Neiker

Val F. Lanza

library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(vegan)

Cargando paquete requerido: permute  
Cargando paquete requerido: lattice  
This is vegan 2.6-4

library(ggsci)  
library(gamlss)

Cargando paquete requerido: splines  
Cargando paquete requerido: gamlss.data  
  
Adjuntando el paquete: 'gamlss.data'  
  
The following object is masked from 'package:datasets':  
  
 sleep  
  
Cargando paquete requerido: gamlss.dist  
Cargando paquete requerido: nlme  
  
Adjuntando el paquete: 'nlme'  
  
The following object is masked from 'package:dplyr':  
  
 collapse  
  
Cargando paquete requerido: parallel  
 \*\*\*\*\*\*\*\*\*\* GAMLSS Version 5.4-22 \*\*\*\*\*\*\*\*\*\*   
For more on GAMLSS look at https://www.gamlss.com/  
Type gamlssNews() to see new features/changes/bug fixes.

library(easystats)

# Attaching packages: easystats 0.7.1  
✔ bayestestR 0.13.2 ✔ correlation 0.8.4   
✔ datawizard 0.10.0 ✔ effectsize 0.8.7   
✔ insight 0.19.10 ✔ modelbased 0.8.7   
✔ performance 0.11.0 ✔ parameters 0.21.6   
✔ report 0.5.8 ✔ see 0.8.4

library(ggstatsplot)

You can cite this package as:  
 Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach.  
 Journal of Open Source Software, 6(61), 3167, doi:10.21105/joss.03167

library(dplyr)

# Loading data

files <- dir(".", pattern = "res")  
  
tabla <- data.frame()  
  
for (i in files)  
{  
 tmp <- read\_tsv(i)  
 sample <- gsub(".res","",i)  
 tmp$Sample <- sample  
   
 tabla <- bind\_rows(tabla,tmp)  
}

## Filter and Transform

tabla\_otu <- tabla %>%   
 filter(Template\_Identity > 95 & Template\_Coverage > 95) %>%   
 dplyr::select(Sample,Gene = `#Template`, Depth) %>%  
 pivot\_wider(names\_from = Gene, values\_from = Depth, values\_fill = 0)

## Load Metadata

metadata <- readxl::read\_xlsx("Metadata.xlsx")  
  
metadata <- metadata %>% rename(Sample = Muestra)  
  
full\_tabla <- tabla %>% inner\_join(metadata %>% mutate(Sample = str\_c("Sample\_",Sample)))

Joining with `by = join\_by(Sample)`

stats <- read\_table("stats.tsv")

── Column specification ────────────────────────────────────────────────────────  
cols(  
 file = col\_character(),  
 format = col\_character(),  
 type = col\_character(),  
 num\_seqs = col\_number(),  
 sum\_len = col\_number(),  
 min\_len = col\_double(),  
 avg\_len = col\_double(),  
 max\_len = col\_double()  
)

lista\_run <- read\_tsv("lista.txt", col\_names = c("Run","file")) %>%   
 mutate(file = gsub(".fastq.gz",".clean.fastq.gz",file))

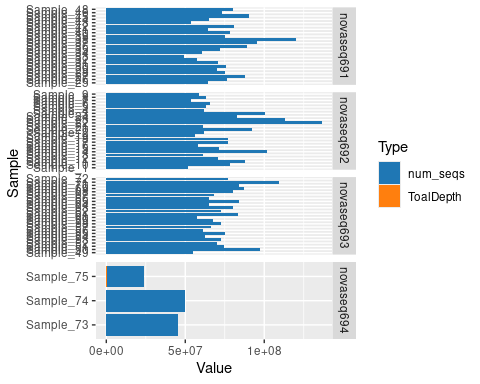
Rows: 376 Columns: 2  
── Column specification ────────────────────────────────────────────────────────  
Delimiter: "\t"  
chr (2): Run, file  
  
ℹ Use `spec()` to retrieve the full column specification for this data.  
ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

stats <- stats %>% inner\_join(lista\_run) %>%   
 separate(file, c("Pool","Sample","Pos","line","Read","kk"), sep ="\_") %>%   
 mutate(Sample = paste0("Sample\_",Sample)) %>%   
 group\_by(Sample,Run) %>% summarise(num\_seqs = sum(num\_seqs), sum\_len = sum(sum\_len))

Joining with `by = join\_by(file)`  
`summarise()` has grouped output by 'Sample'. You can override using the  
`.groups` argument.

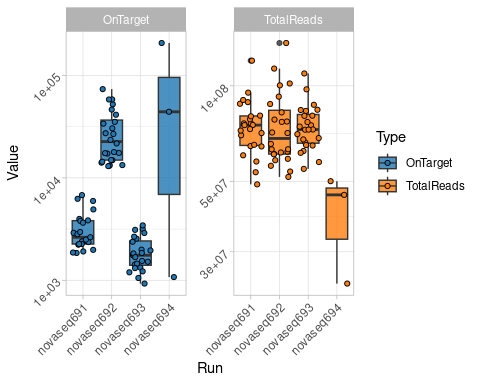
full\_tabla %>% inner\_join(stats) %>%   
 group\_by(Sample, num\_seqs, Run) %>%   
 summarise(ToalDepth = sum(Depth)) %>%   
 ungroup() %>%   
 pivot\_longer(names\_to = "Type", values\_to = "Value", -c(Sample,Run)) %>%   
 ggplot(aes(x = Sample, y = Value, fill =Type)) +   
 geom\_col(position = "stack") +   
 facet\_grid(Run ~., scales = "free\_y") +   
 coord\_flip() +  
 scale\_fill\_d3()

Joining with `by = join\_by(Sample)`  
`summarise()` has grouped output by 'Sample', 'num\_seqs'. You can override  
using the `.groups` argument.



full\_tabla %>% inner\_join(stats) %>%   
 group\_by(Sample, TotalReads = num\_seqs, Run) %>%   
 summarise(OnTarget = sum(Depth)) %>%   
 ungroup() %>%   
 pivot\_longer(names\_to = "Type", values\_to = "Value", -c(Sample,Run)) %>%   
 #filter(Run != "novaseq694") %>%   
 ggplot(aes(x = Run, y = Value, fill =Type)) +   
 geom\_boxplot(alpha = 0.8) +   
 geom\_jitter(shape = 21) +  
 facet\_wrap(Type ~., scales = "free") +   
 scale\_fill\_d3() +  
 theme\_light() +  
 scale\_y\_log10() +  
 theme(axis.text = element\_text(angle = 45, hjust = 1))

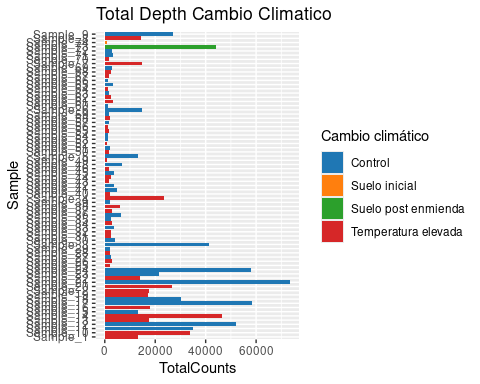
Joining with `by = join\_by(Sample)`  
`summarise()` has grouped output by 'Sample', 'TotalReads'. You can override  
using the `.groups` argument.



# Quality Control

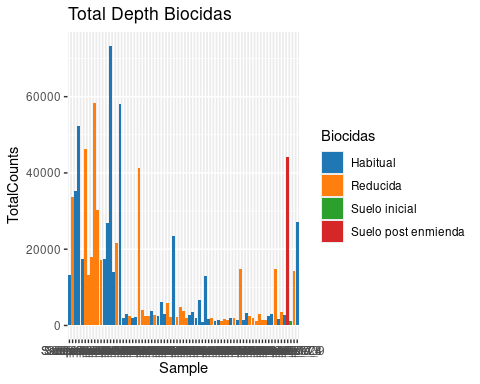
full\_tabla %>% filter(Sample != "Sample\_75") %>%   
 group\_by(Sample,`Cambio climático`,Biocidas) %>%   
 summarise(TotalCounts = sum(Depth)) %>%   
 ggplot(aes(x= Sample, y=TotalCounts, fill = `Cambio climático`)) +   
 geom\_col() +  
 scale\_fill\_d3() +  
 labs(title = "Total Depth Cambio Climatico") +  
 coord\_flip()

`summarise()` has grouped output by 'Sample', 'Cambio climático'. You can  
override using the `.groups` argument.



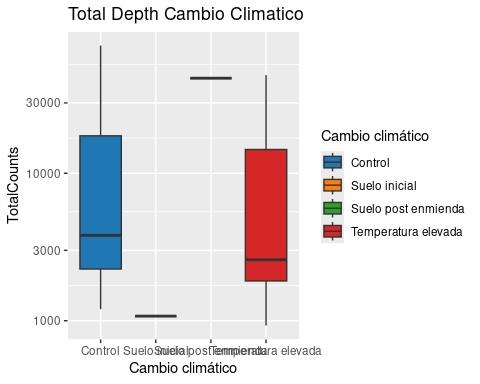
full\_tabla %>% filter(Sample != "Sample\_75") %>%   
 group\_by(Sample,`Cambio climático`,Biocidas) %>%   
 summarise(TotalCounts = sum(Depth)) %>%   
 ggplot(aes(x= Sample, y=TotalCounts, fill = Biocidas)) +   
 geom\_col() +  
 scale\_fill\_d3()+  
 labs(title = "Total Depth Biocidas")

`summarise()` has grouped output by 'Sample', 'Cambio climático'. You can  
override using the `.groups` argument.



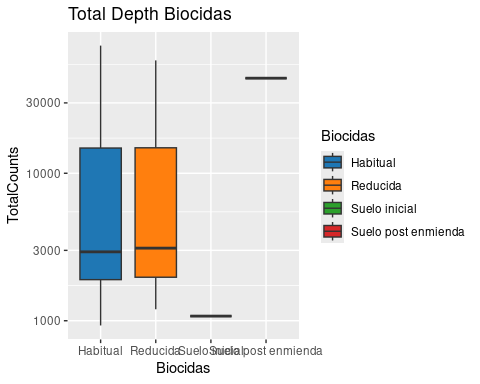
full\_tabla %>% filter(Sample != "Sample\_75") %>%   
 group\_by(Sample,`Cambio climático`,Biocidas) %>%   
 summarise(TotalCounts = sum(Depth)) %>%   
 ggplot(aes(x= `Cambio climático`, y=TotalCounts, fill = `Cambio climático`)) +   
 geom\_boxplot()+  
 scale\_fill\_d3()+  
 scale\_y\_log10()+  
 labs(title = "Total Depth Cambio Climatico")

`summarise()` has grouped output by 'Sample', 'Cambio climático'. You can  
override using the `.groups` argument.



full\_tabla %>% filter(Sample != "Sample\_75") %>%   
 group\_by(Sample,`Cambio climático`,Biocidas) %>%   
 summarise(TotalCounts = sum(Depth)) %>%   
 ggplot(aes(x= Biocidas, y=TotalCounts, fill = Biocidas)) +   
 geom\_boxplot()+  
 scale\_fill\_d3()+  
 scale\_y\_log10()+  
 labs(title = "Total Depth Biocidas")

`summarise()` has grouped output by 'Sample', 'Cambio climático'. You can  
override using the `.groups` argument.



Validos <- full\_tabla %>%   
 group\_by(Sample) %>%   
 summarise(TotalCounts = sum(Depth)) %>%  
 filter(TotalCounts > 10000) %>% dplyr::select(Sample)

# Alpha Diversity

MinSampleSize <- full\_tabla %>%   
 semi\_join(Validos) %>%   
 filter(Template\_Coverage > 85, Template\_Identity > 85) %>%   
 dplyr::select( Sample, `#Template`, Depth) %>%   
 group\_by(Sample) %>%   
 summarise(SampleSize = sum(Depth)) %>%   
 pull(SampleSize) %>% min()

Joining with `by = join\_by(Sample)`

rare\_table <- full\_tabla %>%   
 semi\_join(Validos) %>%   
 filter(Template\_Coverage > 85, Template\_Identity > 85) %>%   
 dplyr::select( Sample, `#Template`, Depth) %>%   
 mutate(Depth = round(Depth)) %>%   
 pivot\_wider(names\_from = `#Template`, values\_from = Depth, values\_fill =0) %>%   
 column\_to\_rownames("Sample") %>%   
 as.matrix() %>%   
 rrarefy( sample = MinSampleSize)

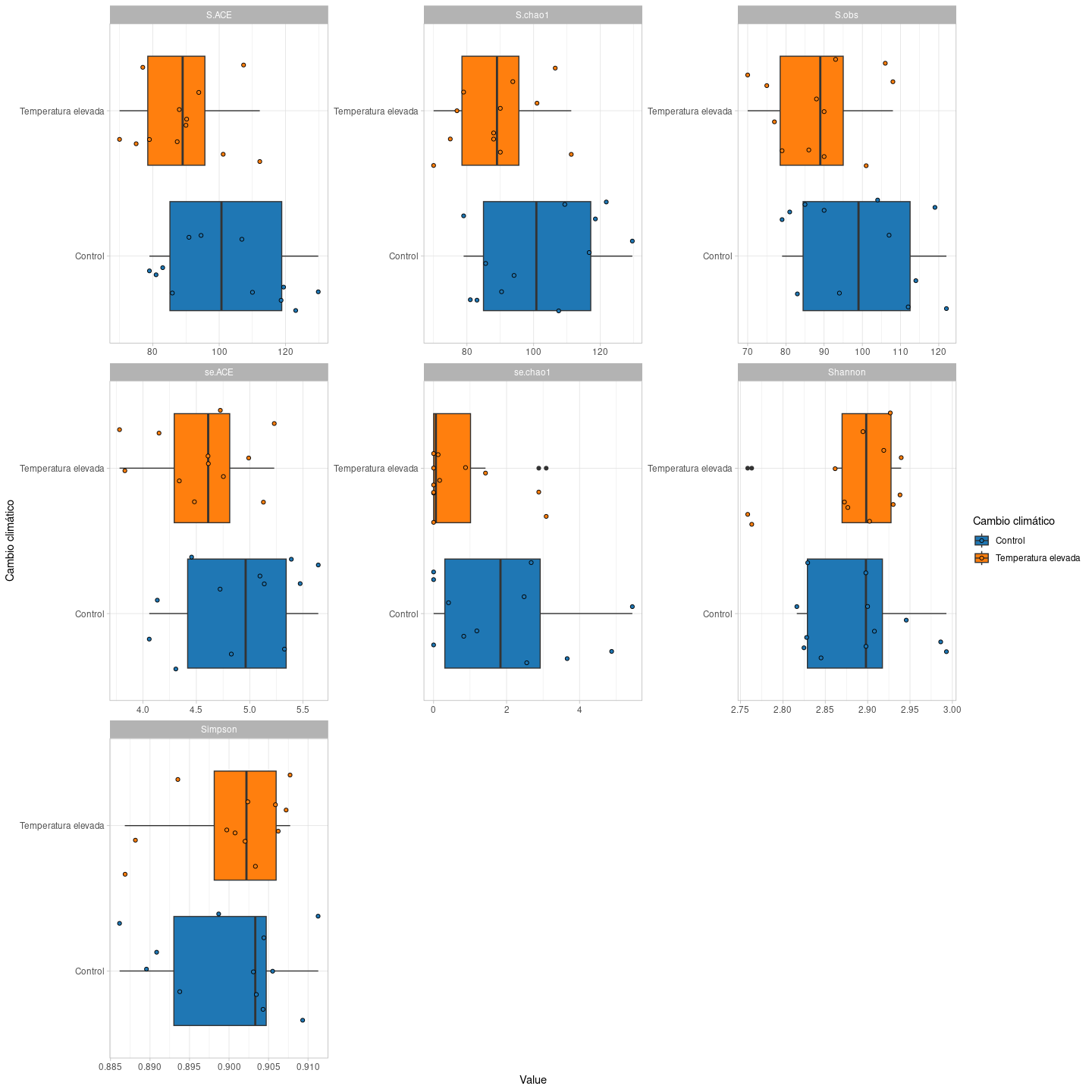
Joining with `by = join\_by(Sample)`

Div <- inner\_join(  
 data.frame(Simpson = diversity(rare\_table, index = "simpson"),  
 Shannon = diversity(rare\_table, index = "shannon")) %>%   
 rownames\_to\_column("Sample"),  
 estimateR(rare\_table) %>% t() %>%as.data.frame() %>% rownames\_to\_column("Sample"))

Joining with `by = join\_by(Sample)`

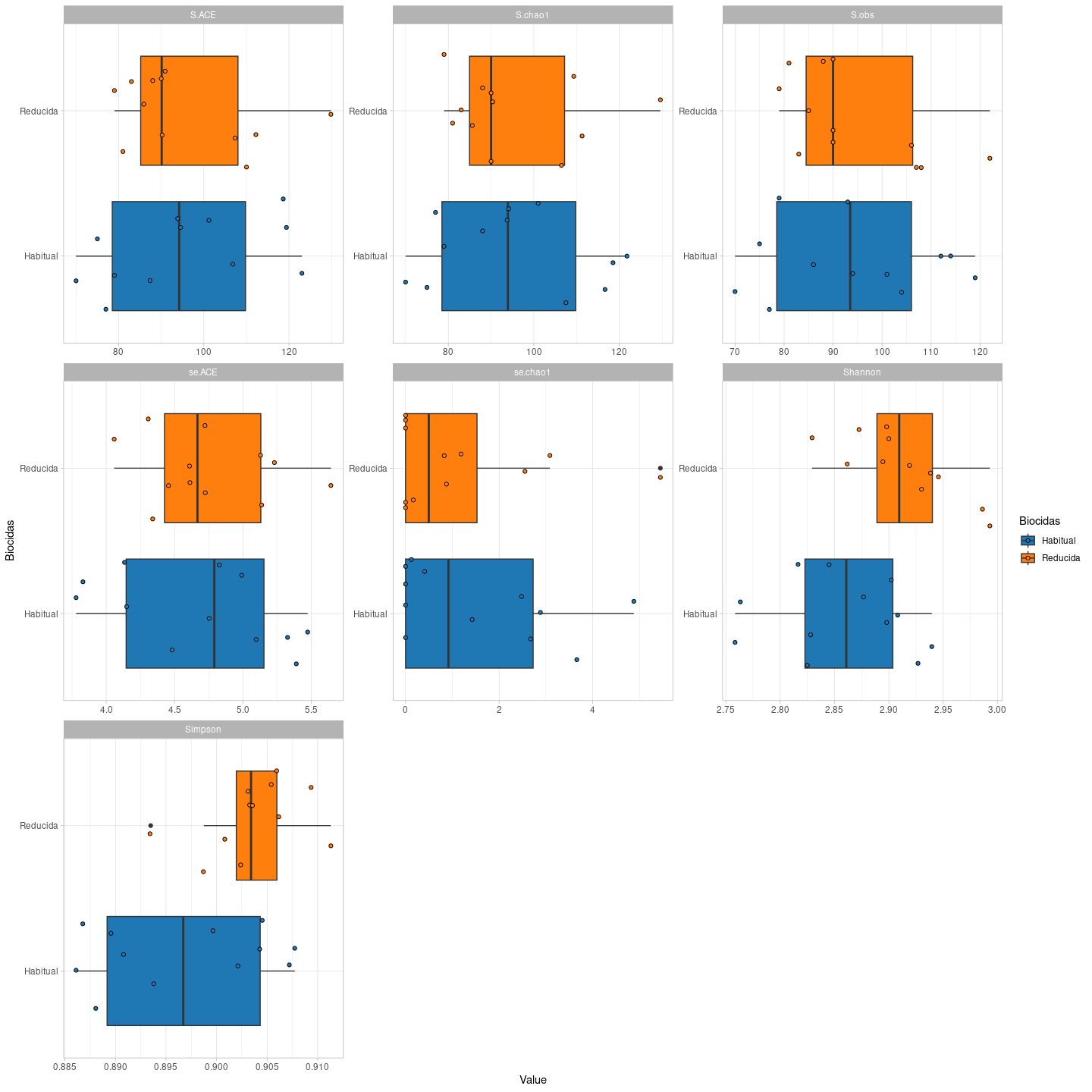
Div %>%   
 pivot\_longer(names\_to = "Index", values\_to = "Value", -Sample) %>%   
 inner\_join(metadata %>%   
 mutate(Sample = str\_c("Sample\_",Sample))) %>%   
 filter(`Cambio climático` == "Control" | `Cambio climático` == "Temperatura elevada") %>%   
 ggplot(aes(x = `Cambio climático`, y = Value, fill = `Cambio climático`)) +   
 geom\_boxplot() +   
 geom\_jitter(shape = 21) +  
 facet\_wrap(~Index, scales = "free") +  
 scale\_fill\_d3() +   
 theme\_light() + coord\_flip()

Joining with `by = join\_by(Sample)`



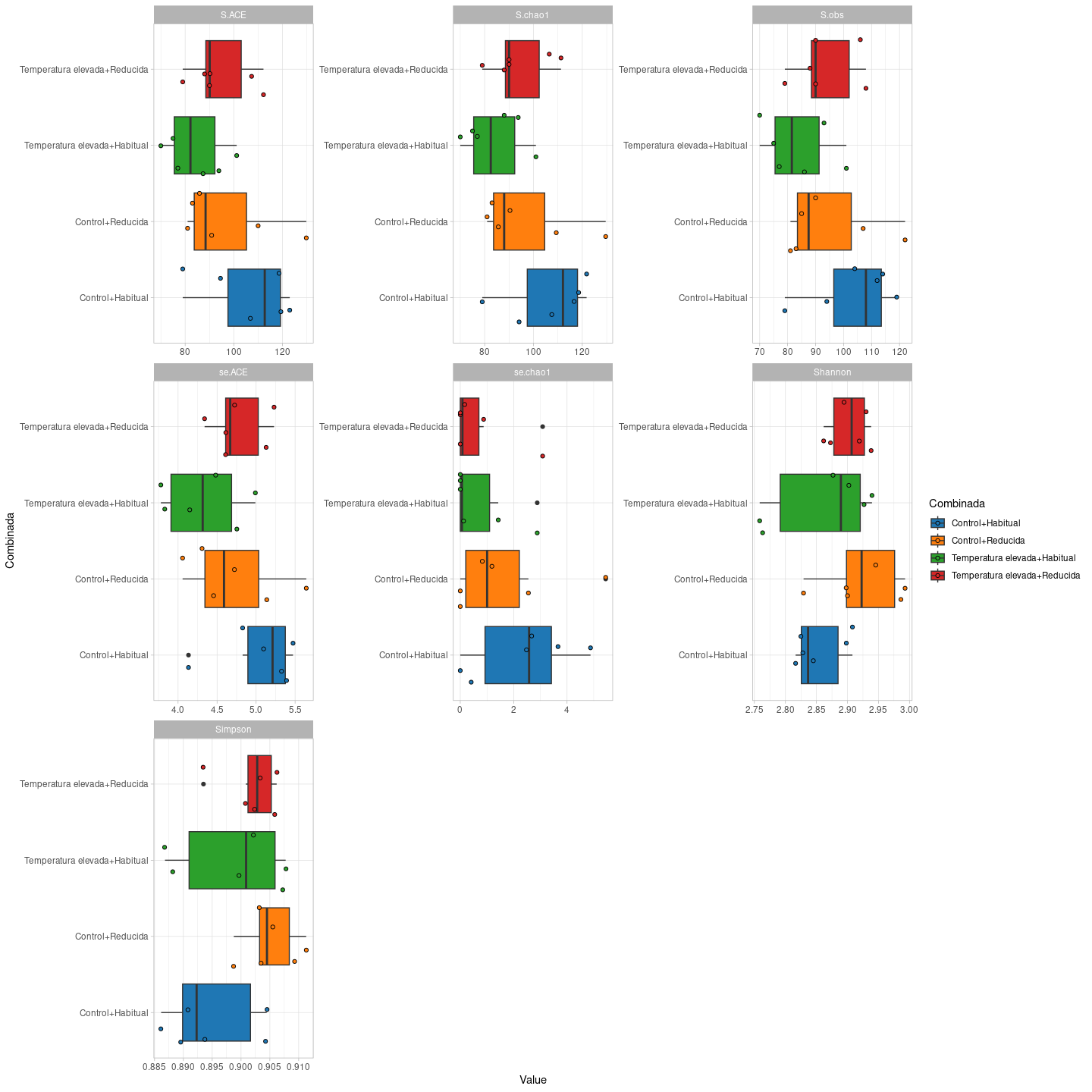
Div %>%   
 pivot\_longer(names\_to = "Index", values\_to = "Value", -Sample) %>%   
 inner\_join(metadata %>%   
 mutate(Sample = str\_c("Sample\_",Sample))) %>%   
 filter(`Cambio climático` == "Control" | `Cambio climático` == "Temperatura elevada") %>%   
 ggplot(aes(x = Biocidas, y = Value, fill = Biocidas)) +   
 geom\_boxplot() +   
 geom\_jitter(shape = 21) +  
 facet\_wrap(~Index, scales = "free") +  
 scale\_fill\_d3() +   
 theme\_light() + coord\_flip()

Joining with `by = join\_by(Sample)`



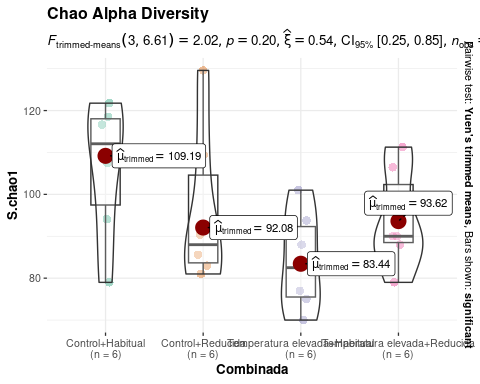
Div %>%   
 pivot\_longer(names\_to = "Index", values\_to = "Value", -Sample) %>%   
 inner\_join(metadata %>%   
 mutate(Sample = str\_c("Sample\_",Sample)) %>%   
 mutate(Combinada = str\_c(`Cambio climático`,"+",Biocidas))  
 ) %>%   
 filter(`Cambio climático` == "Control" | `Cambio climático` == "Temperatura elevada") %>%   
 ggplot(aes(x = Combinada, y = Value, fill = Combinada)) +   
 geom\_boxplot() +   
 geom\_jitter(shape = 21) +  
 facet\_wrap(~Index, scales = "free") +  
 scale\_fill\_d3() +   
 theme\_light() + coord\_flip()

Joining with `by = join\_by(Sample)`



ggbetweenstats(x = Combinada,  
 y = S.chao1,   
 data = Div %>% inner\_join(metadata %>% filter(Sample < 73) %>%   
 mutate(Sample = str\_c("Sample\_",Sample)) %>%   
 mutate(Combinada = str\_c(`Cambio climático`,"+",Biocidas))),  
 type = "robust", title = "Chao Alpha Diversity")

Joining with `by = join\_by(Sample)`



# Beta Diversity

otu\_table\_sorted <- rare\_table %>%   
 as.data.frame() %>%   
 rownames\_to\_column("Sample") %>%   
 inner\_join(metadata %>%   
 filter(Sample < 73) %>%   
 mutate(Sample = str\_c("Sample\_",Sample))) %>%   
 dplyr::select(-(Subparcela:Otros)) %>% column\_to\_rownames("Sample") %>% as.matrix()

Joining with `by = join\_by(Sample)`

metadata\_sorted <- rare\_table %>%   
 as.data.frame() %>%   
 rownames\_to\_column("Sample") %>%   
 inner\_join(metadata %>%   
 filter(Sample < 73) %>%   
 mutate(Sample = str\_c("Sample\_",Sample)) %>%   
 mutate(Combinada = str\_c(`Cambio climático`,"+",Biocidas))) %>%   
 dplyr::select(Sample,Subparcela:Combinada) %>% column\_to\_rownames("Sample")

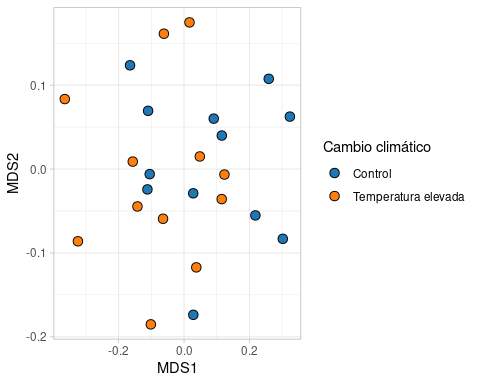
Joining with `by = join\_by(Sample)`

map <- metaMDS(otu\_table\_sorted)

Square root transformation  
Wisconsin double standardization  
Run 0 stress 0.1601272   
Run 1 stress 0.151888   
... New best solution  
... Procrustes: rmse 0.1174547 max resid 0.3422234   
Run 2 stress 0.1735241   
Run 3 stress 0.1740639   
Run 4 stress 0.1550119   
Run 5 stress 0.1655551   
Run 6 stress 0.164601   
Run 7 stress 0.1807836   
Run 8 stress 0.1648439   
Run 9 stress 0.1740658   
Run 10 stress 0.1819888   
Run 11 stress 0.1610556   
Run 12 stress 0.1673521   
Run 13 stress 0.1714218   
Run 14 stress 0.163355   
Run 15 stress 0.1731077   
Run 16 stress 0.163355   
Run 17 stress 0.1709845   
Run 18 stress 0.1649024   
Run 19 stress 0.173817   
Run 20 stress 0.1673524   
\*\*\* Best solution was not repeated -- monoMDS stopping criteria:  
 18: stress ratio > sratmax  
 2: scale factor of the gradient < sfgrmin

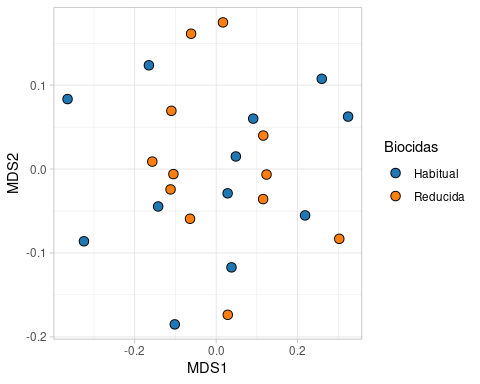
map$points %>%   
 as.data.frame() %>%   
 rownames\_to\_column("Sample") %>%   
 inner\_join(metadata %>%   
 mutate(Sample = str\_c("Sample\_",Sample)) %>%   
 mutate(Combinada = str\_c(`Cambio climático`,"+",Biocidas))) %>%   
 ggplot(aes(x = MDS1, y = MDS2, fill = `Cambio climático`)) +   
 geom\_point(shape = 21, size =3) +  
 theme\_light() +  
 scale\_fill\_d3()

Joining with `by = join\_by(Sample)`



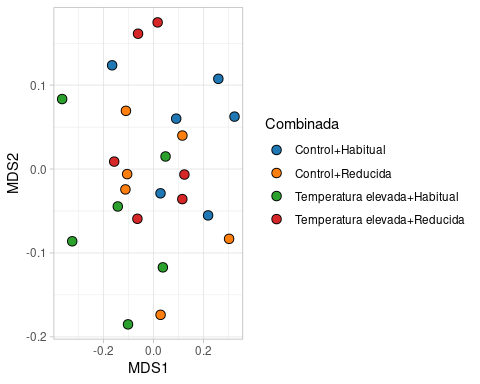
map$points %>%   
 as.data.frame() %>%   
 rownames\_to\_column("Sample") %>%   
 inner\_join(metadata %>%   
 mutate(Sample = str\_c("Sample\_",Sample)) %>%   
 mutate(Combinada = str\_c(`Cambio climático`,"+",Biocidas))) %>%   
 ggplot(aes(x = MDS1, y = MDS2, fill = Biocidas)) +   
 geom\_point(shape = 21, size =3) +  
 theme\_light() +  
 scale\_fill\_d3()

Joining with `by = join\_by(Sample)`



map$points %>%   
 as.data.frame() %>%   
 rownames\_to\_column("Sample") %>%   
 inner\_join(metadata %>%   
 mutate(Sample = str\_c("Sample\_",Sample)) %>%   
 mutate(Combinada = str\_c(`Cambio climático`,"+",Biocidas))) %>%   
 ggplot(aes(x = MDS1, y = MDS2, fill = Combinada)) +   
 geom\_point(shape = 21, size =3) +  
 theme\_light() +  
 scale\_fill\_d3()

Joining with `by = join\_by(Sample)`



adonis2(otu\_table\_sorted ~ Biocidas \* `Cambio climático`, metadata\_sorted,permutations = 5000)

Permutation test for adonis under reduced model  
Terms added sequentially (first to last)  
Permutation: free  
Number of permutations: 5000  
  
adonis2(formula = otu\_table\_sorted ~ Biocidas \* `Cambio climático`, data = metadata\_sorted, permutations = 5000)  
 Df SumOfSqs R2 F Pr(>F)   
Biocidas 1 0.008211 0.04147 1.0051 0.39972   
`Cambio climático` 1 0.008193 0.04139 1.0030 0.40032   
Biocidas:`Cambio climático` 1 0.018194 0.09190 2.2272 0.04419 \*  
Residual 20 0.163378 0.82524   
Total 23 0.197976 1.00000   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

adonis2(otu\_table\_sorted ~ Biocidas + `Cambio climático` + Combinada, metadata\_sorted, permutations = 5000)

Permutation test for adonis under reduced model  
Terms added sequentially (first to last)  
Permutation: free  
Number of permutations: 5000  
  
adonis2(formula = otu\_table\_sorted ~ Biocidas + `Cambio climático` + Combinada, data = metadata\_sorted, permutations = 5000)  
 Df SumOfSqs R2 F Pr(>F)   
Biocidas 1 0.008211 0.04147 1.0051 0.40732   
`Cambio climático` 1 0.008193 0.04139 1.0030 0.40512   
Combinada 1 0.018194 0.09190 2.2272 0.04699 \*  
Residual 20 0.163378 0.82524   
Total 23 0.197976 1.00000   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Differential analysis

test\_table <- rare\_table %>%   
 as.data.frame() %>%   
 rownames\_to\_column("Sample") %>%   
 pivot\_longer(names\_to = "Gene", values\_to = "Abundance", -Sample) %>%   
 inner\_join(metadata %>%   
 filter(Sample < 73) %>%   
 mutate(Sample = str\_c("Sample\_",Sample)) %>%   
 mutate(Combinada = str\_c(`Cambio climático`,"+",Biocidas))) %>%   
 dplyr::select(-Otros) %>%   
 group\_by(Sample) %>%   
 mutate(TotalAbundance = sum(Abundance)) %>%   
 ungroup() %>%   
 mutate(Freq = (Abundance +1) /TotalAbundance) %>%   
 mutate(Ab2 = Abundance +1) %>%   
 group\_by(Gene) %>%   
 mutate(NZeros = sum(Abundance ==0)) %>%   
 filter(NZeros < 10) %>%   
 ungroup() %>%   
 nest\_by(Gene) %>%   
 mutate(model\_full = list(gamlss(Freq ~ Combinada,   
 data = data,  
 family = "BE",   
 control = gamlss.control(n.cyc = 200, trace = F)))) %>%   
 mutate(model\_reduced = list(gamlss(Freq ~ 1,   
 data = data,  
 family = "BE",   
 control = gamlss.control(n.cyc = 200, trace = F)))) %>%  
 mutate(LRT = list(test\_lrt(model\_full,model\_reduced)))

Joining with `by = join\_by(Sample)`

Warning: There were 6 warnings in `mutate()`.  
The first warning was:  
ℹ In argument: `LRT = list(test\_lrt(model\_full, model\_reduced))`.  
ℹ In row 1.  
Caused by warning:  
! reiniciar evaluación premisa interrumpida  
ℹ Run `dplyr::last\_dplyr\_warnings()` to see the 5 remaining warnings.

test\_table %>%   
 dplyr::select(Gene,LRT) %>%   
 unnest(LRT) %>%   
 drop\_na() %>%   
 mutate(padj = p.adjust(p,method = "fdr")) %>%   
 filter(padj < 0.05)

# A tibble: 10 × 8  
# Groups: Gene [10]  
 Gene Name Model df df\_diff Chi2 p padj  
 <chr> <chr> <chr> <int> <int> <dbl> <dbl> <dbl>  
 1 aadA13\_1\_AY713504 model\_re… gaml… 2 -3 15.9 0.00118 0.00118  
 2 aadA2\_1\_NC\_010870 model\_re… gaml… 2 -3 11.3 0.0102 0.0102   
 3 aph(3')-III\_1\_M26832 model\_re… gaml… 2 -3 9.27 0.0259 0.0259   
 4 blaOXA-347\_1\_ACWG01000053 model\_re… gaml… 2 -3 11.7 0.00845 0.00845  
 5 blaZ\_12\_M15195 model\_re… gaml… 2 -3 11.3 0.0101 0.0101   
 6 catB10\_1\_AJ878850 model\_re… gaml… 2 -3 7.98 0.0465 0.0465   
 7 mef(A)\_1\_AJ971089 model\_re… gaml… 2 -3 11.7 0.00866 0.00866  
 8 mef(A)\_3\_AF227520 model\_re… gaml… 2 -3 9.47 0.0237 0.0237   
 9 mef(A)\_4\_HG423652 model\_re… gaml… 2 -3 8.71 0.0334 0.0334   
10 msr(D)\_2\_AF274302 model\_re… gaml… 2 -3 16.1 0.00107 0.00107

significativos <- test\_table %>%   
 dplyr::select(Gene,LRT) %>%   
 unnest(LRT) %>%   
 drop\_na() %>%   
 mutate(padj = p.adjust(p,method = "fdr")) %>%   
 filter(padj < 0.05) %>% pull(Gene)  
  
test\_table %>%   
 filter(Gene %in% significativos) %>%   
 dplyr::select(Gene,data) %>%   
 unnest(data) %>%   
 ggplot(aes(x= Combinada, y = Abundance, fill = Combinada)) +   
 geom\_boxplot() +  
 geom\_jitter(shape=21) +  
 facet\_wrap(~Gene, scales = "free") +  
 scale\_fill\_d3() +   
 theme(axis.text.x = element\_text(angle = 25, hjust = 1))

