

# elephants\_PHYLUM\_ABUNDANCE

Joe Gunn

8/12/2019

## AFRICAN ELEPHANT MICROBIOME - PHYLUM ABUNDANCE

Purpose: Here we are determining the phyla that are most abundant among African Elephant individuals across the entire dataset. We are then considering the three most abundant phyla and determining their respective abundances (raw sums) across all samples.

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

```
##
## *****

## Note: As of version 1.0.0, cowplot does not change the

## default ggplot2 theme anymore. To recover the previous

## behavior, execute:
## theme_set(theme_cowplot())

## *****

## — Attaching packages ————— tidyverse 1.2.1 —

## ✔ ggplot2 3.2.0      ✔ purrr   0.3.2
## ✔ tibble  2.1.3      ✔ dplyr   0.8.3
## ✔ tidyr   0.8.3      ✔ stringr 1.4.0
## ✔ readr   1.3.1      ✔ forcats 0.4.0

## — Conflicts ————— tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()

## Loading required package: permute

## Loading required package: lattice

## This is vegan 2.5-4

## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ

## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':
##
## expand

##
## Attaching package: 'nlme'

## The following object is masked from 'package:lme4':
##
## lmList

## The following object is masked from 'package:dplyr':
##
## collapse

##
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':
##
## nasa

## Loading required package: foreign

## Loading required package: survival

## Loading required package: MASS

##
## Attaching package: 'MASS'

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

## Loading required package: nnet

##
## Attaching package: 'epiDisplay'

## The following object is masked from 'package:lattice':
##
## dotplot

## The following object is masked from 'package:ggplot2':
##
## alpha

## Loading required package: mvtnorm

## Loading required package: TH.data

##
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':
##
## geyser

##
## Attaching package: 'taRifx'

## The following objects are masked from 'package:dplyr':
##
## between, distinct, first, last

## The following object is masked from 'package:purrr':
##
## rep_along
```

### Metadata

### Get Rarefied OTU abundance data

```
options(scipen = 999) #gets rid of scientific notation, 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 number 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")
```

### Get Most Abundant Phyla

```
#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")
```

### Clean the phylum-level dataset

```
#Summarize Rarefied OTU abundance by ALL Phyla
aem_phylum_data <- aem_tax_data[,c(1:2,4:8)]
aem_phylum_data <- as.data.frame(aem_phylum_data)
aem_phylum_data <- aem_phylum_data %>%
  group_by(Phylum) %>%
  summarize_all(funs(sum))

## Warning: Factor `Phylum` contains implicit NA, consider using
## `forcats::fct_explicit_na`

## 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.

aem_phylum_data <- aem_phylum_data %>%
  drop_na()

aem_phylum_data <- column_to_rownames(aem_phylum_data, "Phylum")

aem_phylum_data_t <- t(aem_phylum_data)
aem_phylum_data_t <- as.data.frame(aem_phylum_data_t)
aem_phylum_data_t <- aem_phylum_data_t %>%
  rownames_to_column("sample_id")

aem_phylum <- merge(metadata, aem_phylum_data_t, by = "sample_id")

#Proportions of each order contributing to each african elephant group
aem_order_phylum_data <- aem_tax_data[,c(1:2,4:8)]
aem_order_phylum_data <- as.tibble(aem_order_phylum_data)

## Warning: `as.tibble()` is deprecated, use `as_tibble()` (but mind the new semantics).
## This warning is displayed once per session.

aem_order_phylum_data <- aem_order_phylum_data %>%
  group_by(Phylum, Order) %>%
  summarize_all(funs(sum))

## Warning: Factor `Phylum` contains implicit NA, consider using
## `forcats::fct_explicit_na`

## Warning: Factor `Order` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```