elephants_METABOLISM

Joe Gunn

8/13/2019

AFRICAN ELEPHANT MICROBIOME - METABOLISM

Purpose: Here we are assessing overall differences in metabolic network structure using beta diversity analysis (PERMANOVA) and in individual KEGG metabolic pathways using linear models between African Elephant species, diets, and habitats.

Data used:

```
KEGG Pathway output data from QIIME.
```

Libraries needed for analysis

Metadata

KEGG PATHWAYS

```
#Metadata by species (kegg_l_species) and by habitat and range (kegg_l_africana) for KEGG Pathways at level 1 metabolism
kegg_l_all <- read_excel("../data/excel_data/metabolism/kegg_l.xlsx")

kegg_l_meta <- kegg_l_all[,-1]
kegg_l_meta <- kegg_l_meta %>%
  group_by(Metabolism_l) %>%
  summarize_all(funs(sum))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

```
kegg_l_meta <- as.data.frame(kegg_l_meta)
kegg_l_meta <- column_to_rownames(kegg_l_meta, "Metabolism_l")
kegg_l_meta <- t(kegg_l_meta)
kegg_l_meta <- as.data.frame(kegg_l_meta)
kegg_l_meta <- rownames_to_column(kegg_l_meta, "sample_id")
kegg_l_meta <- merge(metadata, kegg_l_meta, by = "sample_id")

kegg_l_species <- kegg_l_meta %>%
  filter(Habitat == "Forest") %>%
  filter(Raider == "No")

kegg_l_africana <- kegg_l_meta %>%
  filter(Species == "africana")

kegg_l_species <- column_to_rownames(kegg_l_species, "sample_id")

#Preparing Vegan matrix for both datasets
kegg_l <- kegg_l_all[,-1]
kegg_l <- kegg_l %>% group_by(Metabolism_l) %>%
  summarize_all(funs(sum))

kegg_l <- as.data.frame(kegg_l)
kegg_l <- column_to_rownames(kegg_l, "Metabolism_l")
kegg_l <- t(kegg_l)
kegg_l <- as.data.frame(kegg_l)
kegg_l <- rownames_to_column(kegg_l, "sample_id")
kegg_l <- merge(metadata, kegg_l, by = "sample_id")

kegg_l_by_species <- kegg_l %>%
  filter(Habitat == "Forest") %>%
  filter(Raider == "No")

kegg_l_by_species <- kegg_l_by_species[,-c(2:11)]
kegg_l_by_species <- column_to_rownames(kegg_l_by_species, "sample_id")
kegg_l_by_africana <- kegg_l %>%
  filter(Species == "africana")

kegg_l_by_africana <- kegg_l_by_africana[,-c(2:11)]
kegg_l_by_africana <- column_to_rownames(kegg_l_by_africana, "sample_id")

#BETA DIVERSITY

#Run Vegan
kegg_l_veg_species_stand <- decostand(kegg_l_by_species, "total")
kegg_l_veg_species <- vegdist(kegg_l_veg_species_stand, "bray")

kegg_l_veg_africana_stand <- decostand(kegg_l_by_africana, "total")
kegg_l_veg_africana <- vegdist(kegg_l_veg_africana_stand, "bray")

#Beta Diversity with PERMANOVA
ado_kegg_species_age <- adonis(kegg_l_veg_species ~ Species, kegg_l_species, Strata = Age, distance = "bray", permutations = 9999)
ado_kegg_species_sex <- adonis(kegg_l_veg_species ~ Species, kegg_l_species, Strata = Sex, distance = "bray", permutations = 9999)

ado_kegg_diet_hab_age <- adonis(kegg_l_veg_africana ~ Raider*Habitat, kegg_l_africana, Strata = Age, distance = "bray", permutations = 9999)
ado_kegg_diet_hab_sex <- adonis(kegg_l_veg_africana ~ Raider*Habitat, kegg_l_africana, Strata = Sex, distance = "bray", permutations = 9999)
```

Run linear models for each pathway to determine differences among species, diets, and habitats

```
kegg_l_props_species <- read_excel("../data/excel_data/metabolism/kegg_l_props_species.xlsx")
kegg_l_props_species <- as.data.frame(kegg_l_props_species)

kegg_l_props_species <- kegg_l_props_species %>%
  mutate(Species = factor(Species), Raider = factor(Raider), Habitat = factor(Habitat), Age = factor(Age), Sex = factor(Sex), Elephant = factor(Elephant), Group = factor(Group), Description = factor(Description))

kegg_l_props_africana <- read_excel("../data/excel_data/metabolism/kegg_l_props_africana.xlsx")
kegg_l_props_africana <- kegg_l_props_africana %>%
  as.data.frame() %>%
  mutate(Species = factor(Species), Raider = factor(Raider), Habitat = factor(Habitat), Age = factor(Age), Sex = factor(Sex), Elephant = factor(Elephant), Group = factor(Group), Description = factor(Description))


#By Species
p_vals_species <- data.frame("A" = numeric(1),  "B" = numeric(1),  "C" = numeric(1),
                             "D" = numeric(1),  "E" = numeric(1),  "F" = numeric(1),
                             "G" = numeric(1),  "H" = numeric(1),  "I" = numeric(1),
                             "J" = numeric(1),  "K" = numeric(1))

#iterate by column index of the data frame

options(scipen = 999) #this removes scientific notation from p-values, makes them easier to read

    for (ii in 12:ncol(kegg_l_props_species)) {

      col = kegg_l_props_species[,ii] #this tells the for loop that it should iterate through       columns in the data set
      lm_temp <- lm(col ~ Species, data = kegg_l_props_species) #tells the for loop to run        linear model for each column specified above
      p_value <- summary(lm_temp)$coefficients[2,4] #extracts p value
      p_vals_species[,ii-11] <- p_value #stores p value in data frame constructed above. Need to       subtract 11 from the iterator so that the loop            doesn't fill starting at row 12
    }


#By habitat
p_vals_habitat <- data.frame("A" = numeric(1),  "B" = numeric(1),  "C" = numeric(1),
                             "D" = numeric(1),  "E" = numeric(1),  "F" = numeric(1),
                             "G" = numeric(1),  "H" = numeric(1),  "I" = numeric(1),
                             "J" = numeric(1),  "K" = numeric(1))

#iterate by column index of the data frame

options(scipen = 999) #this removes scientific notation from p-values, makes them easier to read

    for (ii in 12:ncol(kegg_l_props_africana)) {

      col = kegg_l_props_africana[,ii] #this tells the for loop that it should iterate through       columns in the data set
      lm_temp <- lm(col ~ Habitat, data = kegg_l_props_africana) #tells the for loop to run        linear model for each column specified above
      p_value <- summary(lm_temp)$coefficients[2,4] #extracts p value
      p_vals_habitat[,ii-11] <- p_value #stores p value in data frame constructed above. Need to       subtract 11 from the iterator so that the loop            doesn't fill starting at row 12
    }

p_vals_habitat_fdr <- p.adjust(p_vals_habitat, method = "fdr")


#By Diet
p_vals_diet <- data.frame("A" = numeric(1),  "B" = numeric(1),  "C" = numeric(1),
                             "D" = numeric(1),  "E" = numeric(1),  "F" = numeric(1),
                             "G" = numeric(1),  "H" = numeric(1),  "I" = numeric(1),
                             "J" = numeric(1),  "K" = numeric(1))

options(scipen = 999) #this removes scientific notation from p-values, makes them easier to read

    for (ii in 12:ncol(kegg_l_props_africana)) {

      col = kegg_l_props_africana[,ii] #this tells the for loop that it should iterate through       columns in the data set
      lm_temp <- lm(col ~ Raider, data = kegg_l_props_africana) #tells the for loop to run        linear model for each column specified above
      p_value <- summary(lm_temp)$coefficients[2,4] #extracts p value
      p_vals_diet[,ii-11] <- p_value #stores p value in data frame constructed above. Need to       subtract 11 from the iterator so that the loop            doesn't fill starting at row 12
    }

p_vals_diet_fdr <- p.adjust(p_vals_diet, method = "fdr")

##Species was the only factor that contained any signficant pathways, so we chose to only plot the metabolic pathways for this analysis.

p_vals_species_fdr <- p.adjust(p_vals_species, method ="fdr")
p_vals_species_fdr <- as.data.frame(p_vals_species_fdr)
p_vals_species_fdr <- t(p_vals_species_fdr)
p_vals_species_fdr <- t(p_vals_species_fdr)
p_vals_species_fdr <- as.data.frame(p_vals_species_fdr)
p_vals_species_fdr <- rownames_to_column(p_vals_species_fdr, "code")

kegg_codes <- read_excel("../data/excel_data/metabolism/kegg_codes.xlsx")
p_vals_species_fdr <- merge(kegg_codes, p_vals_species_fdr, by = "code")
```

Plot metabolic pathways at KEGG Level 1 for species

```
#NOTE: After we compiled data individual metabolic proportions across species, we exported the data into Excel and prepared a file organizing proportion by species and by metabolic pathway to save time (and to avoid long R coding). We then used this re-organized data to caclucate averages and standard deviations between species for all pathways in R, and then exported this data again to prepare it for plotting in R.

kegg_species_met1 <- read_excel("../data/excel_data/metabolism/kegg_l_species_prop_graph.xlsx")
kegg_species_met1 <- kegg_species_met1 %>%
  mutate(species = factor(species))

levels(kegg_species_met1$species) <- c("L. africana", "L. cyclotis")

pdf("/Users/joegunn/Desktop/Grad_School_Stuff/Research/Projects/Elephant_Microbiome/Attempt_2/visualization/metabolism_figures/kegg_level1_species.pdf", width = 20, height = 9)

ggplot(kegg_species_met1, aes(x = pathway, y = mean, fill = species)) +
  geom_bar(position=position_dodge(), stat="identity", show.legend = T) +
  geom_errorbar(aes(ymin=mean-sd, ymax=mean+sd), width=.4, position=position_dodge(.9)) +
  scale_fill_manual(values = c("red2", "blue2")) +
  theme(axis.text.x = element_text(angle = 0, hjust = 1)) +
  theme(axis.text = element_text(size = 20)) +
  theme(axis.title = element_text(size = 30)) +
  theme(legend.position = c(0.6,0.9)) +
  labs(fill = "Elephant Species", x = "Metabolic Pathway", y = "Mean Metabolic Contribution") +
  theme(legend.text = element_text(size = 30)) +
  theme(legend.title = element_text(size = 30)) +
  theme(legend.text = element_text(size = 30)) +
  theme(legend.text = element_text(face = "italic")) +
  coord_flip() +
  theme(axis.text.y = element_text(size = 30)) +
  theme(axis.title = element_text(size = 30))

dev.off()
```

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