3 Datasets

-Dietswapdata:

Fat, fibre and cancer risk in African Americans and rural Africans (2015)

<https://www.nature.com/articles/ncomms7342>

“Rates of colon cancer are much higher in African Americans (65:100,000) than in rural South Africans (<5:100,000). The higher rates are associated with higher animal protein and fat, and lower fibre consumption, higher colonic secondary bile acids, lower colonic short-chain fatty acid quantities and higher mucosal proliferative biomarkers of cancer risk in otherwise healthy middle-aged volunteers. Here we investigate further the role of fat and fibre in this association. We performed 2-week food exchanges in subjects from the same populations, where African Americans were fed a high-fibre, low-fat African-style diet and rural Africans a high-fat, low-fibre western-style diet, under close supervision.”

resulted in remarkable reciprocal changes in mucosal biomarkers of cancer risk and in aspects of the microbiota and metabolome known to affect cancer risk, and reduced secondary bile acid synthesis in the African Americans.

-Atlas1006 data:

Tipping elements in the human intestinal ecosystem(2014)

130 genus-like taxonomic groups across 1006 western adults with no reported health complications

-peerj32

Associations between the human intestinal microbiota, Lactobacillus rhamnosus GG and serum lipids indicated by integrated analysis of high-throughput profiling data (2013)

We performed a comprehensive intestinal microbiota analysis using a phylogenetic microarray before and after *Lactobacillus rhamnosus* GG intervention. While a specific increase in the *L. rhamnosus*-related bacteria was observed during the intervention, no other changes in the composition or stability of the microbiota were detected. After the intervention, lactobacilli returned to their initial levels. As previously reported, also the serum lipid profiles remained unaltered during the intervention.

Datasets from other formats can be loaded:

*# Import output CSV files generated by write\_phyloseq*

pseq1 <- read\_phyloseq(otu.file, taxonomy.file, metadata.file, type = "simple")

*# Import mother .shared and .taxonomy and metadata files*

pseq2 <- read\_phyloseq(otu.file, taxonomy.file, metadata.file, type = "mothur")

*# Import BIOM files*

pseq3 <- read\_phyloseq(otu.file, taxonomy.file, metadata.file, type = "biom")

* Phyloseq can be used to convert other datasets to phyloseq format also:

<http://joey711.github.io/phyloseq/import-data>

* Install kable/kableExtra

Use BiocManager to install kableExtra

BiocManager::install(c("kableExtra"))

Note: trying to install kable in R 3.5/3.6 results in error, therefore install kable extra and modify the kable(head(tab)) command to tab %>%

kable() %>%

kable\_styling()

* Load datasets for the tutorial:
* **library**(microbiome)
* data(dietswap)
* data(data(peerj32))
* data(atlas1006)
* pseq <- dietswap
* pseq2 <- dietswap
* pseq2 <- dietswap

Global indicators

A comprehensive list of global indicators of the ecosystem state can be obtained as follows. This includes various measures of richness, evenness, diversity, dominance, and rarity with default parameters. See the individual functions for more options regarding parameter tuning.

tab <- global(pseq, index = "all")

tab %>%

kable() %>%

kable\_styling()

alpha

tab <- alpha(pseq, index = "all")

tab %>%

kable() %>%

kable\_styling()

%all can be changed to any index you need ex:chao1

tab <- richness(pseq)

tab %>%

kable() %>%

kable\_styling()

tab <- dominance(pseq, index = "all")

tab %>%

kable() %>%

kable\_styling()

tab <- rarity(pseq, index = "all")

tab %>%

kable() %>%

kable\_styling()

tab <- evenness(pseq, "all")

tab %>%

kable() %>%

kable\_styling()

1) load data and packages

library(microbiome)

library(dplyr)

data(dietswap)

data <- dietswap

2) Calculate group divergences between the African American and rural Africans based on dietswap data

divergence(subset\_samples(dietswap, nationality == "AFR"))

am <- divergence(subset\_samples(dietswap, nationality == "AAM"))

3) view differences in sample diversity in a boxplot:

boxplot(list(American = am, African = af ))

* To calculate beta diversity changing across time, example code is:
* betas <- list()
* groups <- as.character(unique(meta(pseq)$group))
* for (g in groups) {
* #df <- meta(subset\_samples(pseq, group == g))
* df <- subset(meta(pseq), group == g)
* beta <- c()
* for (subj in df$subject) {
* # Pick the samples for this subject
* dfs <- subset(df, subject == subj)
* # Check that the subject has two time points
* if (nrow(dfs) == 2) {
* s <- as.character(dfs$sample)
* # Here with just two samples we can calculate the
* # beta diversity directly
* beta[[subj]] <- 1-cor(abundances(pseq)[, s[[1]]],
* abundances(pseq)[, s[[2]]],
* method = "spearman")
* }
* }
* betas[[g]] <- beta
* }
* boxplot(betas)

Example code of Betadiversity changing over time:

* Note this analysis and code is for atlas1006 data

data(atlas1006)

pseq <- atlas1006

# Identify subject with the longest time series (most time points)

s <- names(which.max(sapply(split(meta(pseq)$time, meta(pseq)$subject), function (x) {length(unique(x))})))

# Pick the metadata for this subject and sort the

# samples by time

library(dplyr)

df <- meta(pseq) %>% filter(subject == s) %>% arrange(time)

# Calculate the beta diversity between each time point and

# the baseline (first) time point

beta <- c(0, 0) # Baseline similarity

s0 <- subset(df, time == 0)$sample

for (tp in df$time[-1]) {

# Pick the samples for this subject

# If the same time point has more than one sample,

# pick one at random

st <- sample(subset(df, time == tp)$sample, 1)

a <- abundances(pseq)

b <- 1 - cor(a[, s0], a[, st], method = "spearman")

beta <- rbind(beta, c(tp, b))

}

colnames(beta) <- c("time", "beta")

beta <- as.data.frame(beta)

library(ggplot2)

p <- ggplot(beta, aes(x = time, y = beta)) +

geom\_point() + geom\_line()

print(p)