Figure3A

Emiliano Pereira

2025-08-06

Intro

Compares the taxonomic diversity and richness between the communities from the semester Summer-Autumn and Winter-Spring, to to reproduce the analyses and the Figure 3A from the original publication Seasonal dynamics of the coastal microbiome and its association with environmental factors.

1. Set the environment

```
library(tidyverse)
library(vegan)
```

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")

DATE2SEASON2COMMUITY <- read_tsv("../data/date2season2community.tsv", show_col_types = FALSE)
```

3. Compute diversity

Compute Shannon diversity, observed richness and Chao1 estimator.

```
div <- diversity(ABUND)
rich_obs <- estimateR(ABUND)["S.obs",]
rich_chao1 <- estimateR(ABUND)["S.chao1",]

output_long_df <- data.frame(Shannon = div,
                             Richness_obs = rich_obs,
                             Richness_chao1 = rich_chao1) %>%
                  rownames_to_column(var = "Date") %>%
                  pivot_longer(names_to = "Index", values_to = "value",
                               cols = -Date) %>%
                  left_join(DATE2SEASON2COMMUITY %>%
                              select(Date, Community) %>%
                              mutate(Date= as.character(Date)),
                            by = "Date")
```

**** 5.Run ANOVA Welch test****

Perform ANOVA Welch test to evaluate if the diversity and richness differences are significant.

```
anova_welch_shannon <- output_long_df %>%
  filter(Index == "Shannon") %>%
  oneway.test(value ~ Community,
              data = ., var.equal = FALSE)

anova_welch_shannon
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  value and Community
## F = 0.81893, num df = 1.000, denom df = 15.186, p-value = 0.3796
```

```
anova_welch_rich_obs <- output_long_df %>%
  filter(Index == "Richness_obs") %>%
  oneway.test(value ~ Community,
              data = ., var.equal = FALSE)

anova_welch_rich_obs
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  value and Community
## F = 6.9118, num df = 1.00, denom df = 17.99, p-value = 0.01703
```

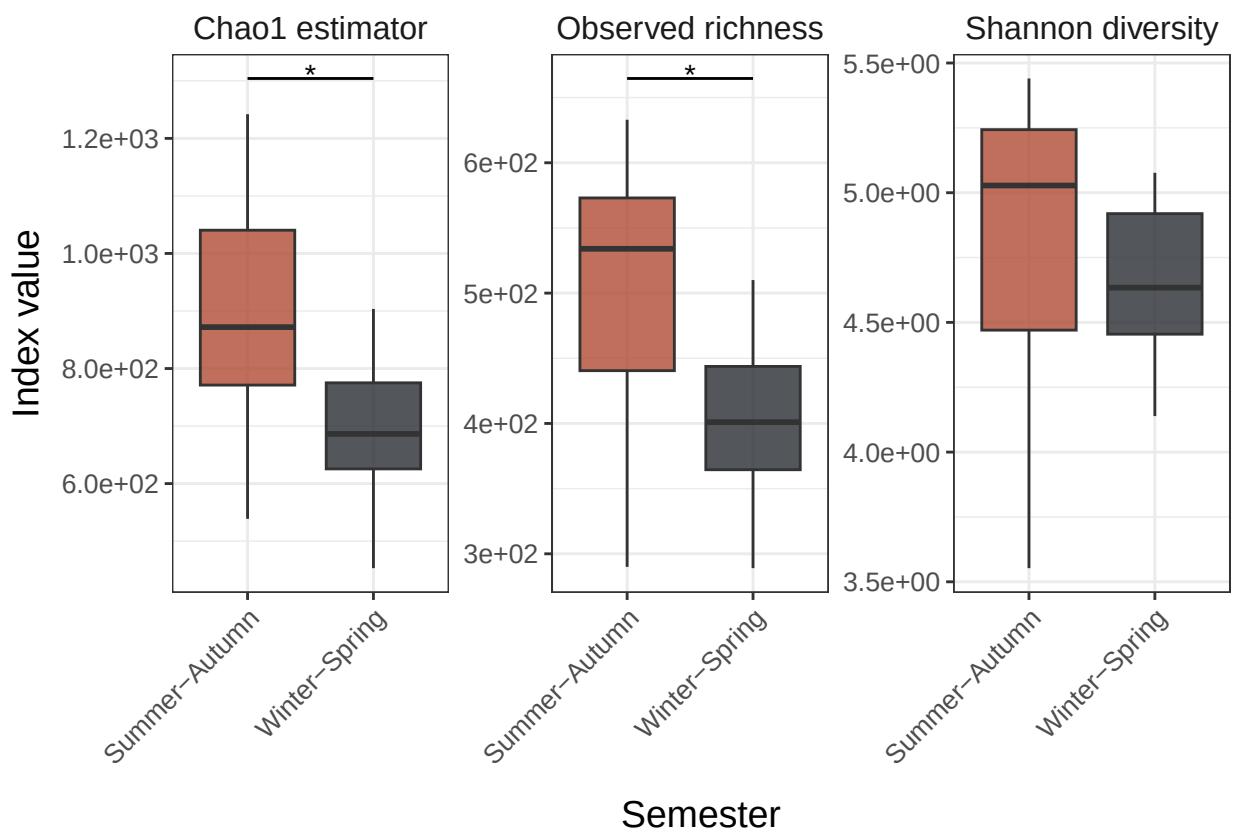
```
anova_welch_rich_chao <- output_long_df %>%
  filter(Index == "Richness_chao1") %>%
  oneway.test(value ~ Community,
              data = ., var.equal = FALSE)

anova_welch_rich_chao
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  value and Community
## F = 7.7405, num df = 1.000, denom df = 17.595, p-value = 0.01248
```

5. Plot diversity and richness value distributions

```
text_size <- 12

plot_labels <- c("Shannon" = "Shannon diversity",
                 "Richness_obs" = "Observed richness",
                 "Richness_chao1" = "Chao1 estimator")

div_plot <- output_long_df %>%
  filter(Index %in% c("Shannon", "Richness_obs", "Richness_chao1")) %>%
  ggplot(aes(x = Community, y = value, fill = Community)) +
  facet_wrap(~Index, scales = "free", labeller = as_labeller(plot_labels)) +
  xlab("Semester") +
  ylab("Index value") +
```

```r
  scale_y_continuous(labels = scales::scientific_format()) +
  scale_x_discrete(labels = c("S1" = "Summer-Autumn", "S2" = "Winter-Spring")) +
  # scale_fill_manual(values = c("#e28743","#063970")) +
  scale_fill_manual(values = c("#b14b34","#292c33")) +
  geom_boxplot(alpha = 0.8) +
  theme_bw() +
  theme(
    axis.text.x = element_text(size = text_size-2, angle = 45, hjust = 1),
    axis.text.y = element_text(size = text_size-2),
    axis.title.x = element_text(size = text_size +2, margin = unit(c(4,0,0,0), "mm")),
    axis.title.y = element_text(size = text_size +2, margin = unit(c(0,2,0,0), "mm")),
    strip.text = element_text(size = text_size),
    strip.background = element_blank()
  ) +
  guides(fill = "none")


y_positions <- output_long_df %>%
  filter(Index %in% c("Richness_obs", "Richness_chao1")) %>%
  group_by(Index) %>%
  summarise(y = max(value) * 1.05)

div_plot_ext <- div_plot +
  geom_segment(
    data = y_positions,
    aes(x = 1, xend = 2, y = y, yend = y),
    inherit.aes = FALSE,
    size = 0.5
  ) +
  geom_text(
    data = y_positions,
    aes(x = 1.5, y = y, label = "*"),
    inherit.aes = FALSE,
    vjust = 0.3,
    size = 4
  )

div_plot_ext
```



6. Print session info

```
sessionInfo()
```

```
## R version 4.4.2 (2024-10-31)
## Platform: x86_64-pc-linux-gnu
## Running under: Ubuntu 20.04.6 LTS
##
## Matrix products: default
## BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.9.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.9.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C               LC_TIME=en_US.UTF-8        LC_COLLATE=en_U
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8    LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C             LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATI
##
## time zone: Etc/UTC
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] vegan_2.6-8     lattice_0.22-6  permute_0.9-7   lubridate_1.9.3 forcats_1.0.0   stringr_1.5.1
##  [8] purrr_1.0.2     readr_2.1.5     tidyr_1.3.1     tibble_3.2.1    ggplot2_3.5.1   tidyverse_2.0.0
```

```
##
## loaded via a namespace (and not attached):
##  [1] utf8_1.2.4        generics_0.1.3    stringi_1.8.4     digest_0.6.37     hms_1.1.3         magri
##  [7] evaluate_1.0.1    grid_4.4.2        timechange_0.3.0  fastmap_1.2.0     Matrix_1.7-0      tinyt
## [13] mgcv_1.9-1        fansi_1.0.6       scales_1.3.0      cli_3.6.3         rlang_1.1.4       crayo
## [19] bit64_4.5.2       munsell_0.5.1     splines_4.4.2     yaml_2.3.10       withr_3.0.1       tools
## [25] parallel_4.4.2    tzdb_0.4.0        colorspace_2.1-1  vctrs_0.6.5       R6_2.5.1          lifec
## [31] bit_4.5.0         vroom_1.6.5       MASS_7.3-61       cluster_2.1.6     pkgconfig_2.0.3   pilla
## [37] gtable_0.3.5      glue_1.8.0        highr_0.11        xfun_0.48         tidyselect_1.2.1  rstud
## [43] knitr_1.48        farver_2.1.2      htmltools_0.5.8.1 nlme_3.1-166      labeling_0.4.3    rmark
## [49] compiler_4.4.2
```