Figure1A

Emiliano Pereira

2025-08-06

Contents

Intro

Non-Metric Multidimensional Scaling (NMDS) performed on the Bray-Curtis compositional dissimilarity matrices of Amplicon Sequence Variants (ASVs), to reproduce the analyses and the Figure 1A from the original publication Seasonal dynamics of the coastal microbiome and its association with environmental factors.

1. Set the environment

```
library(tidyverse)
library(vegan)
library(ggpubr)
library(corrgram)
library(ggcorrplot)
getwd()
```

```
## [1] "/home/epereira/workspace/repositories/studies/Coastal-Microbiome-Seasonality-Risks/scripts"
```

2. Load data

`asvs_workable.tsv` contains the rarefied ASV abundance profiles, with samples as rows and ASVs as columns.
`metadata_workable.tsv` contains the metadata for the samples.

```
ABUND <- read_tsv("../data/asvs_workable.tsv.gz", show_col_types = FALSE) %>%
         column_to_rownames("Date")

METADATA <- read_tsv("../data/metadata_workable.tsv", show_col_types = FALSE)
```

3. Run NMDS

NMDS is performed on Bray–Curtis dissimilarities computed from Hellinger-transformed ASV abundance profiles.

```
ABUND_norm <- decostand(ABUND, method = "hellinger")
set.seed(1234)
ABUND_norm_nmds <- metaMDS(ABUND_norm, k = 2)
```

```
## Run 0 stress 0.1158503
## Run 1 stress 0.1169114
## Run 2 stress 0.1158503
## ... Procrustes: rmse 2.042096e-05  max resid 6.076376e-05
```

```
## ... Similar to previous best
## Run 3 stress 0.1554071
## Run 4 stress 0.1312047
## Run 5 stress 0.1169114
## Run 6 stress 0.1309233
## Run 7 stress 0.1436816
## Run 8 stress 0.1158503
## ... Procrustes: rmse 1.242594e-05  max resid 3.887782e-05
## ... Similar to previous best
## Run 9 stress 0.1309233
## Run 10 stress 0.1158503
## ... Procrustes: rmse 7.514949e-06  max resid 1.966508e-05
## ... Similar to previous best
## Run 11 stress 0.1158503
## ... New best solution
## ... Procrustes: rmse 5.469813e-06  max resid 1.781298e-05
## ... Similar to previous best
## Run 12 stress 0.1697951
## Run 13 stress 0.1312047
## Run 14 stress 0.1158503
## ... Procrustes: rmse 1.002897e-05  max resid 3.099001e-05
## ... Similar to previous best
## Run 15 stress 0.1158503
## ... New best solution
## ... Procrustes: rmse 1.688453e-06  max resid 5.43847e-06
## ... Similar to previous best
## Run 16 stress 0.1436816
## Run 17 stress 0.1158503
## ... Procrustes: rmse 4.096718e-06  max resid 1.419209e-05
## ... Similar to previous best
## Run 18 stress 0.1158503
## ... Procrustes: rmse 3.400028e-06  max resid 1.060752e-05
## ... Similar to previous best
## Run 19 stress 0.1158503
## ... Procrustes: rmse 6.879947e-06  max resid 2.637872e-05
## ... Similar to previous best
## Run 20 stress 0.1158503
## ... Procrustes: rmse 1.390654e-05  max resid 4.668607e-05
## ... Similar to previous best
## *** Best solution repeated 5 times
```

4. Merge tables to plot

The coordinates of the samples in the NMDS space are extracted and merged with the metadata for plotting.

```r
ABUND_norm_nmds_ext <- ABUND_norm_nmds$points %>%
                       as.data.frame %>%
                       rownames_to_column("Date") %>%
                       mutate(Date = as.Date(Date)) %>%
                       left_join(x = ., y = METADATA , by = "Date")
```

5. Add centroid coordinates

The centroid coordinates for each season in 2018 are calculated to draw segments connecting the centroids. The year 2018 is selected for this purpose given that is the only one including all four seasons in the dataset.

```r
ctd_coords <- ABUND_norm_nmds_ext %>%
              filter(Year == "2018") %>%
              mutate(Season = factor(Season,
                                     levels = c("Summer", "Autumn",
                                                "Winter", "Spring"))) %>%
              ungroup() %>%
              group_by(Season, Year) %>%
              summarize(mean_nmds1 = mean(MDS1),
                        mean_nmds2 = mean(MDS2)) %>%
              arrange(Year, Season)
```

```
## `summarise()` has grouped output by 'Season'. You can override using the `.groups` argument.
```

```r
ctd_coords$segment_x <- NA
ctd_coords$segment_y <- NA
ctd_coords$segment_xend <- NA
ctd_coords$segment_yend <- NA


for (i in 1:(dim(ctd_coords)[1] -1)) {

  ctd_coords$segment_x[i] <- ctd_coords$mean_nmds1[i]
  ctd_coords$segment_y[i] <- ctd_coords$mean_nmds2[i]

  ctd_coords$segment_xend[i] <- ctd_coords$mean_nmds1[i + 1]
  ctd_coords$segment_yend[i] <- ctd_coords$mean_nmds2[i + 1]

}

# add last values
last_i <- dim(ctd_coords)[1]

ctd_coords$segment_x[last_i] <- ctd_coords$mean_nmds1[last_i]
ctd_coords$segment_y[last_i] <- ctd_coords$mean_nmds2[last_i]
ctd_coords$segment_xend[last_i] <- ABUND_norm_nmds_ext %>% filter(Season == "Summer" & Year  == 2020) %
ctd_coords$segment_yend[last_i] <- ABUND_norm_nmds_ext %>% filter(Season == "Summer" & Year  == 2020) %
```

7. Plot: MDS1 vs MDS2

```r
x_title <- paste("MDS1")
y_title <- paste("MDS2")
x_max <- max(ABUND_norm_nmds_ext$MDS1) + max(ABUND_norm_nmds_ext$MDS1)/10
x_min <- min(ABUND_norm_nmds_ext$MDS1) - abs(min(ABUND_norm_nmds_ext$MDS1)/10)
y_max <- max(ABUND_norm_nmds_ext$MDS2) + max(ABUND_norm_nmds_ext$MDS2)/10
y_min <- min(ABUND_norm_nmds_ext$MDS2) - abs(min(ABUND_norm_nmds_ext$MDS2)/10)

ABUND_norm_nmds_ext$Season <- factor(ABUND_norm_nmds_ext$Season,
                                     levels = c("Summer", "Autumn", "Winter", "Spring"))

text_size <- 13
season_colors <- c("#c93f1b","#98482b", "#154360", "#3c7810")

nmds_plot_mds1_vs_mds2 <- ggplot(ABUND_norm_nmds_ext, aes(x = MDS1, y = MDS2,
                                                          color = Season)) +
```

```
                         geom_point(size = 3, alpha = 0.85) +
                         scale_color_manual(values = season_colors) +
                         geom_vline(xintercept = 0, color = "gray40",linetype='dotted') +
                         geom_hline(yintercept = 0, color = "gray40",linetype='dotted') +
                         scale_size_continuous(name = "Days from start") +
                         xlab(x_title) +
                         ylab(y_title) +
                         theme_bw() +
                         theme(
                           axis.title = element_text(size = text_size +4),
                           axis.text = element_text(size = text_size),
                           legend.text = element_text(size= text_size),
                           legend.title = element_text(size = text_size),
                           plot.margin = unit(c(0.2,1,0.2,0.2), "lines")
                           ) +
                         guides(shape = guide_legend(override.aes = list(size = 3)),
                                color = guide_legend(override.aes = list(shape = 19, size =3))) +
                         xlim(c(x_min, x_max)) +
                         ylim(c(y_min, y_max)) +
                         annotate("text",
                                  label = paste("Stress:", round(ABUND_norm_nmds$stress,3)),
                                  x = 0.7 , y = 0.55, size = 3) +
                         geom_segment(data = ctd_coords, aes(x = segment_x, y = segment_y,
                                                             xend = segment_xend, yend = segment_yend)
                                      arrow = arrow(length = unit(0.1, "inches")),
                                      color = "gray10", alpha = 0.7, linewidth = 1) +
                         geom_text(data = ctd_coords,  aes(x = mean_nmds1, y = mean_nmds2),
                                   label = c("Summer", "Autumn", "Winter", "Spring"),
                                   vjust = c(-0.5,1.5,0.5,2.5), hjust = c(0.7,1,-0.35,-0.2),
                                   size = 2, fontface="bold", show.legend = F) +
                         geom_text(aes(label = Date), vjust = 2, size = 1.5,
                                   check_overlap = T, show.legend = F) +
                         scale_shape(guide = "none")
```

8. Create box plot for MDS1

```
boxplot_margin_pc1 <- c(0.2,7.48,0.2,4.85)
x_title_boxplot_pc1 <- x_title
y_title_boxplot_pc1 <- "Season"
ABUND_norm_nmds_ext$Season_rev <- factor(ABUND_norm_nmds_ext$Season,
                                         levels = c("Spring", "Winter", "Autumn", "Summer"))

boxplot_mds1_vs_season <- ggplot(ABUND_norm_nmds_ext,
                                 aes(x = MDS1, y = Season_rev, fill = Season_rev)) +
                         geom_boxplot() +
                         geom_vline(xintercept = 0, color = "gray40",linetype='dotted') +
                         scale_fill_manual(values = rev(season_colors), name = "Season") +
                         xlab(x_title_boxplot_pc1) +
                         ylab(y_title_boxplot_pc1) +
                         theme_bw() +
                         theme(
                           axis.title.x = element_text(size = text_size +4, margin = unit(c(4,0,0,0),
                           axis.title.y = element_text(size = text_size +1, margin = unit(c(0,2,0,0),
```

```
                              axis.text.x = element_text(size = text_size),
                              axis.text.y = element_text(size = text_size -3),
                              legend.text = element_text(size= text_size),
                              legend.title = element_text(size = text_size),
                              plot.margin = unit(boxplot_margin_pc1, "lines")) +
                           xlim(c(x_min, x_max)) +
                           # add a segment from y = 2 to y = 3 and x = 0.6
                           geom_segment(aes(x = 0.85, y = 1.5, xend = 0.85, yend = 3.5),
                                        color = "gray10", alpha = 0.7, linewidth = 0.1) +
                           annotate("text", x = 0.88, y = 2.5, label = "***", size = 2, angle = 90)
```

9. Create boxplot for MDS2

```
boxplot_margin_pc2 <- c(0.2,0.2,0.2,0.2)
x_title_boxplot_pc2 <- "Season"
y_title_boxplot_pc2 <- y_title

boxplot_mds2_vs_season <- ggplot(ABUND_norm_nmds_ext,
                                 aes(y = MDS2, x = Season_rev, fill = Season_rev)) +
                          geom_boxplot() +
                          geom_hline(yintercept = 0, color = "gray40",linetype='dotted') +
                          scale_fill_manual(values = rev(season_colors), name = "Season") +
                          xlab(x_title_boxplot_pc2) +
                          ylab(y_title_boxplot_pc2) +
                          theme_bw() +
                          theme(
                            axis.title.x = element_text(size = text_size +4, margin = unit(c(4,0,0,0),
                            axis.title.y = element_text(size = text_size +4, margin = unit(c(0,2,0,0),
                            axis.text.y = element_text(size = text_size),
                            legend.text = element_text(size= text_size),
                            legend.title = element_text(size = text_size),
                            plot.margin = unit(boxplot_margin_pc2, "lines")) +
                          ylim(c(y_min, y_max))
```

10. Merge plots

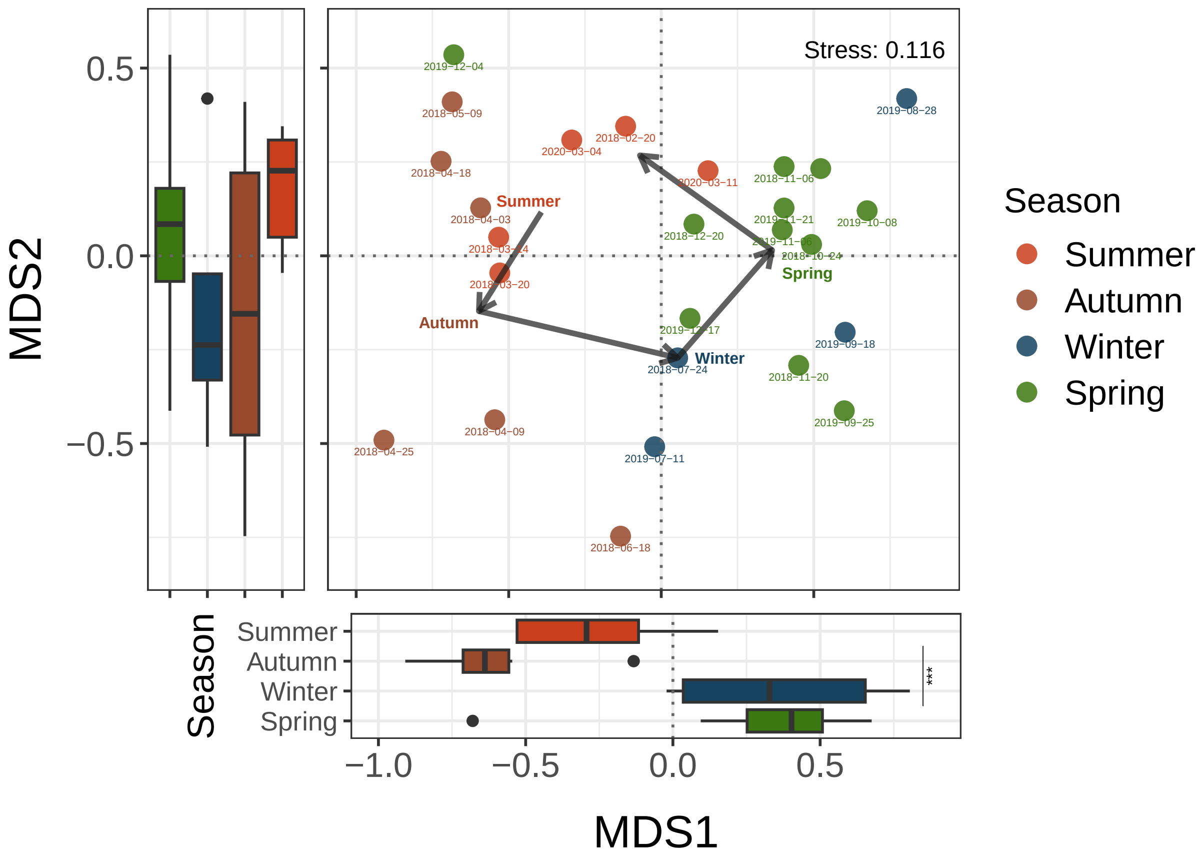
```
nmds_plot_mds1_vs_mds2_and_boxplot <- ggarrange(boxplot_mds2_vs_season + rremove("legend") + rremove("x
                                                nmds_plot_mds1_vs_mds2  + rremove("y.title") + rremove(
                                                  rremove("x.title") + rremove("x.text"),
                                                nrow = 1, ncol = 2,
                                                widths = c(0.25, 0.75))



nmds_plot_mds1_vs_mds2_and_mds1_boxplot <- ggarrange(nmds_plot_mds1_vs_mds2_and_boxplot,
                                                     boxplot_mds1_vs_season + rremove("legend"),
                                                     nrow = 2, ncol = 1,
                                                     heights = c(0.70, 0.3))
```

11. Visualize plot

```
nmds_plot_mds1_vs_mds2_and_mds1_boxplot
```



12. Perform ANOVA Welch test: MDS1

Test the differentiation of semesters Winter-Spring and Autumn-Summer along the MDS1 axis.

```
ABUND_norm_nmds_ext$Semester <- ifelse(ABUND_norm_nmds_ext$Season %in% c("Winter","Spring"),
                                     "Winter-Spring", "Autumn-Summer")

anova_welch_mds1 <- ABUND_norm_nmds_ext %>%
  oneway.test(MDS1 ~ Semester,
              data = ., var.equal = FALSE)

anova_welch_mds1
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  MDS1 and Semester
## F = 33.148, num df = 1.000, denom df = 23.403, p-value = 6.817e-06
```

13. Perform ANOVA Welch test: MDS2

Test the differentiation of semesters Summer-Spring and Autumn-Winter along the MDS1 axis.

```
ABUND_norm_nmds_ext$Semester2 <- ifelse(ABUND_norm_nmds_ext$Season %in% c("Summer","Spring"),
                                     "Summer-Spring", "Autumn-Winter")

anova_welch_mds2 <- ABUND_norm_nmds_ext %>%
```

```r
  oneway.test(MDS2 ~ Semester2,
              data = ., var.equal = FALSE)

anova_welch_mds2
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  MDS2 and Semester2
## F = 2.6298, num df = 1.000, denom df = 12.751, p-value = 0.1293
```