# raber_550_introductory_analysis

*Austin Hammer*

*9/12/2019*

### R Markdown

library(“ggplot2”) library(“phangorn”) library(“phyloseq”) library(“ape”) library(“Rcpp”) library(“readxl”) library(“dplyr”) library(“vegan”)

## Setting ggplot2 theme

theme_set(theme_bw())

## I want to read in the file and just figure out what it looks like after unpacking it from the RDS file format

seqtab <- readRDS(“C:/Users/hammera/Desktop/raber_550/data/F_280_reads_seqtab_final.rds”)

seqtab.taxa <- readRDS(“C:/Users/hammera/Desktop/raber_550/data/F_280_reads_taxa_final.rds”)

View(seqtab.taxa)

## check that number of ASVs and taxa rows is the same

ASV_taxa_equal <- ncol(seqtab) == nrow(seqtab.taxa)

print(ASV_taxa_equal)

## read pdf metadata in as a data_frame

meta_data <- as.data.frame(read_excel(“C:/Users/hammera/Desktop/raber_550/data/550_metadata.xlsx”)) rownames(seqtab) <- meta_data$RaberSID

## set data equal to phyloseq object parameters

samples = sample_data(meta_data) ASV = otu_table(seqtab, taxa_are_rows = FALSE) TAX = tax_table(seqtab.taxa)

## I had an issue getting the sample names the same and into phyloseq

sample_names(samples) <- sample_names(ASV)

## create a phyloseq object

physeq = phyloseq(ASV, TAX, samples)

## if you want to prune a weird sample this is one way

## sample 93 was wonky in terms of # of reads, and diversity, while 59 had some metadata issues

## I ran alpha diversity measures including both of these samples and without, but there was

## no statistically significant association between radiation and alpha diversity when they were

## included vs when they weren't

physeq = subset_samples(physeq, sample_names(physeq) != "93") physeq = subset_samples(physeq, sample_names(physeq) != "59")

## makes a histogram of the reads, and tests for normality using Shapiro-Wilks

## result yields a p-value of 0.44 and suggests that the the data is from a normal distribution

hist(sample_sums(physeq)) shapiro.test(sample_sums(physeq))

## Change the alpha diversity measure, assign it to a matrix, use shapiro.test to test for normality of richness

## only a couple of alpha diversity measures followed a normal distribution

sample_data(physeq1)$Richness <- estimate_richness(physeq1, split=TRUE, measures=c("Shannon"))

alphdiv <- as.matrix(sample_data(physeq1)$Richness)

shapiro.test(alphdiv)

## start visualizing richness and looking for how richness might vary as a function of radiation status

## I categorized each Radation exposure, and used the MWU test to look for connections with alpha diversity

## Simpson, Shannon, ACE, Chao1, InvSimpson, and Fisher were all evaluated

physeq1 <- physeq

sample_data(physeq1)$Radiated <- get_variable(physeq1, "Treatment") %in% c("25", "50", "200")

sample_data(physeq1)$Richness <- estimate_richness(physeq1, split=TRUE, measures=c("Shannon"))

alphadiv <- as.matrix(sample_data(physeq1)$Richness)

radiated <- sample_data(physeq1)$Radiated

## Looking at alpha diversity (Shannon) as a function of radiation

wilcox.test(alphadiv ~ radiated)

## here's a box plot of the MWU test above

## as well as a boxplot that shows alphadiv for each dose of radiation

boxplot(alphadiv ~ radiated, color=sample_data(physeq1)$Treatment$) $stripchart(alphadiv\ radiated, vertical = TRUE, method = "jitter", add = TRUE, pch = 20, col = "blue")boxplot(alphadiv\ sample_data(physeq1)$Treatment)
stripchart(alphadiv ~ sample_data(physeq1)$Treatment, vertical = TRUE, method = "jitter", add = TRUE, pch = 20, col ="blue")

## ordination by radiation

## coloring based on the "Radiated" sample data didn't reveal much

## F vs F+R also doesn't appear to change much visually, but check that

## with some real numbers/analysis

physeq1.ord <- ordinate(physeq1, "NMDS", "bray", weighted=TRUE) physeq1.ord.plot <- plot_ordination(physeq1, physeq1.ord, type="samples", color="Sex", title="ByRadiation") print(physeq1.ord.plot)

**using PERMANOVA(adonis) on a distance matrix, perms=999**

**I looked at the result of running physeq1.dist as a function of each**

**covariate and recorded some the data on Wed 9/18/19 in my lab notebook**

**I also tried to look at multiple variables at once by using the + sign**

**after the independent variable (Treatment + Sex, Treatment + Box, etc)**

physeq1.dist <- vegdist(otu_table(physeq1), method="bray") sampledf <- data.frame(sample_data(physeq1)) set.seed(1) perm.results.treatment <- adonis(physeq1.dist ~ Treatment, data=sampledf, permutations=5000) perm.results.treatment perm.results.sex <- adonis(physeq1.dist ~ Box, data=sampledf) perm.results.sex