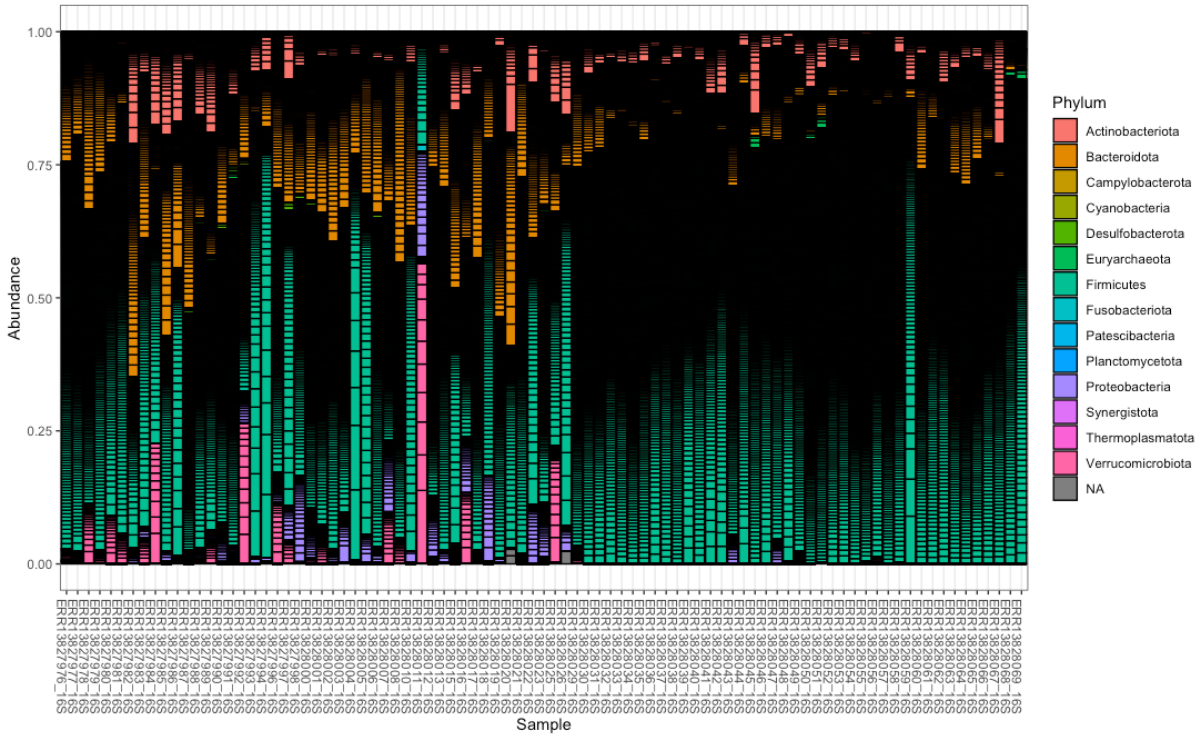
BIOINFORMATIC ANALYSES SUMMARY

This document explains the different bioinformatic analyses performed on the microbiome data. For each step, you will find the name of the analysis, the type of plot, what it shows and what we can learn from it.

I. Microbial composition by Phylum

1. Absolute abundances

```
plot_bar(ps.prop, fill="Phylum")
```

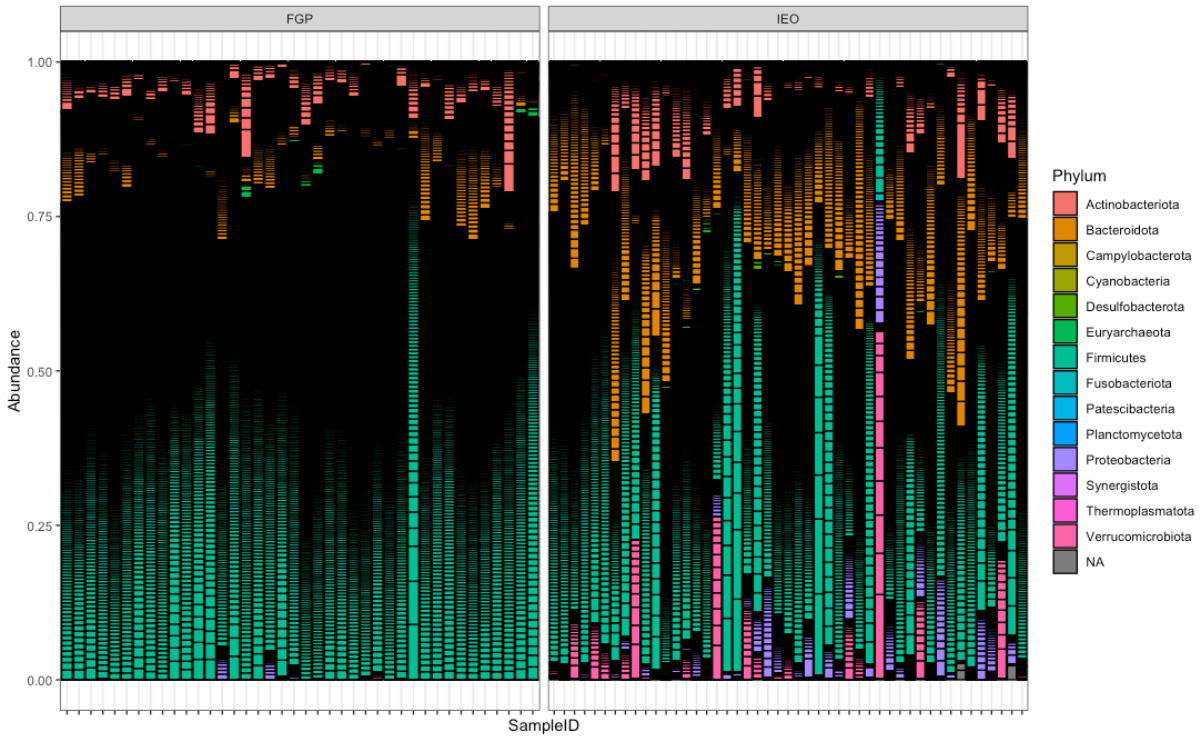


This barplot shows the number of sequences per sample, colored by Phylum. Each bar represents one sample, and the height reflects the absolute abundance of each phylum.

So taller bars mean more sequences, and each color represents a phylum. This gives a general idea of which phyla are present and how many sequences were found but does not allow easy comparison between samples because the total number of reads is different

2. Relative abundances

```
plot_bar(ps.prop, fill="Phylum", x="SampleID") +
 facet_wrap(~ Group, scales = "free_x") +
theme(axis.text.x = element_blank())
```

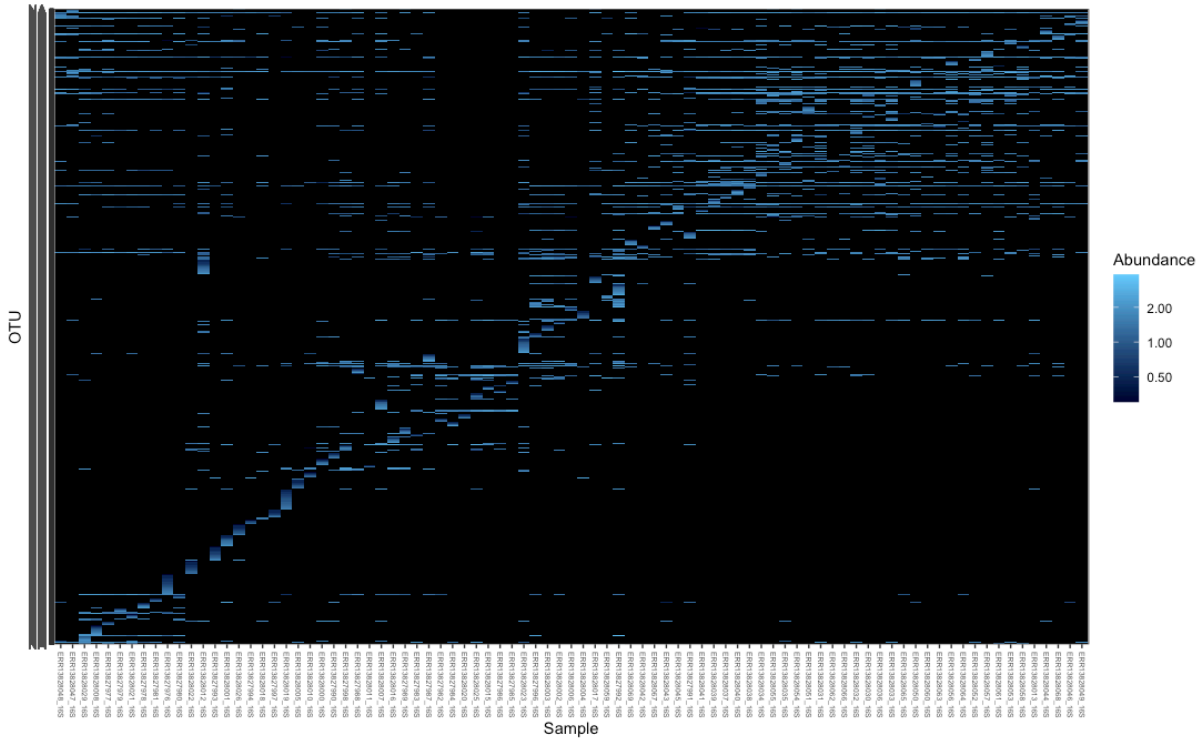


Before this plot, we normalized the data so that each sample has a total of 1 (100%). This shows the proportion of each phylum in each sample.

This representation makes it easier to compare samples and groups (FGP vs IEO). We can see the differences in microbial composition between groups.

II. Heatmap of OUT Abundances (log scale)

```
plot_heatmap(transform_sample_counts(ps, function(x) log10(x+1)))
```

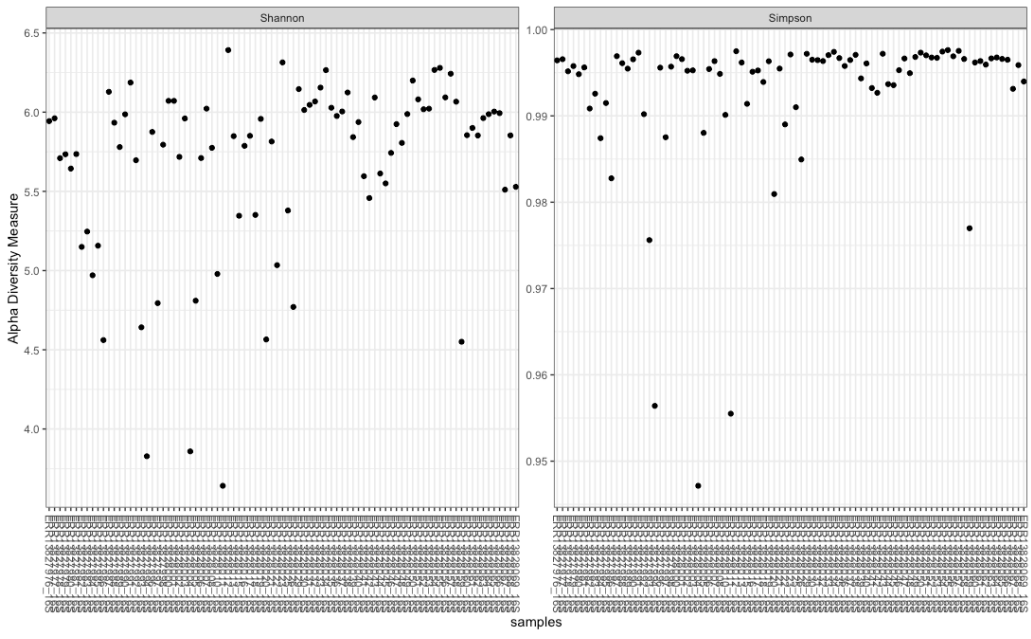


We made a heatmap to see how different OTUs are present in the samples. On the x axis we have the sample and on the y axis we have the OTUs (bacterial types). The color shows how many reads (abundance) we found for each OTU in each sample: darker blue means higher abundance and black means almost no reads. So, this heatmap shows how abundant each OTU is in each sample and the log scale helps show small differences.

This representation helps to find patterns and dominant taxa. Samples with similar color patterns may have similar microbiomes. Most OTUs are rare and a few are found in many samples, but it is hard to see clear group of samples.

III. Alpha diversity (Shannon and Simpson index)

```
plot_richness(ps, measures = c("Shannon", "Simpson"))
```



This plot shows alpha diversity for each sample, using two common diversity indexes: Shannon and Simpson. Alpha diversity tells us how complex or diverse the microbial community is within each individual sample.

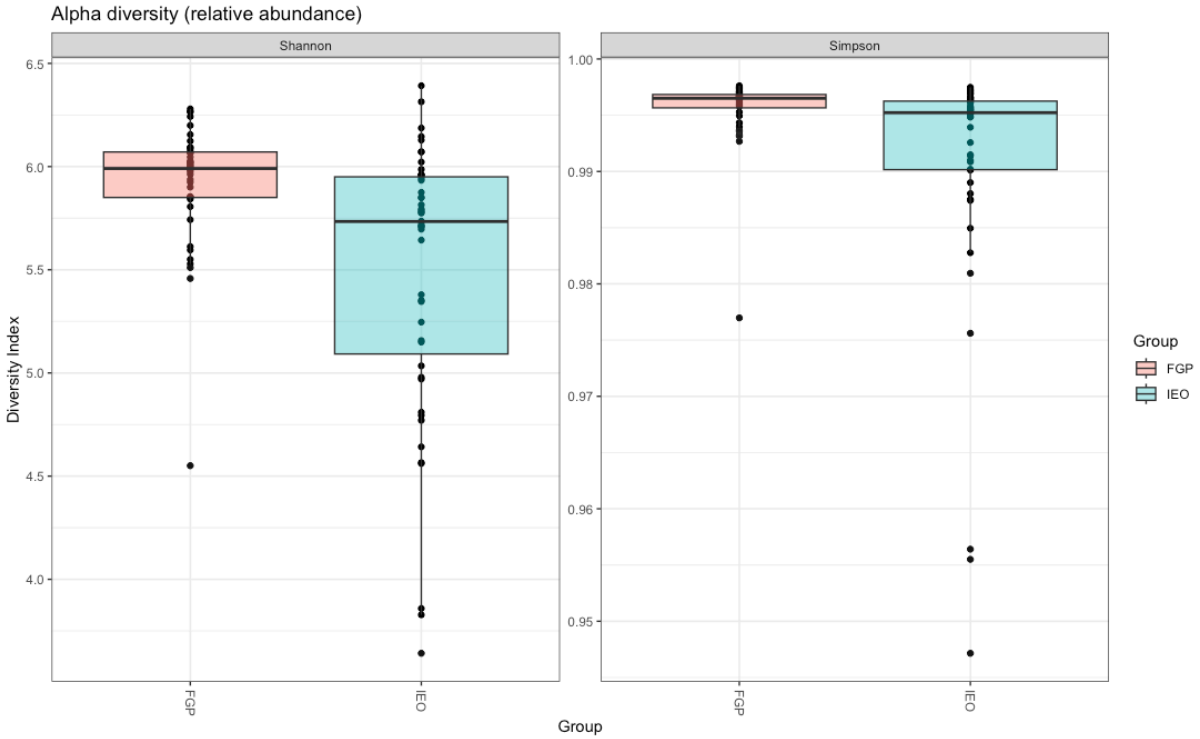
The Shannon index considers both the number of different species (richness) and how evenly they are distributed. A higher Shannon value means the sample has many different OTUs and they are more evenly spread, and, in this plot, most samples have Shannon value between 5.5 and 6.5 which suggest good diversity. A few samples have lower values, possibly because they have fewer dominant OTUs or low richness.

The Simpson index also measures diversity but gives more weight to the most common OTUs. Values are between 0 and 1 and here we see that most are very close to 1 which indicates high diversity. (no OTU dominates too much). Some samples are a little bit lower which may mean that one OTU is dominating in those.

Most samples are diverse in terms of species number and balance.

IV. Alpha diversity by group

```
plot_richness(ps, x="Group", measures=c("Shannon", "Simpson")) +
  geom_boxplo(aes(fill=Group), alpha=0.4) +
  labs(title="Alpha diversity (relative abundance)", y="Diversity Index", x="Group"
```



This plot shows the alpha diversity for each sample grouped by condition: FGP and IEO. Here again we used the Shannon and Simpson indexes. Each point is a sample, and the box represents the distribution of values in each group.

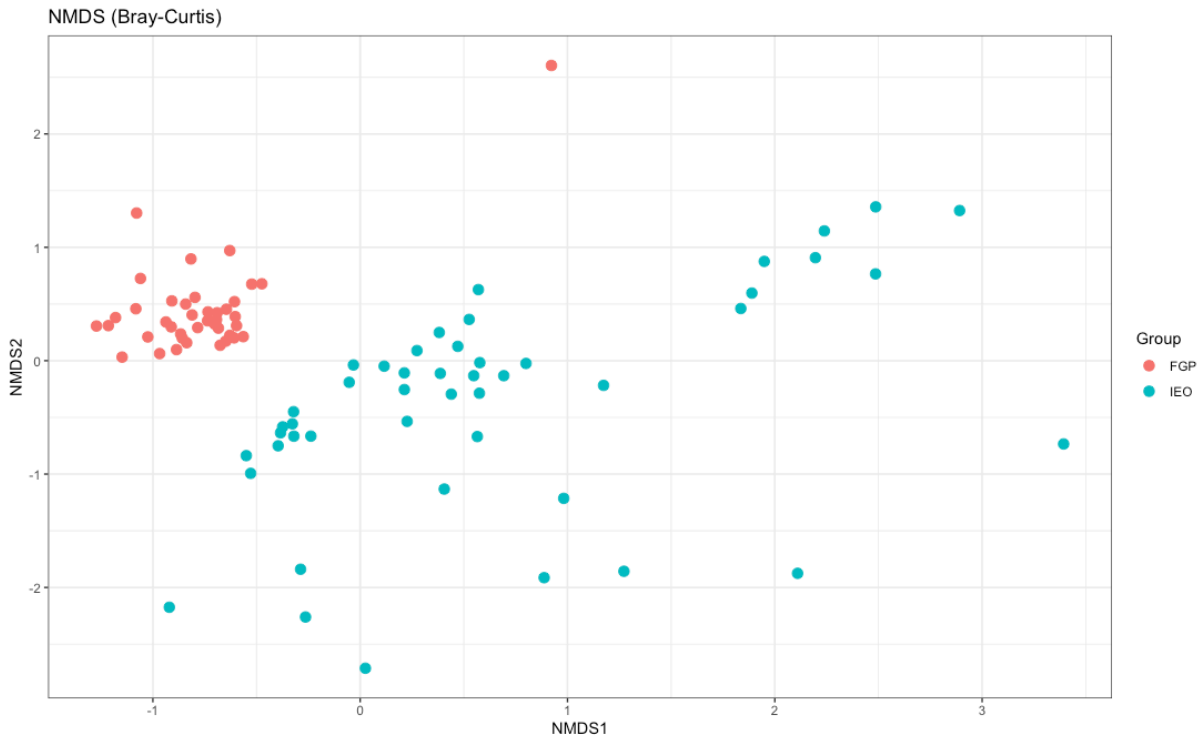
The Shannon index measures how many different OTUs there are and how balanced they are. The FGP group has slightly higher Shannon values, so we can suggest that there is a little bit more diversity and evenness. The IEO group also shows more variability between samples, with some samples having lower diversity.

The Simpson index focuses more on the dominance of abundant OTUs. Again, FGP tends to have little higher value, meaning that its microbial communities are more balanced, with less dominance by just one or two taxa. The IEO group has more samples with few lower values which may suggest that some taxa dominate in certain IEO samples.

Both groups have high diversity, but the FGP samples appear more consistent and slightly more diverse than the IEO ones.

V. NMDS ordination: Bray-Curtis Distance

```
ord.nmds.bray <- ordinate(ps.prop, method="NMDS", distance="bray")
plot_ordination(ps.prop, ord.nmds.bray, color="Group", title="Bray NMDS") +
 geom_point(size=3) +
```



This plot shows the result of a NMDS ordination (Non-Metric Multidimensional Scaling) NMDS helps us to visualize differences between samples based on their microbial composition. We used Bray-Curtis distance which compares how different two samples are in terms