botero_analysis

Sean Lu

2025-05-18

Setup

```r
library("phyloseq")
library("ggplot2")
library("dplyr")
library("tibble")
library("ggpubr")
library("phylosmith")
```

Load in processed Data

```r
ps_dada2 <- readRDS("data_processed/botero_2014/ps_dada2.rds")
ps_ms <- readRDS("data_processed/botero_2014/ps_metascope_priors_trimmed.rds")

ps_dada2_oral <- subset_samples(ps_dada2, Sample_type == "Oropharynx")
ps_ms_oral <- subset_samples(ps_ms, Sample_type == "Oropharynx")
ps_dada2_nasal <- subset_samples(ps_dada2, Sample_type == "Nasal")
ps_ms_nasal <- subset_samples(ps_ms, Sample_type == "Nasal")
```

Plotting relative abundances of MetaScope and DADA2

Species Level Abundances

```r
dada2_df <- psmelt(ps_dada2) |>
  dplyr::mutate(Species = ifelse(is.na(Species), NA, paste0(Genus, " ", Species)))
dada2_df$pipeline = "DADA2"
colnames(dada2_df)
```

```
##  [1] "OTU"             "Sample"          "Abundance"       "Sequencing_Type"
##  [5] "Patient"         "Sample_type"     "status"          "Kingdom"
##  [9] "Phylum"          "Class"           "Order"           "Family"
## [13] "Genus"           "Species"         "pipeline"
```

```r
ms_df <- psmelt(ps_ms)
ms_df$pipeline = "MetaScope"
ms_df$kingdom = "Bacteria"
ms_df <- ms_df |>
  dplyr::relocate(kingdom, .before = phylum)
colnames(ms_df) <- c("OTU", "Sample", "Abundance", "Sequencing_Type", "Patient",
                     "Sample_type", "status", "Kingdom", "Phylum", "Class",
                     "Order", "Family", "Genus", "Species", "pipeline")

merged_df <- rbind(dada2_df, ms_df)


relab_species <- ggplot(merged_df, aes(x = Sample, y = Abundance, fill = Species)) +
  geom_bar(position ="fill",stat = "identity") +
  facet_grid(cols = c(vars(Sample_type), vars(status), vars(pipeline)), scales = "free_x")
relab_species +
  theme(legend.position = "none",
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```
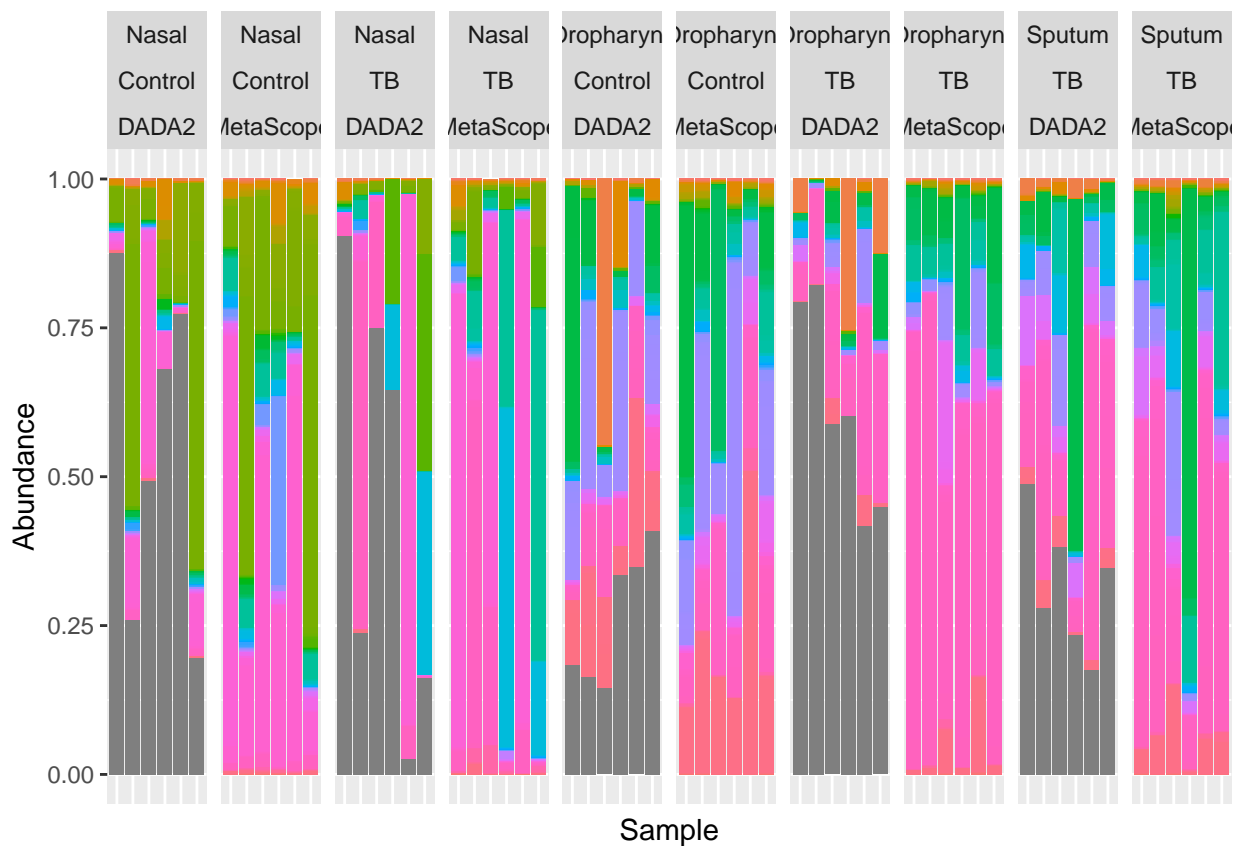


```r
relab_species_legend <- get_legend(relab_species)
as_ggplot(relab_species_legend)
```

s aquiplantarum	▉	Mechercharimyces asporophorigenens	▉	Mesorhizobium hankyongi	▉	Methyloba
s arenicola	▉	Mediterraneibacter faecis	▉	Mesorhizobium huakuii	▉	Methyloba
s balearica	▉	Mediterraneibacter gnavus	▉	Mesorhizobium thiogangeticum	▉	Methyloba
s communis	▉	Megalodesulfovibrio gigas	▉	Metallibacterium scheffleri	▉	Methyloba
s foliarum	▉	Megamonas hypermegale	▉	Metallococcus carri	▉	Methyloba
s mediterranea	▉	Megamonas rupellensis	▉	Metamycoplasma arthritidis	▉	Methyloba
s pollencensis	▉	Megasphaera cerevisiae	▉	Metamycoplasma cloacale	▉	Methyloba
s posidonica	▉	Megasphaera elsdenii	▉	Metamycoplasma orale	▉	Methyloca
s rhizomae	▉	Megasphaera micronuciformis	▉	Metamycoplasma salivarium	▉	Methyloce
s vaga	▉	Megasphaera paucivorans	▉	Methylobacillus flagellatus	▉	Methyloce
cola antarctica	▉	Megasphaera sueciensis	▉	Methylobacillus glycogenes	▉	Methyloco
r fusiformis	▉	Melioribacter roseus	▉	Methylobacterium aquaticum	▉	Methylocy
ctuosa	▉	Melissococcus plutonius	▉	Methylobacterium dankookense	▉	Methylom
onas ophiurae	▉	Melittangium lichenicola	▉	Methylobacterium durans	▉	Methylom
ssiliensis	▉	Mesobacillus campisalis	▉	Methylobacterium fujisawaense	▉	Methylom
italea	▉	Mesobacillus subterraneus	▉	Methylobacterium goesingense	▉	Methylom
vichensis	▉	Mesomycoplasma neurolyticum	▉	Methylobacterium iners	▉	Methyloph
onensis	▉	Mesoplasma corruscae	▉	Methylobacterium isbiliense	▉	Methyloph
uanensis	▉	Mesoplasma seiffertii	▉	Methylobacterium komagatae	▉	Methyloph

Genus Level Abundances

```r
relab_genus <- ggplot(merged_df, aes(x = Sample, y = Abundance, fill = Genus)) +
  geom_bar(position ="fill",stat = "identity") +
  facet_grid(cols = c(vars(Sample_type), vars(status), vars(pipeline)), scales = "free_x")
relab_genus +
  theme(legend.position = "none",
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```
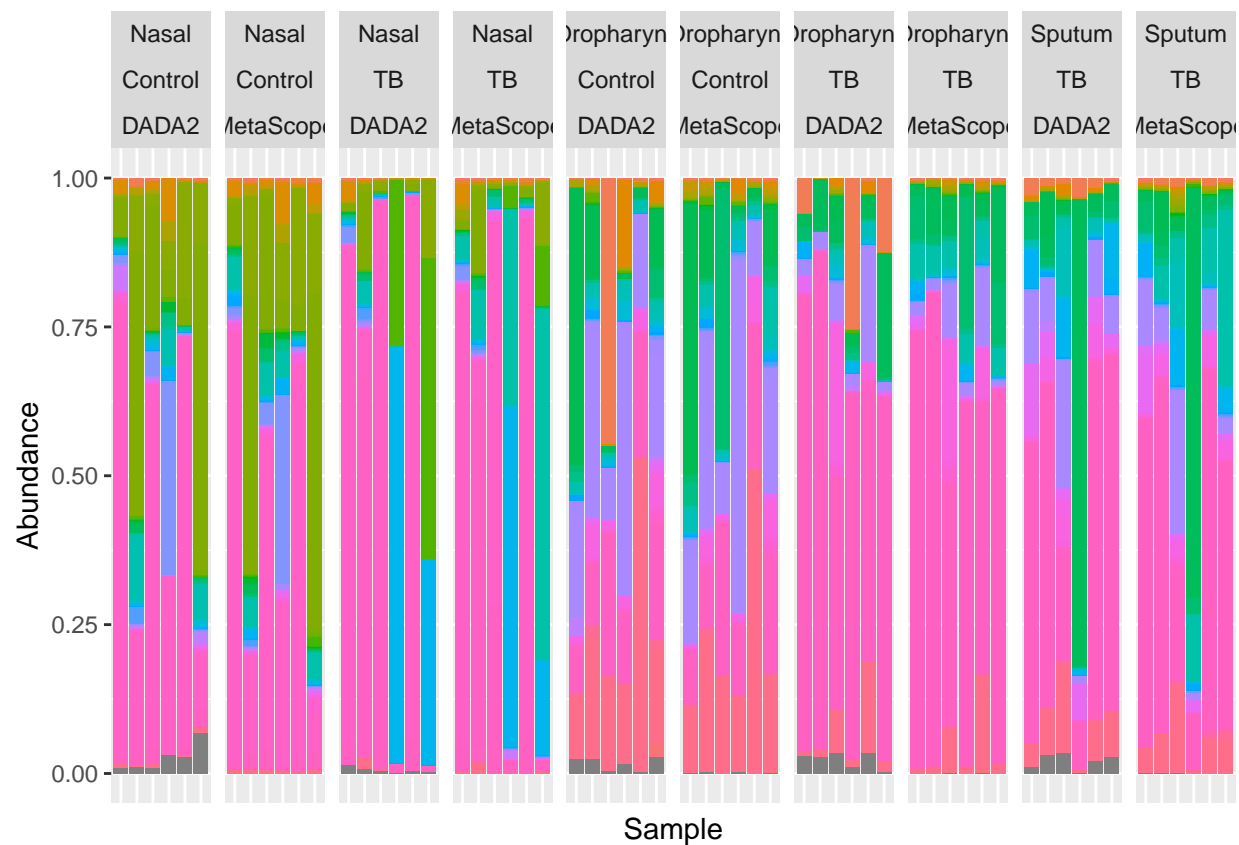
```
relab_genus_legend <- get_legend(relab_genus)
as_ggplot(relab_genus_legend)
```

ebsiella	■	Lacrimispora	■	Lentzea	■	Luteibacter	■	Maribact
enkia	■	Lacticaseibacillus	■	Leptolyngbya PCC–6306	■	Luteibaculum	■	Marichro
uyvera	■	Lacticigenium	■	Leptospira	■	Luteimicrobium	■	Marinicri
ocuria	■	Lactiplantibacillus	■	Leptotrichia	■	Luteimonas	■	Marinilal
omagataeibacter	■	Lactobacillus	■	Leucobacter	■	Luteococcus	■	Marinilac
ordia	■	Lactococcus	■	Leuconostoc	■	Lutibacter	■	Marinine
osakonia	■	Lagierella	■	Levilactobacillus	■	Lutispora	■	Marinoba
ibbella	■	Lancefieldella	■	Levyella	■	Lysinibacillus	■	Marinoba
ibbia	■	Laribacter	■	Lichenibacterium	■	Lysobacter	■	Marinoco
edonobacter	■	Larsenimonas	■	Ligilactobacillus	■	Macrococcoides	■	Marinom
urthia	■	Latilactobacillus	■	Limibacter	■	Macrococcus	■	Marisedi
utzneria	■	Lautropia	■	Limosilactobacillus	■	Mageeibacillus	■	Maritimil
yrpidia	■	Lawsonella	■	Liquorilactobacillus	■	Magnetovibrio	■	Marivirga
ytococcus	■	Lebetimonas	■	Listeria	■	Mailhella	■	Marixant
abrys	■	Lederbergia	■	Litorisediminivivens	■	Malacoplasma	■	Marmori
aceyella	■	Leeia	■	Loigolactobacillus	■	Malikia	■	Marseilla
achnoanaerobaculum	■	Legionella	■	Longimicrobium	■	Mameliella	■	Massilia
achnoclostridium	■	Leifsonia	■	Longispora	■	Mammaliicoccus	■	Mecherc
achnospira	■	Lentilactobacillus	■	Lonsdalea	■	Mangrovimonas	■	Mediterra

Phylum Level Abundances

```r
relab_phylum <- ggplot(merged_df, aes(x = Sample, y = Abundance, fill = Phylum)) +
  geom_bar(position ="fill",stat = "identity") +
  facet_grid(cols = c(vars(Sample_type), vars(status), vars(pipeline)), scales = "free_x")
relab_phylum +
  theme(legend.position = "none",
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```
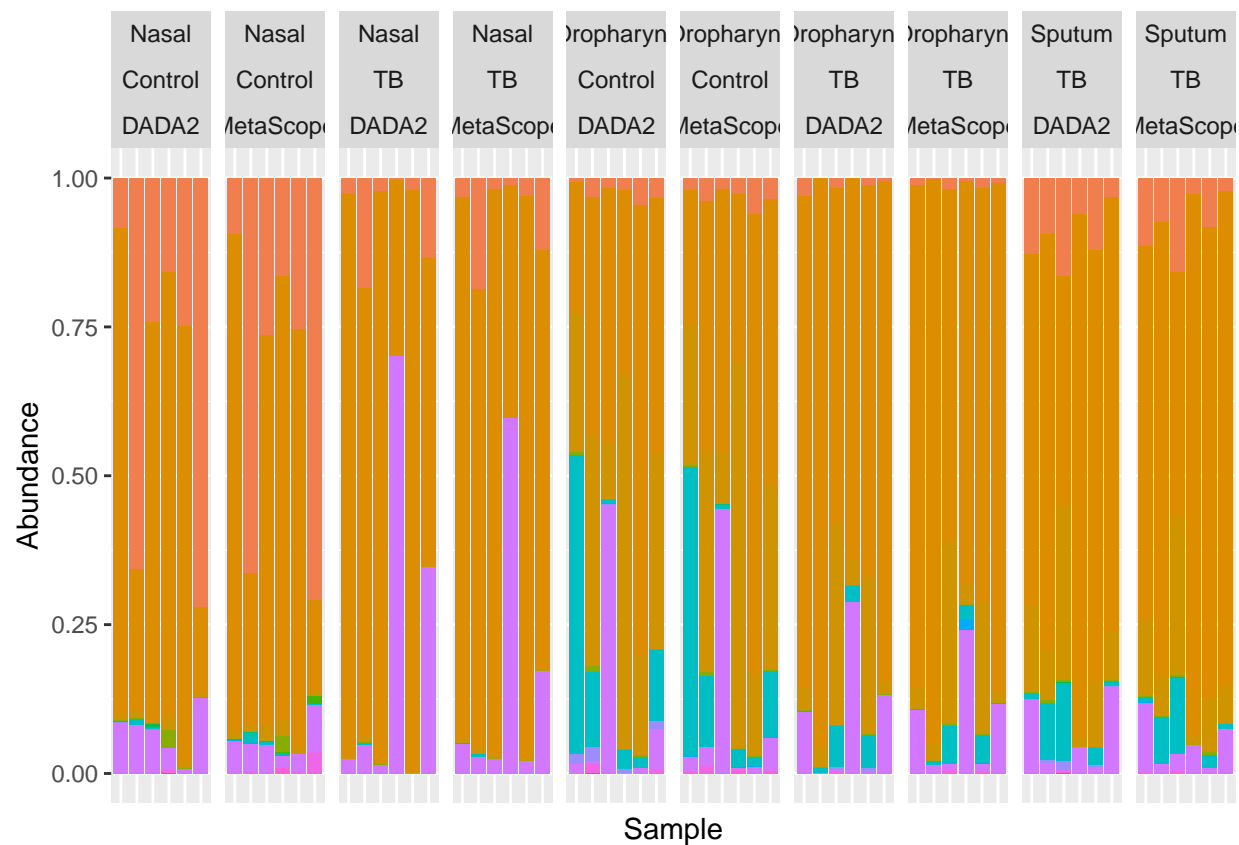
```
relab_phylum_legend <- get_legend(relab_phylum)
as_ggplot(relab_phylum_legend)
```

Phylum

■	Acidobacteriota	■	Gemmatimonadota
■	Actinomycetota	■	Ignavibacteriota
■	Aquificota	■	Lentisphaerota
■	Bacillota	■	Mycoplasmatota
■	Bacteroidota	■	Myxococcota
■	Balneolota	■	Nitrospirota
■	Bdellovibrionota	■	Patescibacteria
■	Caldisericota	■	Planctomycetota
■	Campylobacterota	■	Pseudomonadota
■	Chlorobiota	■	Rhodothermota
■	Chloroflexota	■	Spirochaetota
■	Chrysiogenota	■	Synergistota
■	Cyanobacteriota	■	Thermodesulfobacteriota
■	Deferribacterota	■	Thermomicrobiota
■	Deinococcota	■	Thermotogota
■	Dictyoglomota	■	Verrucomicrobiota
■	Fibrobacterota	■	NA
■	Fusobacteriota		

Heatmaps

```
plot_heatmap(ps_ms, sample.label="Sample_type")
```

```
## Warning in scale_fill_gradient(low = low, high = high, trans = trans, na.value
## = na.value): log-4 transformation introduced infinite values.
```
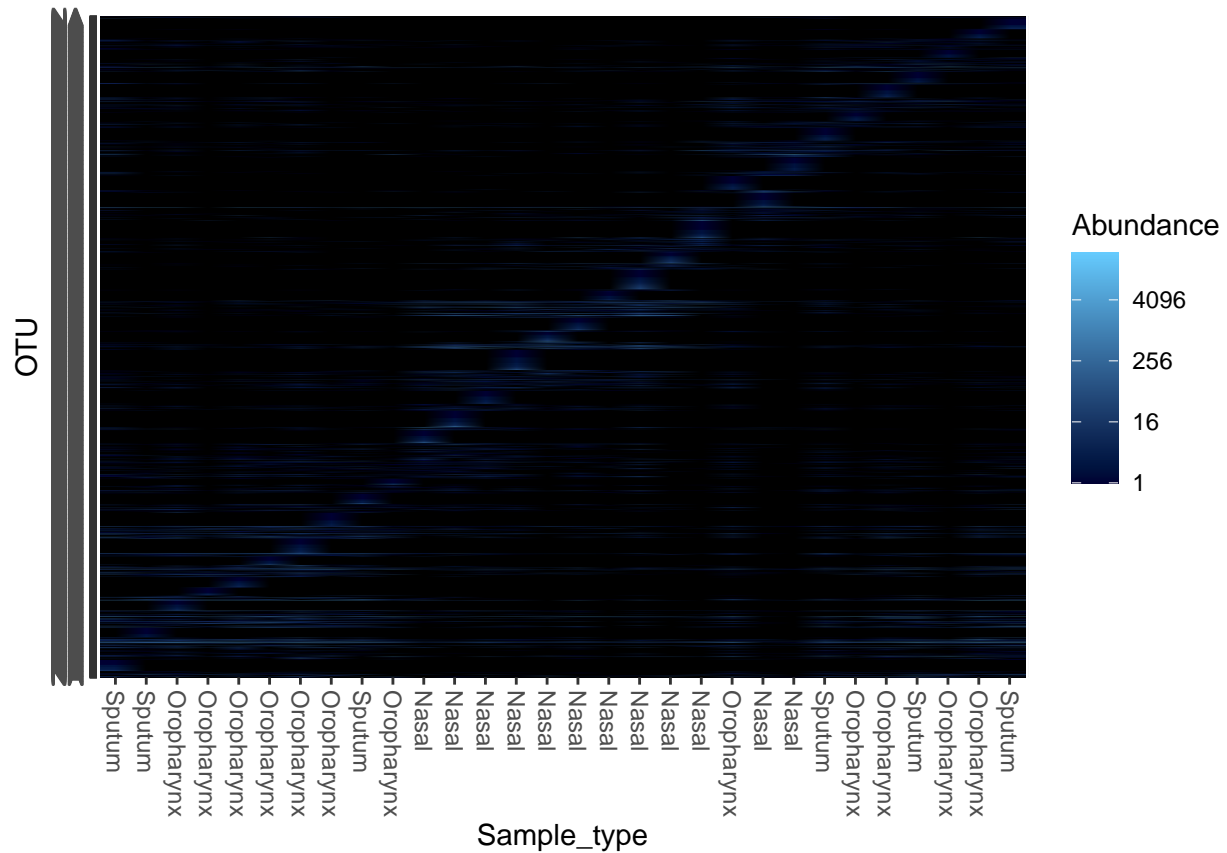
```
plot_heatmap(ps_dada2, sample.label="Sample_type")
```

```
## Warning in metaMDS(veganifyOTU(physeq), distance, ...): stress is (nearly)
## zero: you may have insufficient data
```

```
## Warning in postMDS(out$points, dis, plot = max(0, plot - 1), ...): skipping
## half-change scaling: too few points below threshold
```

```
## Warning in scale_fill_gradient(low = low, high = high, trans = trans, na.value
## = na.value): log-4 transformation introduced infinite values.
```
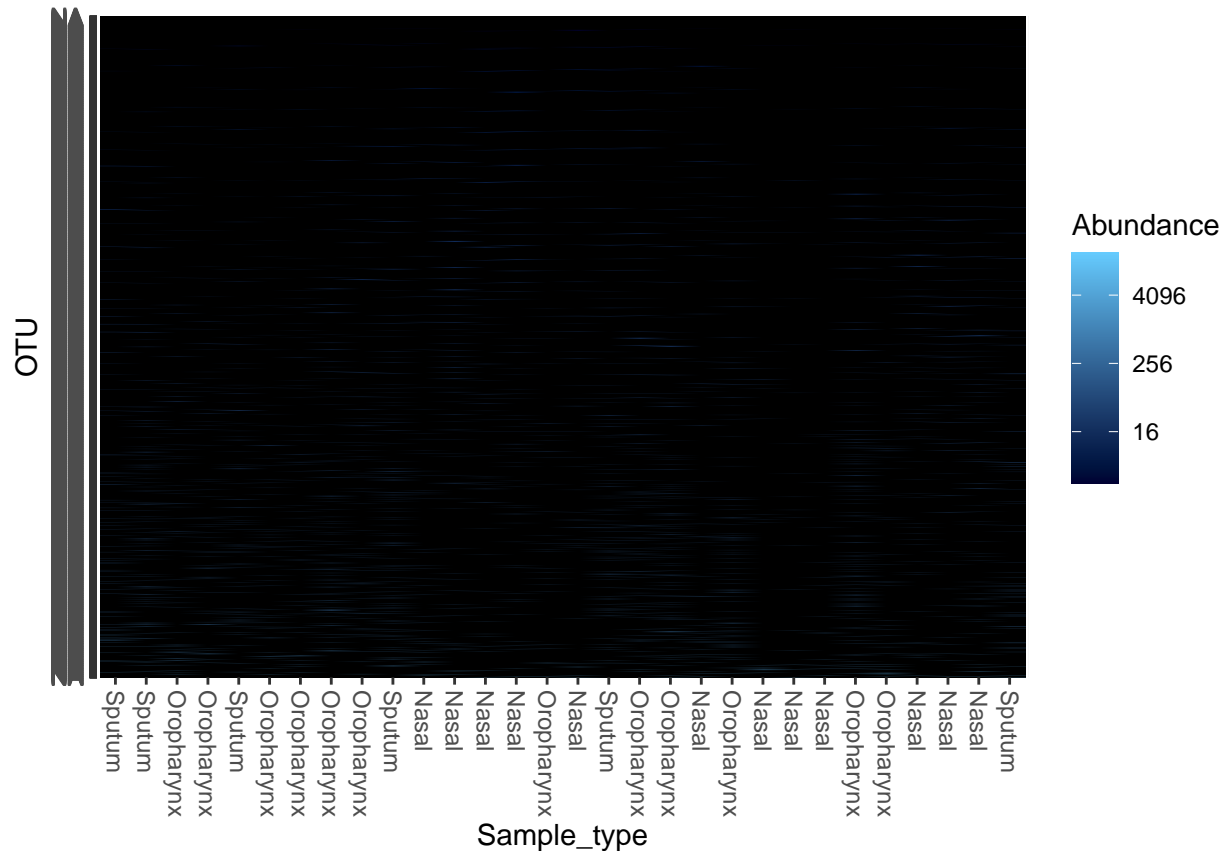
Filtered abundance barplots

```r
high_abund_dada2 <- merged_df |>
  dplyr::filter(Abundance > 1000) |>
  dplyr::filter(pipeline == "DADA2")

high_abund_ms <- merged_df |>
  dplyr::filter(Abundance > 1000) |>
  dplyr::filter(pipeline == "MetaScope")

select_species_both <- unique(high_abund_dada2$Species[high_abund_dada2$Species %in% high_abund_ms$Spec
select_species_dada2 <- unique(high_abund_dada2$Species[!(high_abund_dada2$Species %in% high_abund_ms$S
select_species_ms <- unique(high_abund_ms$Species[!(high_abund_ms$Species %in% high_abund_dada2$Species

ms_unique_species <- merged_df |>
  dplyr::filter(Species %in% select_species_ms) |>
  ggplot(aes(fill=Sample_type, y=Abundance, x=Species)) +
  geom_bar(position="dodge", stat="identity") +
  facet_grid(vars(pipeline), vars(status)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
ms_unique_species
```
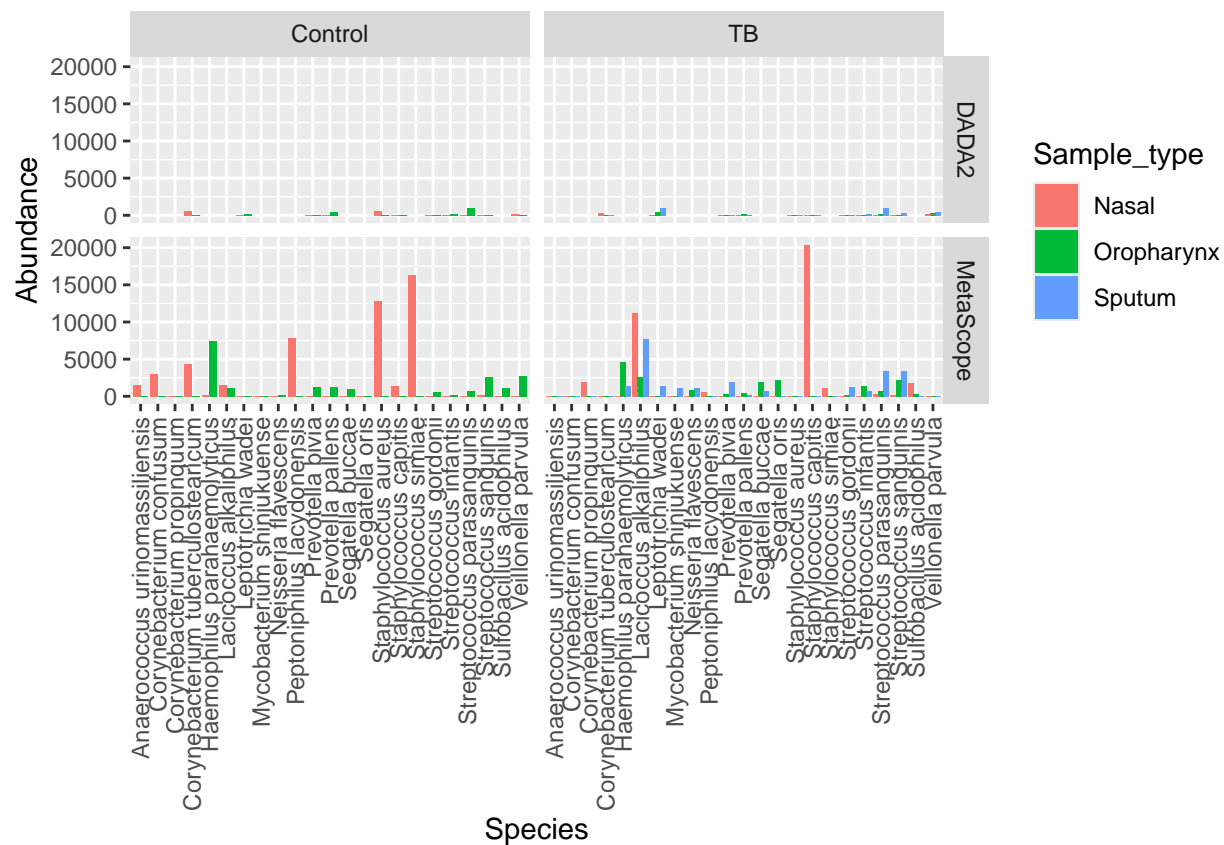
```r
dada2_unique_species <- merged_df |>
  dplyr::filter(Species %in% select_species_dada2) |>
  ggplot(aes(fill=Sample_type, y=Abundance, x=Species)) +
  geom_bar(position="dodge", stat="identity") +
  facet_grid(vars(pipeline), vars(status)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
dada2_unique_species
```
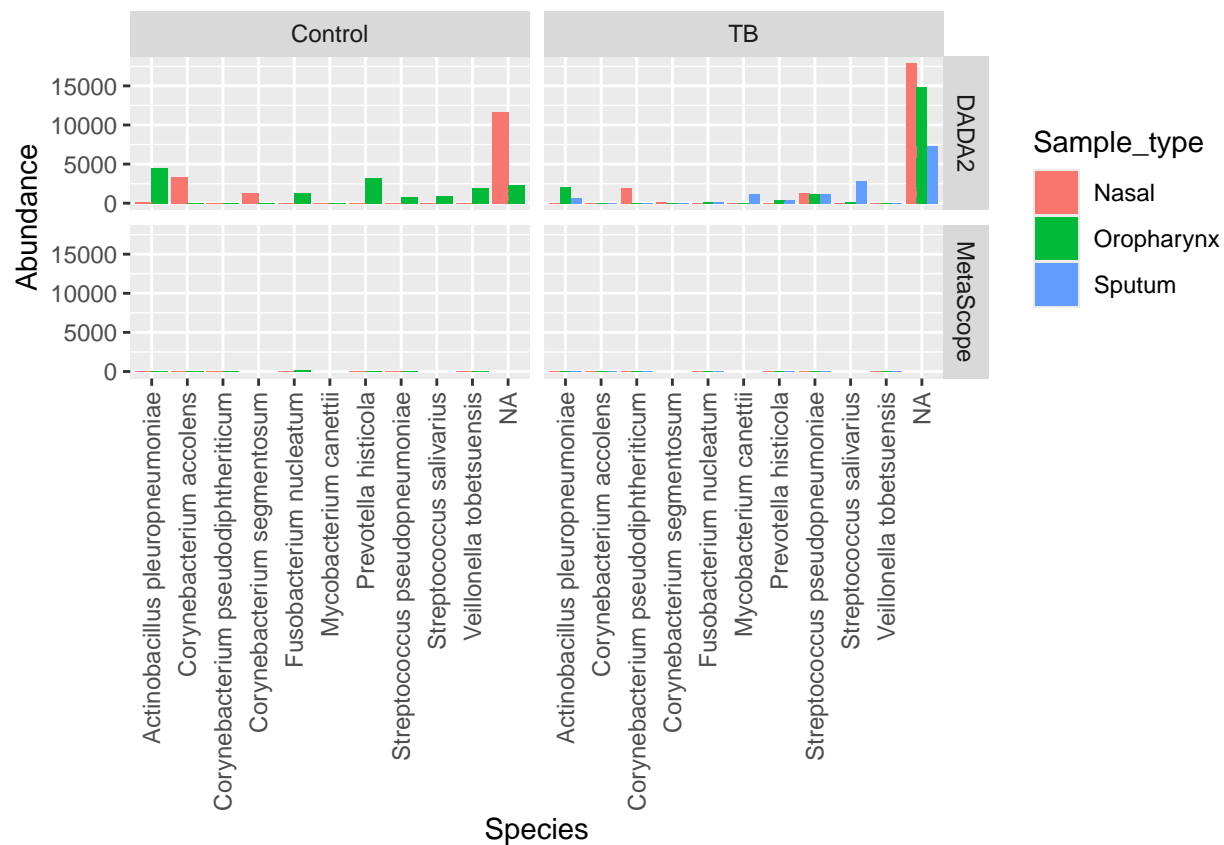
```
high_abund_species <- merged_df |>
  dplyr::filter(Species %in% select_species_both) |>
  ggplot(aes(fill=Sample_type, y=Abundance, x=Species)) +
  geom_bar(position="dodge", stat="identity") +
  facet_grid(vars(pipeline), vars(status)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
high_abund_species
```
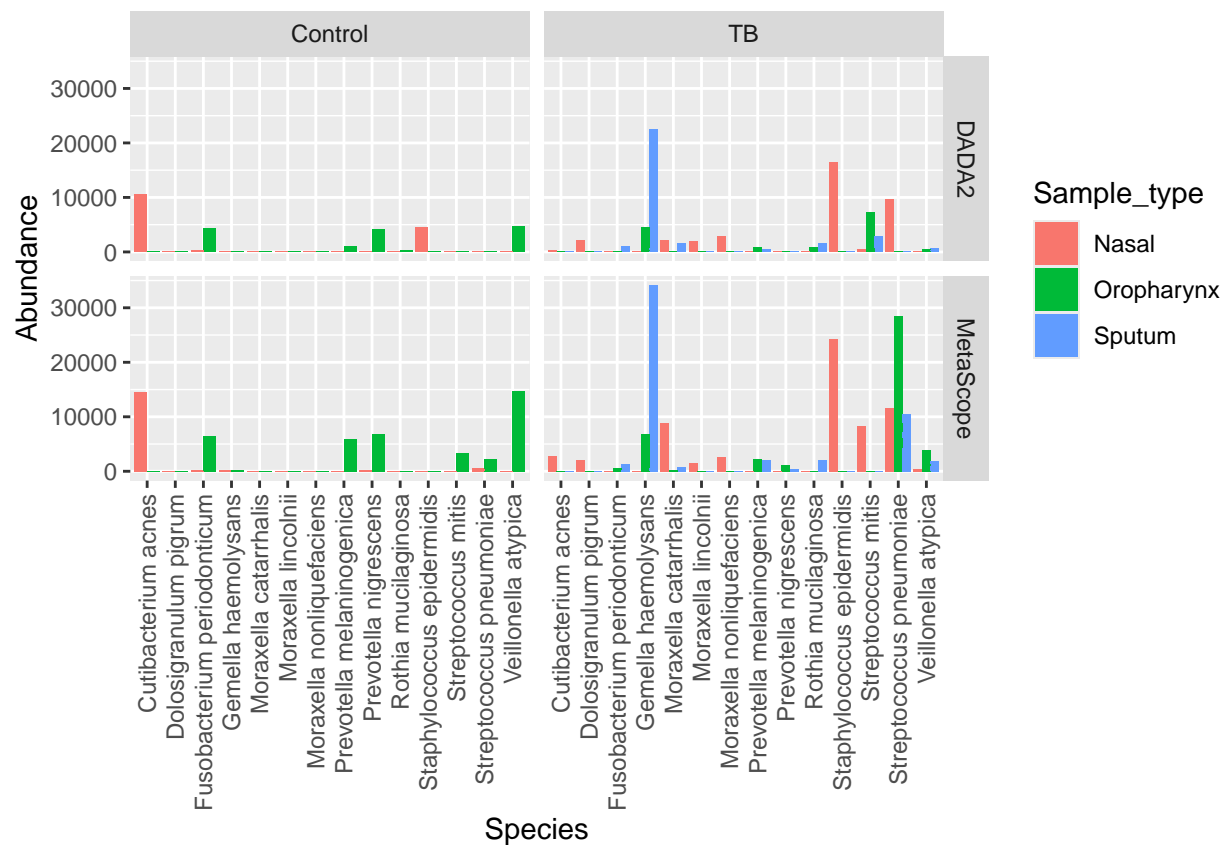
```
abundance_heatmap(ps_ms, classification = 'species',
                  treatment = c("status","Sample_type"), transformation = 'relative_abundance')
```
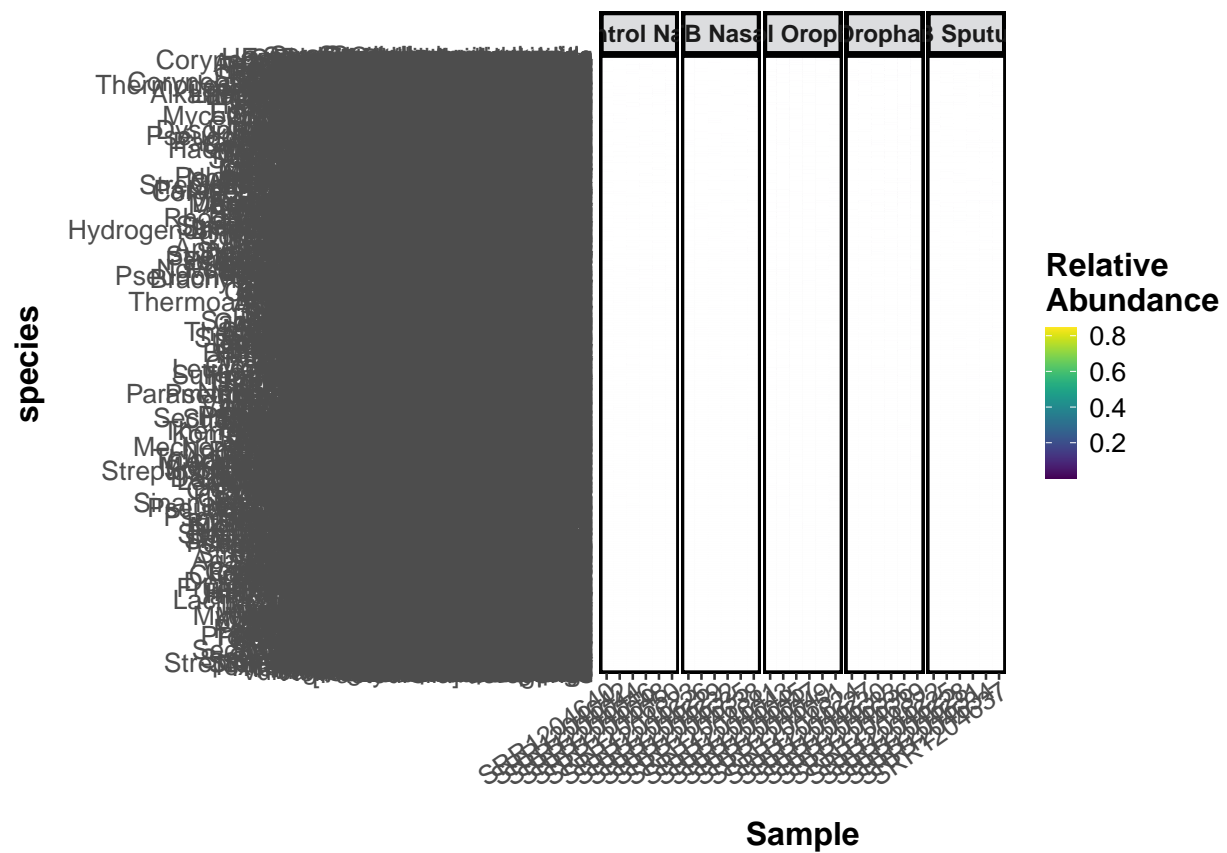
```
abundance_heatmap(ps_dada2, classification = 'Species',
                treatment = c("status","Sample_type"), transformation = 'relative_abundance')
```

```
phylogeny_profile(ps_ms, classification = 'family',
  treatment = c("status","Sample_type"), merge = TRUE,
  relative_abundance = TRUE)
```

Desulfuribacillaceae	Sphaerotilaceae	Halomonadaceae
Thiobacillaceae	Dermabacteraceae	Caminicellaceae
Legionellaceae	Shewanellaceae	Aurantimonadaceae
Eubacteriales Family XII. Incertae Sedis	Nautiliaceae	Sphingobacteriaceae
Deferrisomataceae	Jonesiaceae	Cytophagaceae
Vicinamibacteraceae	Nocardiopsidaceae	Glycomycetaceae
Thioprofundaceae	Moorellaceae	Flammeovirgaceae
Gaiellaceae	Desulfovibrionaceae	Chitinophagaceae
Cellvibrionaceae	Cellulomonadaceae	Roseobacteraceae
Saccharospirillaceae	Brucellaceae	Intrasporangiaceae
Thermomonosporaceae	Desulfotomaculaceae	Nitrobacteraceae
Methylococcaceae	Acidaminococcaceae	Chromobacteriaceae
Acutalibacteraceae	Promicromonosporaceae	Listeriaceae
Spiroplasmataceae	Celerinatantimonadaceae	Planococcaceae
Demequinaceae	Sphaerobacteraceae	Micromonosporaceae
Ahrensiaceae	Geodermatophilaceae	Erythrobacteraceae
Sporolactobacillaceae	Methylophilaceae	Bacillales Family X. In
Rhodovibrionaceae	Desulfitibacteraceae	Dietziaceae
Bartonellaceae	Rhizobiaceae	Nocardioidaceae
Alteromonadaceae	Pectobacteriaceae	Marinilabiliaceae
Blastochloridaceae	Wilmottiaceae	Peptococcaceae
Acidithiobacillaceae	Lewinellaceae	Dysgonomonadaceae
Pseudobdellovibrionaceae	Tsukamurellaceae	Piscirickettsiaceae
Xenococcaceae	Iamiaceae	Eggerthellaceae
Syntrophotaleaceae	Xanthobacteraceae	Unclassified
Leptospiraceae	Francisellaceae	Oscillochloridaceae
Kallotenuaceae	Dermacoccaceae	Hymenobacteraceae
Geovibrionaceae	Blastocatellaceae	Acetobacteraceae
Archangiaceae	Zoogloeaceae	Anaerotignaceae

ae

```r
filtered_ms <- taxa_filter(ps_ms, frequency = 0.8)
taxa_abundance_bars(filtered_ms, classification = 'genus',
  treatment = c("status","Sample_type")) +
  geom_boxplot() +
  geom_point()
```
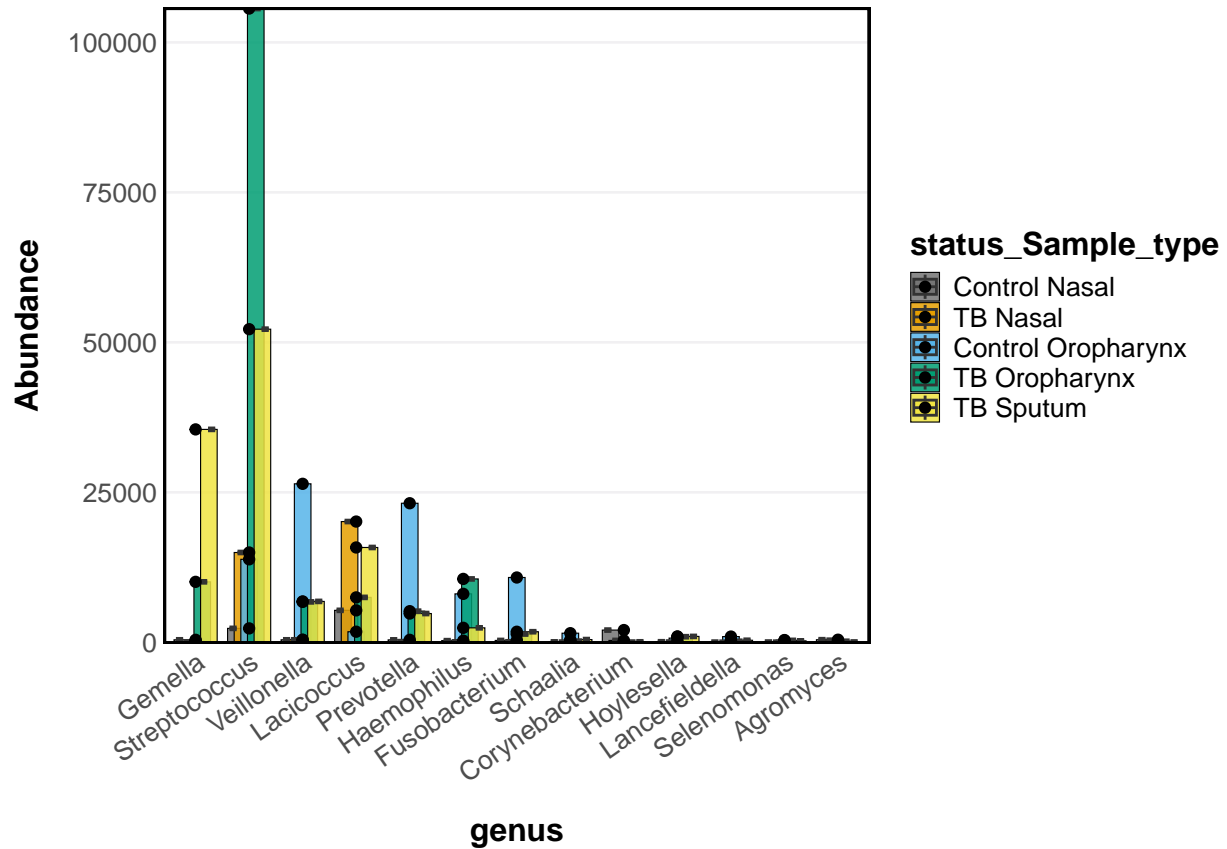
Plotting Alpha Diversity

```r
p2_1 <- plot_richness(ps_dada2_oral, measures=c("Observed", "Chao1", "ACE", "Shannon", "Simpson", "InvS
  stat_compare_means(label = "p.signif", label.x = 1.5, comparisons = list(c("Control", "TB"))) +
  geom_boxplot()

p2_2 <- plot_richness(ps_ms_oral, measures=c("Observed", "Chao1", "ACE", "Shannon", "Simpson", "InvSimps
  stat_compare_means(label = "p.signif", label.x = 1.5, comparisons = list(c("Control", "TB"))) +
  geom_boxplot()

p2_1
```
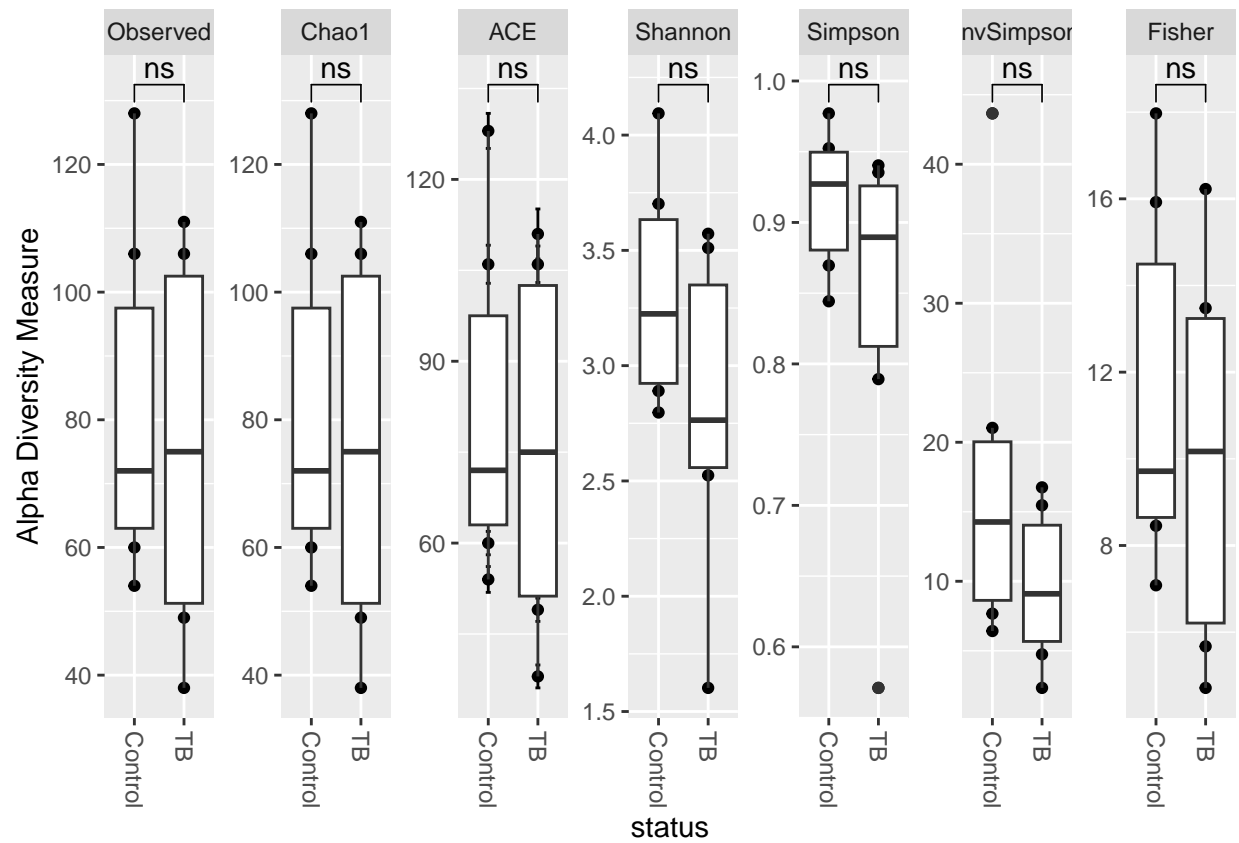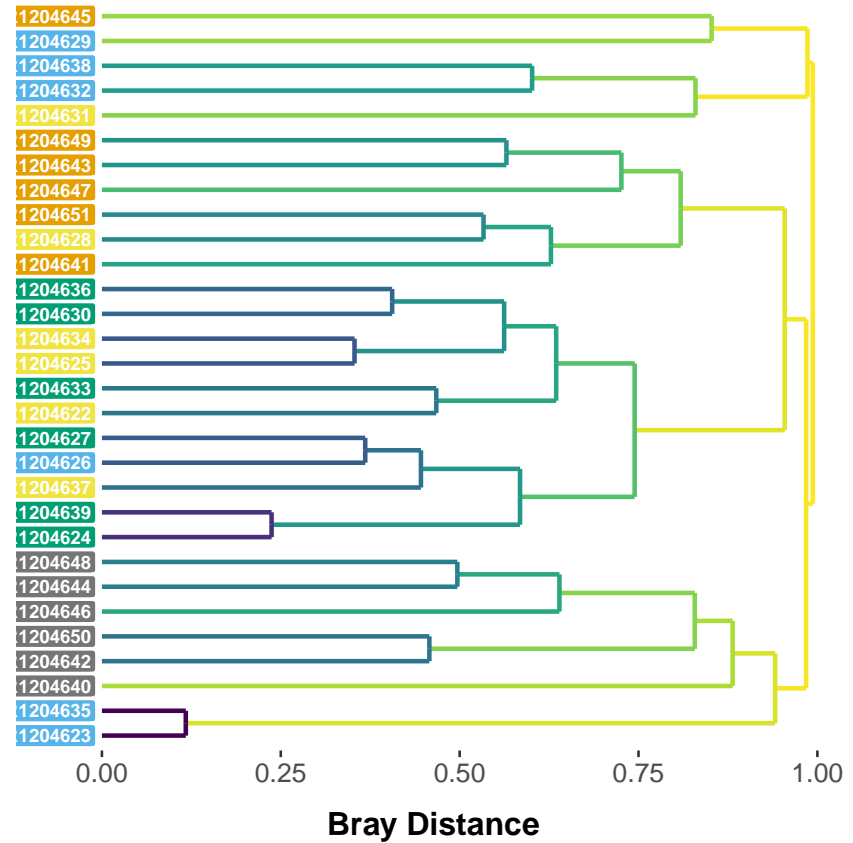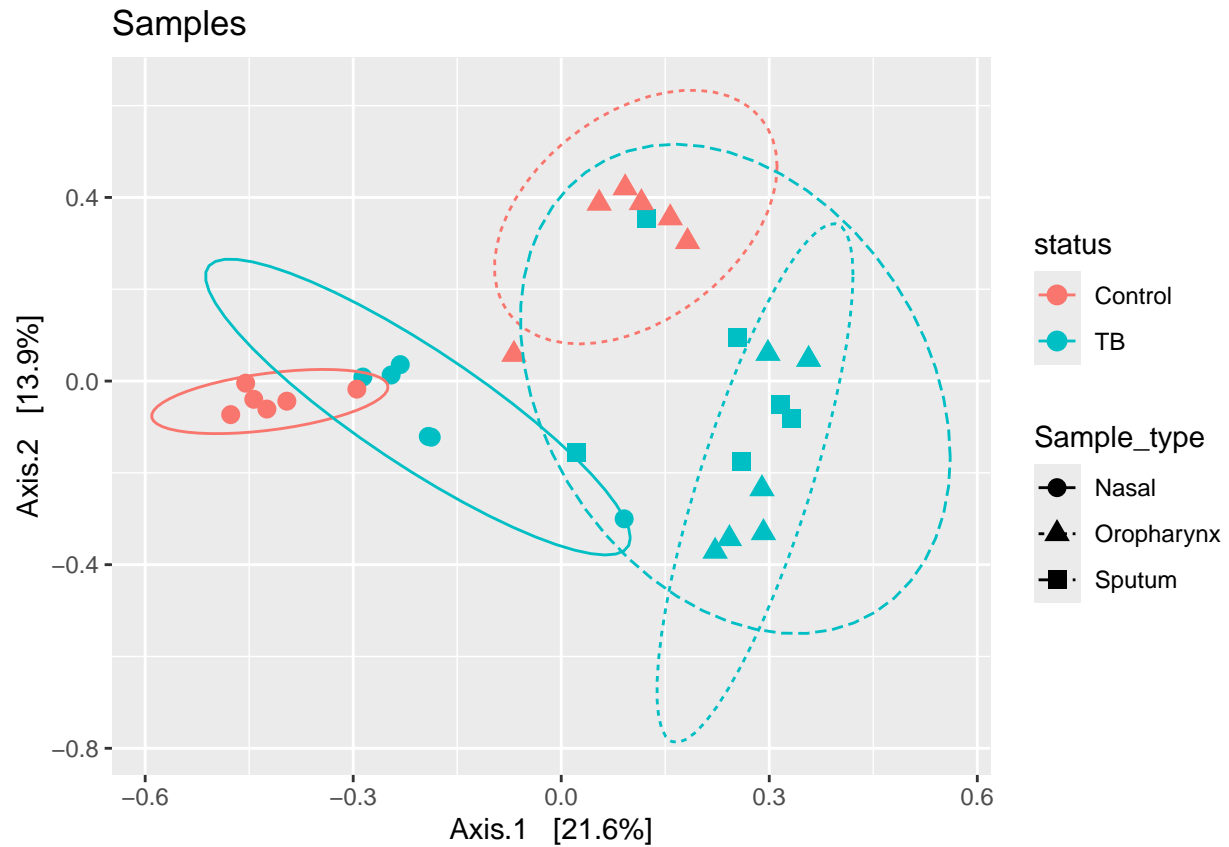
```
dendrogram_phyloseq(ps_ms, c("status","Sample_type"), method = 'bray')
```

```
ps_ms.ord <- ordinate(ps_ms, "PCoA", "bray")
plot_ordination(ps_ms, ps_ms.ord, type="samples", color="status", shape="Sample_type", title="Samples")
  geom_point(size=3) +
  stat_ellipse(
    aes(linetype=Sample_type))
```

```
ps_dada2.ord <- ordinate(ps_dada2, "PCoA", "bray")
plot_ordination(ps_dada2, ps_dada2.ord, type="samples", color="status", shape="Sample_type", title="Samp
  geom_point(size=3) +
  stat_ellipse(
    aes(linetype=Sample_type))
```

Samples