

# Diversidades alfa y beta en datos metagenomicos de shotgun de cultivo de fresa, a diferentes niveles taxonomicos

Camila Silva

2023-04-18

## Observacion en diferentes niveles taxonomicos

```
library("phyloseq")
library("ggplot2")
library("igraph")

##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
## 
##     decompose, spectrum

## The following object is masked from 'package:base':
## 
##     union

library("readr")
library("patchwork")
library("vegan")

## Loading required package: permute

##
## Attaching package: 'permute'

## The following object is masked from 'package:igraph':
## 
##     permute

## Loading required package: lattice

## This is vegan 2.6-4

##
## Attaching package: 'vegan'
```

```

## The following object is masked from 'package:igraph':
##
##     diversity

library("GUniFrac")
library("kableExtra")

## Registered S3 method overwritten by 'httr':
##   method      from
##   print.response rmutil

library("RColorBrewer")

```

## Cargado de datos originales

```

setwd("/home/camila/GIT/Tesis_Maestria/Data/fresa_solena/Data1")
outpath = "/home/camila/GIT/Tesis_Maestria/Analisis_Comparativo/Fresa_Solena/Results_img"

fresa_kraken <- import_biom("fresa_kraken.biom")
colnames(fresa_kraken@tax_table@Data) <- c("Kingdom", "Phylum", "Class", "Order", "Family", "Genus", "Species")
fresa_kraken@tax_table@Data <- substr(fresa_kraken@tax_table@Data, 4, 100)
colnames(fresa_kraken@otu_table@Data) <- substr(colnames(fresa_kraken@otu_table@Data), 1, 6)
metadata_fresa <- read.csv2("/home/camila/GIT/Tesis_Maestria/Data/fresa_solena/Data1/metadata.csv", header=TRUE)
fresa_kraken@sam_data <- sample_data(metadata_fresa)
fresa_kraken@sam_data$Sample <- row.names(fresa_kraken@sam_data)
colnames(fresa_kraken@sam_data) <- c('Treatment', 'Samples')
samples_to_remove <- c("MP2079", "MP2080", "MP2088", "MP2109", "MP2137")
fresa_kraken_fil <- prune_samples(!(sample_names(fresa_kraken) %in% samples_to_remove), fresa_kraken)
percentages_fil <- transform_sample_counts(fresa_kraken_fil, function(x) x*100 / sum(x) )
percentages_df <- psmelt(percentages_fil)

```

De aqui en adelante solo trabajaremos con los datos filtrados por calidad, ya que no vale la pena continuar con los datos que contiene muestras en ceros, ya que esto altera las visualizaciones.

Queremos explorar nuestras muestras a diferentes niveles taxonómicos específicos, empezando por una agrupación de los datos a nivel de reino y filo, Para agrupar todas las OTU que tienen la misma taxonomía en un determinado rango taxonómico, utilizaremos la función *tax\_grom()*.

## A nivel de Kingdom

```

percentages_glom_kingdom <- tax_grom(percentages_fil, taxrank = 'Kingdom')
percentages_glom_kingdom@tax_table@Data

```

|         | Kingdom     | Phylum | Class | Order | Family | Genus | Species |
|---------|-------------|--------|-------|-------|--------|-------|---------|
| ## 1883 | "Bacteria"  | NA     | NA    | NA    | NA     | NA    | NA      |
| ## 5515 | "Eukaryota" | NA     | NA    | NA    | NA     | NA    | NA      |

En reino podemos ver que tenemos solamente dos grupos que son: Bacteria y Eucariota.

## A nivel de Phylum

```
percentages_glom_phylum <- tax_glom(percentages_fil, taxrank = 'Phylum')
percentages_glom_phylum@tax_table@.Data
```

|            | Kingdom     | Phylum                          | Class | Order | Family | Genus |
|------------|-------------|---------------------------------|-------|-------|--------|-------|
| ## 1883    | "Bacteria"  | "Actinobacteria"                | NA    | NA    | NA     | NA    |
| ## 1406    | "Bacteria"  | "Firmicutes"                    | NA    | NA    | NA     | NA    |
| ## 1298    | "Bacteria"  | "Deinococcus-Thermus"           | NA    | NA    | NA     | NA    |
| ## 1173025 | "Bacteria"  | "Cyanobacteria"                 | NA    | NA    | NA     | NA    |
| ## 2057    | "Bacteria"  | "Chloroflexi"                   | NA    | NA    | NA     | NA    |
| ## 1005039 | "Bacteria"  | "Armatimonadetes"               | NA    | NA    | NA     | NA    |
| ## 2132    | "Bacteria"  | "Tenericutes"                   | NA    | NA    | NA     | NA    |
| ## 374     | "Bacteria"  | "Proteobacteria"                | NA    | NA    | NA     | NA    |
| ## 2528023 | "Bacteria"  | "Planctomyces"                  | NA    | NA    | NA     | NA    |
| ## 1882749 | "Bacteria"  | "Verrucomicrobia"               | NA    | NA    | NA     | NA    |
| ## 1307763 | "Bacteria"  | "Kiritimatiellaeota"            | NA    | NA    | NA     | NA    |
| ## 389348  | "Bacteria"  | "Chlamydiae"                    | NA    | NA    | NA     | NA    |
| ## 1930593 | "Bacteria"  | "Candidatus Omnitrophica"       | NA    | NA    | NA     | NA    |
| ## 354356  | "Bacteria"  | "Bacteroidetes"                 | NA    | NA    | NA     | NA    |
| ## 274537  | "Bacteria"  | "Chlorobi"                      | NA    | NA    | NA     | NA    |
| ## 1457365 | "Bacteria"  | "Balneolaeota"                  | NA    | NA    | NA     | NA    |
| ## 1134405 | "Bacteria"  | "Ignavibacteriae"               | NA    | NA    | NA     | NA    |
| ## 861299  | "Bacteria"  | "Gemmatimonadetes"              | NA    | NA    | NA     | NA    |
| ## 833     | "Bacteria"  | "Fibrobacteres"                 | NA    | NA    | NA     | NA    |
| ## 456827  | "Bacteria"  | "Candidatus Cloacimonetes"      | NA    | NA    | NA     | NA    |
| ## 2802971 | "Bacteria"  | "Acidobacteria"                 | NA    | NA    | NA     | NA    |
| ## 42253   | "Bacteria"  | "Nitrospirae"                   | NA    | NA    | NA     | NA    |
| ## 154     | "Bacteria"  | "Spirochaetes"                  | NA    | NA    | NA     | NA    |
| ## 638849  | "Bacteria"  | "Synergistetes"                 | NA    | NA    | NA     | NA    |
| ## 2026885 | "Bacteria"  | "Candidatus Bipolaricaulota"    | NA    | NA    | NA     | NA    |
| ## 1332188 | "Bacteria"  | "Candidatus Saccharibacteria"   | NA    | NA    | NA     | NA    |
| ## 2735562 | "Bacteria"  | "Candidatus Absconditabacteria" | NA    | NA    | NA     | NA    |
| ## 2335    | "Bacteria"  | "Thermotogae"                   | NA    | NA    | NA     | NA    |
| ## 412593  | "Bacteria"  | "Aquificae"                     | NA    | NA    | NA     | NA    |
| ## 2047767 | "Bacteria"  | "Thermodesulfobacteria"         | NA    | NA    | NA     | NA    |
| ## 139438  | "Bacteria"  | "Deferribacteres"               | NA    | NA    | NA     | NA    |
| ## 848     | "Bacteria"  | "Fusobacteria"                  | NA    | NA    | NA     | NA    |
| ## 936456  | "Bacteria"  | "Chrysioigenetes"               | NA    | NA    | NA     | NA    |
| ## 187145  | "Bacteria"  | "Calditrichaeota"               | NA    | NA    | NA     | NA    |
| ## 423605  | "Bacteria"  | "Elusimicrobia"                 | NA    | NA    | NA     | NA    |
| ## 693075  | "Bacteria"  | "Caldiserica"                   | NA    | NA    | NA     | NA    |
| ## 35786   | "Bacteria"  | "Coprothermobacterota"          | NA    | NA    | NA     | NA    |
| ## 2847778 | "Bacteria"  | "Atribacterota"                 | NA    | NA    | NA     | NA    |
| ## 13      | "Bacteria"  | "Dictyoglomi"                   | NA    | NA    | NA     | NA    |
| ## 5515    | "Eukaryota" | "Ascomycota"                    | NA    | NA    | NA     | NA    |
| ## 84751   | "Eukaryota" | "Basidiomycota"                 | NA    | NA    | NA     | NA    |
| ## 6035    | "Eukaryota" | "Microsporidia"                 | NA    | NA    | NA     | NA    |
| ## 4785    | "Eukaryota" | "Oomycota"                      | NA    | NA    | NA     | NA    |
| ##         | Species     |                                 |       |       |        |       |
| ## 1883    | NA          |                                 |       |       |        |       |
| ## 1406    | NA          |                                 |       |       |        |       |
| ## 1298    | NA          |                                 |       |       |        |       |

```

## 1173025 NA
## 2057 NA
## 1005039 NA
## 2132 NA
## 374 NA
## 2528023 NA
## 1882749 NA
## 1307763 NA
## 389348 NA
## 1930593 NA
## 354356 NA
## 274537 NA
## 1457365 NA
## 1134405 NA
## 861299 NA
## 833 NA
## 456827 NA
## 2802971 NA
## 42253 NA
## 154 NA
## 638849 NA
## 2026885 NA
## 1332188 NA
## 2735562 NA
## 2335 NA
## 412593 NA
## 2047767 NA
## 139438 NA
## 848 NA
## 936456 NA
## 187145 NA
## 423605 NA
## 693075 NA
## 35786 NA
## 2847778 NA
## 13 NA
## 5515 NA
## 84751 NA
## 6035 NA
## 4785 NA

```

A nivel de filo, ya podemos ver mas variedad en nuestras muestras, teniendo como mayoria los filo pertenecientes a bacteria.

Usaremos la función de phyloseq ***psmelt()***, que fusiona objetos de phyloseq en un data.frame para manipularlos con paquetes como ggplot2 y vegan. Con esta creamos un dataframe con los porcentajes de “phylum”, este nos dará los porcentajes de cada OTU, perteneciente a cada muestra, a nivel taxonomico de filo.

```

percentages_df_phylum <- psmelt(percentages_glom_phylum)
str(percentages_df_phylum)

```

```

## 'data.frame': 2279 obs. of 7 variables:
## $ OTU      : chr "1883" "1883" "1883" "1883" ...
## $ Sample   : chr "MP2060" "MP2058" "MP2059" "MP2049" ...

```

```

## $ Abundance: num 56.3 52.7 52.7 52.7 52.4 ...
## $ Treatment: chr "healthy" "healthy" "healthy" "healthy" ...
## $ Samples : chr "MP2060" "MP2058" "MP2059" "MP2049" ...
## $ Kingdom : chr "Bacteria" "Bacteria" "Bacteria" "Bacteria" ...
## $ Phylum : chr "Actinobacteria" "Actinobacteria" "Actinobacteria" "Actinobacteria" ...

head(percentages_df_phylum)

##      OTU Sample Abundance Treatment Samples Kingdom          Phylum
## 738 1883 56.28736   healthy MP2060 Bacteria Actinobacteria
## 709 1883 52.71057   healthy MP2058 Bacteria Actinobacteria
## 726 1883 52.70278   healthy MP2059 Bacteria Actinobacteria
## 701 1883 52.70204   healthy MP2049 Bacteria Actinobacteria
## 690 1883 52.35707    wilted MD2055 Bacteria Actinobacteria
## 725 1883 51.64839   healthy MP2120 Bacteria Actinobacteria

```

Ahora, creamos un data frame con los datos originales. Esta estructura nos ayudará a comparar la abundancia absoluta con la relativa

```

absolute_glm_phylum <- tax_glm(physeq = fresa_kraken_fil, taxrank = "Phylum")
absolute_df_phylum <- psmelt(absolute_glm_phylum)
str(absolute_df_phylum)

```

```

## 'data.frame': 2279 obs. of 7 variables:
## $ OTU      : chr "1883" "1883" "1883" "374" ...
## $ Sample    : chr "MP2058" "MP2087" "MP2068" "MP2068" ...
## $ Abundance: num 11028897 10342054 9967507 9190530 8805431 ...
## $ Treatment: chr "healthy" "healthy" "healthy" "healthy" ...
## $ Samples   : chr "MP2058" "MP2087" "MP2068" "MP2068" ...
## $ Kingdom   : chr "Bacteria" "Bacteria" "Bacteria" "Bacteria" ...
## $ Phylum    : chr "Actinobacteria" "Actinobacteria" "Actinobacteria" "Proteobacteria" ...

```

```

absolute_df_phylum$Phylum <- as.factor(absolute_df_phylum$Phylum)
phylum_colors_abs <- colorRampPalette(brewer.pal(8,"Dark2")) (length(levels(absolute_df_phylum$Phylum)))

```

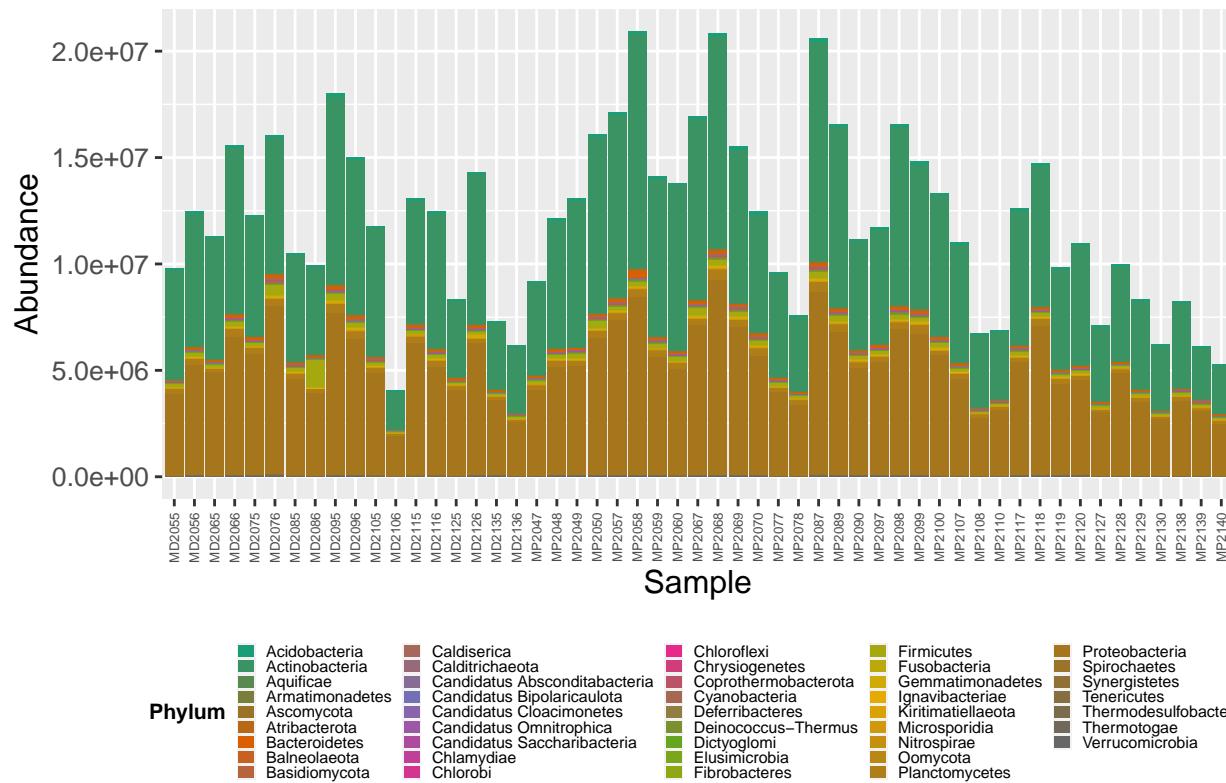
```

absolute_plot <- ggplot(data= absolute_df_phylum, aes(x=Sample, y=Abundance, fill=Phylum))+
  geom_bar(aes(), stat="identity", position="stack")+
  scale_fill_manual(values = phylum_colors_abs) +
  labs(title = "Abundance", x='Sample', y='Abundance') +
  theme(legend.key.size = unit(0.2, "cm"),
        legend.key.width = unit(0.25,"cm"),
        legend.position = "bottom",
        legend.direction = "horizontal",
        legend.title=element_text(size=8, face = "bold"),
        legend.text=element_text(size=6),
        text = element_text(size=12),
        axis.text.x = element_text(angle=90, size=5, hjust=1, vjust=0.5))

```

```
absolute_plot
```

## Abundance



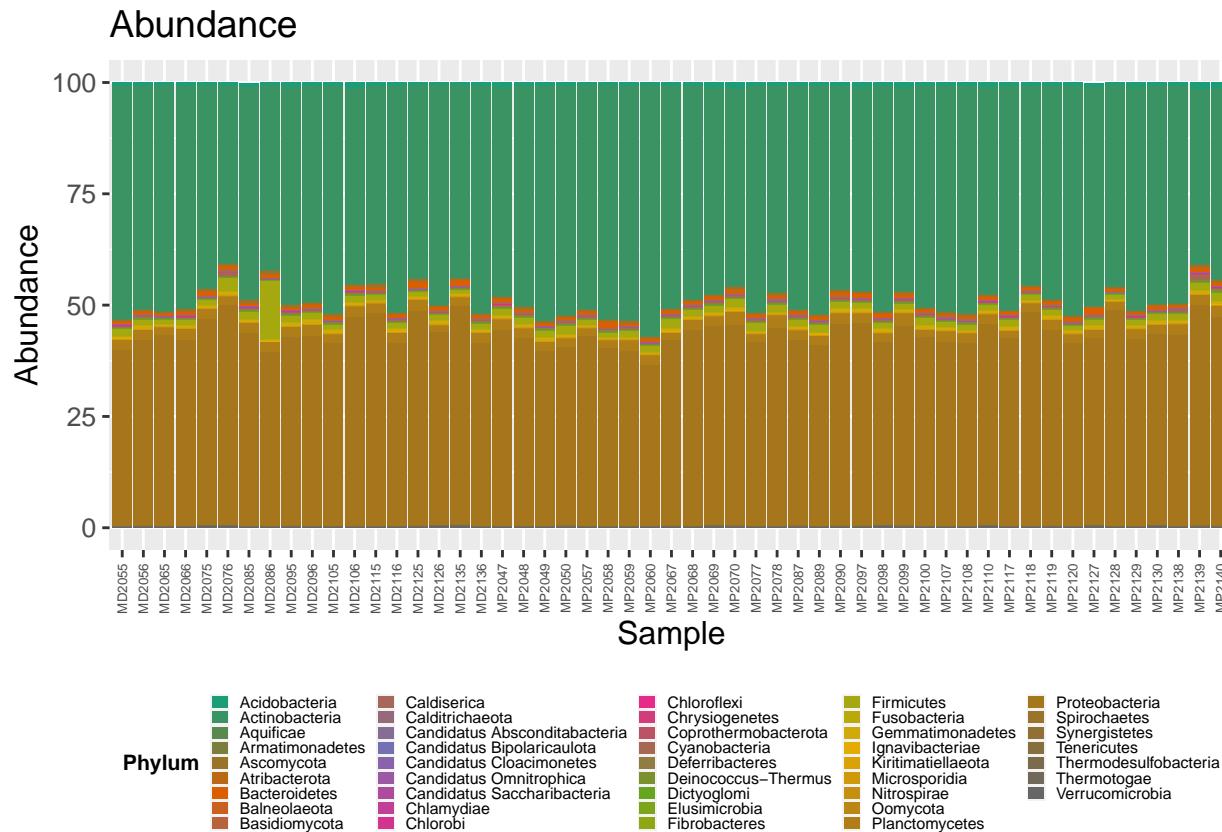
```

percentages_df_phylum$Phylum <- as.factor(percentages_df_phylum$Phylum)
phylum_colors_rel<- colorRampPalette(brewer.pal(8,"Dark2")) (length(levels(percentages_df_phylum$Phylum))
relative_plot <- ggplot(data=percentages_df_phylum, aes(x=Sample, y=Abundance, fill=Phylum))+  

  geom_bar(aes(), stat="identity", position="stack")+
  scale_fill_manual(values = phylum_colors_rel) +
  labs(title = "Abundance", x='Sample', y='Abundance') +
  theme(legend.key.size = unit(0.2, "cm"),
        legend.key.width = unit(0.25,"cm"),
        legend.position = "bottom",
        legend.direction = "horizontal",
        legend.title=element_text(size=8, face = "bold"),
        legend.text=element_text(size=6),
        text = element_text(size=12),
        axis.text.x = element_text(angle=90, size=5, hjust=1, vjust=0.5))

relative_plot

```

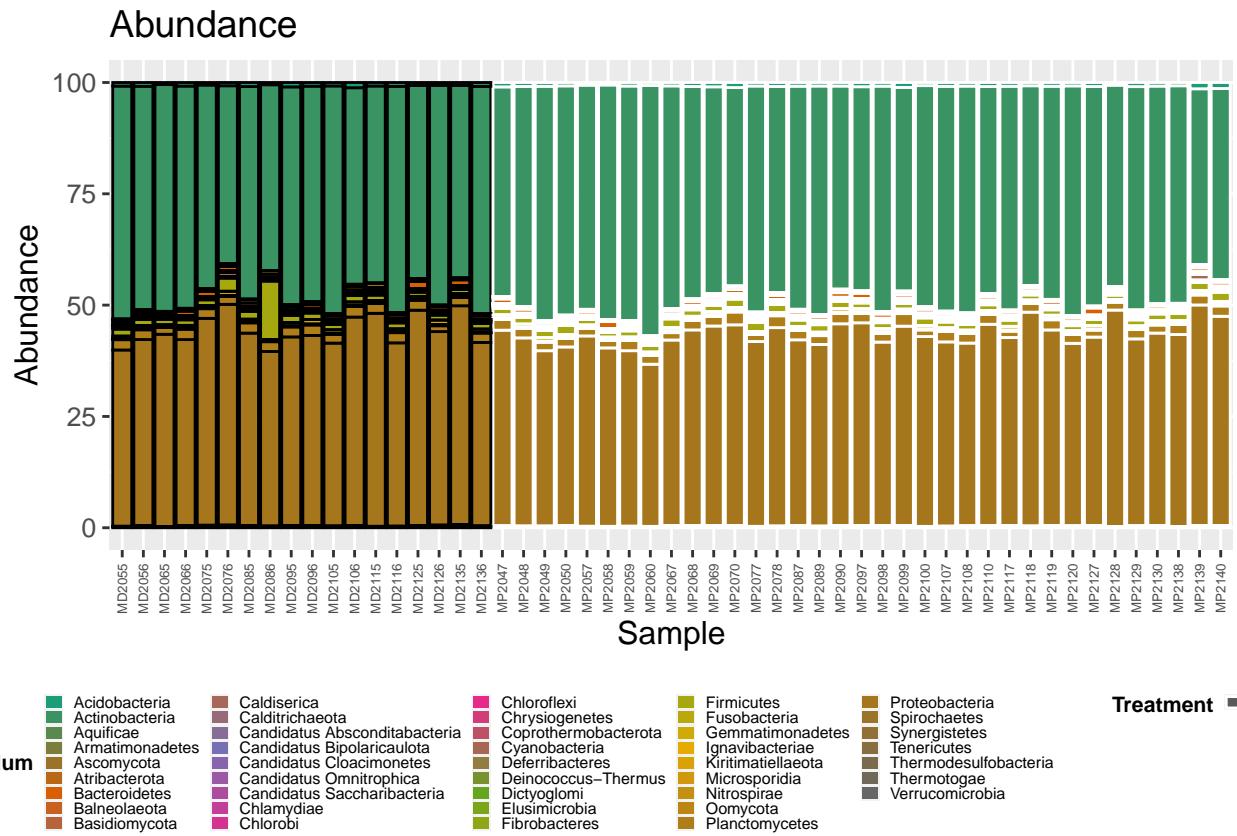


Podemos ver algunas diferencias entre los diferentes filo, sin embargo, se nota que tenemos demasiados taxones para distinguir adecuadamente el color de cada uno, a menos que se tenga una abundancia muy grande.

Como nuestro objetivo es diferenciar entre muestras sanas y enfermas, le agregaremos una diferenciación entre la variable ‘Treatment’, con color blanco las muestras sanas y negras las muestras enfermas.

```
percentages_df_phylum$Phylum <- as.factor(percentages_df_phylum$Phylum)
phylum_colors_rel <- colorRampPalette(brewer.pal(8, "Dark2")) (length(levels(percentages_df_phylum$Phylum))
relative_plot <- ggplot(data=percentages_df_phylum, aes(x=Sample, y=Abundance, fill=Phylum, color=Treatment))
  scale_colour_manual(values=c('white','black')) +
  geom_bar(aes(), stat="identity", position="stack")+
  scale_fill_manual(values = phylum_colors_rel) +
  labs(title = "Abundance", x='Sample', y='Abundance') +
  theme(legend.key.size = unit(0.2, "cm"),
        legend.key.width = unit(0.25,"cm"),
        legend.position = "bottom",
        legend.direction = "horizontal",
        legend.title=element_text(size=8, face = "bold"),
        legend.text=element_text(size=6),
        text = element_text(size=12),
        axis.text.x = element_text(angle=90, size=5, hjust=1, vjust=0.5))

relative_plot
```



Igualmente tenemos demasiados taxones para ver una distinción entre las muestras, más adelante crearemos subconjuntos mas pequeños para tener una mejor observación de nuestros datos.

Por otro lado, podemos usar un comando llamado “unique()” para explorar cuántos filos y reinos tenemos.

```
unique(fresa_kraken_fil@tax_table@Data[, "Kingdom"])
```

```
## [1] "Bacteria" "Eukaryota"
```

```
unique(fresa_kraken_fil@tax_table@Data[, "Phylum"])
```

```
## [1] "Actinobacteria" "Firmicutes"
## [3] "Deinococcus-Thermus" "Cyanobacteria"
## [5] "Chloroflexi" "Armatimonadetes"
## [7] "Tenericutes" "Proteobacteria"
## [9] "Planctomycetes" "Verrucomicrobia"
## [11] "Kiritimatiellaeota" "Chlamydiae"
## [13] "Candidatus Omnitrophica" "Bacteroidetes"
## [15] "Chlorobi" "Balneolaeota"
## [17] "Ignavibacteriae" "Gemmatimonadetes"
## [19] "Fibrobacteres" "Candidatus Cloacimonetes"
## [21] "Acidobacteria" "Nitrospirae"
## [23] "Spirochaetes" "Synergistetes"
## [25] "Candidatus Bipolaricaulota" "Candidatus Saccharibacteria"
## [27] "" "Candidatus Absconditabacteria"
```

```

## [29] "Thermotogae"           "Aequificae"
## [31] "Thermodesulfobacteria"  "Deferrribacteres"
## [33] "Fusobacteria"          "Chrysogenetes"
## [35] "Calditrichaeota"        "Elusimicrobia"
## [37] "Caldisericia"          "Coprothermobacterota"
## [39] "Atribacterota"          "Dictyoglomi"
## [41] "Ascomycota"             "Basidiomycota"
## [43] "Microsporidia"          "Oomycota"

```

Ahora, podemos ver cuantos “Eukaryota” tenemos en “Kingdom”.

```
sum(fresa_kraken_fil@tax_table@Data[, "Kingdom"] == "Eukaryota")
```

```
## [1] 181
```

y cuantos “Bacteria”

```
sum(fresa_kraken_fil@tax_table@Data[, "Kingdom"] == "Bacteria")
```

```
## [1] 8822
```

Verificando así lo que notamos anteriormente con la observación de los subconjuntos aglomerados, tenemos mayor cantidad de muestras clasificadas como bacterias, que las clasificadas como eucariota.

## Diversidad Beta

Como se menciono anteriormente el objetivo de este documento es explorar las visualizaciones de: barras, alfa diversidad y beta diversidad, para subconjuntos seleccionados de nuestros datos.

Veremos aquí solo el reino Eucariota:

```
merge_Eukaryota <- subset_taxa(fresa_kraken_fil, Kingdom == "Eukaryota")
```

sacamos las abundancias relativas,

```
percentages_Eukaryota <- transform_sample_counts(merge_Eukaryota, function(x) x*100 / sum(x) )
percentages_Eukaryota_df <- psmelt(percentages_Eukaryota)
```

### Beta diversidad de Eukaryota

```

meta_ord_Eukaryota <- ordinate(physeq = percentages_Eukaryota, method = "NMDS", distance = "bray")

## Wisconsin double standardization
## Run 0 stress 0.1341389
## Run 1 stress 0.1397091
## Run 2 stress 0.1344767
## ... Procrustes: rmse 0.009222729 max resid 0.06214174
## Run 3 stress 0.150346

```

```

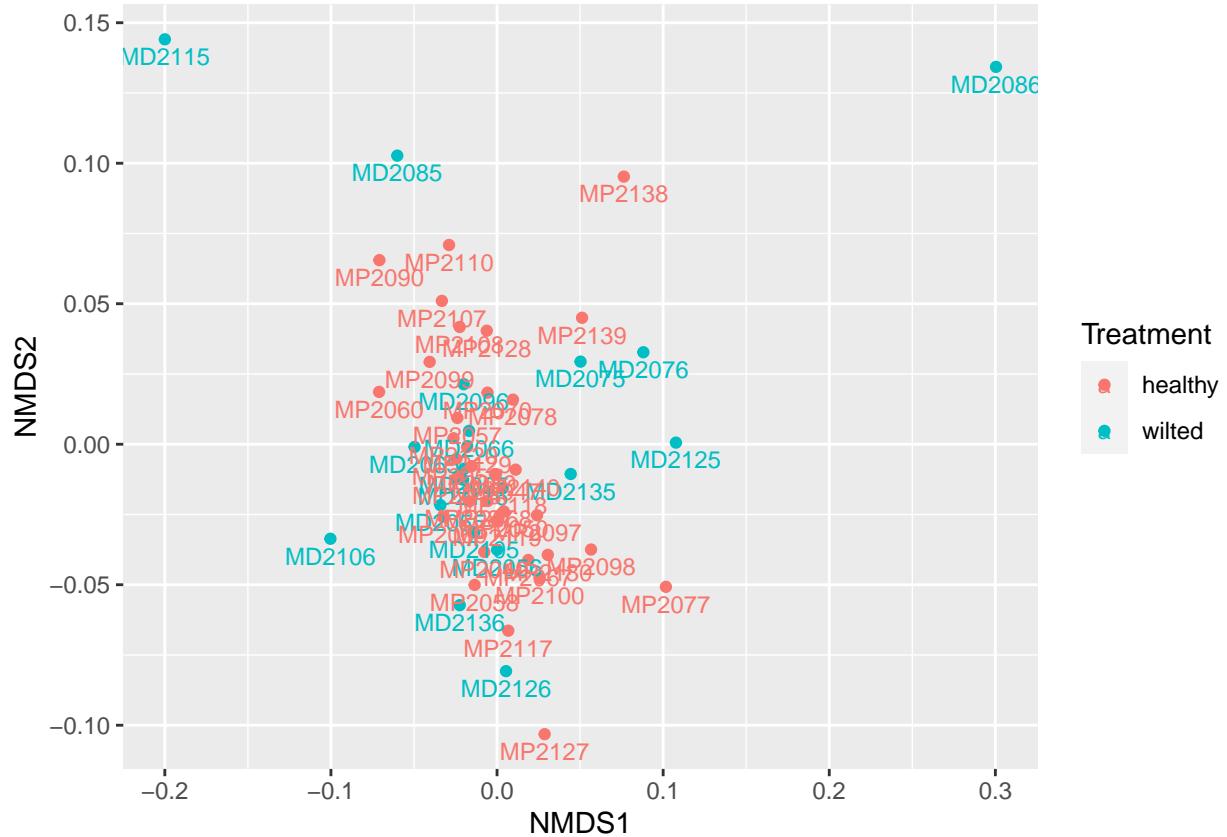
## Run 4 stress 0.1341401
## ... Procrustes: rmse 0.0004496422 max resid 0.002603379
## ... Similar to previous best
## Run 5 stress 0.1349242
## Run 6 stress 0.1492025
## Run 7 stress 0.1543662
## Run 8 stress 0.402454
## Run 9 stress 0.1345727
## ... Procrustes: rmse 0.01473282 max resid 0.08004303
## Run 10 stress 0.1341979
## ... Procrustes: rmse 0.01194142 max resid 0.08064589
## Run 11 stress 0.1368375
## Run 12 stress 0.1652394
## Run 13 stress 0.1397083
## Run 14 stress 0.1696714
## Run 15 stress 0.1341979
## ... Procrustes: rmse 0.01194247 max resid 0.08065211
## Run 16 stress 0.1521023
## Run 17 stress 0.1550757
## Run 18 stress 0.1341966
## ... Procrustes: rmse 0.0118919 max resid 0.08042326
## Run 19 stress 0.1575264
## Run 20 stress 0.1649151
## *** Best solution repeated 1 times

```

```

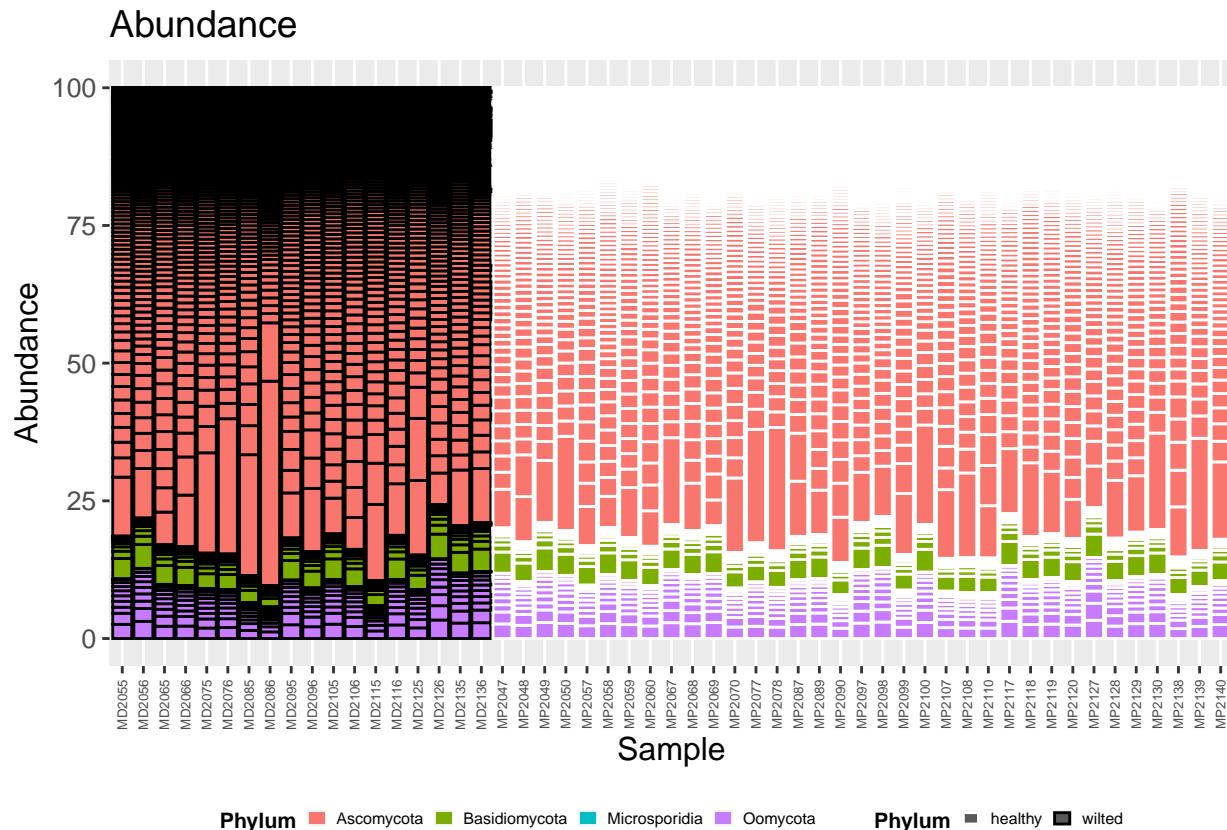
plot_ordination(physeq = percentages_Eukaryota, ordination = meta_ord_Eukaryota, color = "Treatment") +
  geom_text(mapping = aes(label = colnames(fresa_kraken_fil@otu_table@Data)), size = 3, vjust = 1.5)

```



## Eukaryota por Phylum

```
ggplot(data= percentages_Eukaryota_df, aes(x=Sample, y=Abundance, fill=Phylum, color=Treatment)) +
  scale_colour_manual(values=c('white','black')) +
  geom_bar(aes(), stat="identity", position="stack") +
  labs(title = "Abundance", x='Sample', y='Abundance', color = 'Phylum') +
  theme(legend.key.size = unit(0.2, "cm"),
        legend.key.width = unit(0.25,"cm"),
        legend.position = "bottom",
        legend.direction = "horizontal",
        legend.title=element_text(size=8, face = "bold"),
        legend.text=element_text(size=6),
        text = element_text(size=12),
        axis.text.x = element_text(angle=90, size=5, hjust=1, vjust=0.5))
```



```
merge_Eukaryota_Phylum<-tax_glom(merge_Eukaryota,taxrank="Phylum")
```

sacamos las abundancias relativas

```
percentages_Eukaryota_Phylum <- transform_sample_counts(merge_Eukaryota_Phylum, function(x) x*100 / sum(merge_Eukaryota_Phylum))
percentages_Eukaryota_Phylum_df <- psmelt(percentages_Eukaryota_Phylum)
meta_ord_Eukaryota_Phylum <- ordinate(physeq = percentages_Eukaryota_Phylum, method = "NMDS", distance = "euclidean", trace = TRUE)

## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.01148681
## Run 1 stress 0.0001796633
## ... New best solution
## ... Procrustes: rmse 0.03415031 max resid 0.089572
## Run 2 stress 6.362911e-05
## ... New best solution
## ... Procrustes: rmse 0.0004784719 max resid 0.001214989
## ... Similar to previous best
## Run 3 stress 0.0001124546
## ... Procrustes: rmse 0.0002801985 max resid 0.0006937247
## ... Similar to previous best
## Run 4 stress 9.847285e-05
## ... Procrustes: rmse 8.595227e-05 max resid 0.0002665847
## ... Similar to previous best
## Run 5 stress 9.861053e-05
```

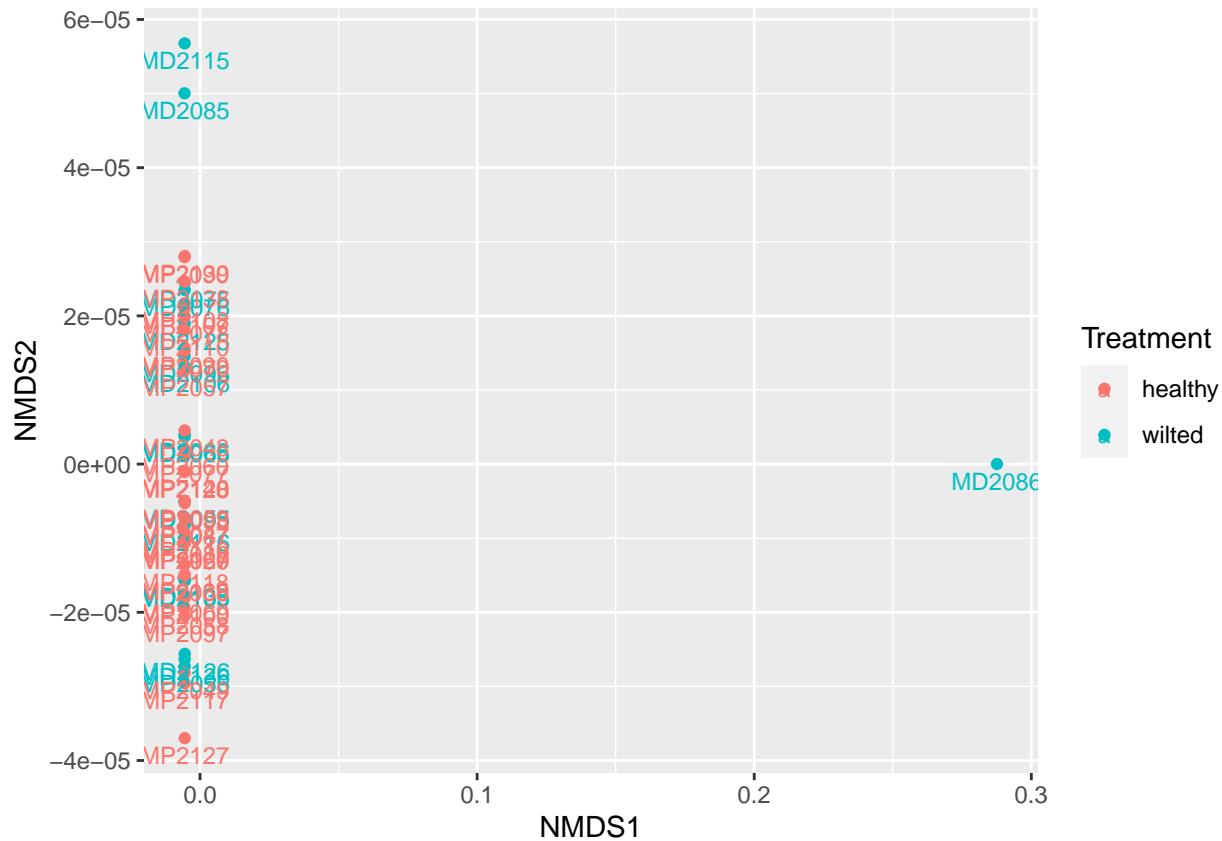
```

## ... Procrustes: rmse 7.10498e-05 max resid 0.0002573164
## ... Similar to previous best
## Run 6 stress 9.890322e-05
## ... Procrustes: rmse 7.53617e-05 max resid 0.0002075598
## ... Similar to previous best
## Run 7 stress 7.922991e-05
## ... Procrustes: rmse 9.421317e-05 max resid 0.0002504548
## ... Similar to previous best
## Run 8 stress 0.0001386841
## ... Procrustes: rmse 0.0003552956 max resid 0.0008951629
## ... Similar to previous best
## Run 9 stress 0.0002508564
## ... Procrustes: rmse 0.0006888503 max resid 0.001765901
## ... Similar to previous best
## Run 10 stress 9.74745e-05
## ... Procrustes: rmse 9.465063e-05 max resid 0.0003113936
## ... Similar to previous best
## Run 11 stress 9.976914e-05
## ... Procrustes: rmse 6.943112e-05 max resid 0.000278123
## ... Similar to previous best
## Run 12 stress 0.0007614602
## Run 13 stress 0.0001109971
## ... Procrustes: rmse 0.0002778267 max resid 0.0006845647
## ... Similar to previous best
## Run 14 stress 7.731697e-05
## ... Procrustes: rmse 7.913996e-05 max resid 0.0002193365
## ... Similar to previous best
## Run 15 stress 9.203413e-05
## ... Procrustes: rmse 6.528499e-05 max resid 0.0001543756
## ... Similar to previous best
## Run 16 stress 9.862459e-05
## ... Procrustes: rmse 0.0001441883 max resid 0.000366917
## ... Similar to previous best
## Run 17 stress 9.576941e-05
## ... Procrustes: rmse 8.512062e-05 max resid 0.0002676717
## ... Similar to previous best
## Run 18 stress 0.0001318963
## ... Procrustes: rmse 0.0003361429 max resid 0.0008443429
## ... Similar to previous best
## Run 19 stress 9.278184e-05
## ... Procrustes: rmse 7.562561e-05 max resid 0.0002002906
## ... Similar to previous best
## Run 20 stress 9.778154e-05
## ... Procrustes: rmse 7.466934e-05 max resid 0.0002727301
## ... Similar to previous best
## *** Best solution repeated 18 times

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

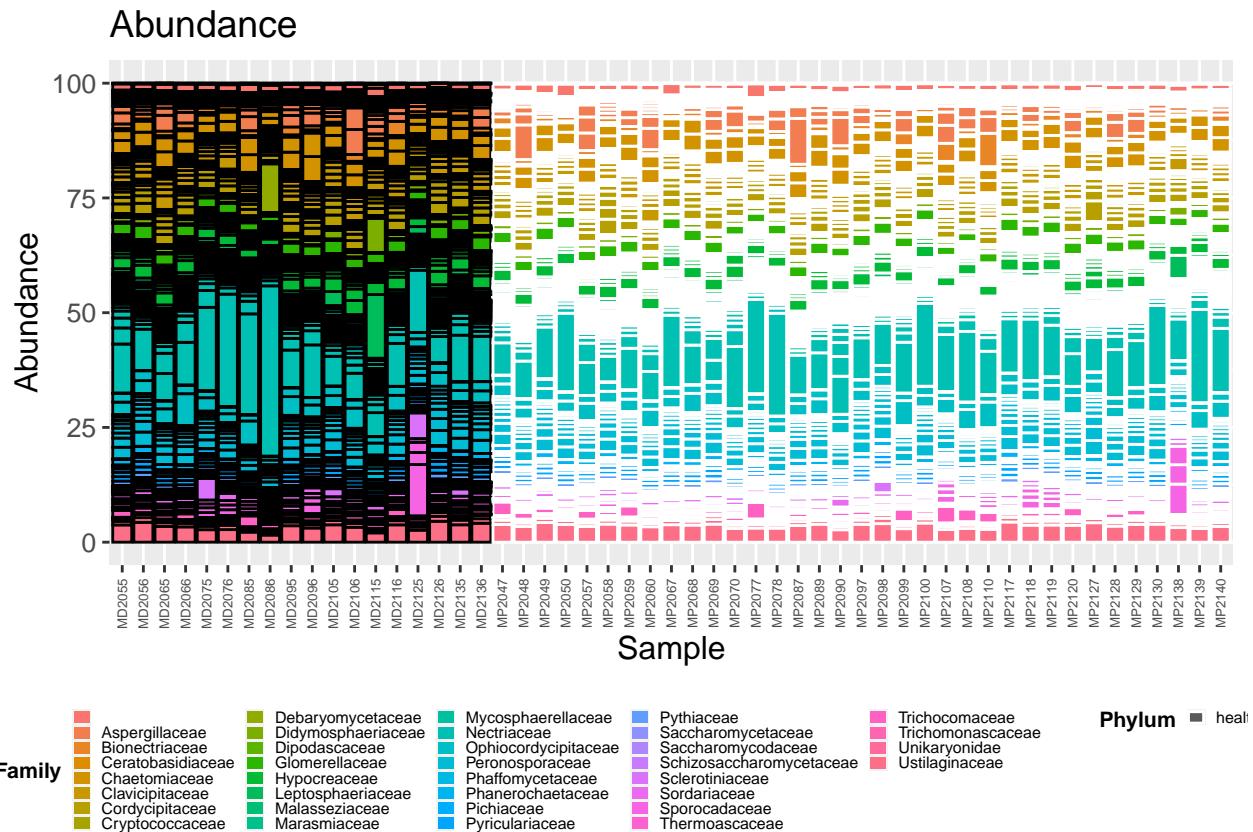
plot_ordination(physeq = percentages_Eukaryota_Phylum, ordination = meta_ord_Eukaryota_Phylum, color =
  geom_text(mapping = aes(label = colnames(fresa_kraken_fil@otu_table@Data)), size = 3, vjust = 1.5)

```



## Eukaryota por Family

```
ggplot(data= percentages_Eukaryota_df, aes(x=Sample, y=Abundance, fill=Family, color=Treatment)) +
  scale_colour_manual(values=c('white','black')) +
  geom_bar(aes(), stat="identity", position="stack") +
  labs(title = "Abundance", x='Sample', y='Abundance', color = 'Phylum') +
  theme(legend.key.size = unit(0.2, "cm"),
        legend.key.width = unit(0.25,"cm"),
        legend.position = "bottom",
        legend.direction = "horizontal",
        legend.title=element_text(size=8, face = "bold"),
        legend.text=element_text(size=6),
        text = element_text(size=12),
        axis.text.x = element_text(angle=90, size=5, hjust=1, vjust=0.5))
```



```
merge_Eukaryota_Family <- tax_glom(merge_Eukaryota, taxrank = "Family")
```

sacamos las abundancias relativas

```
percentages_Eukaryota_Family <- transform_sample_counts(merge_Eukaryota_Family, function(x) x*100 / sum(x))
percentages_Eukaryota_Family_df <- psmelt(percentages_Eukaryota_Family)
meta_ord_Eukaryota_Family <- ordinate(physeq = percentages_Eukaryota_Family, method = "NMDS", distance =
```

```
## Wisconsin double standardization
## Run 0 stress 0.1089081
## Run 1 stress 0.1265906
## Run 2 stress 0.1089074
## ... New best solution
## ... Procrustes: rmse 0.0004720922 max resid 0.002602299
## ... Similar to previous best
## Run 3 stress 0.1088886
## ... New best solution
## ... Procrustes: rmse 0.001906552 max resid 0.0092281
## ... Similar to previous best
## Run 4 stress 0.1263443
## Run 5 stress 0.10889
## ... Procrustes: rmse 0.001752725 max resid 0.01052926
## Run 6 stress 0.1089294
## ... Procrustes: rmse 0.008422603 max resid 0.05463454
## Run 7 stress 0.1088974
```

```

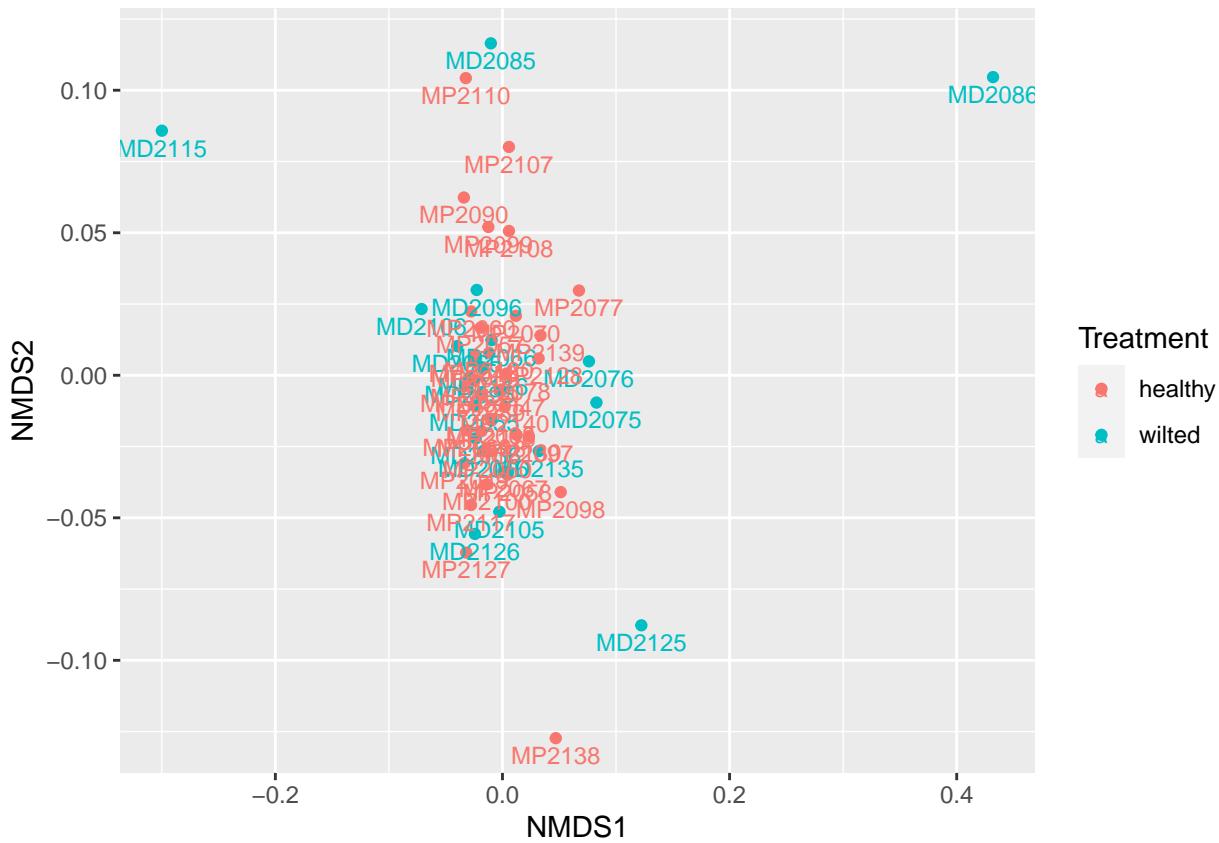
## ... Procrustes: rmse 0.001744503 max resid 0.0090408
## ... Similar to previous best
## Run 8 stress 0.1089487
## ... Procrustes: rmse 0.008056088 max resid 0.05392854
## Run 9 stress 0.1089087
## ... Procrustes: rmse 0.001771024 max resid 0.009318456
## ... Similar to previous best
## Run 10 stress 0.1089294
## ... Procrustes: rmse 0.008422082 max resid 0.05463253
## Run 11 stress 0.1089172
## ... Procrustes: rmse 0.002623485 max resid 0.008865082
## ... Similar to previous best
## Run 12 stress 0.1091256
## ... Procrustes: rmse 0.009991728 max resid 0.05408419
## Run 13 stress 0.1089297
## ... Procrustes: rmse 0.008435116 max resid 0.0546608
## Run 14 stress 0.1089221
## ... Procrustes: rmse 0.004498622 max resid 0.02587995
## Run 15 stress 0.1089291
## ... Procrustes: rmse 0.008372618 max resid 0.05445949
## Run 16 stress 0.108939
## ... Procrustes: rmse 0.003858916 max resid 0.02189402
## Run 17 stress 0.1088964
## ... Procrustes: rmse 0.001799785 max resid 0.009019592
## ... Similar to previous best
## Run 18 stress 0.1265271
## Run 19 stress 0.1089303
## ... Procrustes: rmse 0.008454543 max resid 0.05469472
## Run 20 stress 0.1091256
## ... Procrustes: rmse 0.01000438 max resid 0.05407975
## *** Best solution repeated 5 times

```

```

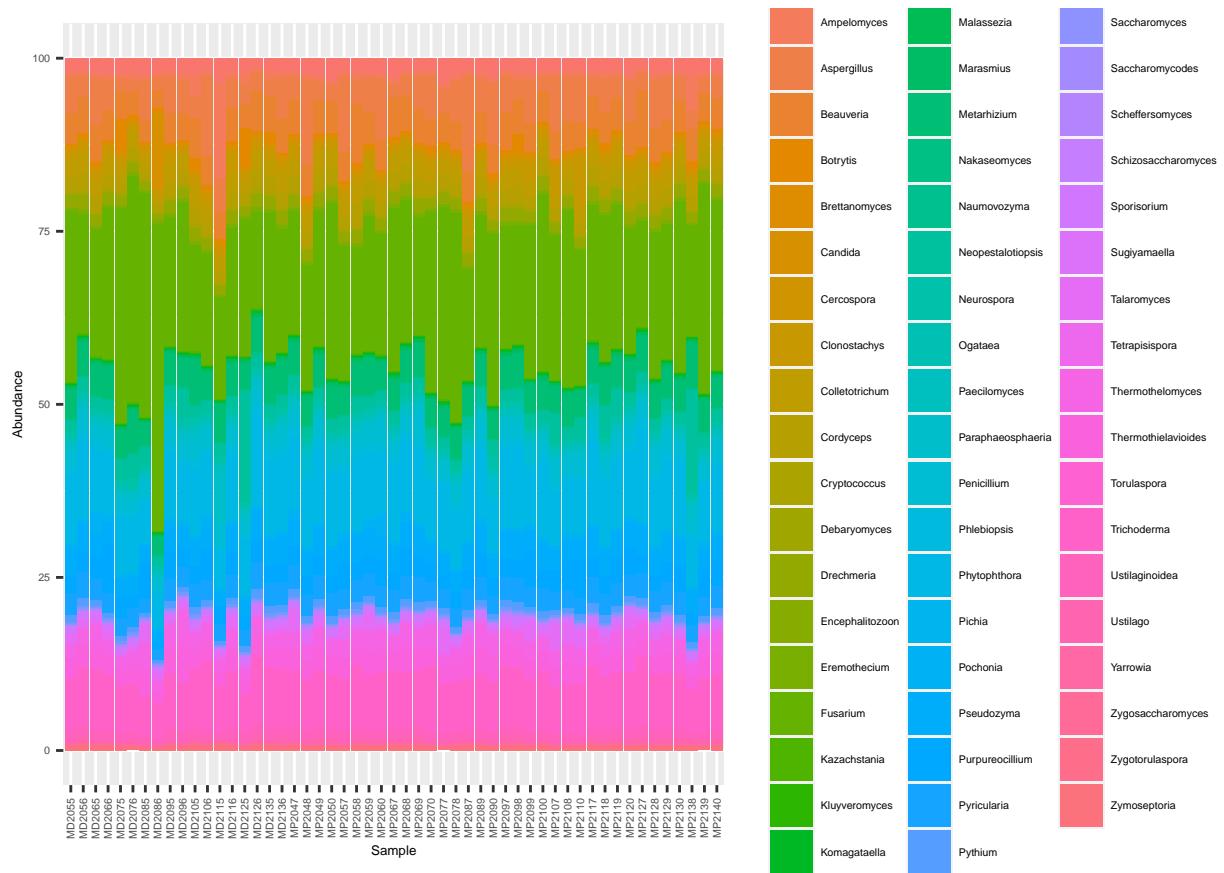
plot_ordination(physeq = percentages_Eukaryota_Family, ordination = meta_ord_Eukaryota_Family, color =
  geom_text(mapping = aes(label = colnames(fresa_kraken_fil@otu_table@Data))), size = 3, vjust = 1.5)

```



## ## Eukaryota por Genus

```
ggplot(data= percentages_Eukaryota_df, aes(x=Sample, y=Abundance, fill=Genus)) +  
  geom_bar(aes(), stat="identity", position="stack") +  
  theme(text = element_text(size = 5),  
        axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```



```
merge_Eukaryota_Genus <- tax_glom(merge_Eukaryota, taxrank = "Genus")
```

sacamos las abundancias relativas

```
percentages_Eukaryota_Genus <- transform_sample_counts(merge_Eukaryota_Genus, function(x) x*100 / sum(x))
percentages_Eukaryota_Genus_df <- psmelt(percentages_Eukaryota_Genus)
meta_ord_Eukaryota_Genus <- ordinate(physeq = percentages_Eukaryota_Genus, method = "NMDS", distance =
```

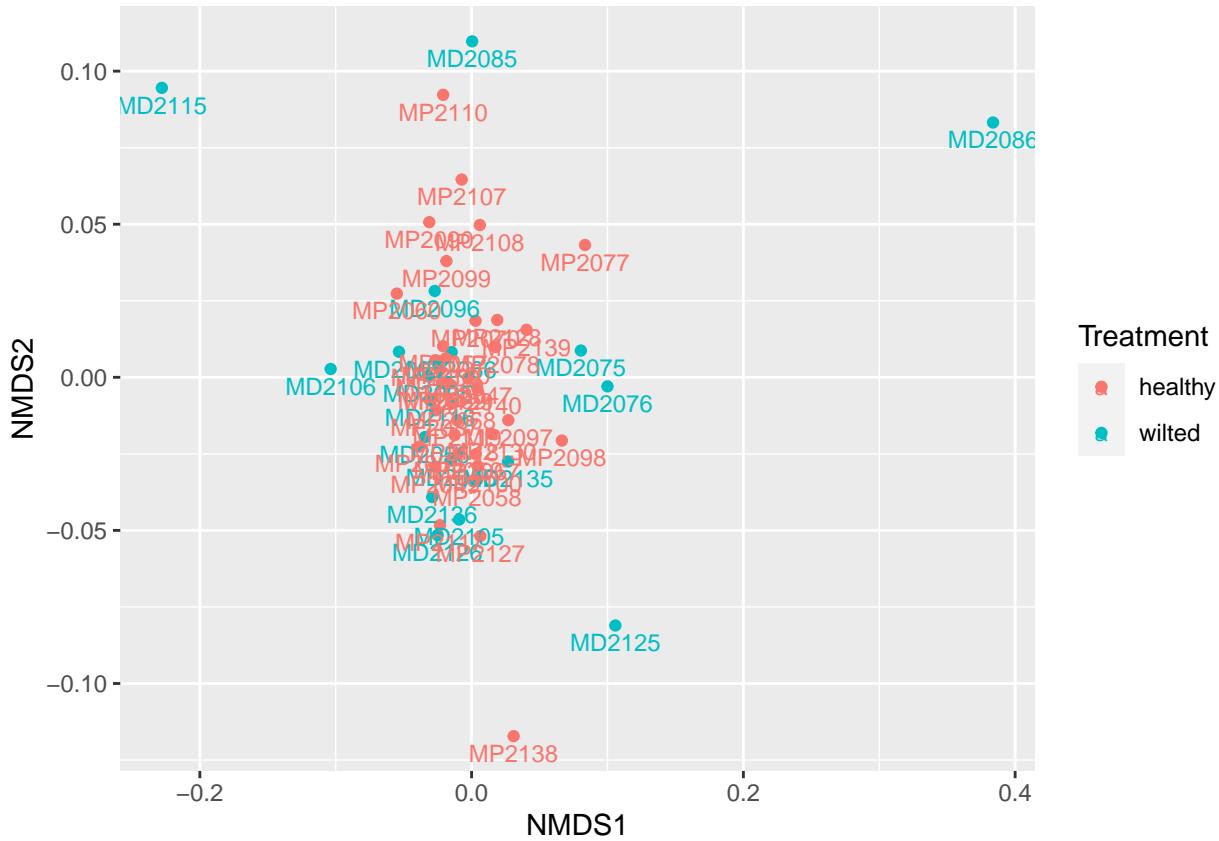
```
## Wisconsin double standardization
## Run 0 stress 0.110709
## Run 1 stress 0.1235029
## Run 2 stress 0.11071
## ... Procrustes: rmse 0.0012056 max resid 0.006894236
## ... Similar to previous best
## Run 3 stress 0.1114076
## Run 4 stress 0.1307224
## Run 5 stress 0.110709
## ... New best solution
## ... Procrustes: rmse 8.588246e-06 max resid 3.642063e-05
## ... Similar to previous best
## Run 6 stress 0.1235031
## Run 7 stress 0.1107093
## ... Procrustes: rmse 8.413263e-05 max resid 0.0004286052
## ... Similar to previous best
## Run 8 stress 0.1222472
```

```

## Run 9 stress 0.1107089
## ... New best solution
## ... Procrustes: rmse 7.955931e-05 max resid 0.0004378354
## ... Similar to previous best
## Run 10 stress 0.1307229
## Run 11 stress 0.1107088
## ... New best solution
## ... Procrustes: rmse 0.0008153885 max resid 0.004664547
## ... Similar to previous best
## Run 12 stress 0.1107159
## ... Procrustes: rmse 0.001157605 max resid 0.006561933
## ... Similar to previous best
## Run 13 stress 0.1119824
## Run 14 stress 0.1235024
## Run 15 stress 0.1107165
## ... Procrustes: rmse 0.001392917 max resid 0.008064705
## ... Similar to previous best
## Run 16 stress 0.1107098
## ... Procrustes: rmse 0.0002969611 max resid 0.001759094
## ... Similar to previous best
## Run 17 stress 0.1107088
## ... New best solution
## ... Procrustes: rmse 0.0008062745 max resid 0.004601496
## ... Similar to previous best
## Run 18 stress 0.1107095
## ... Procrustes: rmse 0.000257015 max resid 0.001650473
## ... Similar to previous best
## Run 19 stress 0.11071
## ... Procrustes: rmse 0.001150729 max resid 0.006594464
## ... Similar to previous best
## Run 20 stress 0.1107091
## ... Procrustes: rmse 0.0008939997 max resid 0.005090997
## ... Similar to previous best
## *** Best solution repeated 4 times

plot_ordination(physeq = percentages_Eukaryota_Genus, ordination = meta_ord_Eukaryota_Genus, color = "T"
  geom_text(mapping = aes(label = colnames(fresa_kraken_fil@otu_table@Data)), size = 3, vjust = 1.5)

```



tomando diferentes porcentajes de abundancia

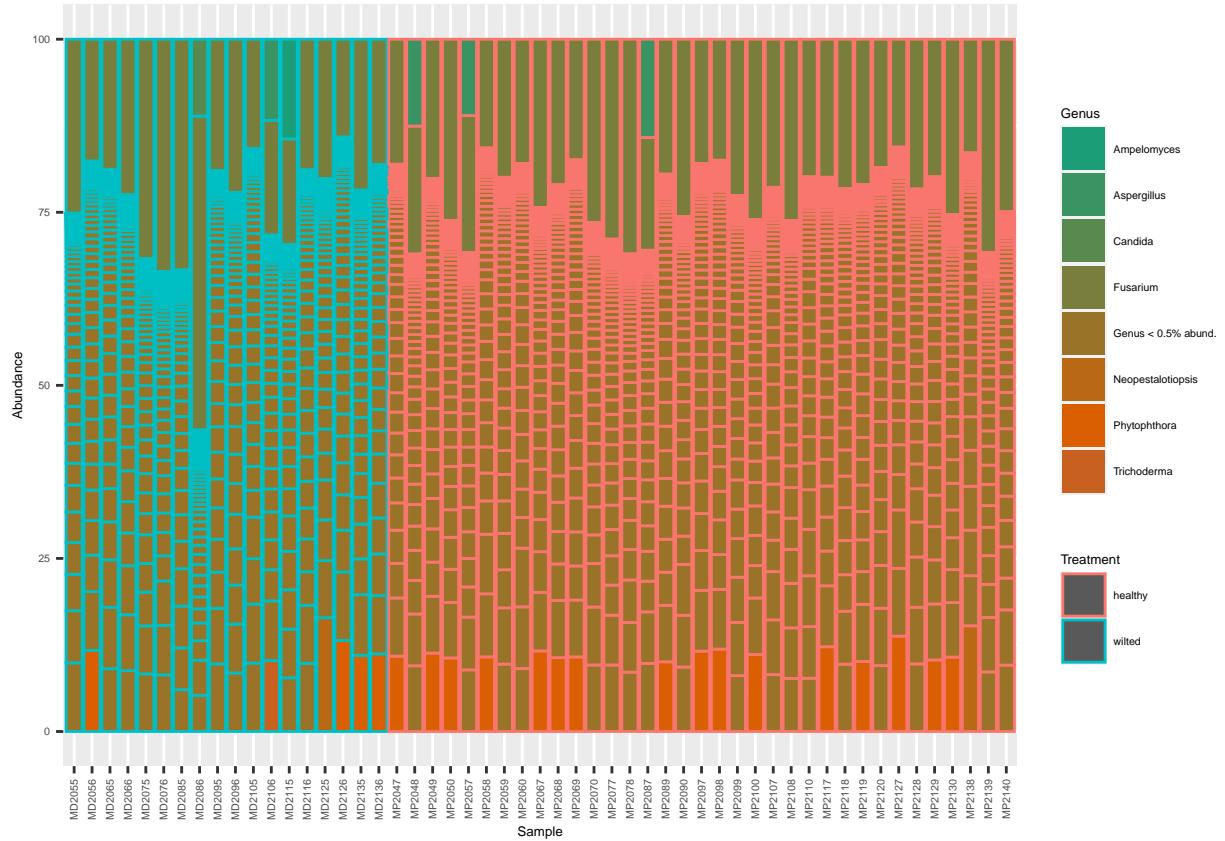
```

percentages_Eukaryota_Genus_df$Genus[percentages_Eukaryota_Genus_df$Abundance < 10.0] <- "Genus < 0.5%"
percentages_Eukaryota_Genus_df$Genus <- as.factor(percentages_Eukaryota_Genus_df$Genus)

genus_colors_rel <- colorRampPalette(brewer.pal(8, "Dark2")) (length(levels(percentages_Eukaryota_Genus_))
relative_plot <- ggplot(data=percentages_Eukaryota_Genus_df, aes(x=Sample, y=Abundance, fill=Genus ,col
  geom_bar(aes(), stat="identity", position="stack") +
  scale_fill_manual(values = phylum_colors_rel) +
  theme(text = element_text(size = 5),
        axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))

relative_plot

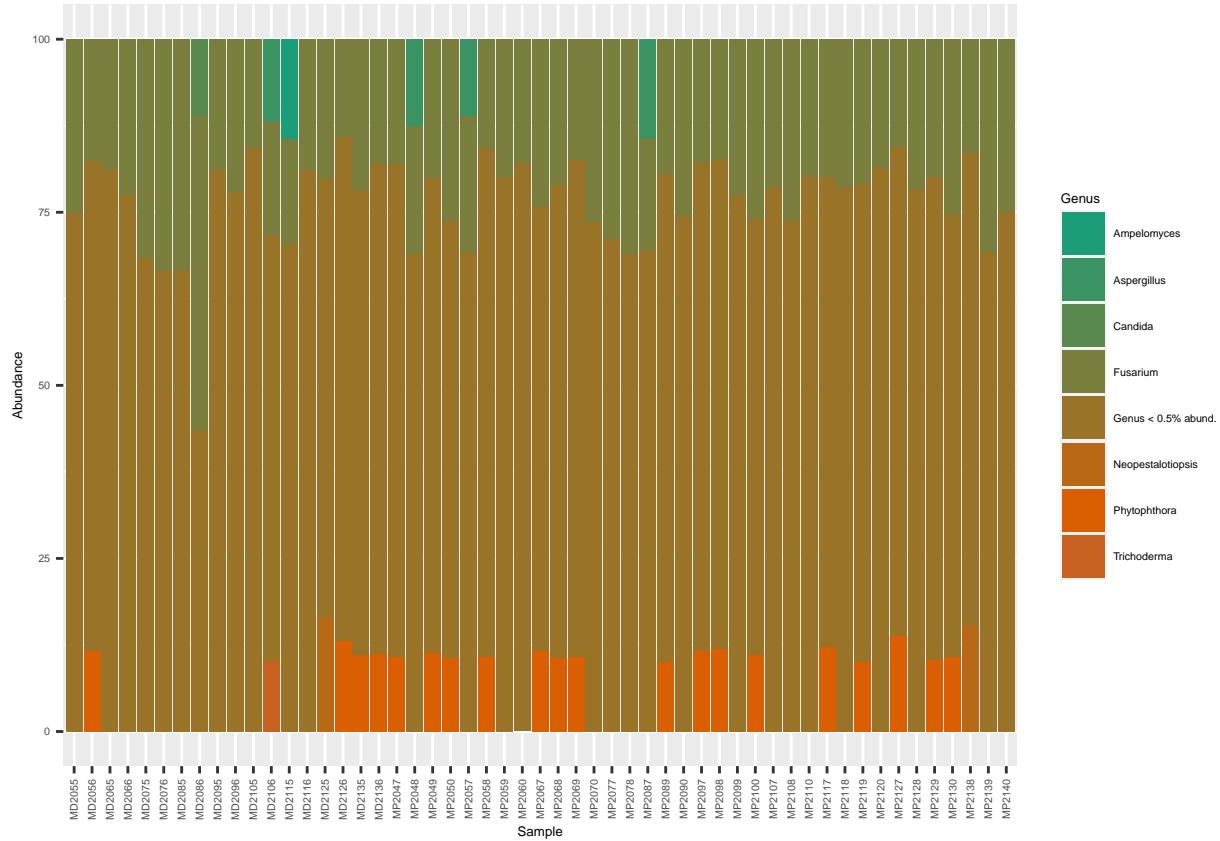
```



```

relative_plot <- ggplot(data=percentages_Eukaryota_Genus_df, aes(x=Sample, y=Abundance, fill=Genus))+
  geom_bar(aes(), stat="identity", position="stack") +
  scale_fill_manual(values = phylum_colors_rel) +
  theme(text = element_text(size = 5),
        axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
relative_plot

```



sacando la beta diversidad con aglomerado de 10%

```
meta_ord_Eukaryota_Genus <- ordinate(physeq = percentages_Eukaryota_Genus, method = "NMDS", distance = 

## Wisconsin double standardization
## Run 0 stress 0.110709
## Run 1 stress 0.1107903
## ... Procrustes: rmse 0.006219316 max resid 0.03417994
## Run 2 stress 0.1119812
## Run 3 stress 0.1119821
## Run 4 stress 0.1107096
## ... Procrustes: rmse 0.001159384 max resid 0.006627418
## ... Similar to previous best
## Run 5 stress 0.1107093
## ... Procrustes: rmse 8.329722e-05 max resid 0.0004493679
## ... Similar to previous best
## Run 6 stress 0.1107091
## ... Procrustes: rmse 0.0009982881 max resid 0.005716501
## ... Similar to previous best
## Run 7 stress 0.1107088
## ... New best solution
## ... Procrustes: rmse 0.0001213406 max resid 0.0007025417
## ... Similar to previous best
## Run 8 stress 0.1222456
## Run 9 stress 0.1119815
## Run 10 stress 0.1107085
```

```

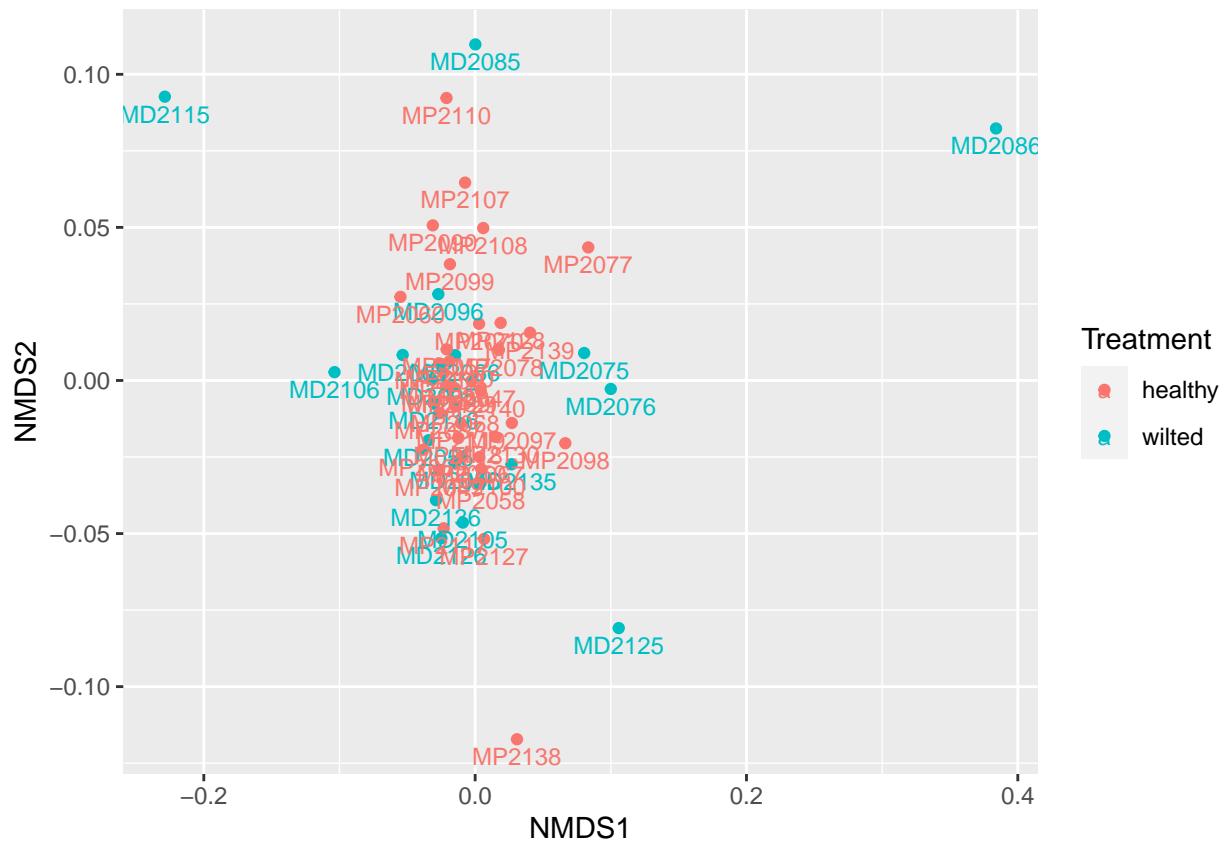
## ... New best solution
## ... Procrustes: rmse 0.0004837029 max resid 0.002759627
## ... Similar to previous best
## Run 11 stress 0.1236303
## Run 12 stress 0.11071
## ... Procrustes: rmse 0.0006731382 max resid 0.003900473
## ... Similar to previous best
## Run 13 stress 0.1119814
## Run 14 stress 0.1107094
## ... Procrustes: rmse 0.0005043254 max resid 0.002888002
## ... Similar to previous best
## Run 15 stress 0.1119795
## Run 16 stress 0.1223836
## Run 17 stress 0.1235028
## Run 18 stress 0.1119822
## Run 19 stress 0.1119815
## Run 20 stress 0.1119813
## *** Best solution repeated 3 times

```

```

plot_ordination(physeq = percentages_Eukaryota_Genus, ordination = meta_ord_Eukaryota_Genus, color = "Treatment"
geom_text(mapping = aes(label = colnames(fresa_kraken_fil@otu_table@Data)), size = 3, vjust = 1.5)

```



Veremos aqui solo el reino Bacteriano

```
merge_Bacteria<-subset_taxa(fresa_kraken_fil,Kingdom=="Bacteria")
```

sacamos las abundancias relativas

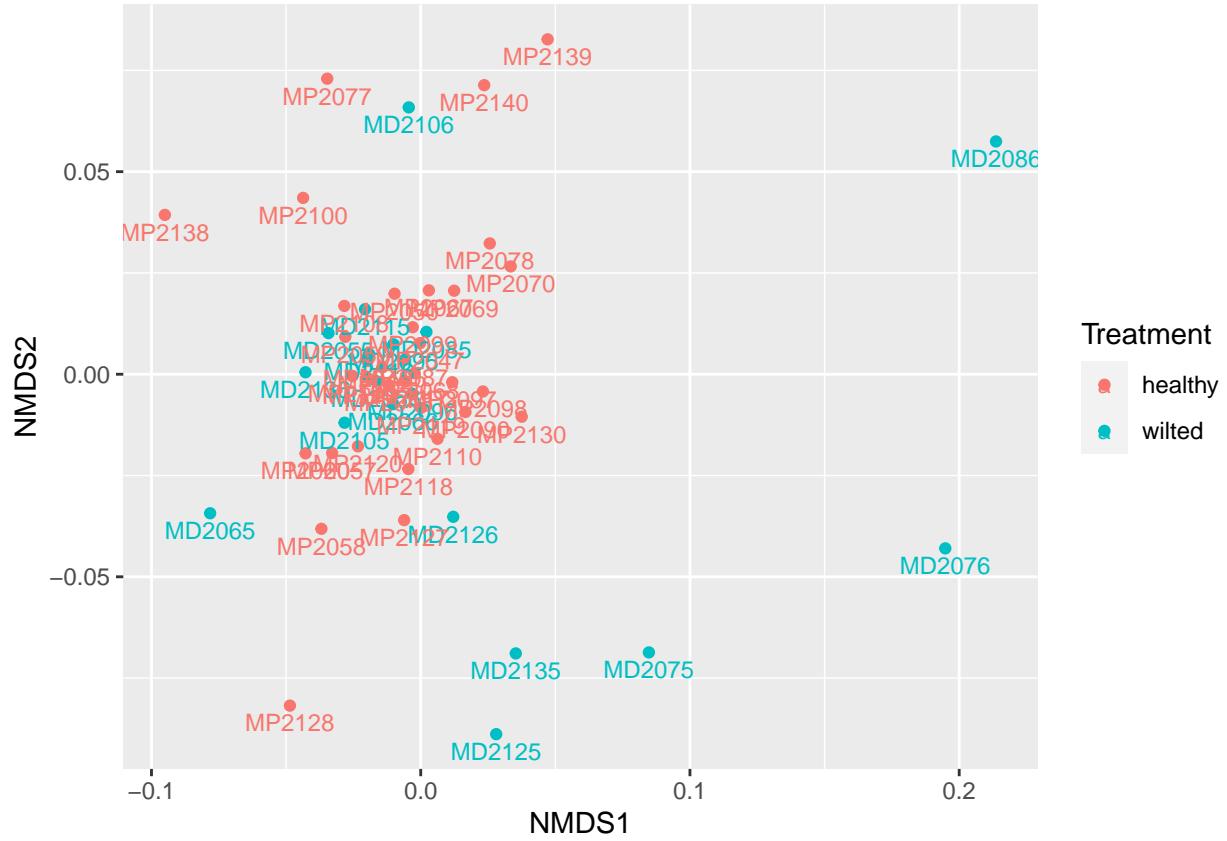
```
percentages_Bacteria <- transform_sample_counts(merge_Bacteria, function(x) x*100 / sum(x) )
percentages_Bacteria_df <- psmelt(percentages_Bacteria)
```

beta diversidad de Bacteria

```
meta_ord_Bacteria <- ordinate(physeq = percentages_Bacteria, method = "NMDS", distance = "bray")
```

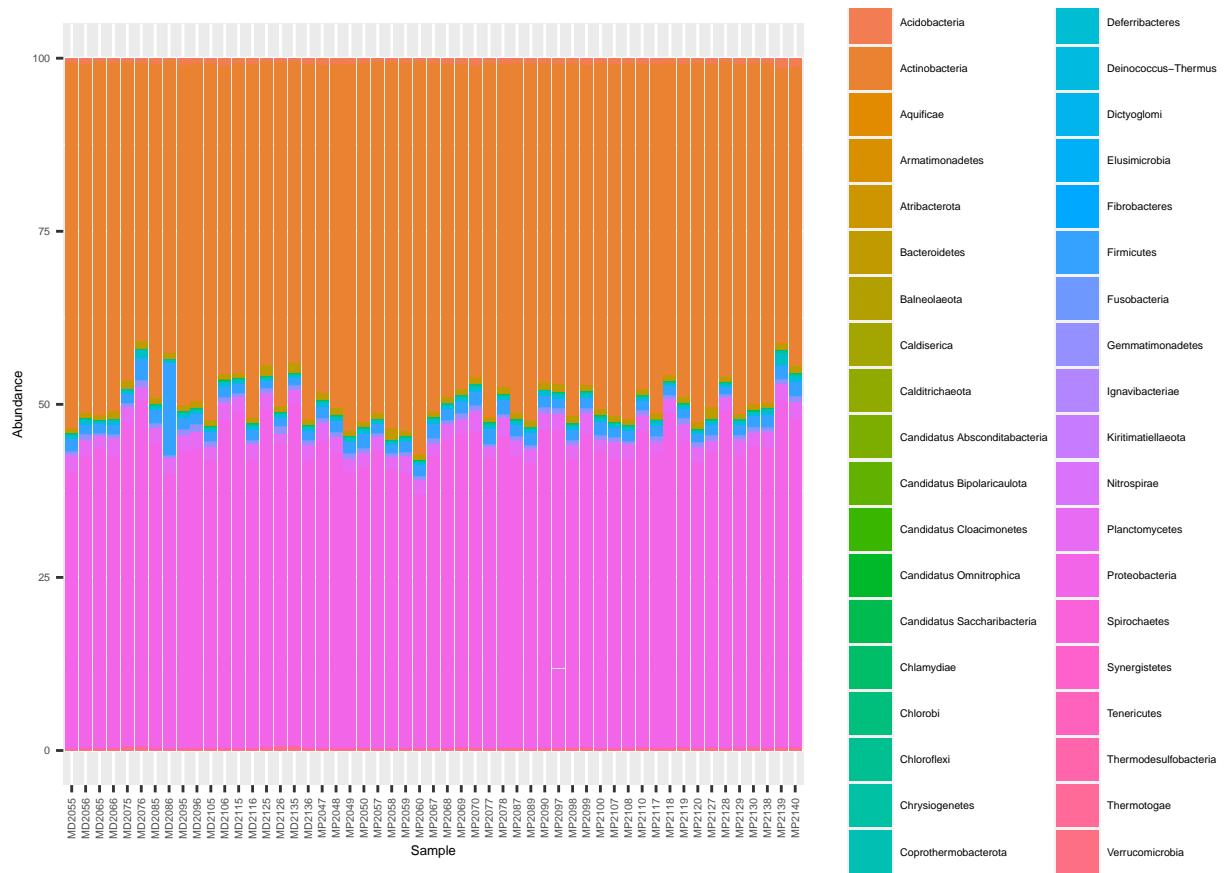
```
## Wisconsin double standardization
## Run 0 stress 0.1680232
## Run 1 stress 0.1915863
## Run 2 stress 0.1791703
## Run 3 stress 0.1669978
## ... New best solution
## ... Procrustes: rmse 0.06782128 max resid 0.3707017
## Run 4 stress 0.1697995
## Run 5 stress 0.1736023
## Run 6 stress 0.1852212
## Run 7 stress 0.1741432
## Run 8 stress 0.1707541
## Run 9 stress 0.1742283
## Run 10 stress 0.1769707
## Run 11 stress 0.163995
## ... New best solution
## ... Procrustes: rmse 0.05197664 max resid 0.2345715
## Run 12 stress 0.1657371
## Run 13 stress 0.1646995
## Run 14 stress 0.1650579
## Run 15 stress 0.1718761
## Run 16 stress 0.1698312
## Run 17 stress 0.1625034
## ... New best solution
## ... Procrustes: rmse 0.03762533 max resid 0.1600756
## Run 18 stress 0.1672217
## Run 19 stress 0.1633436
## Run 20 stress 0.1745444
## *** Best solution was not repeated -- monoMDS stopping criteria:
##      3: no. of iterations >= maxit
##      17: stress ratio > sratmax
```

```
plot_ordination(physeq = percentages_Bacteria, ordination = meta_ord_Bacteria, color = "Treatment") +
  geom_text(mapping = aes(label = colnames(fresa_kraken_fil@otu_table@Data))), size = 3, vjust = 1.5)
```



## Bacterias por Phylum

```
ggplot(data= percentages_Bacteria_df, aes(x=Sample, y=Abundance, fill=Phylum))+  
  geom_bar(aes(), stat="identity", position="stack") +  
  theme(text = element_text(size = 5),  
        axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```



```
merge_Bacteria_Phylum<-tax_glm(merge_Bacteria, taxrank="Phylum")
```

sacamos las abundancias relativas

```
percentages_Bacteria_Phylum <- transform_sample_counts(merge_Bacteria_Phylum, function(x) x*100 / sum(x))
percentages_Bacteria_Phylum_df <- psmelt(percentages_Bacteria_Phylum)
meta_ord_Bacteria_Phylum <- ordinate(physeq = percentages_Bacteria_Phylum, method = "NMDS", distance =
```

```
## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.1520004
## Run 1 stress 0.1766348
## Run 2 stress 0.159459
## Run 3 stress 0.1613697
## Run 4 stress 0.1519725
## ... New best solution
## ... Procrustes: rmse 0.008359202 max resid 0.04727895
## Run 5 stress 0.1592605
## Run 6 stress 0.1622254
## Run 7 stress 0.1520602
## ... Procrustes: rmse 0.04846904 max resid 0.3153058
## Run 8 stress 0.1581348
## Run 9 stress 0.1672981
## Run 10 stress 0.1554202
## Run 11 stress 0.15139
```

```

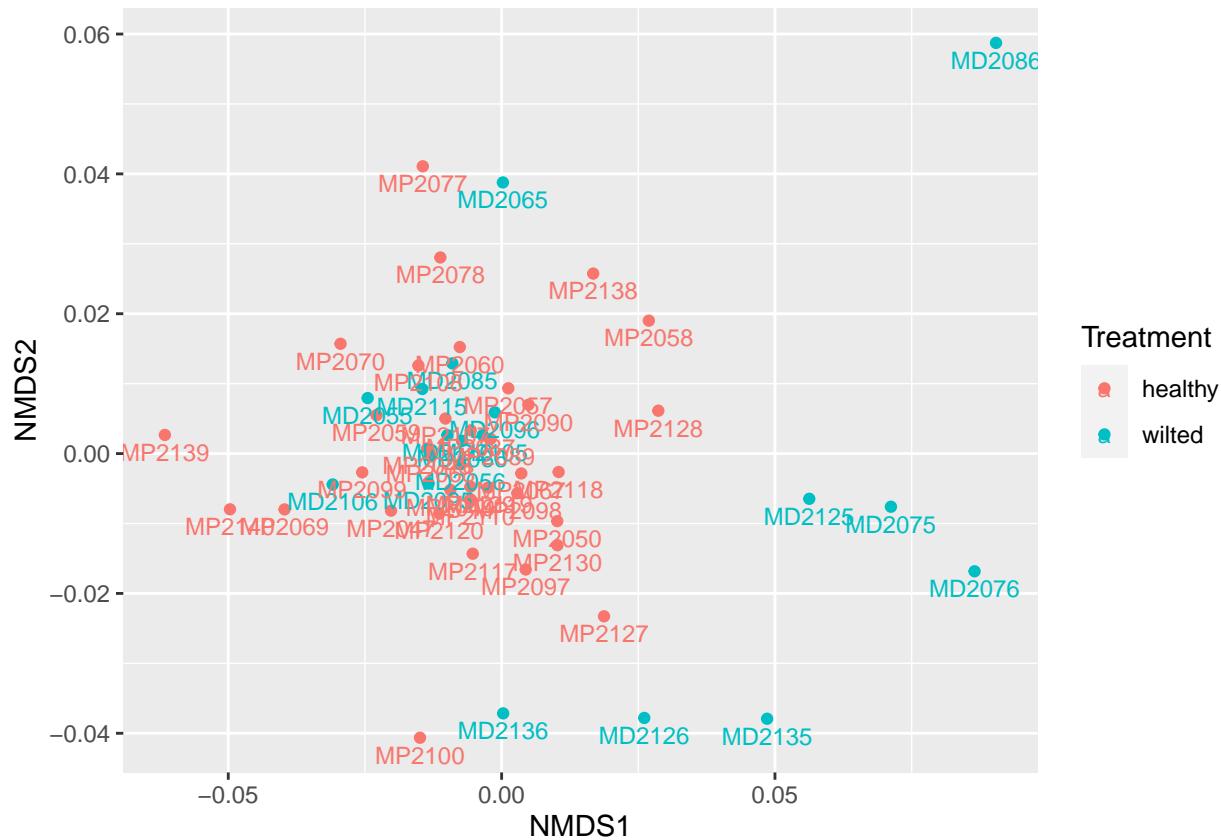
## ... New best solution
## ... Procrustes: rmse 0.01144968 max resid 0.05371022
## Run 12 stress 0.1741911
## Run 13 stress 0.1613696
## Run 14 stress 0.1755042
## Run 15 stress 0.16759
## Run 16 stress 0.177434
## Run 17 stress 0.1597361
## Run 18 stress 0.1663602
## Run 19 stress 0.1572804
## Run 20 stress 0.1705532
## *** Best solution was not repeated -- monoMDS stopping criteria:
##     19: stress ratio > sratmax
##     1: scale factor of the gradient < sfgrmin

```

```

plot_ordination(physeq = percentages_Bacteria_Phylum, ordination = meta_ord_Bacteria_Phylum, color = "Treatment",
geom_text(mapping = aes(label = colnames(fresa_kraken_fil@otu_table@Data))), size = 3, vjust = 1.5)

```



```

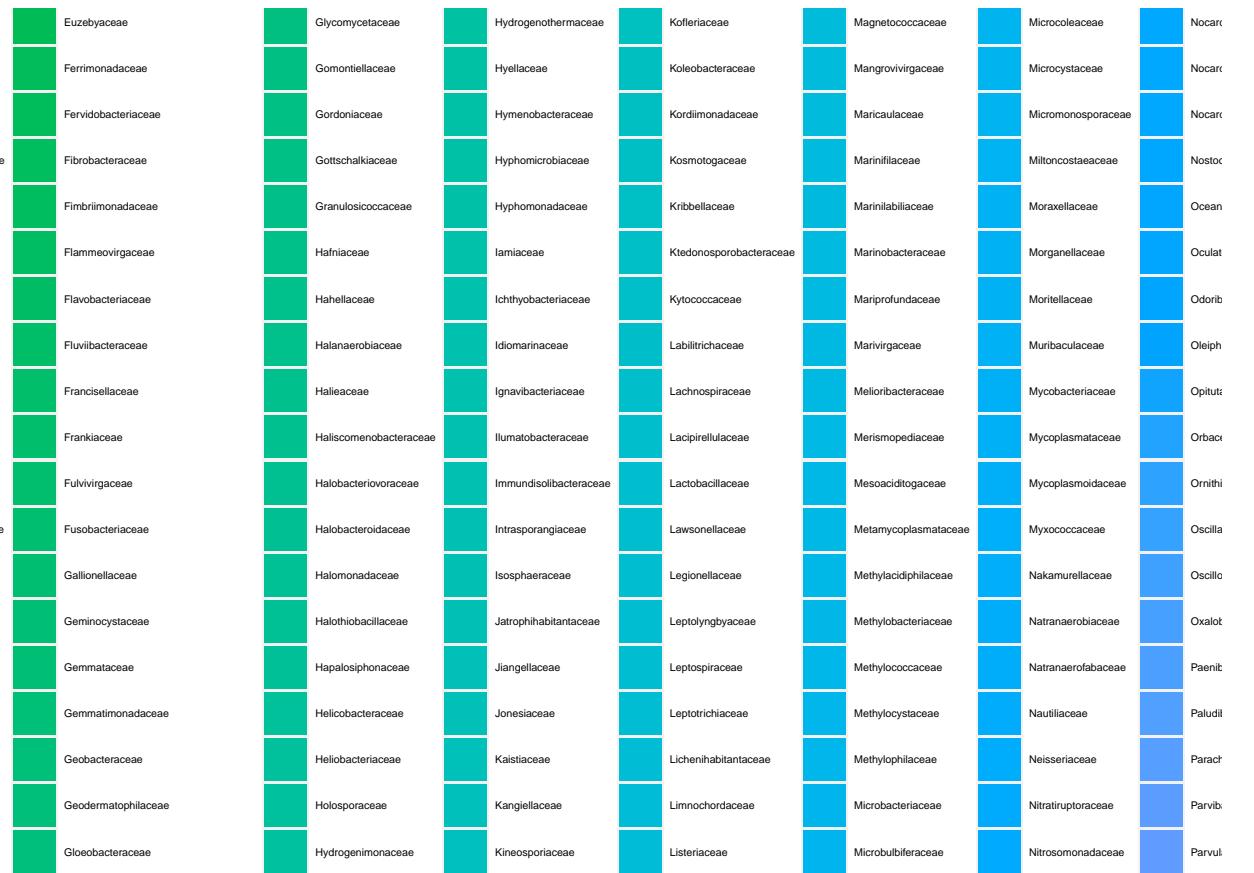
## Bacterias por Family

```

```

ggplot(data= percentages_Bacteria_df, aes(x=Sample, y=Abundance, fill=Family))+  
  geom_bar(aes(), stat="identity", position="stack") +  
  theme(text = element_text(size = 5),  
        axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))

```



```
merge_Bacteria_Family<-tax_glom(merge_Bacteria,taxrank="Family")
```

sacamos las abundancias relativas

```
percentages_Bacteria_Family <- transform_sample_counts(merge_Bacteria_Family, function(x) x*100 / sum(x))
percentages_Bacteria_Family_df <- ps melt(percentages_Bacteria_Family)
meta_ord_Bacteria_Family <- ordinate(phylseq = percentages_Bacteria_Family, method = "NMDS", distance = '
```

```
## Wisconsin double standardization
## Run 0 stress 0.1412073
## Run 1 stress 0.1657312
## Run 2 stress 0.1723767
## Run 3 stress 0.1390459
## ... New best solution
## ... Procrustes: rmse 0.0213499 max resid 0.1335054
## Run 4 stress 0.1521842
## Run 5 stress 0.1370381
## ... New best solution
## ... Procrustes: rmse 0.06280971 max resid 0.3760329
## Run 6 stress 0.15752
## Run 7 stress 0.1369864
## ... New best solution
## ... Procrustes: rmse 0.002352728 max resid 0.01255418
```

```

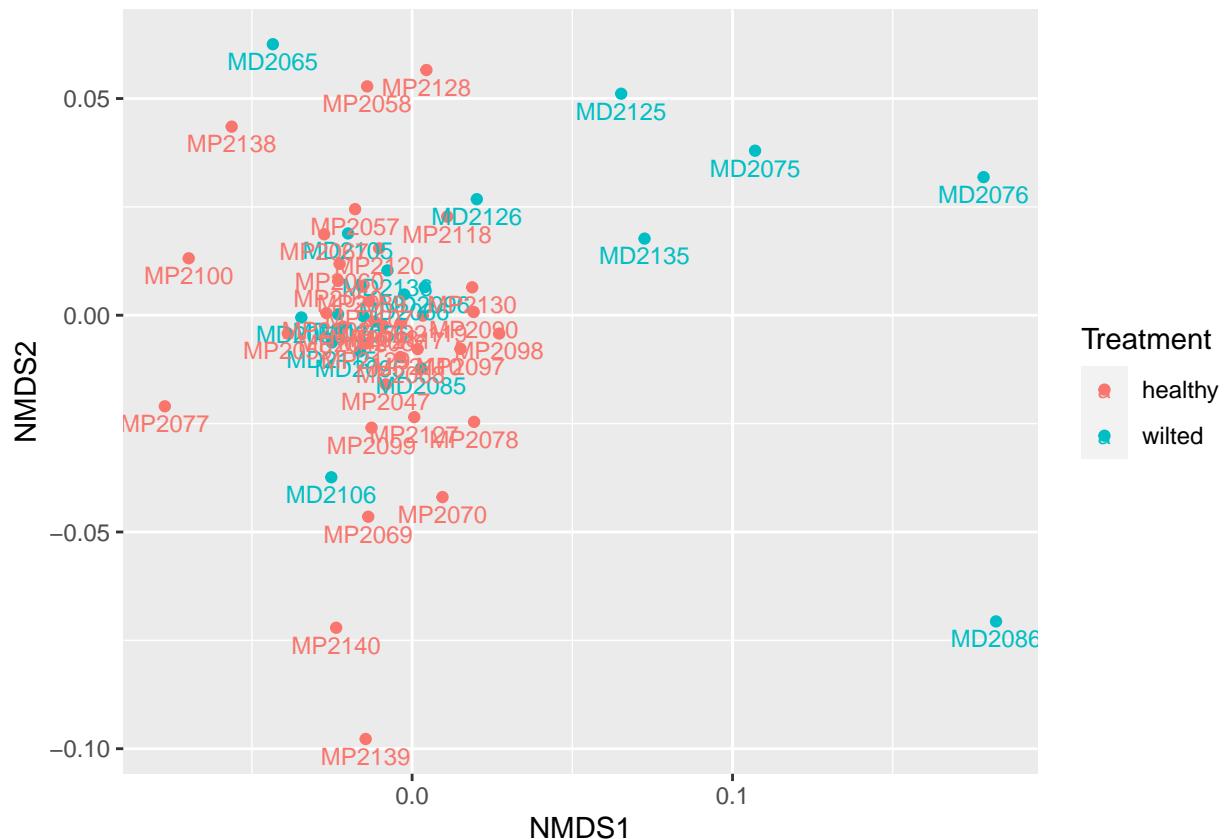
## Run 8 stress 0.1397114
## Run 9 stress 0.140564
## Run 10 stress 0.1723792
## Run 11 stress 0.1392799
## Run 12 stress 0.1396308
## Run 13 stress 0.1370378
## ... Procrustes: rmse 0.00232936 max resid 0.01220736
## Run 14 stress 0.1371725
## ... Procrustes: rmse 0.01190384 max resid 0.05738298
## Run 15 stress 0.1455598
## Run 16 stress 0.1840637
## Run 17 stress 0.1391843
## Run 18 stress 0.1441743
## Run 19 stress 0.1396492
## Run 20 stress 0.1677301
## *** Best solution was not repeated -- monoMDS stopping criteria:
##      3: no. of iterations >= maxit
##      17: stress ratio > sratmax

```

```

plot_ordination(physeq = percentages_Bacteria_Family, ordination = meta_ord_Bacteria_Family, color = "Treatment")
geom_text(mapping = aes(label = colnames(fresa_kraken_fil@otu_table@Data)), size = 3, vjust = 1.5)

```



```

## Bacterias por Genero

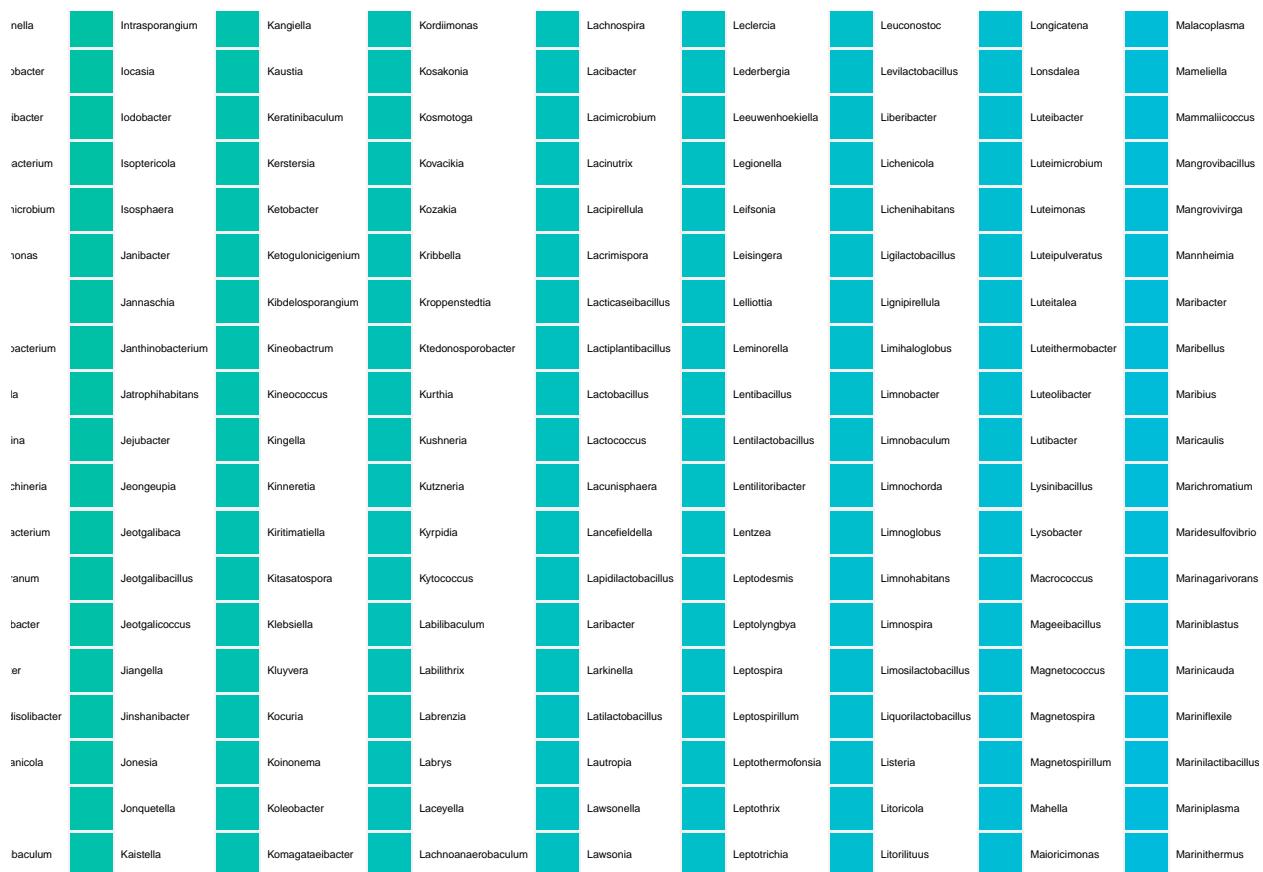
```

```

ggplot(data= percentages_Bacteria_df, aes(x=Sample, y=Abundance, fill=Genus))+
  geom_bar(aes(), stat="identity", position="stack") +

```

```
theme(text = element_text(size = 5),  
      axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```



```
merge_Bacteria_Genus<-tax_glom(merge_Bacteria,taxrank="Genus")
```

sacamos las abundancias relativas

```
percentages_Bacteria_Genus <- transform_sample_counts(merge_Bacteria_Genus, function(x) x*100 / sum(x))
percentages_Bacteria_Genus_df <- psmelt(percentages_Bacteria_Genus)
meta_ord_Bacteria_Genus <- ordinate(physeq = percentages_Bacteria_Genus, method = "NMDS", distance = "bray")
```

```
## Wisconsin double standardization
## Run 0 stress 0.1467945
## Run 1 stress 0.1467426
## ... New best solution
## ... Procrustes: rmse 0.008080137 max resid 0.03208235
## Run 2 stress 0.1438141
## ... New best solution
## ... Procrustes: rmse 0.06685495 max resid 0.2987778
## Run 3 stress 0.1687629
## Run 4 stress 0.1506174
## Run 5 stress 0.1512148
## Run 6 stress 0.1586371
## Run 7 stress 0.1476747
```

```

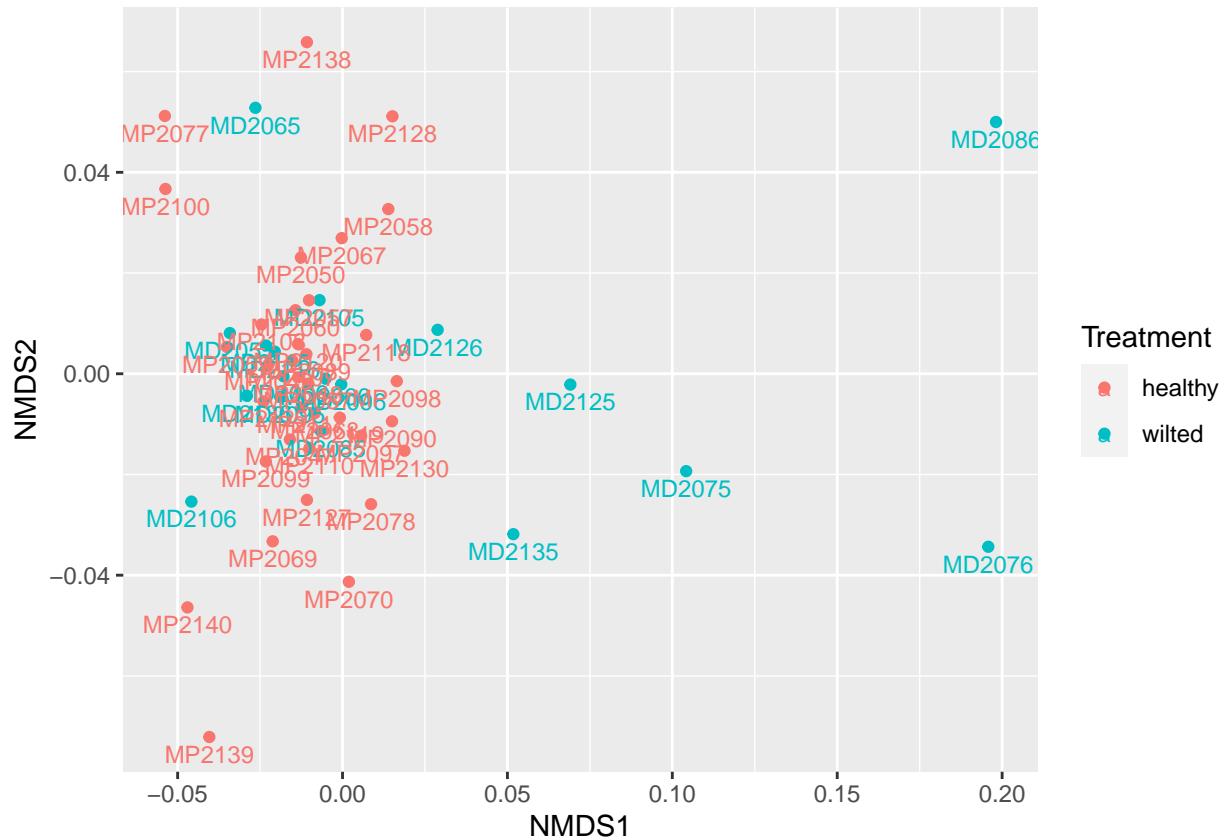
## Run 8 stress 0.1490334
## Run 9 stress 0.1438187
## ... Procrustes: rmse 0.001064456 max resid 0.004049413
## ... Similar to previous best
## Run 10 stress 0.1505679
## Run 11 stress 0.1490742
## Run 12 stress 0.1454657
## Run 13 stress 0.1432836
## ... New best solution
## ... Procrustes: rmse 0.05017339 max resid 0.2980421
## Run 14 stress 0.1527485
## Run 15 stress 0.1455188
## Run 16 stress 0.1438621
## Run 17 stress 0.1586553
## Run 18 stress 0.1510903
## Run 19 stress 0.1506453
## Run 20 stress 0.1512152
## *** Best solution was not repeated -- monoMDS stopping criteria:
##      2: no. of iterations >= maxit
##      18: stress ratio > sratmax

```

```

plot_ordination(physeq = percentages_Bacteria_Genus, ordination = meta_ord_Bacteria_Genus, color = "Treatment",
geom_text(mapping = aes(label = colnames(fresa_kraken_fil@otu_table@Data)), size = 3, vjust = 1.5)

```



tomando diferentes porcentajes de abundancia

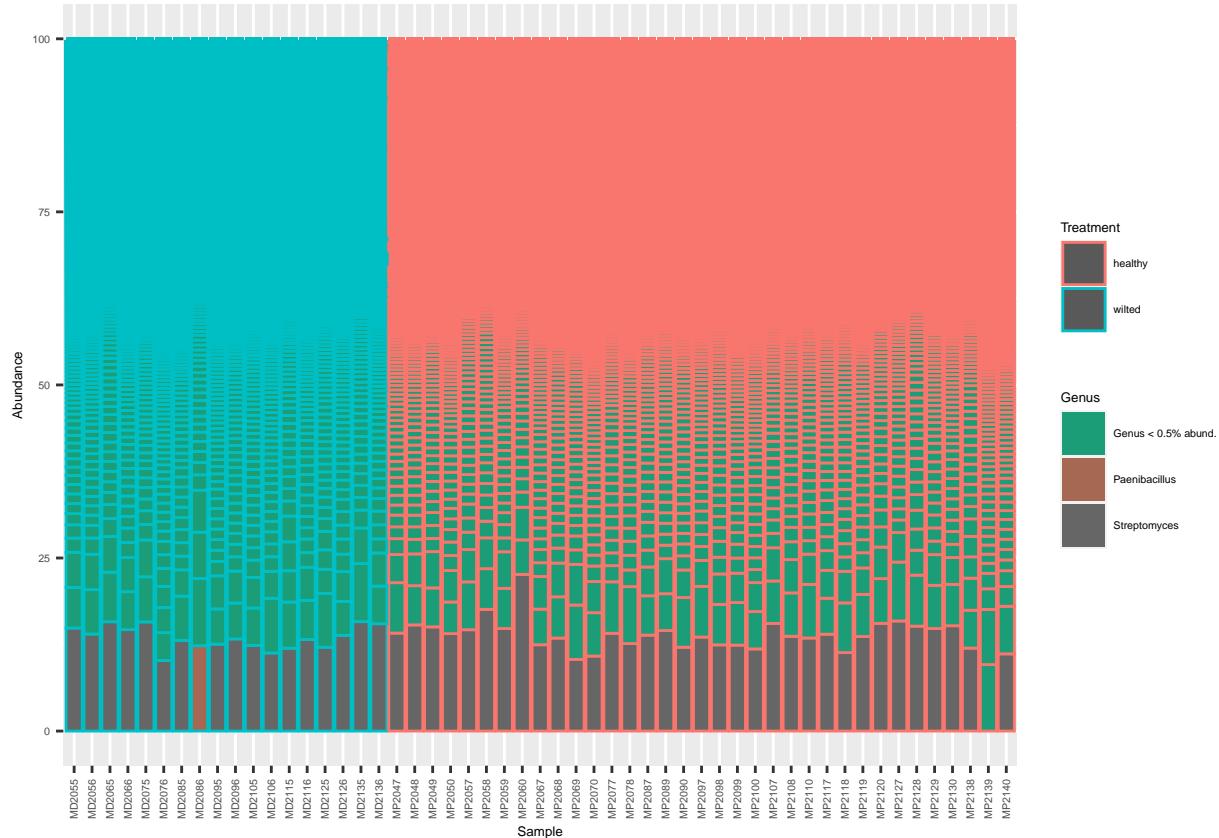
```

percentages_Bacteria_Genus_df$Genus [percentages_Bacteria_Genus_df$Abundance < 10.0] <- "Genus < 0.5% abd"
percentages_Bacteria_Genus_df$Genus <- as.factor(percentages_Bacteria_Genus_df$Genus)

genus_colors_rel <- colorRampPalette(brewer.pal(8, "Dark2")) (length(levels(percentages_Bacteria_Genus_df$Genus))
relative_plot <- ggplot(data=percentages_Bacteria_Genus_df, aes(x=Sample, y=Abundance, fill=Genus ,color=Genus))
  geom_bar(aes(), stat="identity", position="stack") +
  scale_fill_manual(values = genus_colors_rel) +
  theme(text = element_text(size = 5),
        axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))

relative_plot

```

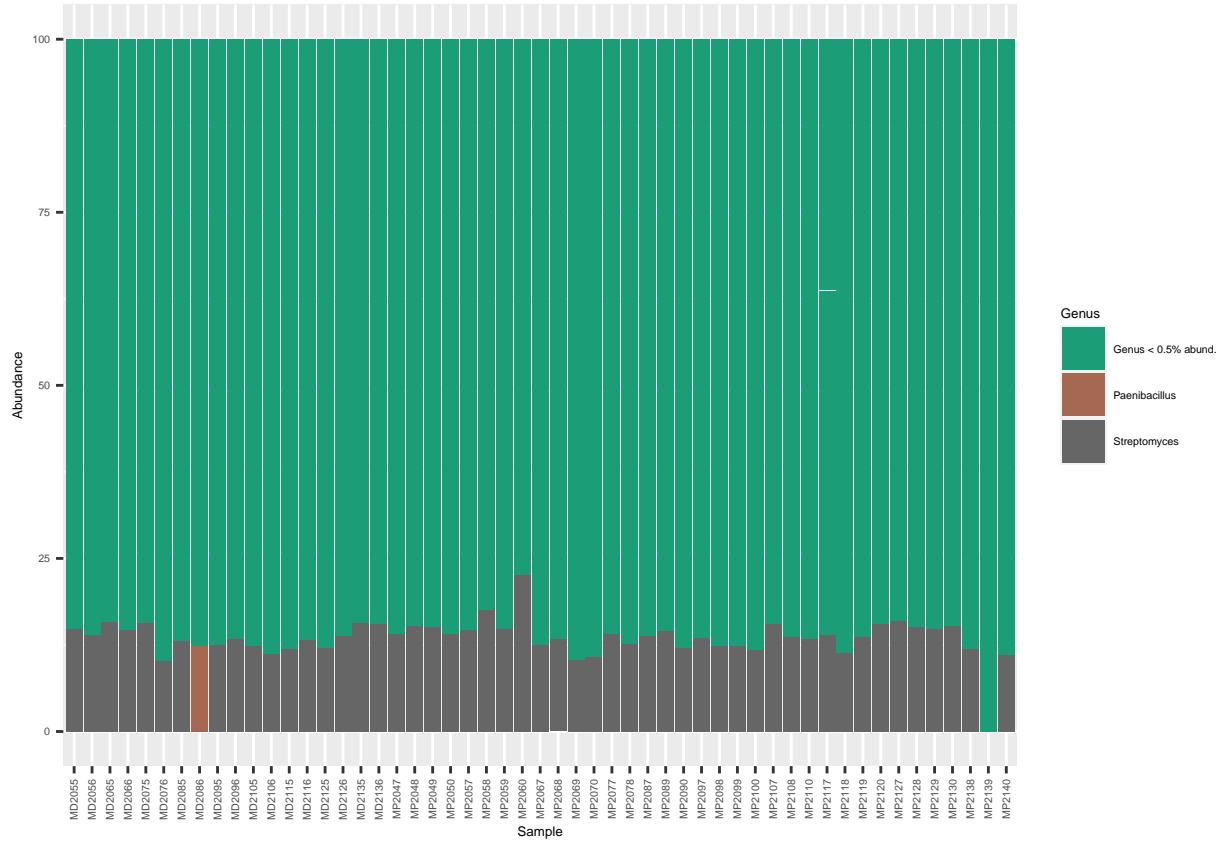


```

relative_plot <- ggplot(data=percentages_Bacteria_Genus_df, aes(x=Sample, y=Abundance, fill=Genus)) +
  geom_bar(aes(), stat="identity", position="stack") +
  scale_fill_manual(values = genus_colors_rel) +
  theme(text = element_text(size = 5),
        axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))

relative_plot

```



sacando la beta diversidad con aglomerado de 10%

```
meta_ord_Bacteria_Genus <- ordinate(physeq = percentages_Bacteria_Genus, method = "NMDS", distance = "b
```

```
## Wisconsin double standardization
## Run 0 stress 0.1467945
## Run 1 stress 0.1525053
## Run 2 stress 0.1438215
## ... New best solution
## ... Procrustes: rmse 0.06870747 max resid 0.2974654
## Run 3 stress 0.143814
## ... New best solution
## ... Procrustes: rmse 0.0008804587 max resid 0.003342016
## ... Similar to previous best
## Run 4 stress 0.1433286
## ... New best solution
## ... Procrustes: rmse 0.04918971 max resid 0.2958043
## Run 5 stress 0.1458461
## Run 6 stress 0.14764
## Run 7 stress 0.1555413
## Run 8 stress 0.149081
## Run 9 stress 0.1467927
## Run 10 stress 0.1484464
## Run 11 stress 0.1493098
## Run 12 stress 0.1437855
## ... Procrustes: rmse 0.01547107 max resid 0.08746434
```

```

## Run 13 stress 0.1489597
## Run 14 stress 0.1440733
## Run 15 stress 0.1477649
## Run 16 stress 0.1485708
## Run 17 stress 0.1531023
## Run 18 stress 0.1728691
## Run 19 stress 0.1541725
## Run 20 stress 0.1534084
## *** Best solution was not repeated -- monoMDS stopping criteria:
##     4: no. of iterations >= maxit
##     16: stress ratio > sratmax

```

```

plot_ordination(physeq = percentages_Bacteria_Genus, ordination = meta_ord_Bacteria_Genus, color = "Treatment"
geom_text(mapping = aes(label = colnames(fresa_kraken_fil@otu_table@Data)), size = 3, vjust = 1.5)

```

