

botero_analysis

Sean Lu

2025-05-13

Setup

```
library("phyloseq")  
library("ggplot2")  
library("dplyr")
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
##   filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library("tibble")  
library("ggpubr")  
library("phylosmith")
```

```
## Registered S3 method overwritten by 'dendextend':  
##   method      from  
##   rev.hclust  vegan
```

Load in data

```
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

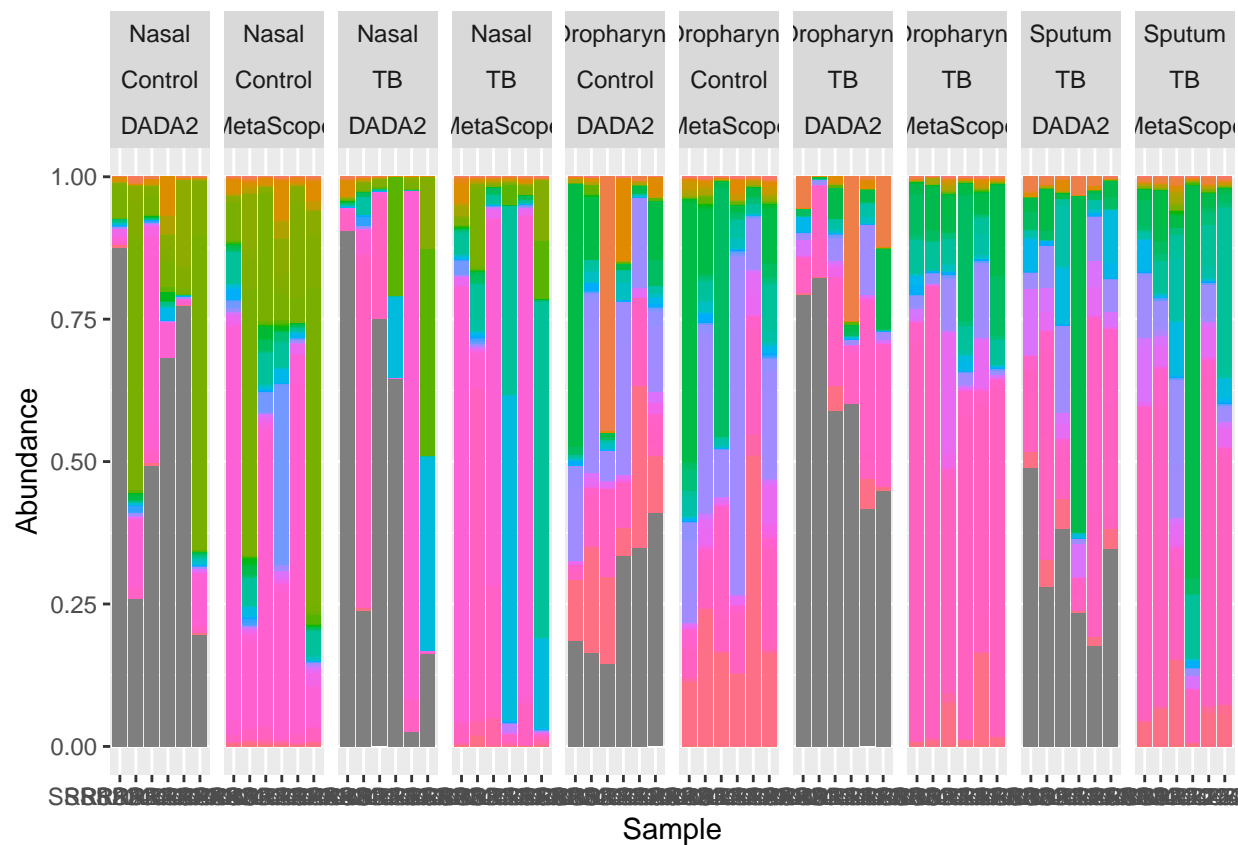
```
dada2_df <- psmelt(ps_dada2) |>
  dplyr::mutate(Species = ifelse(is.na(Species), NA, paste0(Genus, " ", Species)))
dada2_df$pipeline = "DADA2"
colnames(dada2_df)
```

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

```
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)

p1 <- 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")
p1 + theme(legend.position = "none")
```

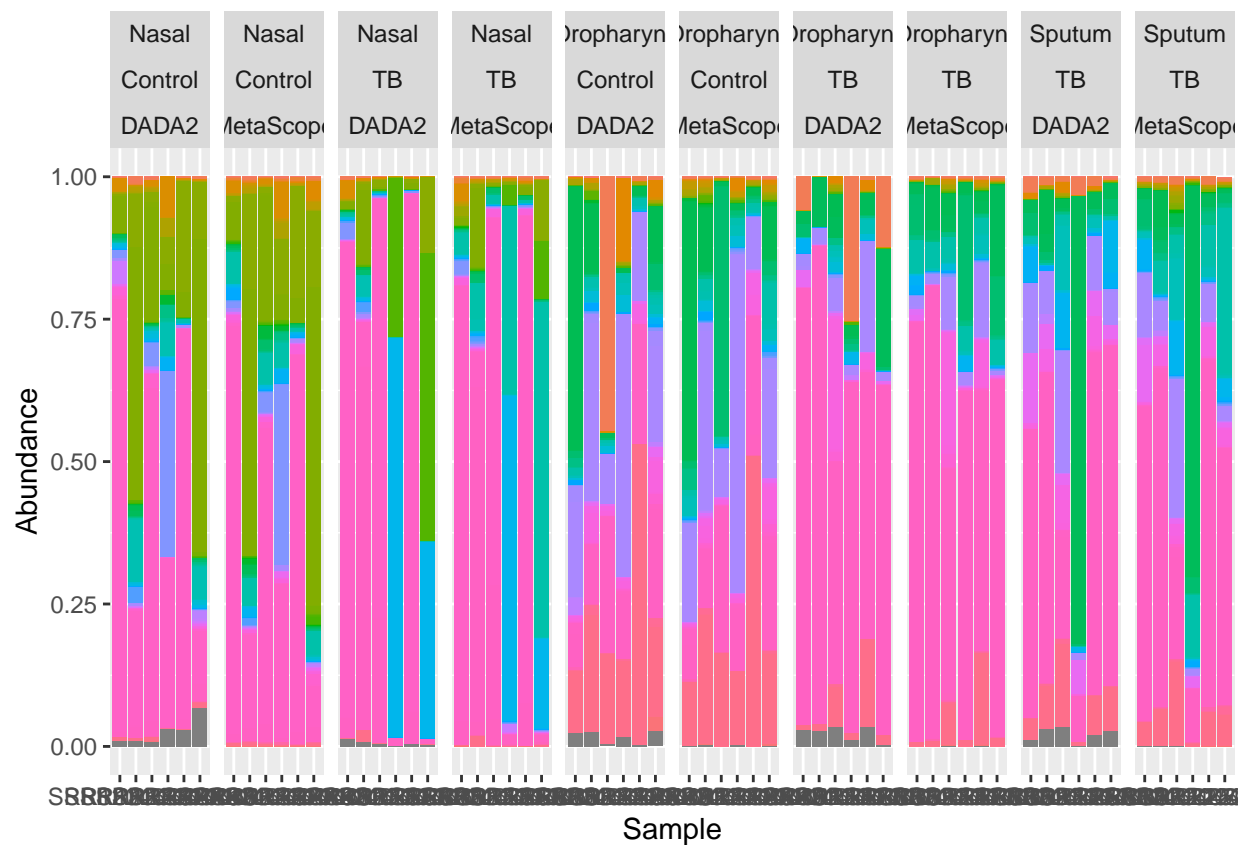


```
p1_legend <- get_legend(p1)
as_ggplot(p1_legend)
```

§ aquiplantarum	Mechercharimyces asporophorigenens	Mesorhizobium hankyongi	Methylob:
§ arenicola	Mediterraneibacter faecis	Mesorhizobium huakuii	Methylob:
§ balearica	Mediterraneibacter gnavus	Mesorhizobium thiogangneticum	Methylob:
§ communis	Megalodesulfovibrio gigas	Metallibacterium scheffleri	Methylob:
§ foliarum	Megamonas hypermegale	Metallococcus carri	Methylob:
§ mediterranea	Megamonas rupellensis	Metamycoplasma arthritidis	Methylob:
§ pollencensis	Megasphaera cerevisiae	Metamycoplasma cloacale	Methylob:
§ posidonica	Megasphaera elsdenii	Metamycoplasma orale	Methyloc:
§ rhizomae	Megasphaera micronuciformis	Metamycoplasma salivarium	Methyloc:
§ vaga	Megasphaera paucivorans	Methylobacillus flagellatus	Methyloc:
cola antarctica	Megasphaera sueciensis	Methylobacillus glycogenes	Methyloc:
ir fusiformis	Melioribacter roseus	Methylobacterium aquaticum	Methyloc:
tuosa	Melissococcus plutonius	Methylobacterium dankookense	Methylom
onas ophiurae	Melittangium lichenicola	Methylobacterium durans	Methylom
ssiliensis	Mesobacillus campisalis	Methylobacterium fujisawaense	Methylom
italea	Mesobacillus subterraneus	Methylobacterium goesingense	Methylom
richensis	Mesomycoplasma neurolyticum	Methylobacterium iners	Methylopl
onensis	Mesoplasma corruscae	Methylobacterium isbiliense	Methylopl
ianensis	Mesoplasma seiffertii	Methylobacterium komagatae	Methylopl

Genus level Abundances

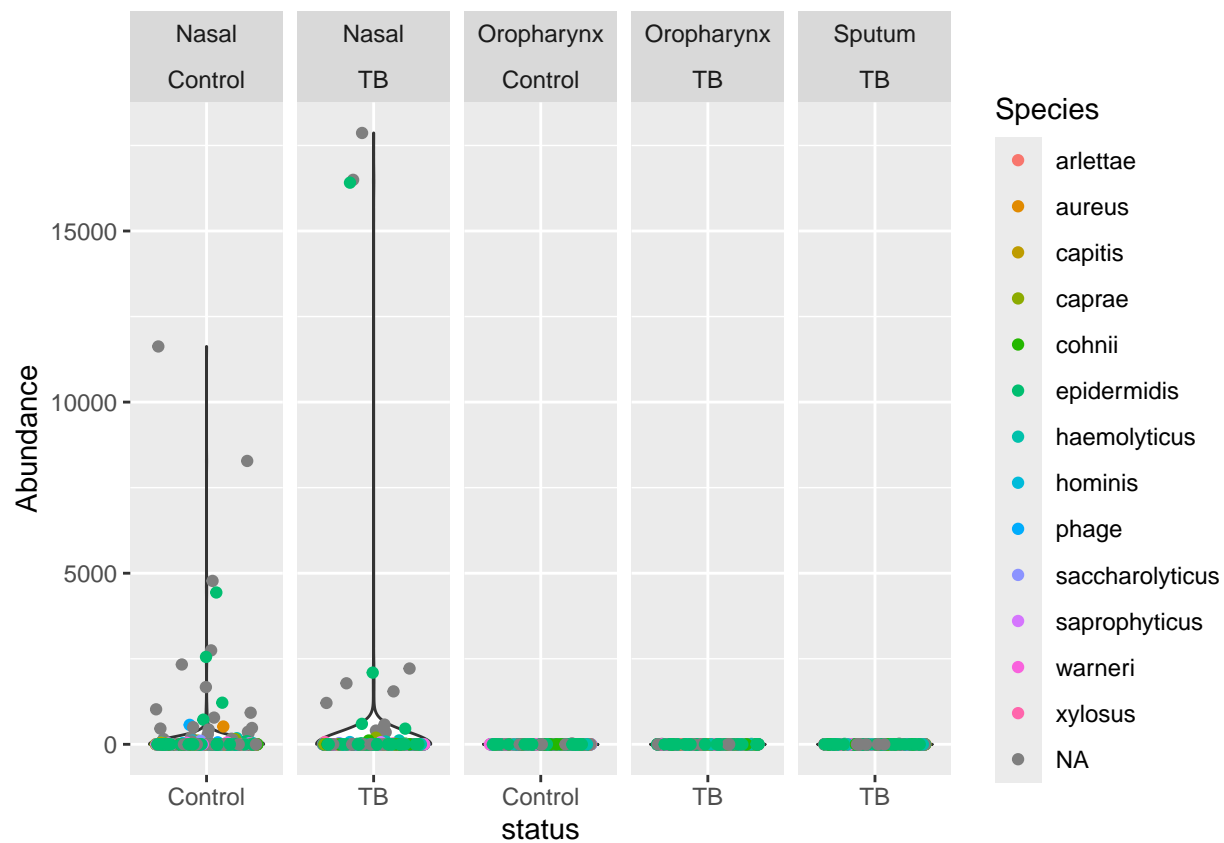
```
p2 <- ggplot(merged_df, aes(x = Sample, y = Abundance, fill = Genus)) +
  geom_bar(position = "fill", stat = "identity") +
  facet_grid(cols = c(vars(Sample_type), vars(status), vars(pipeline)), scales = "free_x")
p2 + theme(legend.position = "none")
```



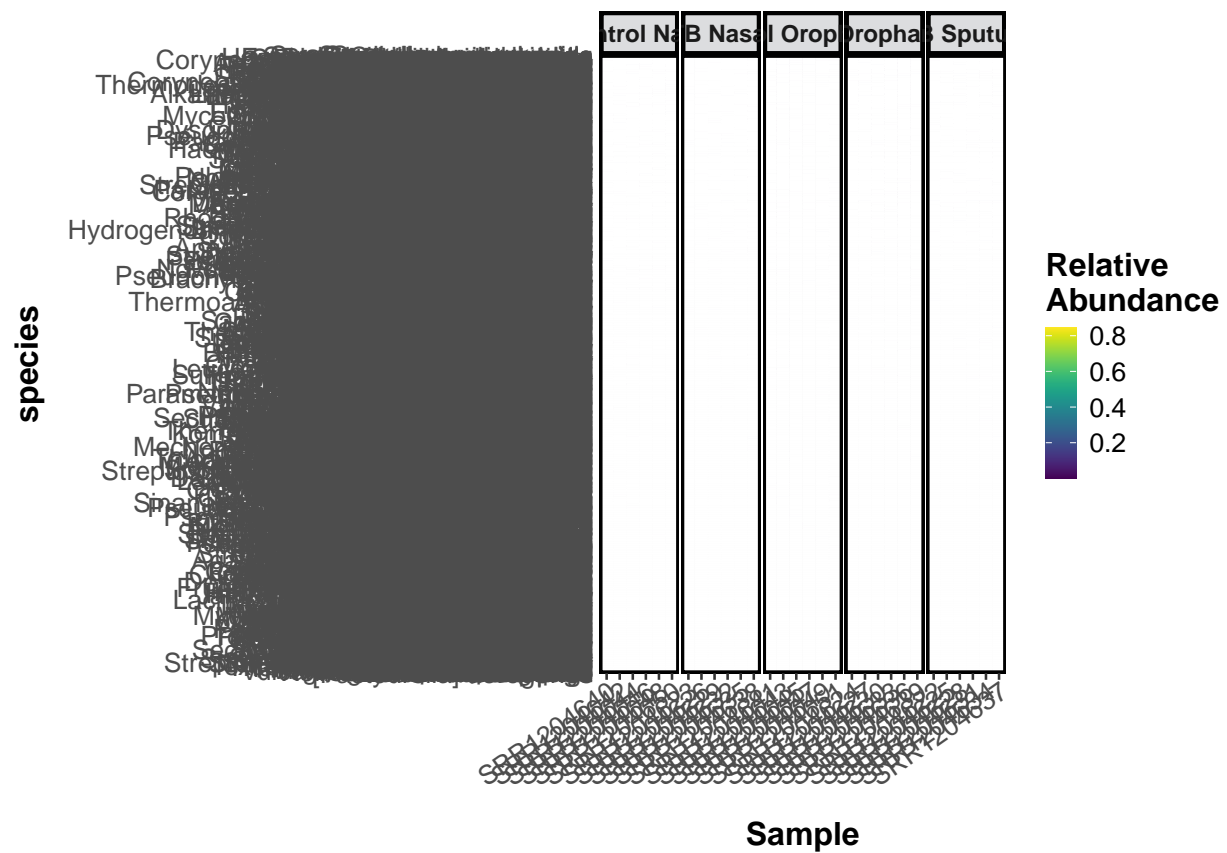
```
p2_legend <- get_legend(p2)
as_ggplot(p2_legend)
```

ebsiella	Lacrimispora	Lentzea	Luteibacter	Maribact
enkia	Lacticaseibacillus	Leptolyngbya PCC-6306	Luteibaculum	Marichro
uyvera	Lacticigenium	Leptospira	Luteimicrobium	Marinicri
ocuria	Lactiplantibacillus	Leptotrichia	Luteimonas	Marinilat
omagataeibacter	Lactobacillus	Leucobacter	Luteococcus	Marinilac
ordia	Lactococcus	Leuconostoc	Lutibacter	Marinine
sakonia	Lagierella	Levilactobacillus	Lutispora	Marinob:
ibbella	Lancefieldella	Levyella	Lysinibacillus	Marinob:
ibbia	Laribacter	Lichenibacterium	Lysobacter	Marinoc:
edonobacter	Larsenimonas	Ligilactobacillus	Macrococcoides	Marinom
urthia	Latilactobacillus	Limibacter	Macrococcus	Marisedi
utzneria	Lautropia	Limosilactobacillus	Mageeibacillus	Maritimik
rpidia	Lawsonella	Liquorilactobacillus	Magnetovibrio	Marivirg:
tococcus	Lebetimonas	Listeria	Mailhella	Marixant
ibrys	Lederbergia	Litorisediminivivens	Malacoplasma	Marmori
aceyella	Leeia	Loigolactobacillus	Malikia	Marseille
ichnoanaerobaculum	Legionella	Longimicrobium	Mameliella	Massilia
ichnoclostridium	Leifsonia	Longispora	Mammaliococcus	Mecherc
ichnospora	Lentilactobacillus	Lonsdalea	Mangrovimonas	Mediterr:

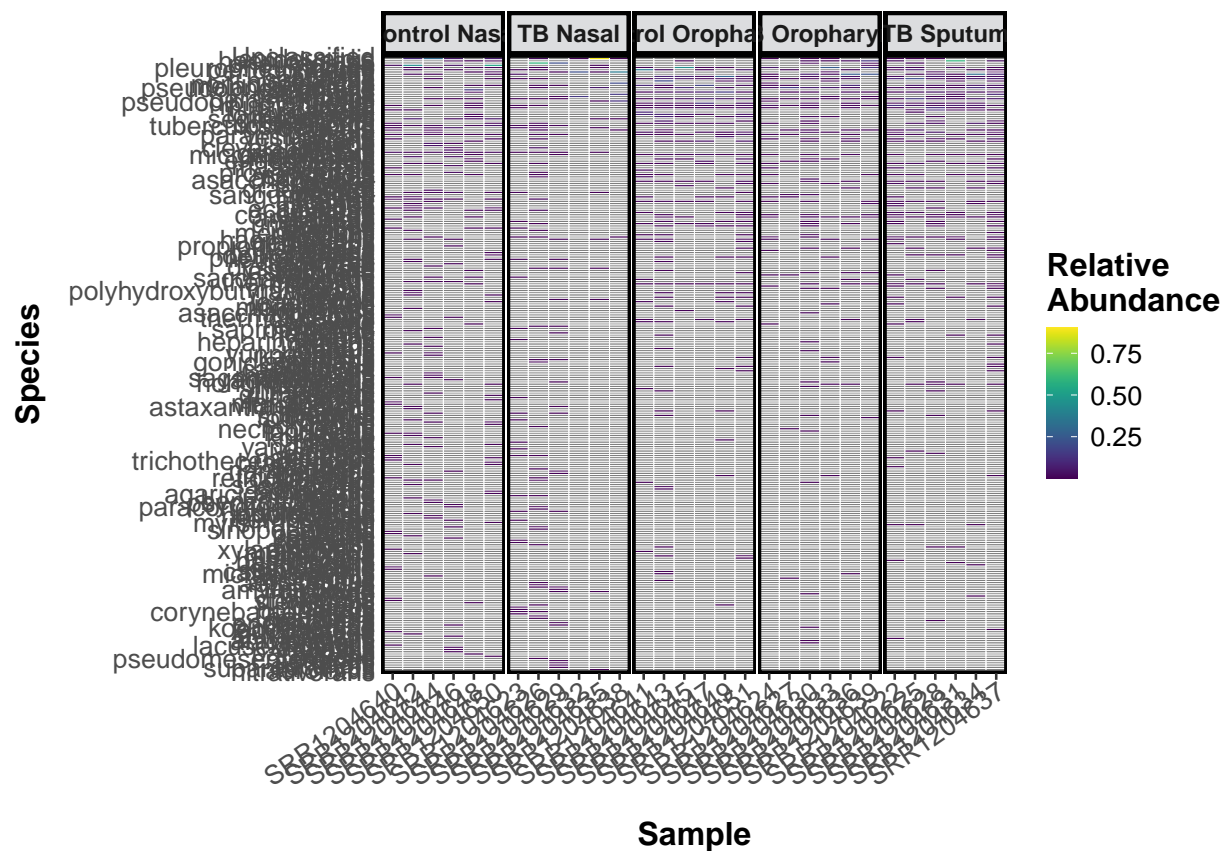
```
psmelt(ps_dada2) |> dplyr::filter(
  Genus == "Staphylococcus") |>
  ggplot(aes(x = status, y = Abundance)) +
  geom_violin() +
  geom_jitter(aes(color = Species)) +
  facet_grid(cols = c(vars(Sample_type), vars(status)), scales = "free_x")
```



```
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')
```

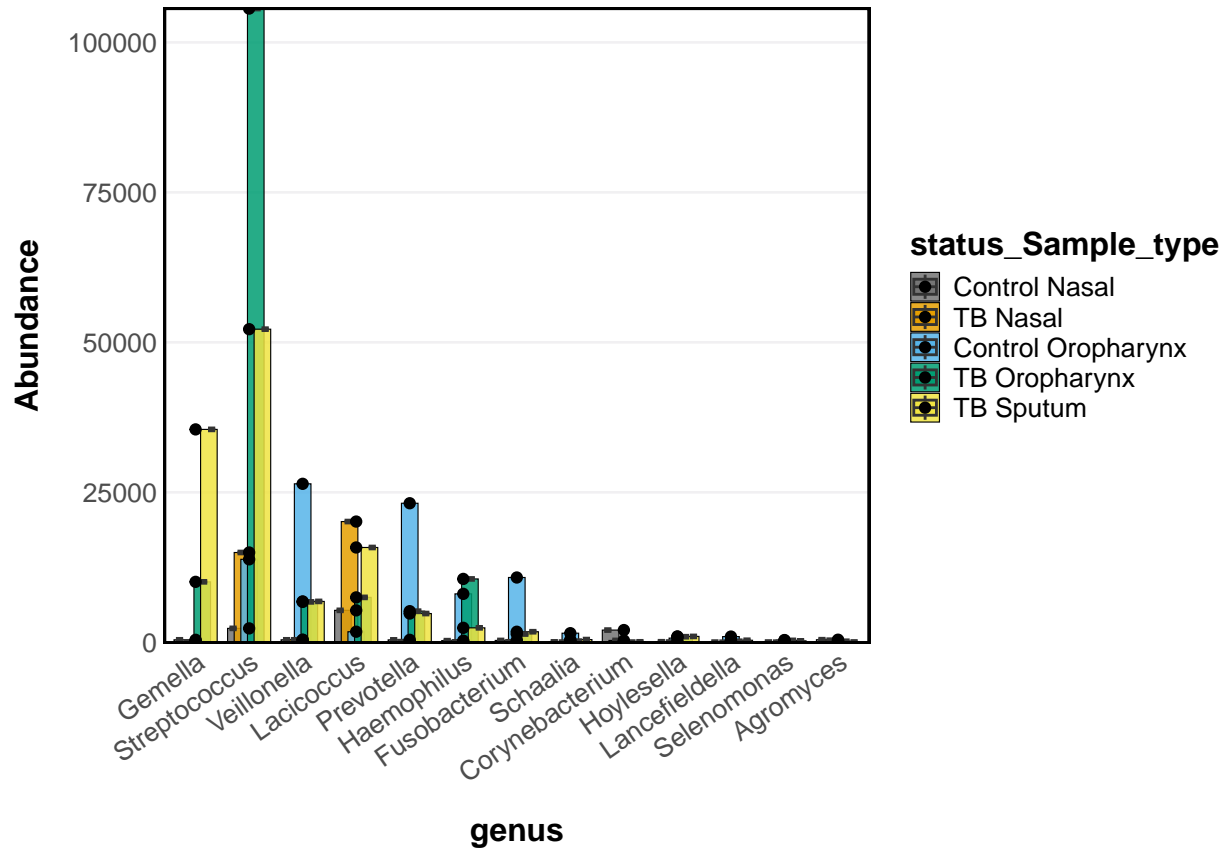



ae	Desulfuribacillaceae	Sphaerotilaceae	Halomonadaceae
	Thiobacillaceae	Dermabacteraceae	Caminiaceae
	Legionellaceae	Shewanellaceae	Aurantimonadaceae
	Eubacteriales Family XII. Incertae Sedis	Nautiliaceae	Sphingobacteriaceae
	Deferriisomataceae	Jonesiaceae	Cytophagaceae
	Vicinamibacteraceae	Nocardiopsidaceae	Glycomycetaceae
	Thiopfundaceae	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
	Ahrensiaaceae	Geodermatophilaceae	Erythrobacteraceae
	Sporolactobacillaceae	Methylophilaceae	Bacillales Family X. Ir
	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
	Archandriaceae	Zoodioaceae	Anaerotrionaceae

```

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()

```

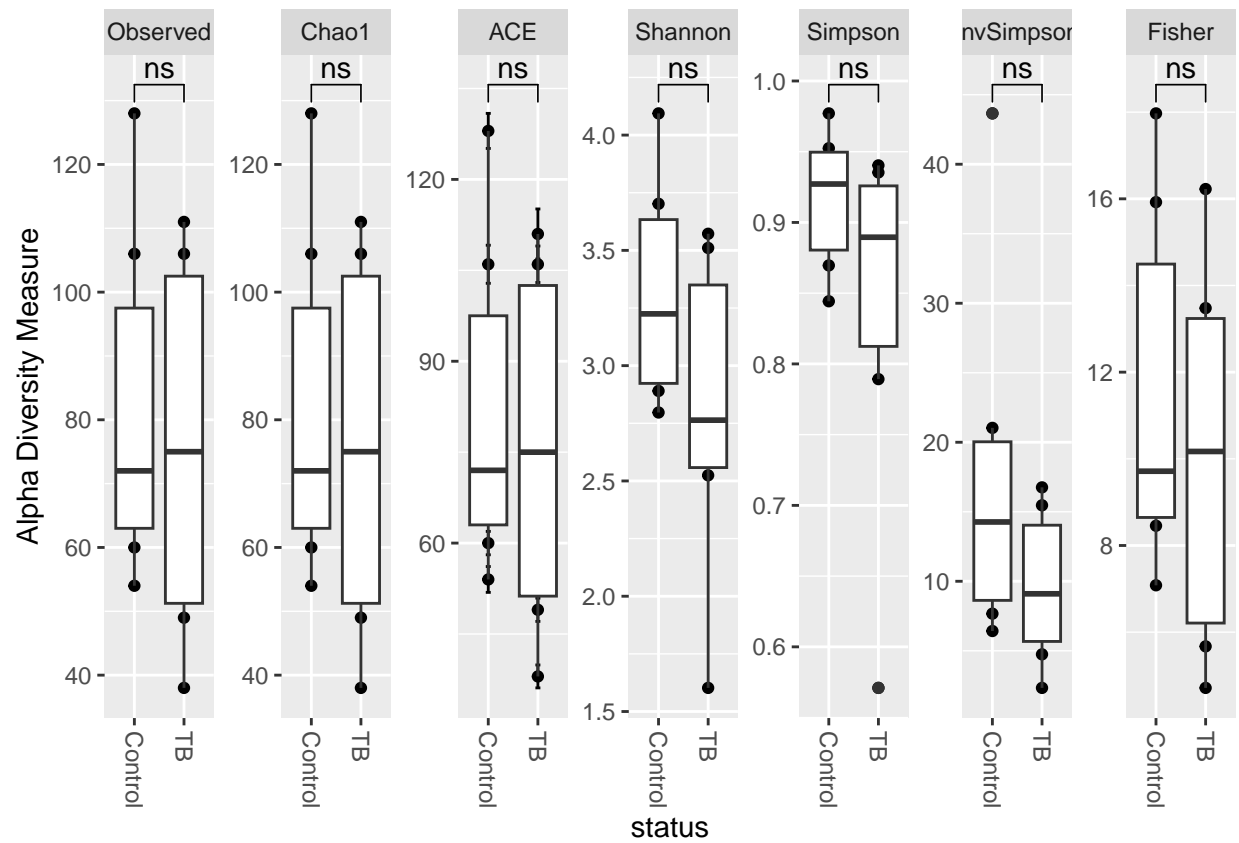


Plotting Alpha Diversity

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

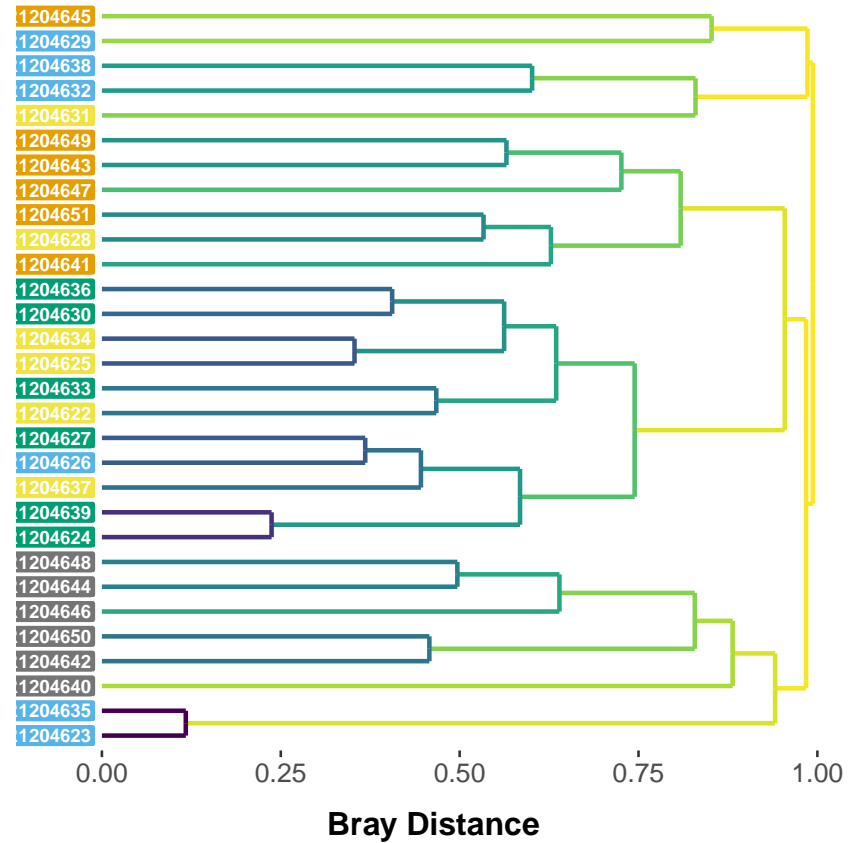
p2_1
```



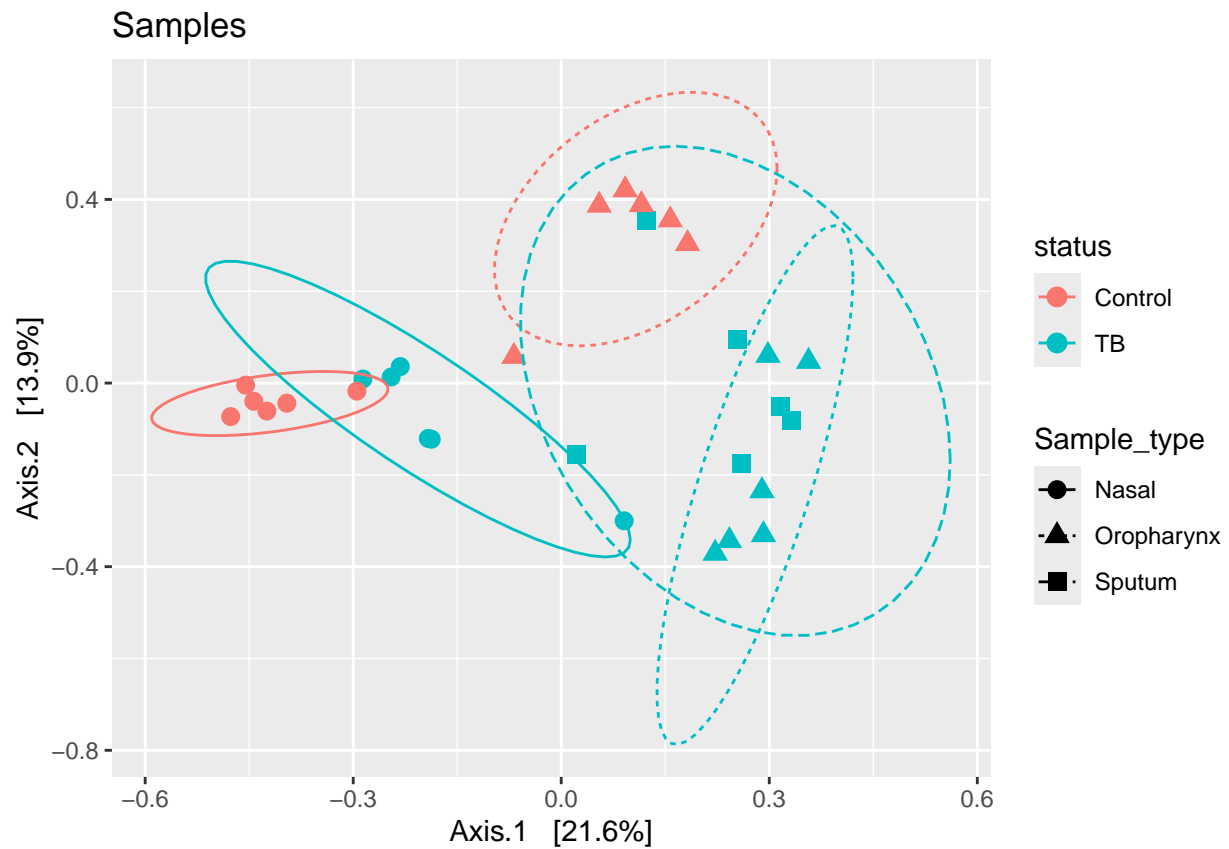
```
dendrogram_phyloseq(ps_ms, c("status", "Sample_type"), method = 'bray')
```

status x Sample_type

- Control Nasal
- Control Oropharynx
- TB Nasal
- TB Oropharynx
- TB Sputum



```
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="Sam
  geom_point(size=3) +
  stat_ellipse(
    aes(linetype=Sample_type))
```

