Microbiome Analysis of the Gut Microbiota of 1006 Western and Traditional Adults

Faith Oyewale OLABISI

Table of Content

Introduction

This study analysed the gut microbiota of 1006 western adults and traditional adults show casing how dietary changes impact microbial communities (O'Keefe et al., 2015). We explored the essential diversity metrics, including alpha diversity (using Shannon diversity and Observed Species) and beta diversity using Principal Coordinates Analysis (PCoA).

In this project we:

Calculated the alpha diversity metrics measuring the species richness and eveness with the total species count.

Investigated the beta diversity using the principal coordinates analysis to visualize the differences in microbial communties.

Visualize the microbial composition before and after diet swap using principal coordinates analysis plots and

```
library(tidyverse)
library(ggpubr)
library(microbiome)
library(RColorBrewer)
library(ComplexHeatmap)
library(knitr)
library(phyloseq)
```

Data Extraction and Preparation

```
# obtain the data from the microbiome package
data('dietswap')
dietswap
```

```
phyloseq-class experiment-level object
otu_table()   OTU Table:         [ 130 taxa and 222 samples ]
sample_data() Sample Data:       [ 222 samples by 8 sample variables ]
tax_table()   Taxonomy Table:    [ 130 taxa by 3 taxonomic ranks ]
```

```
# First 10 microbes for first 7 sample
dietswap@otu_table@.Data[1:10, 1:7] |>
  as.data.frame() #Convert the data into a dataframe
```

	Sample-1	Sample-2	Sample-3	Sample-4	Sample-5
Actinomycetaceae	0	1	0	1	0
Aerococcus	0	0	0	0	0
Aeromonas	0	0	0	0	0
Akkermansia	18	97	67	256	21
Alcaligenes faecalis et rel.	1	2	3	2	2
Allistipes et rel.	336	63	36	96	49
Anaerobiospirillum	0	0	0	0	0
Anaerofustis	0	1	0	0	0
Anaerostipes caccae et rel.	244	137	27	36	23
Anaerotruncus colihominis et rel.	12	108	203	68	15

	Sample-6	Sample-7
Actinomycetaceae	0	0
Aerococcus	0	0
Aeromonas	0	0
Akkermansia	16	26

```
Alcaligenes faecalis et rel.             2        2
Allistipes et rel.                      17       47
Anaerobiospirillum                       0        0
Anaerofustis                             0        0
Anaerostipes caccae et rel.             29       58
Anaerotruncus colihominis et rel.       36       31
```

Convert metadata into a tidy dataframe

```
# Converting metadata into a tidy dataframe
metadata <- meta(dietswap)

metadata |>
  head()
```

```
         subject    sex nationality group   sample timepoint
Sample-1     byn   male         AAM    DI Sample-1         4
Sample-2     nms   male         AFR    HE Sample-2         2
Sample-3     olt   male         AFR    HE Sample-3         2
Sample-4     pku female         AFR    HE Sample-4         2
Sample-5     qjy female         AFR    HE Sample-5         2
Sample-6     riv female         AFR    HE Sample-6         2
         timepoint.within.group  bmi_group
Sample-1                      1      obese
Sample-2                      1       lean
Sample-3                      1 overweight
Sample-4                      1      obese
Sample-5                      1 overweight
Sample-6                      1      obese
```

```
## Correcting the data types
metadata <- metadata |>
  mutate(
    timepoint = as.factor(timepoint),
    timepoint.within.group = as.factor(timepoint.within.group)
  )

metadata |>
  head()
```

```
         subject    sex nationality group   sample timepoint
Sample-1     byn   male         AAM    DI Sample-1         4
Sample-2     nms   male         AFR    HE Sample-2         2
Sample-3     olt   male         AFR    HE Sample-3         2
Sample-4     pku female         AFR    HE Sample-4         2
Sample-5     qjy female         AFR    HE Sample-5         2
Sample-6     riv female         AFR    HE Sample-6         2
         timepoint.within.group  bmi_group
Sample-1                      1      obese
Sample-2                      1       lean
Sample-3                      1 overweight
Sample-4                      1      obese
Sample-5                      1 overweight
Sample-6                      1      obese
```

```
## Summary of the data
metadata |>
  summary()
```

```
    subject        sex      nationality group      sample          timepoint
 azh    :  6   female:102   AAM:123     DI:72   Length:222         1:38
 azl    :  6   male  :120   AFR: 99     ED:75   Class :character   2:37
 byn    :  6                           HE:75   Mode  :character   3:38
 cxj    :  6                                                      4:37
 dwc    :  6                                                      5:35
 eve    :  6                                                      6:37
 (Other):186
 timepoint.within.group      bmi_group
 1:112                  lean      :56
 2:110                  overweight:76
                        obese     :90
```

```
taxon_table <- tax_table(dietswap)

taxon_table |>
  head(n = 10)
```

```
Taxonomy Table:     [10 taxa by 3 taxonomic ranks]:
```

```
                                Phylum            Family
Actinomycetaceae                "Actinobacteria"  "Actinobacteria"
Aerococcus                      "Firmicutes"      "Bacilli"
Aeromonas                       "Proteobacteria"  "Proteobacteria"
Akkermansia                     "Verrucomicrobia" "Verrucomicrobia"
Alcaligenes faecalis et rel.    "Proteobacteria"  "Proteobacteria"
Allistipes et rel.              "Bacteroidetes"   "Bacteroidetes"
Anaerobiospirillum              "Proteobacteria"  "Proteobacteria"
Anaerofustis                    "Firmicutes"      "Clostridium cluster XV"
Anaerostipes caccae et rel.     "Firmicutes"      "Clostridium cluster XIVa"
Anaerotruncus colihominis et rel. "Firmicutes"    "Clostridium cluster IV"
                                Genus
Actinomycetaceae                "Actinomycetaceae"
Aerococcus                      "Aerococcus"
Aeromonas                       "Aeromonas"
Akkermansia                     "Akkermansia"
Alcaligenes faecalis et rel.    "Alcaligenes faecalis et rel."
Allistipes et rel.              "Allistipes et rel."
Anaerobiospirillum              "Anaerobiospirillum"
Anaerofustis                    "Anaerofustis"
Anaerostipes caccae et rel.     "Anaerostipes caccae et rel."
Anaerotruncus colihominis et rel. "Anaerotruncus colihominis et rel."
```

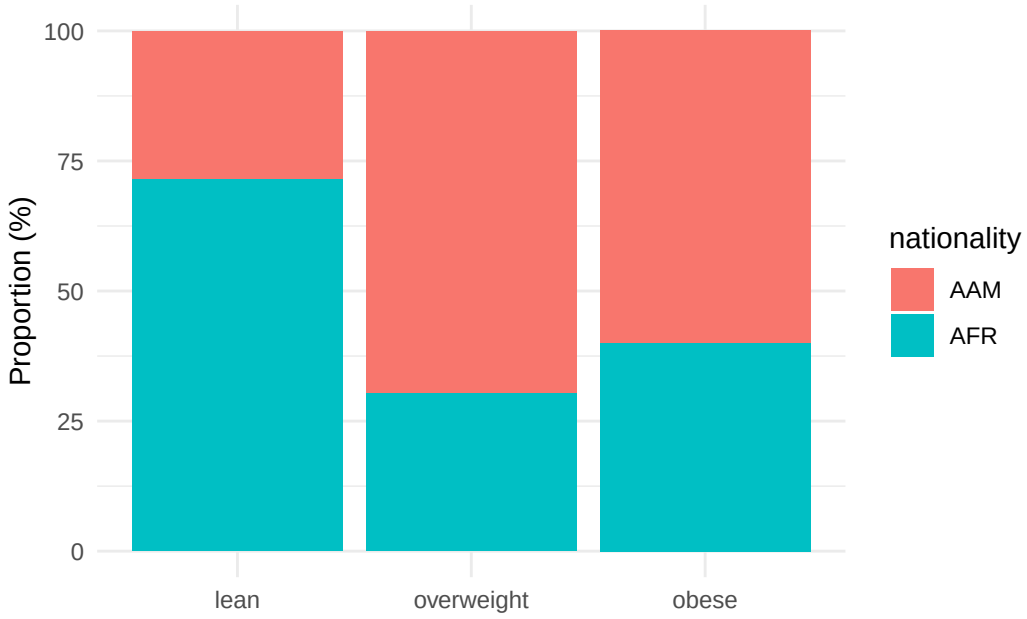
```
#Join all three component into a dataframe
diet <- psmelt(dietswap)

diet |>
  head()
```

```
                                    OTU     Sample Abundance subject    sex
21328 Prevotella melaninogenica et rel. Sample-208     14270     olt   male
21418 Prevotella melaninogenica et rel. Sample-212     13889     shj female
21457 Prevotella melaninogenica et rel.  Sample-11     13640     olt   male
21481 Prevotella melaninogenica et rel. Sample-125     13509     nmz   male
21438 Prevotella melaninogenica et rel. Sample-210     13402     qjy female
21319 Prevotella melaninogenica et rel. Sample-107     13289     byu   male
      nationality group     sample timepoint timepoint.within.group  bmi_group
21328         AFR    ED Sample-208         1                      1 overweight
21418         AFR    ED Sample-212         1                      1      obese
21457         AFR    HE  Sample-11         3                      2 overweight
21481         AAM    HE Sample-125         3                      2      obese
21438         AFR    ED Sample-210         1                      1 overweight
```

```
21319         AFR    HE Sample-107         3                     2       lean
             Phylum        Family                              Genus
21328 Bacteroidetes Bacteroidetes Prevotella melaninogenica et rel.
21418 Bacteroidetes Bacteroidetes Prevotella melaninogenica et rel.
21457 Bacteroidetes Bacteroidetes Prevotella melaninogenica et rel.
21481 Bacteroidetes Bacteroidetes Prevotella melaninogenica et rel.
21438 Bacteroidetes Bacteroidetes Prevotella melaninogenica et rel.
21319 Bacteroidetes Bacteroidetes Prevotella melaninogenica et rel.
```

Exploratory Data Analysis

```
# EDA for bmi group proportion for each nationality

plot_frequencies(x = sample_data(dietswap),
                 Groups = 'bmi_group', Factor = 'nationality')+
  labs(fill = 'nationality')+
  theme_minimal()+
  theme(axis.text.x = element_text (angle = 0, hjust = 0.5))
```



Filter the Prevalent Taxa from the data

```
# filtering prevalent data
core_diet <- core(x = dietswap,
                  detection = 50,
                  prevalence = 50/100)

core_diet
```

```
phyloseq-class experiment-level object
otu_table()   OTU Table:         [ 17 taxa and 222 samples ]
sample_data() Sample Data:       [ 222 samples by 8 sample variables ]
tax_table()   Taxonomy Table:    [ 17 taxa by 3 taxonomic ranks ]
```

```
diet_log <- transform_sample_counts(core_diet, function(x) log(1+x))
```

Alpha Diversity Analysis

```
# Calculate alpha diversity metrics
rich <- estimate_richness(core_diet, measures = c('Observed', 'shannon'))
```

```
Warning in estimate_richness(core_diet, measures = c("Observed", "shannon")): The data you ha
any singletons. This is highly suspicious. Results of richness
estimates (for example) are probably unreliable, or wrong, if you have already
trimmed low-abundance taxa from the data.

We recommended that you find the un-trimmed data and retry.
```

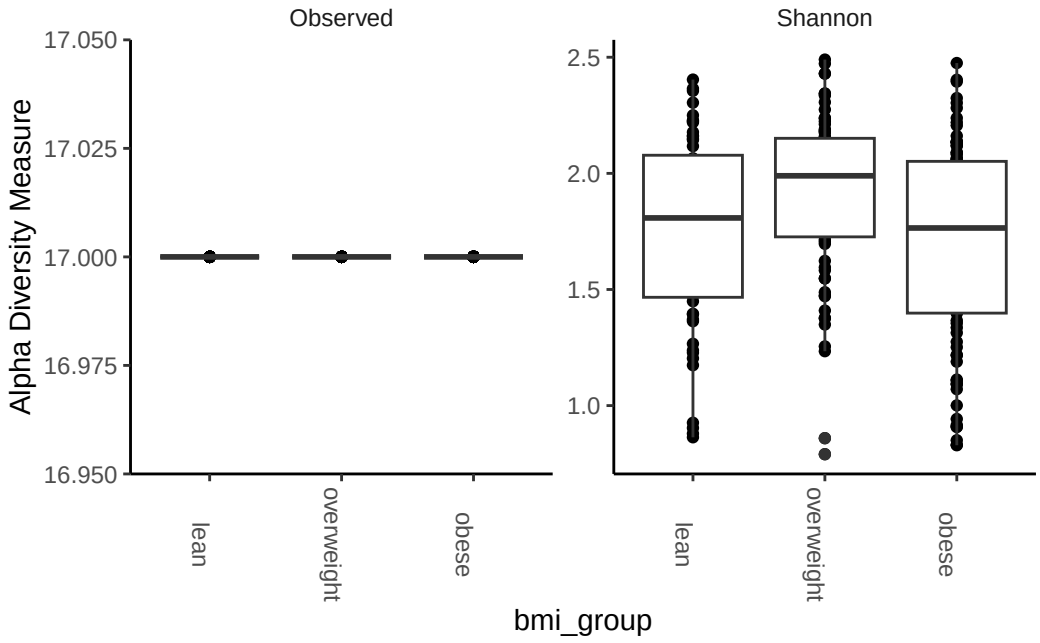
```
rich |>
  head() |>
  kable()
```

	Observed	Shannon
Sample.1	17	1.907579
Sample.2	17	2.065016
Sample.3	17	1.696573
Sample.4	17	1.910158
Sample.5	17	1.547898
Sample.6	17	1.402061

Data Visualization of Alpha Diversity

```
# Visualize alpha diversity using boxplot
plot_richness(core_diet, x = 'bmi_group', measures = c('Observed', 'shannon'))+
  geom_boxplot()+
  theme_classic()+
  theme(strip.background = element_blank(),
        axis.text.x.bottom = element_text(angle = -90))
```

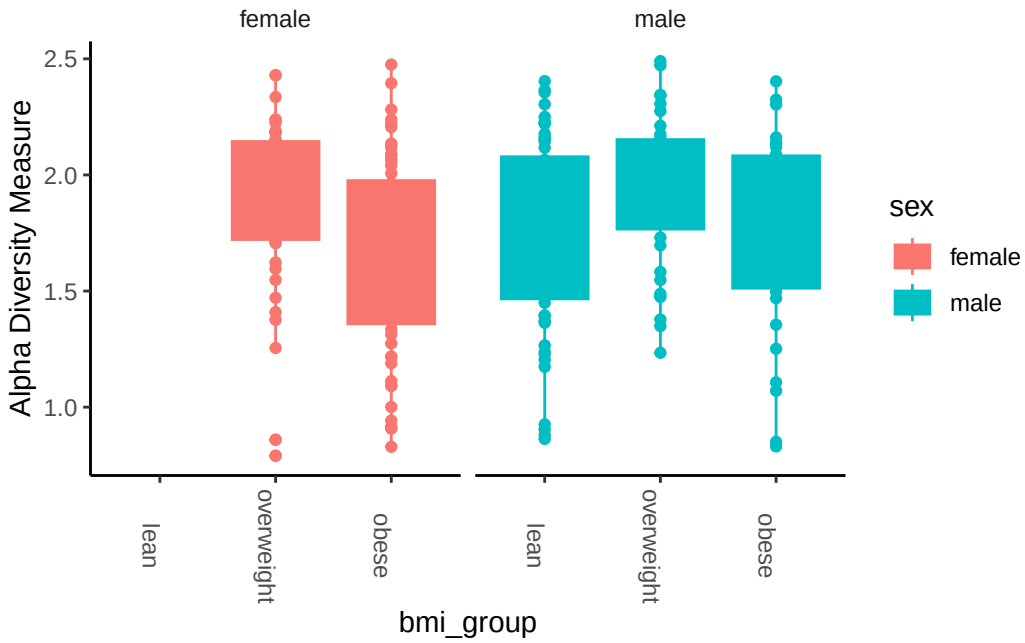
```
Warning in estimate_richness(physeq, split = TRUE, measures = measures): The data you have pr
any singletons. This is highly suspicious. Results of richness
estimates (for example) are probably unreliable, or wrong, if you have already
trimmed low-abundance taxa from the data.

We recommended that you find the un-trimmed data and retry.
```




```
## Visualize alpha diversity using boxplots facted by sex
plot_richness(core_diet, x = 'bmi_group', measures = c('shannon'), color = 'sex')+
  geom_boxplot(aes(fill = sex))+
  theme_classic()+
  facet_wrap(~sex)+
```

```
  theme(strip.background = element_blank(),
        axis.text.x.bottom = element_text(angle = -90))
```

```
Warning in estimate_richness(physeq, split = TRUE, measures = measures): The data you have pr
any singletons. This is highly suspicious. Results of richness
estimates (for example) are probably unreliable, or wrong, if you have already
trimmed low-abundance taxa from the data.

We recommended that you find the un-trimmed data and retry.
```



Data Manipulation of the Richness data

```
# Convert richness data into  dataframe for easy handling
rich_df <- as.data.frame(rich)

# Add the bmi_group to the richness data frame
rich_df$bmi_group <- sample_data(core_diet)$bmi_group

# Wilcoxon rank-sum text for observed richness
wilcox_observed <- pairwise.wilcox.test(
```

```
  rich_df$Observed,
  rich_df$bmi_group,
  p.adjust.method = 'fdr'
)

# Wilcoxon rank-sum test for shannon diversity
wilcox_shannon <- pairwise.wilcox.test(
  rich_df$Shannon,
  rich_df$bmi_group,
  p.adjust.method = 'fdr'
)

# Print results
print(wilcox_observed)
```

```
    Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data:  rich_df$Observed and rich_df$bmi_group

           lean overweight
overweight -    -
obese      -    -

P value adjustment method: fdr
```

```
print(wilcox_shannon)
```

```
    Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data:  rich_df$Shannon and rich_df$bmi_group

           lean   overweight
overweight 0.0487 -
obese      0.4385 0.0021

P value adjustment method: fdr
```

Shannon Diversity Results

Lean vs Overweight: The p-value is 0.0487. This indicates a statistically significant difference between the lean and overweight groups at the conventional alpha level of 0.05. This means that the diversity in microbial communities between these two groups is likely different.

Overweight vs Obese: The p-value is 0.0021. This is also statistically significant, indicating that there is a significant difference in microbial diversity between overweight and obese groups.

Lean vs Obese: The p-value is 0.4385, suggesting there is no statistically significant difference in Shannon diversity between the lean and obese groups.

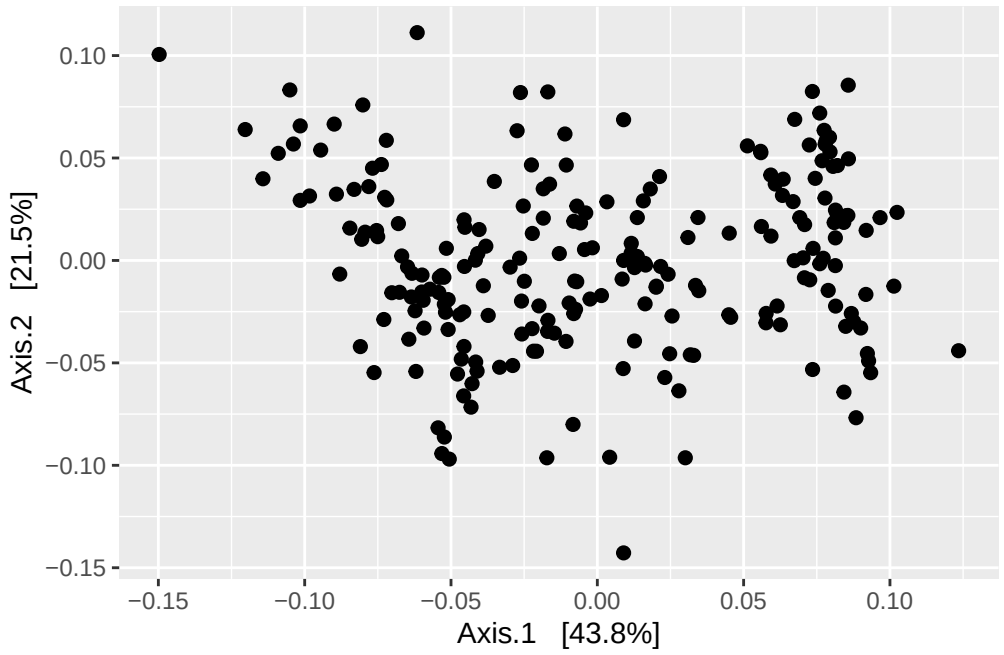
Beta Diversity Analysis

Beta diversity metrics assess the dissimilarity between ecosystem, telling us to what extent one community is different from another.

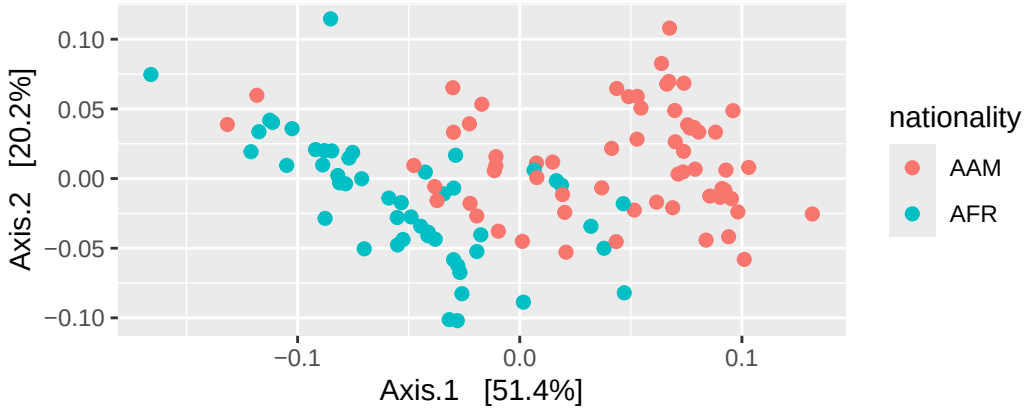
Exploring Pattern with PCA Plot

The first thing to be explored is to see if there are specific pattern that explains our data, we can try to perform multivariate projection of our sample data, specifically using the OTU Table or the microbial abundance data for each sample. Before plotting, its advisable we transform our abundance data into its log value as an approximate variance stabilizing transformation (Ben J. Callahan, 2016).

```
# Ordinate the data
set.seed(100)

ord <-  ordinate(physeq = diet_log,
                 method = 'MDS',
                 distance = 'bray')

# Prepare eigen values to adjust axis
evals <- ord$values$Eigenvalues

# Plotting
plot_ordination(physeq = diet_log,
                ordination = ord) +
  geom_point(size = 2)+
  coord_fixed(sqrt(evals[2] / evals[1]))
```




```
diet_log_prior <- subset_samples(diet_log, timepoint.within.group == 1)
# Ordinate the data
set.seed(100)

ord <-  ordinate(physeq = diet_log_prior,
                 method = 'MDS',
                 distance = 'bray')

# Prepare eigen values to adjust axis
evals <- ord$values$Eigenvalues

# Plotting
plot_ordination(physeq = diet_log,
                ordination = ord,
                color = 'nationality') +
  geom_point(size = 2)+
  labs(col = 'nationality')+
  coord_fixed(sqrt(evals[2] / evals[1]))
```



The study determined whether dietswap can change the composition of the initial microbial composition between African American and Native African. To better visualize the difference of microbial composition between subjects with different nationality, we can also plot the microbial abundance which will be discussed next.

Microbial Abundance Plot

Microbial communities are often viusualized in the form of stacked barplot. Each bar representing one sample community and harbours different abundance and diversity of microbes. One can analyse the dominant micrfobes for each sample community or each group sample.

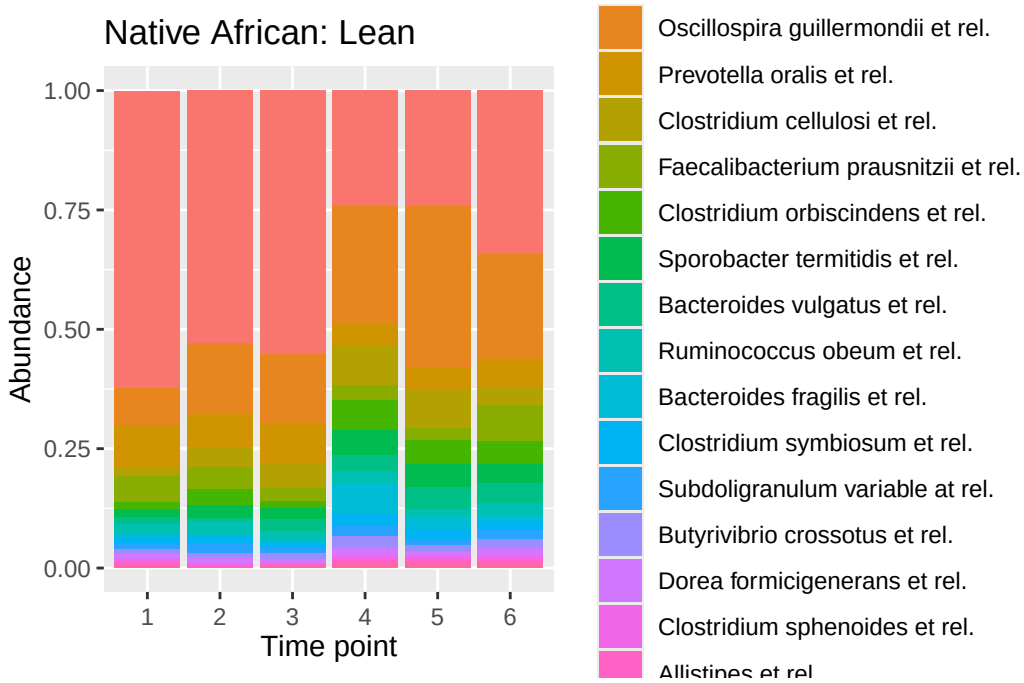
Before plotting, it is better to transform our original microbial abundance data into its relative abundance (frequencies of microbes present per total count for each sample)

```
# transform the data
diet_relav <- microbiome::transform(core_diet, 'compositional')

# Inspect the data
diet_relav@otu_table@.Data[1:3, 1:3]
```

```
                              Sample-1    Sample-2    Sample-3
Allistipes et rel.           0.05710401 0.003699354 0.001502003
Bacteroides fragilis et rel. 0.07528892 0.001233118 0.003045728
Bacteroides vulgatus et rel. 0.47144799 0.002818555 0.018983645
```

```r
# filtering data based on label/condition
afr_lean <- subset_samples(diet_relav, nationality == 'AFR' & bmi_group == 'lean')
afr_over <- subset_samples(diet_relav,
                           nationality == 'AFR' & bmi_group == 'overweight')
afr_obese <- subset_samples(diet_relav,
                            nationality == 'AFR' & bmi_group == 'obese')

aam_lean <- subset_samples(diet_relav, nationality == 'AAM' & bmi_group == 'lean')
aam_over <- subset_samples(diet_relav,
                           nationality == 'AAM' & bmi_group == 'overweight')
aam_obese <- subset_samples(diet_relav,
                            nationality == 'AAM' & bmi_group == 'obese')
```

```r
# Plot average of Sample
plot_composition(afr_lean,
                 taxonomic.level = 'Genus',
                 average_by = 'timepoint',
                 otu.sort = 'abundance',
                 x.label = 'timepoint')+
  labs(
    x = 'Time point',
    y = 'Abundance',
    title = 'Native African: Lean'
  )+
  theme(
    axis.text.x = element_text(angle = 0, hjust = 0.5)
  )
```


Native African: Lean
Abundance
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
6
Time point
Oscillospira guillermondii et rel.
Prevotella oralis et rel.
Clostridium cellulosi et rel.
Faecalibacterium prausnitzii et rel.
Clostridium orbiscindens et rel.
Sporobacter termitidis et rel.
Bacteroides vulgatus et rel.
Ruminococcus obeum et rel.
Bacteroides fragilis et rel.
Clostridium symbiosum et rel.
Subdoligranulum variable at rel.
Butyrivibrio crossotus et rel.
Dorea formicigenerans et rel.
Clostridium sphenoides et rel.
Allistipes et rel