elephants_ALPHA_DIVERSITY

Joe Gunn

8/13/2019

AFRICAN ELEPHANT MICROBIOME - ALPHA DIVERSITY

Purpose: Here we are assessing differences in alpha diversity between African Elephant species, diets, and habitats by comparing Shannon Diversity within treatment groups. We used linear mixed effect models to, including elephant sex as a random effect, to compare Shannon diversity.

Data used:

```
  TABLE output from QIIME: all OTUs detected across African Elephant sampels, taxonomic level, and raw abundance
for each individual <- 8,248 OTUs (after rarefaction)
```

Libraries needed for analysis

Metadata

Rarefied OTU abundance data

```
options(scipen = 999) #gets rid of scientific notatation, useful for reading p-values

#read in taxonomic designations and OTU abundance
aem_physeq <- qza_to_phyloseq(features = "../data/qiime_data/table.qza", taxonomy = "../data/qiime_data/taxonomy.
qza",  tree = "../data/qiime_data/rooted-tree.qza", metadata = "../data/metadata/METADATA.tsv")

#omit sample OB182 due to low read count. We don't want to rarefy to such a low number of reads across the whole
dataset
aem_physeq <- subset_samples(aem_physeq, Elephant != "OB182")

#Rarefy and clean the data set for use in downstream analysis
aem_physeq <- rarefy_even_depth(aem_physeq, rngseed = 5) #must set seed in order to maintain the same rarefied num
ber of otus
```

```
## `set.seed(5)` was used to initialize repeatable random subsampling.
```

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

```
## Try `set.seed(5); .Random.seed` for the full vector
```

```
## ...
```

```
## 818OTUs were removed because they are no longer
## present in any sample after random subsampling
```

```
## ...
```

```
    #Number of OTUS: 8248
    #Number of reads per sample: 11460

#Extract data of all types (meta data, all taxonomy data)
aem_meta <- as.data.frame(sample_data(aem_physeq))
aem_tax <- as.data.frame(tax_table(aem_physeq))
aem_tax <- aem_tax %>%
  rownames_to_column("otu")

aem_data <- as.data.frame(otu_table(aem_physeq))
aem_data <- aem_data %>%
  rownames_to_column("otu")

aem_tax_data <- merge(aem_tax, aem_data, by = "otu")
```

Subset the data into only phylum Bacterioidetes, Firmicutes, and Proteobacteria

```
#Subset the data by the three most abundant phyla: Bacteroidetes, firmicutes, proteobacteria
aem_bacteroidetes <- aem_tax_data %>%
  filter(Phylum == "p__Bacteroidetes")

aem_firmicutes <- aem_tax_data %>%
  filter(Phylum == "p__Firmicutes")

aem_proteobacteria <- aem_tax_data %>%
  filter(Phylum == "p__Proteobacteria")
```

Prepare and clean data for two datasets

```
#Prepare data at the OTU level for analysis with both datasets (just forest non-cropraiders from both species, an
d just africana)
aem_physeq <- as.data.frame(otu_table(aem_physeq))
aem_physeq_t <- t(aem_physeq)
aem_physeq <- as.data.frame(aem_physeq_t)
aem_physeq <- aem_physeq %>%
  rownames_to_column("sample_id")

aem_physeq <- merge(metadata, aem_physeq, by = "sample_id")
aem_physeq_forest_ncr <- aem_physeq %>%
  filter(Habitat == "Forest") %>%
  filter(Raider == "No")

aem_physeq_africana <- aem_physeq %>%
  filter(Species == "africana")

# Prepare data for Shannon Diversity FOR SPECIES
aem_veg_data <- aem_physeq_forest_ncr[,-c(2:11)]
aem_veg_data <- column_to_rownames(aem_veg_data, "sample_id")

aem_veg2_data <- aem_physeq_africana[,-c(2:11)]
aem_veg2_data <- column_to_rownames(aem_veg2_data, "sample_id")
```

ALPHA DIVERSITY

Alpha Diversity - Shannon Diversity Index

```
#Analysis of Shannon Diversity
shan_div_species <- diversity(aem_veg_data, "shannon")
shan_div_diet <- diversity(aem_veg2_data, "shannon")

#Prepare data for linear models
shan_div_species <- as.data.frame(shan_div_species)
shan_div_species <- shan_div_species %>%
  rownames_to_column("sample_id")

colnames(shan_div_species) <- c("sample_id", "shannon")
shan_div_species <- merge(metadata, shan_div_species, by = "sample_id")

#Prepare data for linear models
shan_div_diet <- as.data.frame(shan_div_diet)
shan_div_diet <- shan_div_diet %>%
  rownames_to_column("sample_id")

colnames(shan_div_diet) <- c("sample_id", "shannon")
shan_div_diet <- merge(metadata, shan_div_diet, by = "sample_id")

#Averages and standard deviations of Shannon Diversity for species, diet, and habitat
shan_div_habitat_aves <- shan_div_diet %>%
  group_by(Habitat) %>%
  summarize(mean = mean(shannon))

shan_div_habitat_sd <- shan_div_diet %>%
  group_by(Habitat) %>%
  summarize(sd = sd(shannon))

shan_div_species_aves <- shan_div_species %>%
  group_by(Species) %>%
  summarize(mean = mean(shannon))

shan_div_species_sd <- shan_div_species %>%
  group_by(Species) %>%
  summarize(sd = sd(shannon))

shan_div_diet_aves <- shan_div_diet %>%
  group_by(Raider) %>%
  summarize(mean = mean(shannon))

shan_div_diet_sd <- shan_div_diet %>%
  group_by(Raider) %>%
  summarize(sd = sd(shannon))
```

Linear Models of alpha diversity with Species data set

```
#one way fixed effect for sex using species data set
lm_sex_species <- lm(shannon ~ Sex, data = shan_div_species)  #p = 0.821

#Various linear mixed effect models with different random effects just for fun
lme_species_sex <- lme(shannon ~ Species, random = ~1|Sex, data = shan_div_species)
lme_int_diet_hab_sex <- lme(shannon ~ Raider*Habitat, random = ~1|Sex, data = shan_div_diet)
lme_diet_sex <- lme(shannon ~ Raider, random = ~1|Sex, data = shan_div_diet)
lme_habitat_sex <- lme(shannon ~ Habitat, random = ~1|Sex, data = shan_div_diet)
```

Visualize Alpha Diversity. We can add a lot of graphs here, and we can compare all kinds of groups within groups.

```
levels(shan_div_species$Species) <- c("L. africana", "L. cyclotis") #This ensures that the labels are in the corr
ect order

alpha_species <- ggplot(shan_div_species, aes(x = Species, y = shannon, fill = Species)) +
  geom_boxplot(outlier.shape = NA, show.legend = F, alpha = 0.8) +
  geom_point(position = position_dodge(width = 0.75), show.legend = F, size = 2) +
  theme_set(theme_cowplot(12)) +
  labs(x = "Species", y = "Shannon Diversity") +
  scale_fill_manual(values = c("red", "blue")) +
  theme(axis.text.x = element_text(face = "italic")) +
  theme(axis.text = element_text(size = 20)) +
  theme(axis.title = element_text(size = 20))

levels(shan_div_diet$Raider) <- c("Non-Crop-Raider", "Crop-Raider")

alpha_diet <- ggplot(shan_div_diet, aes(x = Raider, y = shannon, fill = Raider)) +
  geom_boxplot(outlier.shape = NA, show.legend = F, alpha = 0.8) +
  geom_point(position = position_dodge(width = 0.75), show.legend = F, size = 2) +
  theme_set(theme_cowplot(12)) +
  labs(x = "Diet", y = "Shannon Diversity") +
  scale_fill_manual(values = c("midnightblue", "deeppink1")) +
  theme(axis.text = element_text(size = 20)) +
  theme(axis.title = element_text(size = 20))

alpha_habitat <- ggplot(shan_div_diet, aes(x = Habitat, y = shannon, fill = Habitat)) +
  geom_boxplot(outlier.shape = NA, show.legend = F, alpha = 0.8) +
  geom_point(position = position_dodge(width = 0.75), show.legend = F, size = 2) +
  theme_set(theme_cowplot(12)) +
  labs(x = "Habitat", y = "Shannon Diversity") +
  scale_fill_manual(values = c("forestgreen", "coral1")) +
  theme(axis.text = element_text(size = 20)) +
  theme(axis.title = element_text(size = 20))

#Combine all plots into one figure
pdf("/Users/joegunn/Desktop/Grad_School_Stuff/Research/Projects/Elephant_Microbiome/Attempt_2/visualization/alpha
_diversity_figures/alpha_diversity.pdf", width=8, height=14)

plot_grid(alpha_species, alpha_diet, alpha_habitat, nrow = 3, labels = c("A", "B", "C"), label_size = 20)

dev.off()
```

```
## quartz_off_screen
##                 2
```