Bacterial Diversity Across Trade Biomes

Moegamad Uzair Jack

2025-06-25

# Loading Packages

library(phyloseq)  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.2 ✔ tibble 3.3.0  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.4   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(plyr)

## ------------------------------------------------------------------------------  
## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)  
## ------------------------------------------------------------------------------  
##   
## Attaching package: 'plyr'  
##   
## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize  
##   
## The following object is masked from 'package:purrr':  
##   
## compact

library(vegan)

## Loading required package: permute

library(ggplot2)  
library(grid)  
library(RColorBrewer)  
library(viridis)

## Loading required package: viridisLite

library(pals)

##   
## Attaching package: 'pals'  
##   
## The following objects are masked from 'package:viridis':  
##   
## cividis, inferno, magma, plasma, turbo, viridis  
##   
## The following objects are masked from 'package:viridisLite':  
##   
## cividis, inferno, magma, plasma, turbo, viridis

library(MASS)

##   
## Attaching package: 'MASS'  
##   
## The following object is masked from 'package:dplyr':  
##   
## select

# Set Working Directory

setwd("C:/Users/Uzair/Documents/R/Marine\_Microbial\_Ecology")

# Import OTU and Taxonomy Tables

otu\_TO <- read.table("TO\_OTU.txt", header=TRUE)  
tax\_TO <- read.delim("TO\_TAX.txt", header=TRUE)  
rownames(tax\_TO) <- tax\_TO$X  
tax\_TO$X <- NULL  
tax\_TO <- as.matrix(tax\_TO)  
otu\_TO <- otu\_table(otu\_TO, taxa\_are\_rows = TRUE)  
tax\_TO <- tax\_table(tax\_TO)  
TO <- merge\_phyloseq(otu\_TO, tax\_TO)

# Load and Merge Metadata

META2 <- read.delim("META2.txt", header=TRUE)  
META <- read.delim("ALL\_META.txt", header=TRUE)  
META\_TO <- merge(META2, META, by="PANGAEA")  
META\_TO <- column\_to\_rownames(META\_TO, "PANGAEA")  
MAP <- sample\_data(META\_TO)  
TO1 <- merge\_phyloseq(TO, MAP)

# Subset Bacteria in Trades Biome and DCM

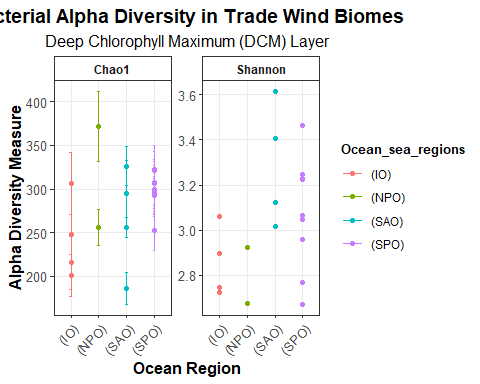
TO.bacteria <- subset\_taxa(TO1, Kingdom == "Bacteria")  
TO.trades.bac <- subset\_samples(TO.bacteria, Marine\_pelagic\_biomes\_Longhurst\_2007 == "Trades")  
TO.trades.bac.DCM <- subset\_samples(TO.trades.bac, EnvironmentalFeature== "(DCM)")  
TO.trades.bac.DCM <- subset\_samples(TO.trades.bac.DCM, !is.na(Mean\_Salinity\_PSU))

# Normalize to Relative Abundance

TO.trades.bac.DCM.rel <- transform\_sample\_counts(TO.trades.bac.DCM, function(x) x / sum(x))  
TO.trades.bac.DCM.rel.mean <- filter\_taxa(TO.trades.bac.DCM.rel, function(x) mean(x) > 0, TRUE)

# Alpha Diversity Comparison

# Calculate alpha diversity metrics  
plot\_richness(TO.trades.bac.DCM, x = "Ocean\_sea\_regions",   
 color = "Ocean\_sea\_regions",   
 measures = c("Shannon", "Chao1")) +  
 theme\_bw() +  
 theme(  
 axis.text.x = element\_text(angle = 45, hjust = 1, size = 10),  
 axis.text.y = element\_text(size = 10),  
 axis.title = element\_text(size = 12, face = "bold"),  
 plot.title = element\_text(size = 14, face = "bold", hjust = 0.5),  
 plot.subtitle = element\_text(size = 12, hjust = 0.5),  
 legend.title = element\_text(size = 10, face = "bold"),  
 panel.grid.minor = element\_blank(),  
 strip.background = element\_rect(fill = "white"),  
 strip.text = element\_text(face = "bold")  
 ) +  
 labs(  
 title = "Bacterial Alpha Diversity in Trade Wind Biomes",  
 subtitle = "Deep Chlorophyll Maximum (DCM) Layer",  
 x = "Ocean Region",  
 y = "Alpha Diversity Measure"  
 )



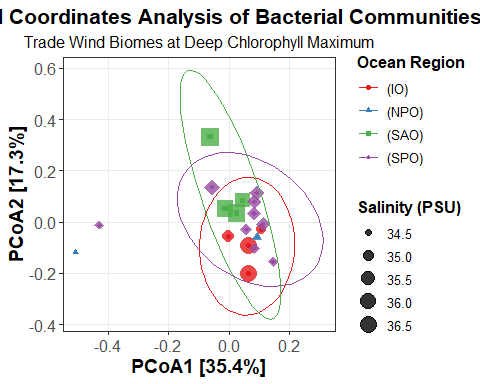
# Ordination: PCoA

# Perform PCoA ordination  
GP.ord <- ordinate(TO.trades.bac.DCM.rel.mean, method = "PCoA", distance = "bray")  
  
# Extract percent variance explained for axes 1 and 2  
pcoa\_var <- GP.ord$values$Relative\_eig \* 100  
x\_lab <- paste0("PCoA1 [", round(pcoa\_var[1], 1), "%]")  
y\_lab <- paste0("PCoA2 [", round(pcoa\_var[2], 1), "%]")  
  
# Create enhanced plot with ellipses and salinity  
plot\_ordination(TO.trades.bac.DCM.rel.mean, GP.ord, type = "samples",   
 color = "Ocean\_sea\_regions", shape = "Ocean\_sea\_regions") +  
 geom\_point(aes(size = Mean\_Salinity\_PSU), alpha = 0.8) +  
 stat\_ellipse(aes(group = Ocean\_sea\_regions), type = "t", level = 0.95) +  
 scale\_color\_brewer(palette = "Set1", name = "Ocean Region") +  
 scale\_shape\_manual(values = c(16, 17, 15, 18, 19), name = "Ocean Region") +  
 scale\_size\_continuous(name = "Salinity (PSU)") +  
 theme\_bw() +  
 theme(  
 axis.text = element\_text(size = 12),  
 axis.title = element\_text(size = 14, face = "bold"),  
 plot.title = element\_text(size = 16, face = "bold", hjust = 0.5),  
 plot.subtitle = element\_text(size = 12, hjust = 0.5),  
 legend.title = element\_text(size = 12, face = "bold"),  
 legend.text = element\_text(size = 10),  
 panel.grid.minor = element\_blank()  
 ) +  
 labs(  
 title = "Principal Coordinates Analysis of Bacterial Communities",  
 subtitle = "Trade Wind Biomes at Deep Chlorophyll Maximum",  
 x = x\_lab,  
 y = y\_lab  
 )

## Too few points to calculate an ellipse

## Warning in MASS::cov.trob(data[, vars]): Probable convergence failure

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_path()`).

 # Beta Diversity Comparison (Ocean\_sea\_regions)

# Calculate beta diversity using Bray-Curtis dissimilarity  
gp\_dist <- distance(TO.trades.bac.DCM.rel.mean, method = "bray")  
# Extract metadata for PERMANOVA  
metadata <- data.frame(sample\_data(TO.trades.bac.DCM.rel.mean))  
# Perform PERMANOVA  
gp\_permanova <- adonis2(gp\_dist ~ Ocean\_sea\_regions,  
 data = metadata,  
 permutations = 999)  
# Print PERMANOVA results  
print(gp\_permanova)

## Permutation test for adonis under reduced model  
## Permutation: free  
## Number of permutations: 999  
##   
## adonis2(formula = gp\_dist ~ Ocean\_sea\_regions, data = metadata, permutations = 999)  
## Df SumOfSqs R2 F Pr(>F)   
## Model 3 0.36256 0.23466 1.5331 0.084 .  
## Residual 15 1.18248 0.76534   
## Total 18 1.54504 1.00000   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# NMDS Ordination

# Convert phyloseq object to matrix for NMDS  
otu\_matrix <- as(otu\_table(TO.trades.bac.DCM.rel.mean), "matrix")  
if(taxa\_are\_rows(TO.trades.bac.DCM.rel.mean)) {  
 otu\_matrix <- t(otu\_matrix)  
}  
  
# Perform NMDS ordination  
nmds\_result <- metaMDS(otu\_matrix, distance = "bray", k = 2, trymax = 100, trace = FALSE)  
  
# Create ordination object compatible with phyloseq  
nmds\_coords <- data.frame(nmds\_result$points)  
rownames(nmds\_coords) <- sample\_names(TO.trades.bac.DCM.rel.mean)  
ord\_obj <- ordinate(TO.trades.bac.DCM.rel.mean, "NMDS", distance = "bray")

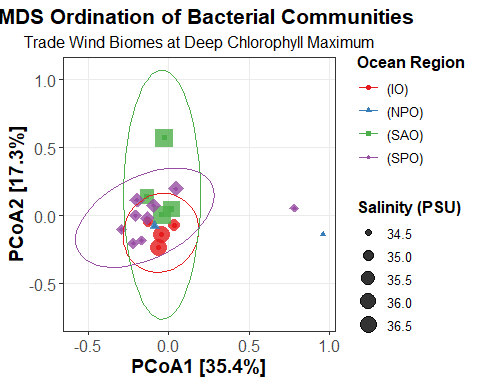
## Run 0 stress 0.1023887   
## Run 1 stress 0.0981361   
## ... New best solution  
## ... Procrustes: rmse 0.04079133 max resid 0.1194201   
## Run 2 stress 0.101471   
## Run 3 stress 0.09744074   
## ... New best solution  
## ... Procrustes: rmse 0.01171933 max resid 0.03805675   
## Run 4 stress 0.09744076   
## ... Procrustes: rmse 0.0001465055 max resid 0.0002750751   
## ... Similar to previous best  
## Run 5 stress 0.09744083   
## ... Procrustes: rmse 0.0002372733 max resid 0.0005152296   
## ... Similar to previous best  
## Run 6 stress 0.09744073   
## ... New best solution  
## ... Procrustes: rmse 6.400704e-05 max resid 0.0001726045   
## ... Similar to previous best  
## Run 7 stress 0.1014708   
## Run 8 stress 0.09744106   
## ... Procrustes: rmse 0.0003699771 max resid 0.001217456   
## ... Similar to previous best  
## Run 9 stress 0.09744077   
## ... Procrustes: rmse 0.0001249258 max resid 0.0003568676   
## ... Similar to previous best  
## Run 10 stress 0.09813584   
## Run 11 stress 0.1014744   
## Run 12 stress 0.09813585   
## Run 13 stress 0.0981358   
## Run 14 stress 0.09813596   
## Run 15 stress 0.3744249   
## Run 16 stress 0.09744073   
## ... New best solution  
## ... Procrustes: rmse 6.726693e-05 max resid 0.000232733   
## ... Similar to previous best  
## Run 17 stress 0.09744091   
## ... Procrustes: rmse 0.000285516 max resid 0.0006936628   
## ... Similar to previous best  
## Run 18 stress 0.1014748   
## Run 19 stress 0.1014742   
## Run 20 stress 0.09744073   
## ... New best solution  
## ... Procrustes: rmse 1.697926e-05 max resid 3.241768e-05   
## ... Similar to previous best  
## \*\*\* Best solution repeated 1 times

# Create NMDS plot with ellipses and salinity  
plot\_ordination(TO.trades.bac.DCM.rel.mean, ord\_obj, type = "samples",  
 color = "Ocean\_sea\_regions", shape = "Ocean\_sea\_regions") +  
 geom\_point(aes(size = Mean\_Salinity\_PSU), alpha = 0.8) +  
 stat\_ellipse(aes(group = Ocean\_sea\_regions), type = "t", level = 0.95) +  
 scale\_color\_brewer(palette = "Set1", name = "Ocean Region") +  
 scale\_shape\_manual(values = c(16, 17, 15, 18, 19), name = "Ocean Region") +  
 scale\_size\_continuous(name = "Salinity (PSU)") +  
 theme\_bw() +  
 theme(  
 axis.text = element\_text(size = 12),  
 axis.title = element\_text(size = 14, face = "bold"),  
 plot.title = element\_text(size = 16, face = "bold", hjust = 0.5),  
 plot.subtitle = element\_text(size = 12, hjust = 0.5),  
 legend.title = element\_text(size = 12, face = "bold"),  
 legend.text = element\_text(size = 10),  
 panel.grid.minor = element\_blank()  
 ) +  
 labs(  
 title = "NMDS Ordination of Bacterial Communities",  
 subtitle = "Trade Wind Biomes at Deep Chlorophyll Maximum",  
 x = x\_lab,  
 y = y\_lab  
 )

## Too few points to calculate an ellipse

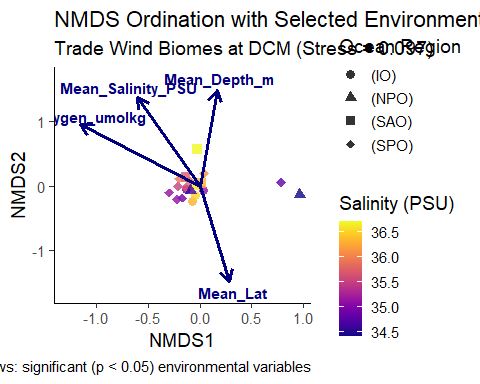
## Warning in MASS::cov.trob(data[, vars]): Probable convergence failure

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_path()`).



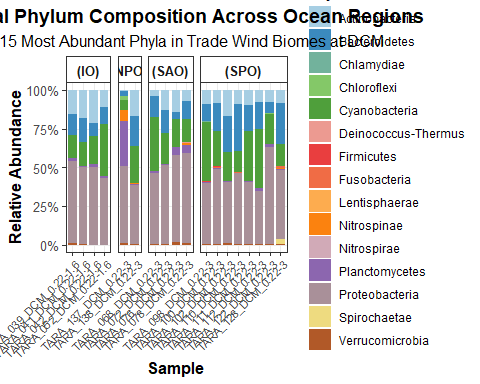
# Environmental Fitting (NMDS + envfit)

# Select only the desired environmental variables for envfit  
env\_vars <- c(  
 "Mean\_Lat", "Mean\_Long", "Mean\_Depth\_m", "Mean\_Temperature\_degC",  
 "Mean\_Salinity\_PSU", "Mean\_Oxygen\_umolkg", "Mean\_Nitrates\_umolL"  
)  
psv <- data.frame(sample\_data(TO.trades.bac.DCM.rel.mean))  
psv\_selected <- psv[, env\_vars]  
  
# NMDS ordination as before  
psotu2veg <- function(physeq) {  
 OTU <- otu\_table(physeq)  
 if (taxa\_are\_rows(OTU)) OTU <- t(OTU)  
 as(OTU, "matrix")  
}  
ps.otu.veg <- psotu2veg(TO.trades.bac.DCM.rel.mean)  
vare.mds <- metaMDS(ps.otu.veg, trace = FALSE)  
  
# Environmental fitting with only selected variables  
ef <- envfit(vare.mds, psv\_selected, permu = 999, na.rm = TRUE)  
  
# Extract NMDS scores and metadata for plotting  
nmds\_scores <- as.data.frame(scores(vare.mds, display = "sites"))  
nmds\_scores$Salinity <- psv$Mean\_Salinity\_PSU  
nmds\_scores$Region <- psv$Ocean\_sea\_regions  
  
# Only plot significant vectors (p < 0.05)  
sig\_vec <- ef$vectors$pvals < 0.05  
vecs <- as.data.frame(ef$vectors$arrows[sig\_vec, , drop = FALSE])  
vecs$var <- rownames(vecs)  
arrow\_mult <- 1.5  
vecs$NMDS1 <- vecs$NMDS1 \* arrow\_mult  
vecs$NMDS2 <- vecs$NMDS2 \* arrow\_mult  
  
# Create NMDS plot with environmental vectors  
ggplot(nmds\_scores, aes(x = NMDS1, y = NMDS2, color = Salinity, shape = Region)) +  
 geom\_point(size = 3, alpha = 0.8) +  
 scale\_color\_viridis(option = "plasma", name = "Salinity (PSU)") +  
 scale\_shape\_manual(values = c(16, 17, 15, 18, 19), name = "Ocean Region") +  
 geom\_segment(  
 data = vecs,  
 aes(x = 0, y = 0, xend = NMDS1, yend = NMDS2),  
 arrow = arrow(length = unit(0.35, "cm")), color = "navy", linewidth = 1.1, inherit.aes = FALSE  
 ) +  
 geom\_text(  
 data = vecs,  
 aes(x = NMDS1 \* 1.12, y = NMDS2 \* 1.12, label = var),  
 color = "navy", size = 4, fontface = "bold", hjust = 0.5, inherit.aes = FALSE  
 ) +  
 theme\_classic(base\_size = 14) +  
 labs(  
 title = "NMDS Ordination with Selected Environmental Vectors",  
 subtitle = paste0("Trade Wind Biomes at DCM (Stress = ", round(vare.mds$stress, 3), ")"),  
 x = "NMDS1",  
 y = "NMDS2",  
 caption = "Arrows: significant (p < 0.05) environmental variables"  
 )



# Barplot of Dominant Bacterial Phyla by Ocean

# First identify samples that are truly DCM and exclude TARA\_100\_MES\_0.22-3  
truly\_dcm <- subset\_samples(TO.trades.bac.DCM.rel.mean, EnvironmentalFeature == "(DCM)" &   
 !Sample\_label %in% c("TARA\_100\_MES\_0.22-3"))  
  
# Aggregate at Phylum level and get top 10 phyla using the filtered dataset  
PhylumGlommed <- tax\_glom(truly\_dcm, "Phylum")  
top\_phyla <- names(sort(taxa\_sums(PhylumGlommed), decreasing = TRUE)[1:15])  
PhylumGlommed\_filtered <- prune\_taxa(top\_phyla, PhylumGlommed)  
  
# Create the plot with the filtered dataset  
ggplot(data = psmelt(PhylumGlommed\_filtered),   
 aes(x = Sample\_label, y = Abundance, fill = Phylum)) +  
 geom\_bar(stat = "identity", position = "stack", width = 0.8) +  
 facet\_grid(~Ocean\_sea\_regions, scales = "free\_x", space = "free") +  
 scale\_fill\_manual(values = colorRampPalette(brewer.pal(12, "Paired"))(15)) +  
 scale\_y\_continuous(labels = scales::percent\_format()) +  
 theme\_bw() +  
 theme(  
 axis.text.x = element\_text(angle = 45, hjust = 1, size = 8),  
 axis.text.y = element\_text(size = 10),  
 axis.title = element\_text(size = 12, face = "bold"),  
 plot.title = element\_text(size = 14, face = "bold", hjust = 0.5),  
 plot.subtitle = element\_text(size = 12, hjust = 0.5),  
 legend.title = element\_text(size = 10, face = "bold"),  
 legend.text = element\_text(size = 9),  
 legend.position = "right",  
 panel.grid.minor = element\_blank(),  
 strip.text = element\_text(size = 10, face = "bold"),  
 strip.background = element\_rect(fill = "white")  
 ) +  
 labs(  
 title = "Bacterial Phylum Composition Across Ocean Regions",  
 subtitle = "Top 15 Most Abundant Phyla in Trade Wind Biomes at DCM",  
 x = "Sample",  
 y = "Relative Abundance"  
 )



# Save Plots

# Create directory for plots if it doesn't exist  
dir.create("plots", showWarnings = FALSE)  
  
# Alpha diversity plot  
ggsave("plots/alpha\_diversity.png",   
 plot = last\_plot(),   
 width = 12,   
 height = 8,   
 dpi = 300)  
  
# PCoA plot  
pcoa\_plot <- plot\_ordination(TO.trades.bac.DCM.rel.mean, GP.ord, type = "samples",   
 color = "Ocean\_sea\_regions", shape = "Ocean\_sea\_regions")  
ggsave("plots/pcoa\_plot.png",   
 plot = pcoa\_plot,   
 width = 10,   
 height = 8,   
 dpi = 300)  
  
# NMDS plot  
nmds\_plot <- plot\_ordination(TO.trades.bac.DCM.rel.mean, ord\_obj, type = "samples",  
 color = "Ocean\_sea\_regions", shape = "Ocean\_sea\_regions")  
ggsave("plots/nmds\_plot.png",   
 plot = nmds\_plot,   
 width = 10,   
 height = 8,   
 dpi = 300)  
  
# Environmental fitting plot  
ggsave("plots/envfit\_plot.png",   
 plot = last\_plot(),   
 width = 10,   
 height = 8,   
 dpi = 300)  
  
# Phylum composition barplot  
ggsave("plots/phylum\_barplot.png",   
 plot = last\_plot(),   
 width = 15,   
 height = 8,   
 dpi = 300)