar2b <- round(pcoa$eig[2] / sum(pcoa$eig), 3) * 100
sum.eigb <- sum(explainvar1b, explainvar2b)

explainvar1b #30.2
```

```
## [1] 30.2
```

```
explainvar2b #21.6
```

```
## [1] 21.6
```

```
pcoa.groups <- paste(new.data$Location, new.data$Transect, sep = "_")
pcoa.points <- data.frame(pcoa$points, group = pcoa.groups)

# Calculate Centroids (mean and SE)
pcoa.L.centroids <- melt(pcoa.points, id="group", measure.vars = c("X1", "X2"))
pcoa.centroids <- acast(pcoa.L.centroids, variable ~ group, mean)
pcoa.centroids.se <- acast(pcoa.L.centroids, variable ~ group, se)
pcoa.centroids.sd <- acast(pcoa.L.centroids, variable ~ group, sd)

# Combine
pcoa.cent.dataframe <- cbind(t(pcoa.centroids), t(pcoa.centroids.se))
colnames(pcoa.cent.dataframe) <- c("V1", "V2", "V1e", "V2e")
pcoa.cent.trts <- rownames(pcoa.cent.dataframe)
```

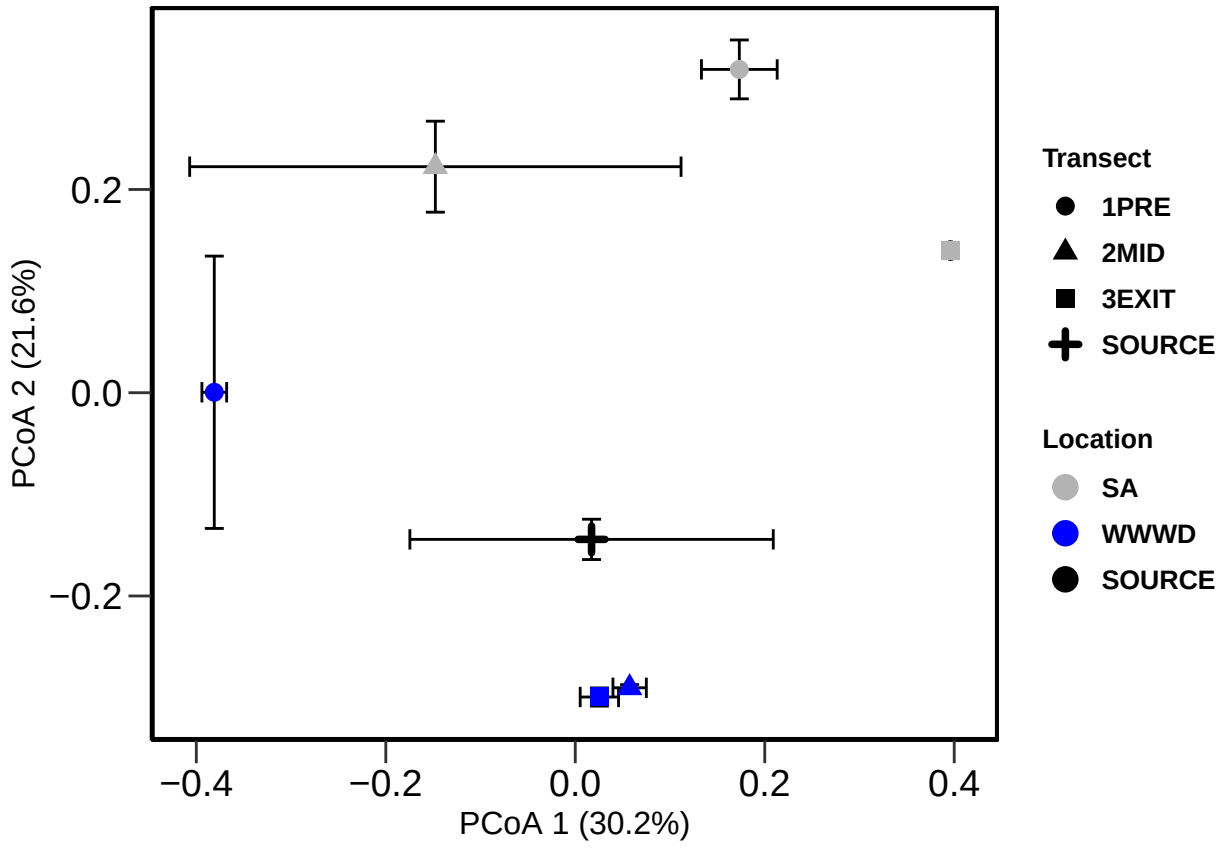
```
pcoa.cent.dataframe.trts <- as.data.frame(pcoa.cent.dataframe)
dim(pcoa.cent.dataframe.trts)
```

```
## [1] 7 4
```

```
#pcoa.col <- as.factor(sapply(strsplit(pcoa.cent.treats, "_"), `[`, 2)) # Transect
#pcoa.shape <- as.factor(sapply(strsplit(pcoa.cent.treats, "_"), `[`, 1))  # Location

Location <- c("1SA", "1SA", "1SA","3SOURCE","2WWWD", "2WWWD", "2WWWD")
Transect <- c("1PRE", "2MID", "3EXIT","SOURCE","1PRE", "2MID", "3EXIT")

pcoa.cent.dataframe.trts$Location <- as.factor(Location)
pcoa.cent.dataframe.trts$Transect <- as.factor(Transect)
dim(pcoa.cent.dataframe.trts) #28 7
```

```
## [1] 7 6
```

```
# Principal Coordinates Analysis
#Plot
df1a <- as.data.frame(pcoa.cent.dataframe.trts)
SH.bact <- ggplot(df1a, aes(x=V1, y=V2), group = interaction(Location,Transect))+
theme_bw() +
#Set error bars for geom_point
geom_errorbarh(aes(xmax=V1+V1e, xmin=V1-V1e, height=0.02), colour="black") +
geom_errorbar(aes(ymax=V2+V2e, ymin=V2-V2e, width=0.02), colour="black") +
geom_point(aes(shape = Transect, colour = Location), stroke = 2, size=2) +
#removes gridlines from plot
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
axis.line = element_line(colour = "black"))+
#Set colors for treatments
scale_colour_manual(labels = c("SA","WWWD","SOURCE"),
                    values = c("gray70","blue", "black")) +
theme(axis.title = element_text(size=12),# face="bold"),
axis.text.x = element_text(size=14, color="black"),  axis.text.y = element_text(size=14, color="black")
panel.border = element_rect(colour = "black", size=1.25)) +
#Set plot title text size
theme(plot.title=element_text(size=12)) +
#Set legend text size
theme(legend.text=element_text(size=10, face="bold"), legend.title = element_text(size=10, face="bold"))
#Sets size of tick marks on axis
theme(axis.ticks.length=unit(0.3,"cm")) +
#Sets labels for plot title, axis titles, and legend headings
xlab("PCoA 1 (30.2%)") + ylab("PCoA 2 (21.6%)")

SH.bact
```




```
ggsave("../figures/SH.bact.ordination.png", plot=last_plot(), device=NULL, path=NULL, scale=1, width=7,
```

```
require(phyloseq)
```

```
## Loading required package: phyloseq
```

```
SH_16s <- import_mothur(mothur_shared_file = "../data/SH.opti_mcc.shared", mothur_constaxonomy_file = "
SH_16s
```

```
## phyloseq-class experiment-level object
## otu_table()   OTU Table:         [ 7992 taxa and 19 samples ]
## tax_table()   Taxonomy Table:    [ 7992 taxa by 6 taxonomic ranks ]
```

```
design$Location <- as.factor(design$Location)
design$Transect <- as.factor(design$Transect)
design$Location_ID <- as.factor(design$Location_ID)
str(design)
```

```
## 'data.frame':    19 obs. of  6 variables:
##  $ Location         : Factor w/ 3 levels "SA","SOURCE",..: 3 3 3 3 3 3 3 3 1 1 ...
##  $ Transect         : Factor w/ 4 levels "1PRE","2MID",..: 1 1 1 2 2 2 3 3 1 1 ...
##  $ Location_ID      : Factor w/ 7 levels "SA","SA_EXIT",..: 7 7 7 5 5 5 6 6 3 3 ...
##  $ Location_ID_order: chr  "1WWWD_PRE" "1WWWD_PRE" "1WWWD_PRE" "2WWWD_BIRD" ...
##  $ Description      : chr  "EMU" "EMU" "EMU" "INFLOW" ...
##  $ Rep              : int  1 2 3 1 2 3 1 2 1 2 ...
```

```
sample <- sample_data(design)

sample_data(SH_16s)<- sample

colnames(tax_table(SH_16s))
```

```
## [1] "Rank1" "Rank2" "Rank3" "Rank4" "Rank5" "Rank6"
```

```
colnames(tax_table(SH_16s)) <- c("Kingdom", "Phylum", "Class",
  "Order", "Family", "Genus")

after_remove_low_depth <- prune_samples(sample_sums(SH_16s) >= 6000, SH_16s)
head(sample_sums(after_remove_low_depth))
```

```
##  SH_1 SH_10 SH_11 SH_12 SH_13 SH_14
## 18904 19828 14099 18365 12250 16743
```

```
set.seed(1)
rare <- rarefy_even_depth(after_remove_low_depth, sample.size = 8000,rngseed=TRUE)
```

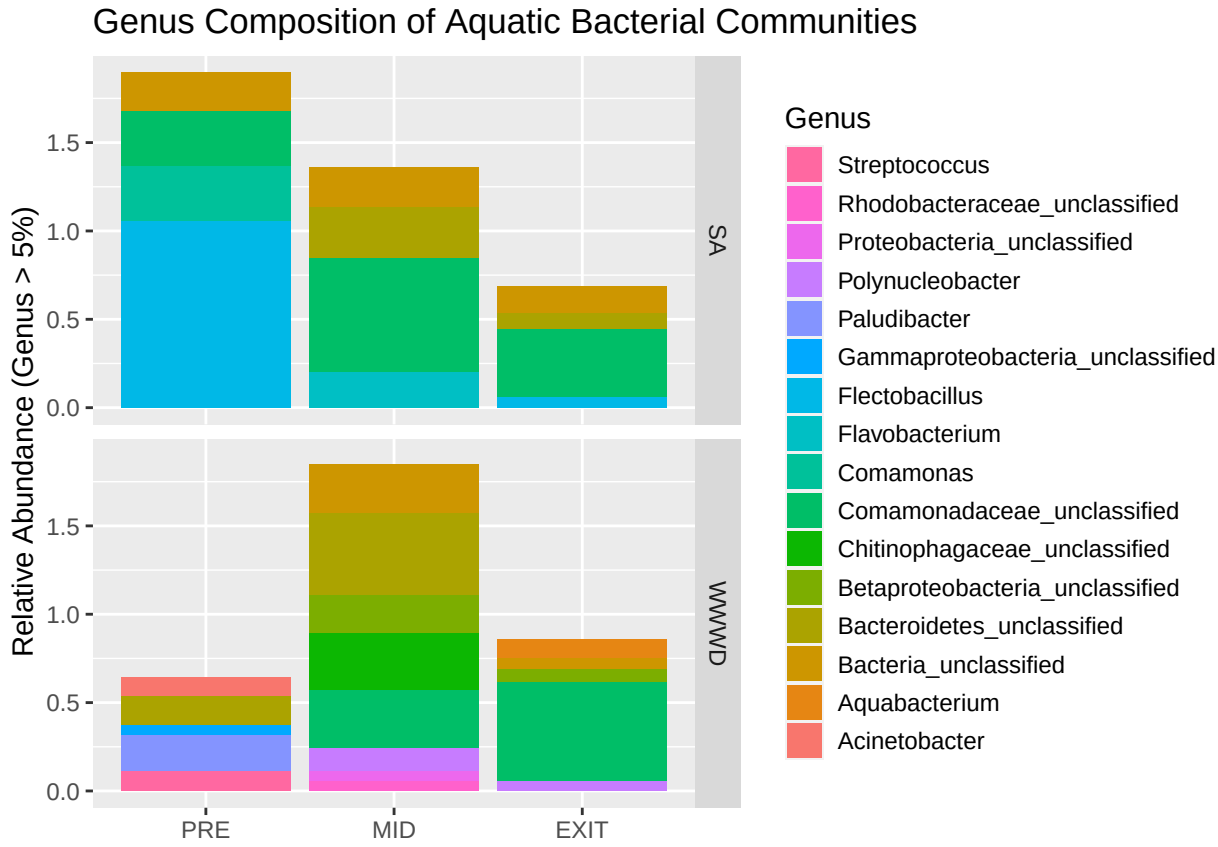
```
## `set.seed(TRUE)` was used to initialize repeatable random subsampling.

## Please record this for your records so others can reproduce.

## Try `set.seed(TRUE); .Random.seed` for the full vector

## ...

## 2633OTUs were removed because they are no longer
## present in any sample after random subsampling

## ...
```

```
#2633OTUs were removed because they are no longer present in any sample after random subsampling

head(sample_sums(rare))
```

```
##  SH_1 SH_10 SH_11 SH_12 SH_13 SH_14
##  8000  8000  8000  8000  8000  8000
```

```
#remove the NTC sample. Check to make sure it doesn't have too many sequences before you through it awa
#to_remove <- c("NTC")
#pruned <- prune_samples(!(rownames(sample_data(rare)) %in% to_remove), rare)

#filter out OTUs less than 10
#darte_ed_16s_filter <- filter_taxa(pruned, function(x) sum(x) > 10, TRUE)

 #relative abundance
SH_16s_filter_re <- transform_sample_counts(rare, function(x) x /sum(x))
```

```r
#Get rid of small taxa
SH_16s_filter2 <- filter_taxa(SH_16s_filter_re, function(x) sum(x) > .001, TRUE)

#Combine OTUs with common taxa
SH_16s_filter_re_g = tax_glom(SH_16s_filter2, "Phylum")
SH_16s_filter_re_g2 = tax_glom(SH_16s_filter2, "Genus")


SH_genus <- SH_16s %>%
  tax_glom(taxrank = "Genus") %>%                      # agglomerate at phylum level
  transform_sample_counts(function(x) {x/sum(x)} ) %>% # Transform to rel. abundance
  psmelt() %>%                                          # Melt to long format
  filter(Abundance > 0.05) %>%                          # Filter out low abundance taxa
  arrange(Genus)                                       # Sort data frame alphabetically by phylum
# Set colors for plotting
genus_colors <- c(
  "salmon", "darkseagreen", "gold", "magenta", "slateblue", "bisque", "darkred", "cadetblue", "darkoran
)


# Plot
a <- list(
  font = list(size = 14),
  xref = "paper",
  yref = "paper",
  yanchor = "bottom",
  xanchor = "center",
  align = "center",
  x = 0.5,
  y = 1,
  showarrow = FALSE)

SH_genus_v2 <- subset(SH_genus, Location == "SA" | Location == "WWWD")

ggplot(transform(SH_genus_v2, Transect=factor(Transect, levels=c("1PRE","2MID","3EXIT"))), aes(x = Trans
  geom_bar(stat = "identity") +
  #scale_fill_manual(values = genus_colors) +
  scale_x_discrete(labels = c("PRE","MID","EXIT"), drop = TRUE) +
  # Remove x axis title
  theme(axis.title.x = element_blank()) +
  #
  guides(fill = guide_legend(reverse = TRUE, keywidth = 1, keyheight = 1)) +
  ylab("Relative Abundance (Genus > 5%)") +
  ggtitle("Genus Composition of Aquatic Bacterial Communities")
```




```
ggsave("../figures/genuscomp_updated.png", plot=last_plot(), device=NULL, path=NULL, scale=1, width=7, h
```

OTU graph

```
#NOTE:
#Otu0113 (451 reads)
#Otu1461 (12 reads)
#Otu2496 (6 reads)
#Otu3114 (4 reads)

#not graphing source
new.data.nosource <-new.data[c(1:16),]
str(new.data.nosource)
```

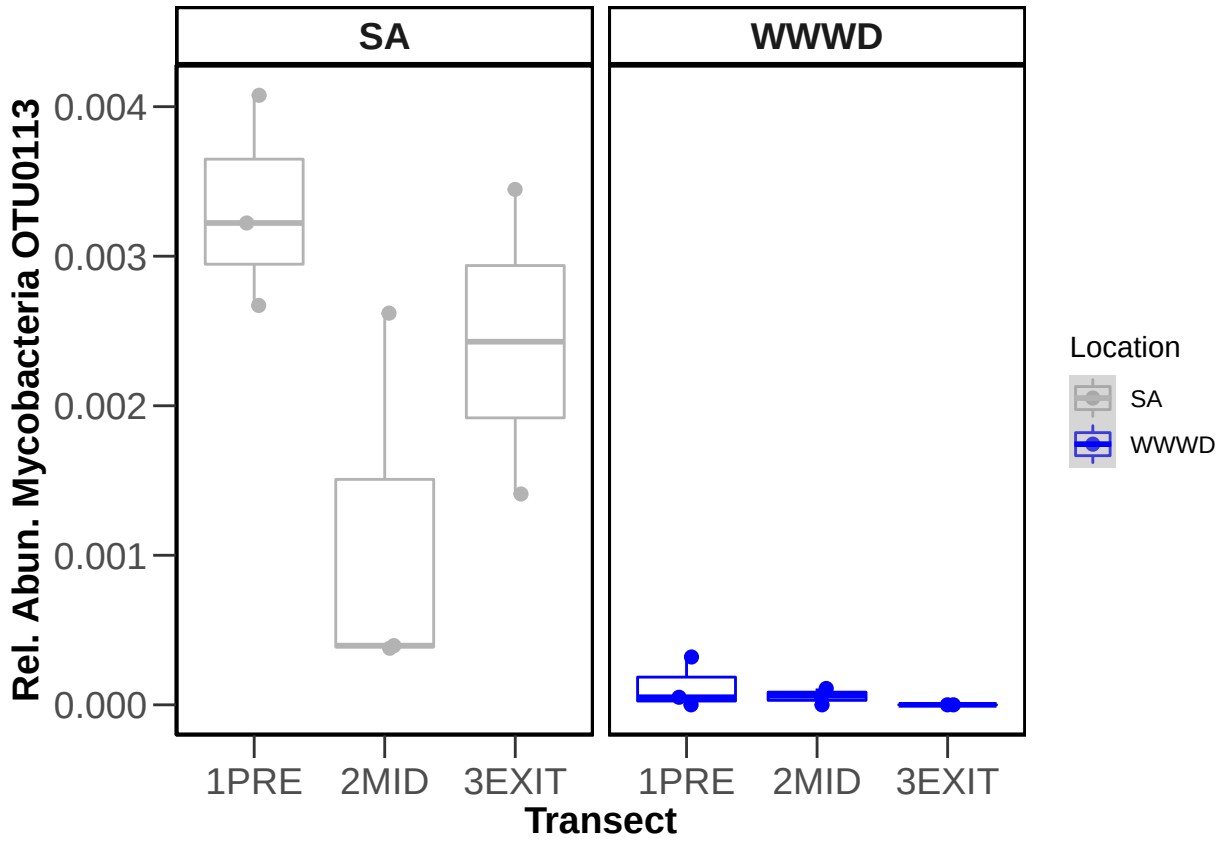
```
## 'data.frame':    16 obs. of  5754 variables:
##  $ Location         : Factor w/ 3 levels "SA","SOURCE",..: 3 3 3 3 3 3 3 3 1 1 ...
##  $ Transect         : Factor w/ 4 levels "1PRE","2MID",..: 1 1 1 2 2 2 3 3 1 1 ...
##  $ Location_ID      : Factor w/ 7 levels "SA","SA_EXIT",..: 7 7 7 5 5 5 6 6 3 3 ...
##  $ Location_ID_order: chr  "1WWWD_PRE" "1WWWD_PRE" "1WWWD_PRE" "2WWWD_BIRD" ...
##  $ Description      : chr  "EMU" "EMU" "EMU" "INFLOW" ...
##  $ Rep              : int  1 2 3 1 2 3 1 2 1 2 ...
##  $ Otu0001          : num  0.00203 0.29996 0.35299 0.03207 0.00314 ...
##  $ Otu0002          : num  0.00304 0.00197 0.00157 0.0388 0.0095 ...
```

```
##  $ Otu0003          : num  0.0119 0.0189 0.0168 0.0333 0.0289 ...
##  $ Otu0004          : num  0.00 0.00 0.00 5.47e-05 0.00 ...
##  $ Otu0005          : num  0.00773 0.11323 0.10673 0.01472 0.00454 ...
##  $ Otu0006          : num  0.0048 0.00319 0.00171 0.09116 0.06731 ...
##  $ Otu0007          : num  0 0 0 0 0 ...
##  $ Otu0008          : num  0.0112 0.00319 0.00271 0.04591 0.04385 ...
##  $ Otu0009          : num  0.00 5.06e-05 5.70e-04 2.79e-02 2.46e-02 ...
##  $ Otu0010          : num  0 0 0 0 0 ...
##  $ Otu0011          : num  0.05109 0.00607 0.00784 0.00914 0.00586 ...
##  $ Otu0012          : num  0 0 0 0.00979 0.00603 ...
##  $ Otu0013          : num  0.00251 0.05087 0.041 0.03869 0.02337 ...
##  $ Otu0014          : num  8.69e-03 1.52e-04 1.43e-04 3.83e-04 8.26e-05 ...
##  $ Otu0015          : num  0 0.0266 0.0216 0.0149 0.0102 ...
##  $ Otu0016          : num  0.04112 0.01589 0.0251 0.00892 0.00586 ...
##  $ Otu0017          : num  0.000747 0.003493 0.002281 0.011765 0.008919 ...
##  $ Otu0018          : num  0.05445 0.00375 0.00456 0.00843 0.00545 ...
##  $ Otu0019          : num  0.042184 0.007542 0.008342 0.002189 0.000908 ...
##  $ Otu0020          : num  0 0.07795 0.039 0.00679 0.03799 ...
##  $ Otu0021          : num  0.00128 0.000405 0.000285 0.021067 0.015608 ...
##  $ Otu0022          : num  0 0 0 0 0 ...
##  $ Otu0023          : num  0.000693 0.003543 0.002567 0.003885 0.003634 ...
##  $ Otu0024          : num  0.00117 0.00967 0.01198 0.04027 0.01718 ...
##  $ Otu0025          : num  0.01664 0.01463 0.01668 0.00679 0.00727 ...
##  $ Otu0026          : num  2.74e-02 4.05e-04 1.43e-04 2.74e-04 8.26e-05 ...
##  $ Otu0027          : num  0 0 0 0.032 0.0231 ...
##  $ Otu0028          : num  0.00096 0.00132 0.0015 0.02484 0.00933 ...
##  $ Otu0029          : num  0.021439 0.002733 0.00164 0.000219 0.001486 ...
##  $ Otu0030          : num  0.00384 0.00693 0.00577 0.0035 0.02692 ...
##  $ Otu0031          : num  0 0 0 0.001149 0.000908 ...
##  $ Otu0032          : num  5.33e-05 0.00 0.00 0.00 0.00 ...
##  $ Otu0033          : num  0.022559 0 0 0.000164 0.001734 ...
##  $ Otu0034          : num  0.000267 0.000101 0.000214 0.000821 0.000495 ...
##  $ Otu0035          : num  0.00816 0.01357 0.01226 0.00394 0.00859 ...
##  $ Otu0036          : num  0 0 0 0.0237 0.0123 ...
##  $ Otu0037          : num  0.016212 0.004505 0.003779 0.000711 0.000413 ...
##  $ Otu0038          : num  0 0.000101 0.000143 0.02446 0.010323 ...
##  $ Otu0039          : num  0.000427 0 0.000285 0.033817 0.01247 ...
##  $ Otu0040          : num  0 0 0 0.0242 0.0126 ...
##  $ Otu0041          : num  0 0.000962 0.001996 0.000219 0.000661 ...
##  $ Otu0042          : num  0.011359 0.002632 0.00164 0.000328 0.00033 ...
##  $ Otu0043          : num  0.00251 0.00223 0.00121 0.00328 0.00182 ...
##  $ Otu0044          : num  0.01141 0.00537 0.00749 0.00268 0.00116 ...
##  $ Otu0045          : num  0 0.02328 0.03051 0.00175 0.00215 ...
##  $ Otu0046          : num  0 0 0 0 0 ...
##  $ Otu0047          : num  1.18e-02 2.53e-04 7.13e-05 6.02e-04 2.15e-03 ...
##  $ Otu0048          : num  0.007306 0.003493 0.003066 0.002845 0.000495 ...
##  $ Otu0049          : num  0.006773 0.004505 0.004206 0.000821 0.001156 ...
##  $ Otu0050          : num  5.87e-04 5.06e-05 3.56e-04 8.32e-03 5.95e-03 ...
##  $ Otu0051          : num  0 0 0 0.0187 0.0114 ...
##  $ Otu0052          : num  0 0.001114 0.000713 0.020082 0.010488 ...
##  $ Otu0053          : num  0.00875 0.00213 0.00235 0.00274 0.00124 ...
##  $ Otu0054          : num  0.010239 0.000304 0.000428 0.002736 0.002065 ...
##  $ Otu0055          : num  0.002507 0.001468 0.000998 0.004542 0.018581 ...
##  $ Otu0056          : num  0 0.000101 0 0.001259 0.000661 ...
```

```
##  $ Otu0057          : num  0.00501 0.00557 0.00934 0.00159 0.00107 ...
##  $ Otu0058          : num  9.81e-03 5.06e-04 7.13e-04 2.19e-04 8.26e-05 ...
##  $ Otu0059          : num  0.0108 0 0 0 0 ...
##  $ Otu0060          : num  5.33e-05 0.00 0.00 1.32e-02 9.58e-03 ...
##  $ Otu0061          : num  0.001867 0.000253 0.000356 0.001423 0.001321 ...
##  $ Otu0062          : num  0.002507 0.000202 0.000214 0.005746 0.005616 ...
##  $ Otu0063          : num  0.0048 0.00526 0.00635 0.00257 0.0038 ...
##  $ Otu0064          : num  0.00016 0 0 0 0 ...
##  $ Otu0065          : num  1.07e-04 0.00 0.00 0.00 8.26e-05 ...
##  $ Otu0066          : num  0.00112 0.000152 0 0.006129 0.006441 ...
##  $ Otu0067          : num  0.00661 0.00385 0.00442 0.00175 0.00264 ...
##  $ Otu0068          : num  4.80e-03 4.05e-04 3.56e-04 5.47e-05 0.00 ...
##  $ Otu0069          : num  0.00544 0.001012 0.000856 0.00093 0.000165 ...
##  $ Otu0070          : num  0.00848 0.001164 0.000713 0.000219 0.000661 ...
##  $ Otu0071          : num  0 0 0 0.000383 0 ...
##  $ Otu0072          : num  0 0 0 0 0 ...
##  $ Otu0073          : num  4.27e-04 1.01e-04 1.43e-04 0.00 8.26e-05 ...
##  $ Otu0074          : num  0.00603 0.00162 0.00171 0.00186 0.00206 ...
##  $ Otu0075          : num  0.000213 0 0 0.002681 0.026674 ...
##  $ Otu0076          : num  5.33e-05 0.00 0.00 9.30e-04 7.43e-04 ...
##  $ Otu0077          : num  0.000373 0 0 0.007332 0.004212 ...
##  $ Otu0078          : num  0.000107 0.011591 0.007201 0.001587 0.010901 ...
##  $ Otu0079          : num  1.07e-04 1.01e-04 7.13e-05 8.32e-03 6.85e-03 ...
##  $ Otu0080          : num  0.005866 0.001468 0.000713 0.00104 0.001486 ...
##  $ Otu0081          : num  0.00544 0.00167 0.001283 0.000438 0.00033 ...
##  $ Otu0082          : num  0 0 0 0 0 ...
##  $ Otu0083          : num  0.00555 0.00218 0.00264 0.00192 0.00116 ...
##  $ Otu0084          : num  2.13e-04 4.56e-04 7.13e-05 0.00 0.00 ...
##  $ Otu0085          : num  0 0 0 0 0 ...
##  $ Otu0086          : num  0.00528 0.000253 0.000143 0.000109 0.000495 ...
##  $ Otu0087          : num  0.00576 0.001721 0.001925 0.000657 0.000413 ...
##  $ Otu0088          : num  0.00016 0 0 0.000109 0 ...
##  $ Otu0089          : num  0.006133 0.000152 0.000143 0.000109 0 ...
##  $ Otu0090          : num  0.002453 0.00162 0.001497 0.000602 0.000578 ...
##  $ Otu0091          : num  5.71e-03 1.52e-04 7.13e-05 2.19e-04 2.48e-04 ...
##  $ Otu0092          : num  0.005813 0.001265 0.001355 0.000547 0.000165 ...
##  $ Otu0093          : num  0 0 0 0.000164 0 ...
##   [list output truncated]
```

```
p <- ggplot(new.data.nosource, aes(x=Transect, y=Otu0113, color=Location))+ geom_boxplot() +
     geom_point(aes(color=Location), size=2, position = position_jitterdodge()) + scale_color_manual(na
p1=p+geom_smooth(method="lm")+facet_wrap(~Location)
Otu0113<-p1 + theme_bw() +
    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(), axis.line
          =element_line(colour = "black")) +
    theme(axis.title=element_text(vjust=1,size=14,face="bold"),
          axis.text=element_text(size=14), axis.text.x = element_text(vjust=0.65, hjust=0.5,
          size=14), panel.border = element_rect(colour = "black",size=1)) +
    theme(axis.ticks.length=unit(0.3,"cm")) + labs(x = "Transect", y = "Rel. Abun. Mycobacteria OTU0113
    theme(strip.text.x = element_text(size=14, face="bold"), strip.text.y =
          element_text(size=14, face="bold"), strip.background = element_rect(colour="black",
          fill="white", size=1))
Otu0113
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ggsave("../figures/OTU0113.png", plot=last_plot(), device=NULL, path=NULL, scale=1, width=7, height=5, c
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Bacterial community indicator species analysis

```
new.data <-cbind(design,dataREL)
library("labdsv")
```

```
## Loading required package: mgcv
```

```
## This is mgcv 1.8-40. For overview type 'help("mgcv-package")'.
```

```
## Registered S3 method overwritten by 'labdsv':
##   method       from
##   summary.dist ade4
```

```
## This is labdsv 2.0-1
## convert existing ordinations with as.dsvord()
```

```
##
## Attaching package: 'labdsv'
```

```
## The following object is masked from 'package:stats':
##
##     density
```

```
group = interaction(new.data$Location_ID)
design.type <- group

dataREL.ind <- dataREL[, colSums(dataREL) > 0.05]
bac.ind <- indval(dataREL.ind, design.type)
levels(design.type)
```

```
## [1] "SA"        "SA_EXIT"   "SA_PRE"    "SOURCE"    "WWWD"      "WWWD_EXIT"
## [7] "WWWD_PRE"
```

```
#"SA"        "SA_EXIT"   "SA_PRE"    "SOURCE"    "WWWD"      "WWWD_EXIT" "WWWD_PRE"
summary(bac.ind)
```

```
##         cluster indicator_value probability
## Otu0029       1           0.5444       0.044
## Otu0022       2           0.7033       0.021
## Otu0046       2           0.6942       0.025
## Otu0032       2           0.6724       0.024
## Otu0007       2           0.6717       0.022
## Otu0031       2           0.6608       0.009
## Otu0012       2           0.5736       0.019
## Otu0041       2           0.5699       0.023
## Otu0034       2           0.5689       0.024
## Otu0009       2           0.3438       0.031
## Otu0003       2           0.2964       0.038
## Otu0010       3           0.9998       0.003
## Otu0004       3           0.9996       0.002
## Otu0014       3           0.8551       0.003
## Otu0043       3           0.4844       0.008
## Otu0015       4           0.6221       0.012
## Otu0052       5           0.6841       0.004
## Otu0060       5           0.6750       0.003
## Otu0039       5           0.5879       0.031
## Otu0024       5           0.4928       0.030
## Otu0038       5           0.4921       0.036
## Otu0066       5           0.3694       0.037
## Otu0061       6           0.8681       0.001
## Otu0027       6           0.5291       0.037
## Otu0055       6           0.4659       0.040
## Otu0062       6           0.4477       0.029
## Otu0028       6           0.4276       0.032
## Otu0025       7           0.3743       0.019
## Otu0035       7           0.3464       0.016
##
## Sum of probabilities                   =  6.027
```

```
##
## Sum of Indicator Values              =  30.99
##
## Sum of Significant Indicator Values  =  17.02
##
## Number of Significant Indicators     =  29
##
## Significant Indicator Distribution
##
##  1  2  3  4  5  6  7
##  1 10  4  1  6  5  2
```

```
inds <- which(bac.ind$pval <= 0.05)
bac.indicators <- as.data.frame(matrix(NA, nrow = length(inds), ncol = 4))
colnames(bac.indicators) <- c("OTU", "Cluster", "IndVal", "Prob")

bac.indicators$OTU <- names(inds)
bac.indicators$Cluster <- bac.ind$maxcls[inds]
bac.indicators$IndVal <- bac.ind$indcls[inds]
bac.indicators$Prob <- bac.ind$pval[inds]

ind.tax <- otu.tax[which(as.character(otu.tax$OTU) %in% bac.indicators$OTU), ]
ind.tax <- ind.tax[match(ind.tax$OTU, bac.indicators$OTU), ]

indicator.bac <- cbind(bac.indicators, ind.tax[, -c(1)])

indicator.bac <- indicator.bac[order(as.numeric(indicator.bac$Cluster)), ]

table(indicator.bac$Cluster)
```

```
##
##  1  2  3  4  5  6  7
##  1 10  4  1  6  5  2
```

```
table(indicator.bac$Phylum)
```

```
##
##        Actinobacteria Bacteria_unclassified         Bacteroidetes
##                     1                     2                     7
##            Firmicutes        Proteobacteria
##                     2                    17
```

```
table(indicator.bac$Cluster)
```

```
##
##  1  2  3  4  5  6  7
##  1 10  4  1  6  5  2
```

```
levels(design.type)
```

```
## [1] "SA"        "SA_EXIT"   "SA_PRE"    "SOURCE"    "WWWD"      "WWWD_EXIT"
## [7] "WWWD_PRE"
```

```
# Export Bacteria Indicator Table
write.table(indicator.bac, "../data/BacterialIndicators_Location.txt",
            sep="\t", row.names = F, quote = F)
```