The coral Oculina patagonica holobiont and its response to confinement, temperature and Vibrio infections: Statistical analysis

A.-B. Martin-Cuadrado, E. Rubio-Portillo, F. Rosselló, J. Antón

2024-08-07 08:38:56

```
rm(list = ls())
library(readxl)
library(tidyverse)
library(stats)
library(kableExtra)
library(DescTools)
library(ecolTest)
library(vegan)
```

For reproducibility, we fix a random seed

```
#initial_seed=as.integer(Sys.time())
#print(initial_seed)
## [1] 1722266529
#initial_seed%%10000
## [1] 6529

set.seed(6529)
```

```
Datos.Muestras.Generos=data.frame(read_excel("tablas_taxonomy.xlsx",sheet="genu
s"))
Datos.Muestras.Familias=data.frame(read_excel("tablas_taxonomy.xlsx",sheet="famil
y"))

Samples=names(Datos.Muestras.Generos)[-1]
Generos=Datos.Muestras.Generos[,1]
Familias=Datos.Muestras.Familias[,1]

DFG=t(Datos.Muestras.Generos[,-1])
colnames(DFG)=Generos
DFG.prop=DFG/rowSums(DFG)

DFF=t(Datos.Muestras.Familias[,-1])
colnames(DFF)=Familias
DFF.prop=DFF/rowSums(DFF)


TamañosG=rowSums(DFG)
TamañosF=rowSums(DFF)

rm(Datos.Muestras.Generos)
rm(Datos.Muestras.Familias)
```

Which samples are significantly different?

Through Bray-Curtis distances

We want to compare the following pairs of samples:

- oculina with c20 and c28
- c20 with ms20, mx20 and i20
- c28 with ms28, mx28 and i28

To do that we use a resampling approach:

- For each pair of samples:
    - We combine them into a single sample.
    - We repeat 1000 times the process of separating that sample into two random samples of the same size as the original ones and calculating the B-C distance (between proportions) of these two samples.
    - We use the fraction of these distances that are greater than or equal to the distance between the two original samples as an estimate of the p-value for the contrast with null hypothesis, “Both samples come from the same community” and alternative hypothesis “The two samples come from different communities.”
- In all cases, we find 0 resamplings that yield a B-C distance greater than or equal to the original. This provides statistically significant evidence that the B-C distance between the original samples is greater than if the two were from the same community.

0 out of 1000 gives an estimate for the p-value of <0.001. After adjusting these p-values using Bonferroni separately for genera and families (tat is, multiplying them by 8, the only adjustement that makes sense when we don't have the actual values) they remain statistically significant: <0.01.

- Actually, in all cases the original B-C distance is two orders of magnitude greater than the simulated ones, except for c20 vs mx20 and ms20, where the original B-C distance is one order of magnitude greater than the simulated ones

```
# X1 and X2 the original samples
Simulacion=function(X1,X2,taxa){
XX=c(X1,X2)
n1=length(X1)
ind=sample(length(XX),n1,replace=FALSE)
Y1=factor(XX[ind],levels=taxa)
Y2=factor(XX[-ind],levels=taxa)
Y1.p=prop.table(table(Y1))
Y2.p=prop.table(table(Y2))
YY=rbind(Y1.p,Y2.p)
vegan::vegdist(YY, method="bray")
}
```

Genera

```
oculina.muestra=factor(rep(colnames(DFG),DFG[1,]),levels=colnames(DFG))
c20.muestra=factor(rep(colnames(DFG),DFG[2,]),levels=colnames(DFG))
c28.muestra=factor(rep(colnames(DFG),DFG[3,]),levels=colnames(DFG))
i20.muestra=factor(rep(colnames(DFG),DFG[4,]),levels=colnames(DFG))
i28.muestra=factor(rep(colnames(DFG),DFG[5,]),levels=colnames(DFG))
ms20.muestra=factor(rep(colnames(DFG),DFG[6,]),levels=colnames(DFG))
ms28.muestra=factor(rep(colnames(DFG),DFG[7,]),levels=colnames(DFG))
mx20.muestra=factor(rep(colnames(DFG),DFG[8,]),levels=colnames(DFG))
mx28.muestra=factor(rep(colnames(DFG),DFG[9,]),levels=colnames(DFG))
Tamaños=TamañosG
```

```
Sim.base.c20=replicate(1000,Simulacion(oculina.muestra,c20.muestra,Generos))
Sim.base.c28=replicate(1000,Simulacion(oculina.muestra,c28.muestra,Generos))

Sim.c20.ms20=replicate(1000,Simulacion(c20.muestra,ms20.muestra,Generos))
Sim.c20.i20=replicate(1000,Simulacion(c20.muestra,i20.muestra,Generos))
Sim.c20.mx20=replicate(1000,Simulacion(c20.muestra,mx20.muestra,Generos))

Sim.c28.ms28=replicate(1000,Simulacion(c28.muestra,ms28.muestra,Generos))
Sim.c28.i28=replicate(1000,Simulacion(c28.muestra,i28.muestra,Generos))
Sim.c28.mx28=replicate(1000,Simulacion(c28.muestra,mx28.muestra,Generos))
```

```
BC.base.c20=vegan::vegdist(DFG.prop[1:2,], method="bray")
BC.base.c28=vegan::vegdist(DFG.prop[c(1,3),], method="bray")

BC.c20.ms20=vegan::vegdist(DFG.prop[c(2,6),], method="bray")
BC.c20.i20=vegan::vegdist(DFG.prop[c(2,4),], method="bray")
BC.c20.mx20=vegan::vegdist(DFG.prop[c(2,8),], method="bray")

BC.c28.ms28=vegan::vegdist(DFG.prop[c(3,7),], method="bray")
BC.c28.i28=vegan::vegdist(DFG.prop[c(3,5),], method="bray")
BC.c28.mx28=vegan::vegdist(DFG.prop[c(3,9),], method="bray")
```

The p-values estimations:

```
length(which(Sim.base.c20>=BC.base.c20))
```

```
## [1] 0
```

```
length(which(Sim.base.c28>=BC.base.c28))
```

```
## [1] 0
```

```
length(which(Sim.c20.ms20>=BC.c20.ms20))
```

```
## [1] 0
```

```
length(which(Sim.c20.i20>=BC.c20.i20))
```

```
## [1] 0
```

```
length(which(Sim.c20.mx20>=BC.c20.mx20))
```

```
## [1] 0
```

```
length(which(Sim.c28.ms28>=BC.c28.ms28))
```

```
## [1] 0
```

```
length(which(Sim.c28.i28>=BC.c28.i28))
```

```
## [1] 0
```

```
length(which(Sim.c28.mx28>=BC.c28.mx28))
```

```
## [1] 0
```

In the following table, for each comparison, the B-C distance between original samples and the maximum and average BC distances between samples from a simulation:

```
max.sim=data.frame(
comparaciones=c("oculina vs c20",
"oculina vs c28",
"c20 vs ms20",
"c20 vs i20",
"c20 vs mx20",
"c28 vs ms28",
"c28 vs i28",
"c28 vs mx28"),
originals=round(c(BC.base.c20,BC.base.c28,BC.c20.ms20,BC.c20.i20,BC.c20.mx20,BC.c2
8.ms28,BC.c28.i28,BC.c28.mx28),4)
,
maximos=round(c(max(Sim.base.c20),
max(Sim.base.c28),
max(Sim.c20.ms20),
max(Sim.c20.i20),
max(Sim.c20.mx20),
max(Sim.c28.ms28),
max(Sim.c28.i28),
max(Sim.c28.mx28)),4),
medias=round(c(mean(Sim.base.c20),
mean(Sim.base.c28),
mean(Sim.c20.ms20),
mean(Sim.c20.i20),
mean(Sim.c20.mx20),
mean(Sim.c28.ms28),
mean(Sim.c28.i28),
mean(Sim.c28.mx28)),4)
)

names(max.sim)=c("Comparison","Original B-C","Largest B-C in simulation","Average
B-C in simulation")

max.sim %>%
  kbl() %>%
  kable_styling()
```

Comparison	Original B-C	Largest B-C in simulation	Average B-C in simulation
oculina vs c20	0.1064	0.0028	0.0025
oculina vs c28	0.3180	0.0035	0.0030
c20 vs ms20	0.0436	0.0028	0.0024
c20 vs i20	0.1225	0.0030	0.0026

c20 vs mx20	0.0311	0.0028	0.0023
c28 vs ms28	0.2164	0.0038	0.0033
c28 vs i28	0.1579	0.0039	0.0033
c28 vs mx28	0.1053	0.0040	0.0035

Families

```
oculina.muestra.F=factor(rep(colnames(DFF),DFF[1,]),levels=colnames(DFF))
c20.muestra.F=factor(rep(colnames(DFF),DFF[2,]),levels=colnames(DFF))
c28.muestra.F=factor(rep(colnames(DFF),DFF[3,]),levels=colnames(DFF))
i20.muestra.F=factor(rep(colnames(DFF),DFF[4,]),levels=colnames(DFF))
i28.muestra.F=factor(rep(colnames(DFF),DFF[5,]),levels=colnames(DFF))
ms20.muestra.F=factor(rep(colnames(DFF),DFF[6,]),levels=colnames(DFF))
ms28.muestra.F=factor(rep(colnames(DFF),DFF[7,]),levels=colnames(DFF))
mx20.muestra.F=factor(rep(colnames(DFF),DFF[8,]),levels=colnames(DFF))
mx28.muestra.F=factor(rep(colnames(DFF),DFF[9,]),levels=colnames(DFF))
Tamaños=TamañosF
```

```
Sim.base.c20.F=replicate(1000,Simulacion(oculina.muestra.F,c20.muestra.F,Familia
s))
Sim.base.c28.F=replicate(1000,Simulacion(oculina.muestra.F,c28.muestra.F,Familia
s))

Sim.c20.ms20.F=replicate(1000,Simulacion(c20.muestra.F,ms20.muestra.F,Familias))
Sim.c20.i20.F=replicate(1000,Simulacion(c20.muestra.F,i20.muestra.F,Familias))
Sim.c20.mx20.F=replicate(1000,Simulacion(c20.muestra.F,mx20.muestra.F,Familias))

Sim.c28.ms28.F=replicate(1000,Simulacion(c28.muestra.F,ms28.muestra.F,Familias))
Sim.c28.i28.F=replicate(1000,Simulacion(c28.muestra.F,i28.muestra.F,Familias))
Sim.c28.mx28.F=replicate(1000,Simulacion(c28.muestra.F,mx28.muestra.F,Familias))
```

```
BC.base.c20.F=vegan::vegdist(DFF.prop[1:2,], method="bray")
BC.base.c28.F=vegan::vegdist(DFF.prop[c(1,3),], method="bray")

BC.c20.ms20.F=vegan::vegdist(DFF.prop[c(2,6),], method="bray")
BC.c20.i20.F=vegan::vegdist(DFF.prop[c(2,4),], method="bray")
BC.c20.mx20.F=vegan::vegdist(DFF.prop[c(2,8),], method="bray")

BC.c28.ms28.F=vegan::vegdist(DFF.prop[c(3,7),], method="bray")
BC.c28.i28.F=vegan::vegdist(DFF.prop[c(3,5),], method="bray")
BC.c28.mx28.F=vegan::vegdist(DFF.prop[c(3,9),], method="bray")
```

The p-values estimations:

```
length(which(Sim.base.c20.F>=BC.base.c20.F))
```

```
## [1] 0
```

```
length(which(Sim.base.c28.F>=BC.base.c28.F))
```

```
## [1] 0
```

```
length(which(Sim.c20.ms20.F>=BC.c20.ms20.F))
```

```
## [1] 0
```

```
length(which(Sim.c20.i20.F>=BC.c20.i20.F))
```

```
## [1] 0
```

```
length(which(Sim.c20.mx20.F>=BC.c20.mx20.F))
```

```
## [1] 0
```

```
length(which(Sim.c28.ms28.F>=BC.c28.ms28.F))
```

```
## [1] 0
```

```
length(which(Sim.c28.i28.F>=BC.c28.i28.F))
```

```
## [1] 0
```

```
length(which(Sim.c28.mx28.F>=BC.c28.mx28.F))
```

```
## [1] 0
```

The table

```
max.sim=data.frame(
comparaciones=c("oculina vs c20",
"oculina vs c28",
"c20 vs ms20",
"c20 vs i20",
"c20 vs mx20",
"c28 vs ms28",
"c28 vs i28",
"c28 vs mx28"),
originals=round(c(BC.base.c20.F,BC.base.c28.F,BC.c20.ms20.F,BC.c20.i20.F,BC.c20.mx
20.F,BC.c28.ms28.F,BC.c28.i28.F,BC.c28.mx28.F),4)
,
maximos=round(c(max(Sim.base.c20.F),
max(Sim.base.c28.F),
max(Sim.c20.ms20.F),
max(Sim.c20.i20.F),
max(Sim.c20.mx20.F),
max(Sim.c28.ms28.F),
max(Sim.c28.i28.F),
max(Sim.c28.mx28.F)),4),
medias=round(c(mean(Sim.base.c20.F),
mean(Sim.base.c28.F),
mean(Sim.c20.ms20.F),
mean(Sim.c20.i20.F),
mean(Sim.c20.mx20.F),
mean(Sim.c28.ms28.F),
mean(Sim.c28.i28.F),
mean(Sim.c28.mx28.F)),4)
)

names(max.sim)=c("Comparison","Original B-C","Largest B-C in simulation","Average
B-C in simulation")

max.sim %>%
  kbl() %>%
  kable_styling()
```

Comparison	Original B-C	Largest B-C in simulation	Average B-C in simulation
oculina vs c20	0.1036	0.0017	0.0013
oculina vs c28	0.3152	0.0019	0.0015
c20 vs ms20	0.0417	0.0017	0.0013
c20 vs i20	0.1156	0.0018	0.0014
c20 vs mx20	0.0293	0.0016	0.0012
c28 vs ms28	0.2122	0.0021	0.0016
c28 vs i28	0.1357	0.0022	0.0016

c28 vs mx28	0.1050	0.0023	0.0017

Biodiversities

For each sample and for each biodiversity index (Gini-Simpson, Shannon)

- We compute the corresponding index and an estimation of its standard deviation (sd)
- We compute the interval (index-3sd,index+3sd). The reason for the factor 3 is that there are 36 pairs of samples to compare, and hence we use as factor for sd the $0.95^{1/36}$ quantile of the standard normal distribution. In this way, disjoint intervals give statistical evidence (at the 5% signification level) that the corresponding two indices are different.
- Moreover, for the Shannon index we perform pairwise Hutcheson t-tests: see below

```
# Gini-Simpson

GS=function(x){
N=sum(x)
1-sum(x*(x-1))/(N*(N-1))
}

pT=function(x){
N=sum(x)
sum(x*(x-1)*(x-2))/(N*(N-1)*(N-2))
}

#Variance unbiased estimation of A. Tiffeas-Mayer (arXiv:2310.03439)
Var.GS=function(x){
N=sum(x)
a=4*(N-2)/(N*(N-1))
b=2*(2*N-3)/(N*(N-1))
c=2/(N*(N-1))
(a/(1-b))*pT(x)-(b/(1-b))*(1-GS(x))^2+(c/(1-b))*(1-GS(x))
}

# Shannon
SH=function(x){
  vegan::diversity(x,index = "shannon")
}


p.lnp2=function(p){
  if (p==0){return(0)}
  else {
    return(p*(log(p)^2))
  }
}

# Hutcheson estimator
Var.SH=function(x){
N=sum(x)
y=sapply(x/sum(x),FUN=p.lnp2)
bit=sapply(x,FUN=function(x){min(x,1)})
S=sum(bit)
vv=(sum(y)-SH(x)^2)/N+(S-1)/(2*N^2)
return(vv)
}
```

Genera

Gini-Simpson

```
GS.Generos=data.frame(
  GS=apply(DFG,MARGIN=1,GS),
  Varianza=apply(DFG,MARGIN=1,Var.GS))
```

```
Resultados.GS.G=data.frame(Muestras=Samples,IGS=GS.Generos$GS,SD=sqrt(GS.Generos$V
arianza),ICEinf=GS.Generos$GS-3*sqrt(GS.Generos$Varianza),ICEsup=GS.Generos$GS+3*s
qrt(GS.Generos$Varianza)
                         )
Resultados.GS.G[,2:5]=round(Resultados.GS.G[,2:5],4)
names(Resultados.GS.G)=c("Samples","Gini-Simpson","SD","GS-3SD","GS+3SD")

Resultados.GS.G %>%
  kbl() %>%
  kable_styling()
```

Samples	Gini-Simpson	SD	GS-3SD	GS+3SD
oculina	0.2901	0.0002	0.2895	0.2906
c20	0.3921	0.0002	0.3915	0.3927
c28	0.6459	0.0001	0.6454	0.6463
l20	0.5210	0.0002	0.5205	0.5215
l28	0.7267	0.0001	0.7264	0.7271
MS20	0.4293	0.0002	0.4287	0.4298
MS28	0.7661	0.0001	0.7658	0.7664
MX20	0.3943	0.0002	0.3938	0.3949
MX28	0.6185	0.0002	0.6179	0.6190

All intervals are pairwise disjoint.

Shannon

```
SH.Generos=data.frame(
  SH=apply(DFG,MARGIN=1,SH),
  Varianza=apply(DFG,MARGIN=1,Var.SH))
```

```
Resultados.SH.G=data.frame(Muestras=Samples,IGS=SH.Generos$SH,SD=sqrt(SH.Generos$V
arianza),ICEinf=SH.Generos$SH-3*sqrt(SH.Generos$Varianza),ICEsup=SH.Generos$SH+3*s
qrt(SH.Generos$Varianza)
                         )
Resultados.SH.G[,2:5]=round(Resultados.SH.G[,2:5],4)
names(Resultados.SH.G)=c("Samples","Shannon","SD","SH-3SD","SH+3SD")

Resultados.SH.G %>%
  kbl() %>%
  kable_styling()
```

Samples	Shannon	SD	SH-3SD	SH+3SD
oculina	1.2099	0.0009	1.2073	1.2125
c20	1.3379	0.0008	1.3355	1.3403
c28	2.2987	0.0009	2.2959	2.3015
I20	1.7588	0.0009	1.7562	1.7614
I28	2.5600	0.0009	2.5573	2.5627
MS20	1.4184	0.0008	1.4159	1.4208
MS28	2.6903	0.0009	2.6876	2.6930
MX20	1.2594	0.0008	1.2571	1.2617
MX28	2.1778	0.0011	2.1745	2.1810

All intervals are pairwise disjoint.

We also use Hutcheson t-test to decide whether Shannon indices are significantly different. We perform pairwise comparisons and adjust the p-values using the Bonferroni method. All p-values are 0 (up to 16 decimal places), which is of course consistent with the previous table..

```
pvalSH.G=matrix(NA,nrow=9,ncol=9)
for (i in 1:8){
  for (j in (i+1):9){
   pvalSH.G[i,j]= 36*Hutcheson_t_test(DFG[i,],DFG[j,])$p.value
  }
}
colnames(pvalSH.G)=Samples
row.names(pvalSH.G)=Samples

pvalSH.G %>%
  kbl() %>%
  kable_styling()
```

	oculina	c20	c28	I20	I28	MS20	MS28	MX20	MX28
oculina	0	0	0	0	0	0	0	0	0
c20		0	0	0	0	0	0	0	0
c28			0	0	0	0	0	0	0
I20				0	0	0	0	0	0
I28					0	0	0	0	0

MS20	0	0	0
MS28		0	0
MX20			0
MX28			

Families

Gini-simpson

```
GS.Familias=data.frame(
  GS=apply(DFF,MARGIN=1,GS),
  Varianza=apply(DFF,MARGIN=1,Var.GS))
```

```
Resultados.GS.F=data.frame(Muestras=Samples,IGS=GS.Familias$GS,SD=sqrt(GS.Familias
$Varianza),ICEinf=GS.Familias$GS-3*sqrt(GS.Familias$Varianza),ICEsup=GS.Familias$G
S+3*sqrt(GS.Familias$Varianza)
                         )
Resultados.GS.F[,2:5]=round(Resultados.GS.F[,2:5],4)
names(Resultados.GS.F)=c("Samples","Gini-Simpson","SD","GS-3SD","GS+3SD")

Resultados.GS.F %>%
  kbl() %>%
  kable_styling()
```

Samples	Gini-Simpson	SD	GS-3SD	GS+3SD
oculina	0.2906	0.0002	0.2900	0.2911
c20	0.3922	0.0002	0.3916	0.3928
c28	0.6573	0.0002	0.6568	0.6577
l20	0.5268	0.0002	0.5263	0.5273
l28	0.7445	0.0001	0.7442	0.7449
MS20	0.4301	0.0002	0.4295	0.4306
MS28	0.7719	0.0001	0.7717	0.7722
MX20	0.3946	0.0002	0.3941	0.3952
MX28	0.6314	0.0002	0.6308	0.6319

Shannon

```
SH.Familias=data.frame(
  SH=apply(DFF,MARGIN=1,SH),
  Varianza=apply(DFF,MARGIN=1,Var.SH))
```

```
Resultados.SH.F=data.frame(Muestras=Samples,IGS=SH.Familias$SH,SD=sqrt(SH.Familias
$Varianza),ICEinf=SH.Familias$SH-3*sqrt(SH.Familias$Varianza),ICEsup=SH.Familias$S
H+3*sqrt(SH.Familias$Varianza)
                          )
Resultados.SH.F[,2:5]=round(Resultados.SH.F[,2:5],4)
names(Resultados.SH.F)=c("Samples","Shannon","SD","SH-3SD","SH+3SD")

Resultados.SH.F %>%
  kbl() %>%
  kable_styling()
```

Samples	Shannon	SD	SH-3SD	SH+3SD
oculina	1.1106	0.0008	1.1083	1.1129
c20	1.2356	0.0007	1.2335	1.2378
c28	2.0382	0.0007	2.0360	2.0404
I20	1.6492	0.0007	1.6470	1.6515
I28	2.2711	0.0007	2.2691	2.2732
MS20	1.3206	0.0007	1.3185	1.3228
MS28	2.2841	0.0007	2.2821	2.2861
MX20	1.1769	0.0007	1.1749	1.1790
MX28	1.9971	0.0009	1.9945	1.9998

```
pvalSH.F=matrix(NA,nrow=9,ncol=9)
for (i in 1:8){
  for (j in (i+1):9){
   pvalSH.F[i,j]= 36*Hutcheson_t_test(DFF[i,],DFF[j,])$p.value
  }
}
colnames(pvalSH.F)=Samples
row.names(pvalSH.F)=Samples

pvalSH.F %>%
  kbl() %>%
  kable_styling()
```

	oculina	c20	c28	I20	I28	MS20	MS28	MX20	MX28

oculina	0	0	0	0	0	0	0	0
c20		0	0	0	0	0	0	0
c28			0	0	0	0	0	0
I20				0	0	0	0	0
I28					0	0	0	0
MS20						0	0	0
MS28							0	0
MX20								0
MX28								

Which genera have statistically significant different proportions?

- For each pair of samples of interest and for each genus, we perform a proportion comparison test:
    - If it is theoretically sound (5 or more hits), a parametric chi-square test.
    - Otherwise, a Monte Carlo chi-square test.
- We also calculate a 95% confidence interval for the ratio of proportions of each genus using the appropriate technique (the normal approximation if the parametric chi-square test could be used, bootstrap if not).
- Next, we adjust the p-values for each pair of samples using the Benjamini-Yekutieli method, which is the best for controlling the false discovery rate when the variables (in this case, the proportions) are not independent. We round the adjusted p-values to 4 decimal places. So, 0 means < 0.00005.
- For each pair of samples, we consider that a genus represents significantly different proportions in them when its adjusted p-value is <0.05.
- We generate a CSV file for each pair of samples that contains, for each genus, the ratio of proportions (for example, p_oculina/p_c20 means the ratio between the proportion in the oculina sample and the sample c20), the adjusted p-value, and the 95% confidence interval.

  We show in this report only the genera of interest mentioned in the paper.

```
Test.Props=function(E,conf.level=0.95){
  n=rowSums(E)
  e=E[,2]
  RR=RelRisk(E[2:1,2:1],conf.level=conf.level)[1]
  #
  if (any(rowSums(E)==0)){
    Res=rbind(c(NA,NA,NA,NA,1))
  } else{
  if (any(E < 5)|any(n<50)){
    p=E[,2]/rowSums(E)
    PVAL=chisq.test(E, simulate.p.value=TRUE,B=5000)$p.value
Sim0=function(E){
  n=rowSums(E)
e.sim=c(rbinom(1,n[1],p[1]),
          rbinom(1,n[2],p[2]))
  p.sim=e.sim/n
  p.sim[2]/p.sim[1]
}
X0=replicate(10000,Sim0(E))
IC=c(quantile(X0,(1-conf.level)/2,na.rm=TRUE),quantile(X0,1-(1-conf.level)/2,na.rm
=TRUE))
  }  else {
    PVAL=chisq.test(E)$p.value
    IC=RelRisk(E[2:1,2:1],conf.level=0.95)[2:3]
  }
      Res=rbind(c(RR,PVAL,IC))
}
 return(Res)
}

Generos.Sig=function(x,y,etiqueta){
DF=DFG[c(x,y),]
n=rowSums(DF)
Res=c()
for (i in 1:dim(DF)[2]){
E=matrix(c(n[1]-DF[1,i],DF[1,i],n[2]-DF[2,i],DF[2,i]),nrow=2,byrow=TRUE)
Res=rbind(Res,Test.Props(E))
}
Resultado=data.frame(Género=Generos,RR=Res[,1],p.val=Res[,2],IC1=Res[,3],IC2=Res[,
4])

Resultado[,c(2,4,5)]=round(Resultado[,c(2,4,5)],3)

Resultado[,6]=round(p.adjust(Resultado[,3],method="BY"),4)
Resultado=Resultado[,c(1,2,6,4,5)]
names(Resultado)=c("Genus", etiqueta, "adjusted p-value", "95% CI lower end", "95%
CI upper end")
return(Resultado)
}
```

```
Interesting=c("Aspergillus",
"Fusarium",
"Saccharomycodes",
"Rhizophagus",
"Symbiodinium",
"Nitrosopumilus",
"Streptomyces",
"Acinetobacter",
"Bordetella",
"Salmonella",
"Neisseria",
"Klebsiella",
"Desulfovibrio",
"Halodesulfovibrio",
"Loktanella",
"Marinifilum",
"Marivita",
"Roseobacter",
"Ruegeria",
"Yoonia",
"Vibrio")
```

oculina vs c20

```
Resultado.sig=Generos.Sig(1,2,"P_c20/P_oculina")
write.csv2(Resultado.sig,"Generos_Base_vs_c20.csv",row.names=FALSE)
```

There are 1839 genera with significantly different proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	P_c20/P_oculina	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	0.413	0.0000	0.399	0.427
Fusarium	0.521	0.0000	0.494	0.549
Saccharomycodes	203.138	0.0000	168.901	244.315
Rhizophagus	0.512	0.0000	0.490	0.535
Symbiodinium	0.454	0.0000	0.448	0.459

Nitrosopumilus	1.140	0.0000	1.097	1.185
Streptomyces	0.584	0.0000	0.574	0.595
Acinetobacter	11.194	0.0000	10.899	11.496
Bordetella	21.951	0.0000	20.455	23.556
Salmonella	13.049	0.0000	12.070	14.108
Neisseria	24.010	0.0000	22.222	25.942
Klebsiella	4.407	0.0000	4.225	4.597
Desulfovibrio	1.573	0.0000	1.487	1.664
Halodesulfovibrio	72.502	0.0000	61.376	85.646
Loktanella	1.307	0.0612	1.101	1.552
Marinifilum	20.861	0.0000	19.415	22.415
Marivita	43.882	0.0000	35.818	53.761
Roseobacter	1.450	0.0000	1.251	1.681
Ruegeria	5.253	0.0000	4.975	5.547
Yoonia	1.683	0.0000	1.388	2.041
Vibrio	0.711	0.0000	0.686	0.736

oculina vs c28

```
Resultado.sig=Generos.Sig(1,3,"P_c28/P_oculina")
write.csv2(Resultado.sig,"Generos_Base_vs_c28.csv",row.names=FALSE)
```

There are 2282 genera with significantly different proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	P_c28/P_oculina	adjusted p-value	95% CI lower end	95% CI upper end

Aspergillus	0.344	0	0.332	0.356
Fusarium	0.381	0	0.359	0.404
Saccharomycodes	229.591	0	190.906	276.116
Rhizophagus	0.241	0	0.227	0.255
Symbiodinium	0.257	0	0.253	0.261
Nitrosopumilus	1.123	0	1.080	1.168
Streptomyces	0.803	0	0.790	0.816
Acinetobacter	6.050	0	5.885	6.219
Bordetella	27.688	0	25.809	29.703
Salmonella	7.420	0	6.848	8.038
Neisseria	10.107	0	9.335	10.944
Klebsiella	6.439	0	6.181	6.708
Desulfovibrio	7.288	0	6.956	7.637
Halodesulfovibrio	568.200	0	481.483	670.536
Loktanella	130.717	0	114.856	148.770
Marinifilum	46.121	0	42.963	49.511
Marivita	3082.635	0	2521.876	3768.085
Roseobacter	80.882	0	72.168	90.648
Ruegeria	38.810	0	36.897	40.822
Yoonia	160.091	0	137.391	186.542
Vibrio	9.300	0	9.080	9.526

c20 vs i20

```
Resultado.sig=Generos.Sig(2,4,"P_i20/P_c20")
write.csv2(Resultado.sig,"Generos_c20_vs_i20.csv",row.names=FALSE)
```

There are 1409 genera with significantly different proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	P_i20/P_c20	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	1.045	0.6922	1.004	1.087
Fusarium	0.824	0.0000	0.773	0.878
Saccharomycodes	0.749	0.0000	0.735	0.764
Rhizophagus	0.715	0.0000	0.676	0.755
Symbiodinium	0.248	0.0000	0.243	0.254
Nitrosopumilus	1.025	1.0000	0.987	1.064
Streptomyces	1.312	0.0000	1.287	1.337
Acinetobacter	2.185	0.0000	2.165	2.206
Bordetella	0.686	0.0000	0.670	0.702
Salmonella	0.531	0.0000	0.512	0.550
Neisseria	0.379	0.0000	0.368	0.390
Klebsiella	3.259	0.0000	3.192	3.328
Desulfovibrio	4.082	0.0000	3.925	4.245
Halodesulfovibrio	6.751	0.0000	6.611	6.894
Loktanella	0.953	1.0000	0.810	1.121
Marinifilum	2.203	0.0000	2.162	2.244
Marivita	0.109	0.0000	0.099	0.120
Roseobacter	1.175	0.3814	1.033	1.337
Ruegeria	0.599	0.0000	0.578	0.620
Yoonia	0.737	0.0402	0.615	0.884
Vibrio	6.693	0.0000	6.501	6.891

c20 vs ms20

```
Resultado.sig=Generos.Sig(2,6,"P_ms20/P_c20")
write.csv2(Resultado.sig,"Generos_c20_vs_ms20.csv",row.names=FALSE)
```

There are 718 genera with significantly different proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	P_ms20/P_c20	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	1.137	0.0000	1.093	1.181
Fusarium	0.896	0.0589	0.842	0.953
Saccharomycodes	0.857	0.0000	0.841	0.874
Rhizophagus	0.943	1.0000	0.896	0.993
Symbiodinium	0.792	0.0000	0.780	0.804
Nitrosopumilus	0.914	0.0008	0.880	0.950
Streptomyces	0.925	0.0000	0.906	0.944
Acinetobacter	1.113	0.0000	1.102	1.125
Bordetella	0.693	0.0000	0.677	0.709
Salmonella	1.499	0.0000	1.459	1.540
Neisseria	0.779	0.0000	0.761	0.797
Klebsiella	0.991	1.0000	0.965	1.017
Desulfovibrio	0.902	0.0097	0.857	0.949
Halodesulfovibrio	1.155	0.0000	1.125	1.186
Loktanella	0.554	0.0000	0.458	0.669
Marinifilum	1.316	0.0000	1.290	1.344
Marivita	0.111	0.0000	0.101	0.122

Roseobacter	0.890	1.0000	0.775	1.022
Ruegeria	0.413	0.0000	0.397	0.430
Yoonia	0.511	0.0000	0.417	0.626
Vibrio	1.261	0.0000	1.216	1.308

c20 vs mx20

```
Resultado.sig=Generos.Sig(2,8,"P_mx20/P_c20")
write.csv2(Resultado.sig,"Generos_c20_vs_mx20.csv",row.names=FALSE)
```

There are 877 genera with significantly different proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	**P_mx20/P_c20**	**adjusted p-value**	**95% CI lower end**	**95% CI upper end**
Aspergillus	0.974	1.0000	0.935	1.014
Fusarium	0.881	0.0073	0.828	0.938
Saccharomycodes	1.472	0.0000	1.447	1.497
Rhizophagus	1.062	0.8416	1.010	1.116
Symbiodinium	0.585	0.0000	0.576	0.595
Nitrosopumilus	0.090	0.0000	0.082	0.099
Streptomyces	0.913	0.0000	0.895	0.932
Acinetobacter	1.296	0.0000	1.282	1.309
Bordetella	0.616	0.0000	0.602	0.631
Salmonella	0.714	0.0000	0.691	0.737
Neisseria	0.525	0.0000	0.512	0.540
Klebsiella	3.528	0.0000	3.456	3.601

Desulfovibrio	0.480	0.0000	0.451	0.511
Halodesulfovibrio	0.011	0.0000	0.009	0.014
Loktanella	0.577	0.0000	0.479	0.696
Marinifilum	0.133	0.0000	0.127	0.139
Marivita	0.064	0.0000	0.056	0.072
Roseobacter	0.631	0.0000	0.542	0.735
Ruegeria	0.301	0.0000	0.288	0.316
Yoonia	0.569	0.0000	0.468	0.693
Vibrio	0.739	0.0000	0.709	0.771

c28 vs i28

```
Resultado.sig=Generos.Sig(3,5,"P_i28/P_c28")
write.csv2(Resultado.sig,"Generos_c28_vs_i28.csv",row.names=FALSE)
```

There are 2071 genera with significantly different proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	P_i28/P_c28	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	0.784	0	0.748	0.822
Fusarium	0.656	0	0.606	0.710
Saccharomycodes	0.474	0	0.464	0.484
Rhizophagus	0.561	0	0.515	0.613
Symbiodinium	1.353	0	1.330	1.377
Nitrosopumilus	0.096	0	0.088	0.105
Streptomyces	0.719	0	0.706	0.733

Acinetobacter	0.412	0	0.404	0.420
Bordetella	0.279	0	0.271	0.287
Salmonella	1.031	1	0.991	1.071
Neisseria	1.646	0	1.596	1.697
Klebsiella	0.462	0	0.450	0.474
Desulfovibrio	2.260	0	2.216	2.305
Halodesulfovibrio	1.985	0	1.968	2.002
Loktanella	0.186	0	0.181	0.191
Marinifilum	9.449	0	9.346	9.552
Marivita	1.334	0	1.328	1.341
Roseobacter	0.333	0	0.324	0.341
Ruegeria	0.259	0	0.255	0.264
Yoonia	0.168	0	0.163	0.174
Vibrio	0.737	0	0.729	0.746

c28 vs ms28

```
Resultado.sig=Generos.Sig(3,7,"P_ms28/P_c28")
write.csv2(Resultado.sig,"Generos_c28_vs_ms28.csv",row.names=FALSE)
```

There are 1365 genera with significantly different proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	P_ms28/P_c28	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	0.673	0	0.641	0.706
Fusarium	0.696	0	0.644	0.752

Saccharomycodes	0.705	0	0.691	0.718
Rhizophagus	0.387	0	0.350	0.427
Symbiodinium	0.612	0	0.599	0.625
Nitrosopumilus	0.402	0	0.383	0.423
Streptomyces	0.701	0	0.688	0.714
Acinetobacter	0.759	0	0.747	0.771
Bordetella	0.494	0	0.483	0.506
Salmonella	1.341	0	1.293	1.391
Neisseria	1.947	0	1.891	2.006
Klebsiella	0.521	0	0.508	0.535
Desulfovibrio	0.671	0	0.654	0.689
Halodesulfovibrio	0.433	0	0.428	0.439
Loktanella	0.378	0	0.370	0.386
Marinifilum	3.476	0	3.435	3.517
Marivita	3.624	0	3.610	3.639
Roseobacter	0.639	0	0.626	0.652
Ruegeria	1.213	0	1.200	1.227
Yoonia	0.453	0	0.444	0.463
Vibrio	0.915	0	0.905	0.925

c28 vs mx28

```
Resultado.sig=Generos.Sig(3,9,"P_mx28/P_c28")
write.csv2(Resultado.sig,"Generos_c28_vs_mx28.csv",row.names=FALSE)
```

There are 1329 genera with significantly different proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	P_mx28/P_c28	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	0.827	0.0000	0.786	0.871
Fusarium	0.986	1.0000	0.912	1.067
Saccharomycodes	0.396	0.0000	0.386	0.407
Rhizophagus	0.893	0.2888	0.821	0.972
Symbiodinium	2.928	0.0000	2.881	2.975
Nitrosopumilus	0.279	0.0000	0.261	0.298
Streptomyces	0.722	0.0000	0.706	0.737
Acinetobacter	1.896	0.0000	1.870	1.922
Bordetella	0.215	0.0000	0.207	0.223
Salmonella	0.725	0.0000	0.691	0.761
Neisseria	0.726	0.0000	0.697	0.757
Klebsiella	0.674	0.0000	0.656	0.692
Desulfovibrio	0.884	0.0000	0.860	0.907
Halodesulfovibrio	0.728	0.0000	0.719	0.737
Loktanella	1.178	0.0000	1.159	1.199
Marinifilum	3.765	0.0000	3.719	3.811
Marivita	0.007	0.0000	0.006	0.007
Roseobacter	1.206	0.0000	1.184	1.229
Ruegeria	0.217	0.0000	0.212	0.222
Yoonia	1.586	0.0000	1.560	1.613
Vibrio	0.488	0.0000	0.481	0.496

Interactions

- For each pair of pairs of samples of interest and for each genus, we perform a comparison test of the ratios of proportions:
    - When it is theoretically sound, (5 or more hits in all involved samples), the natural test based on lognormal approximations.
    - Otherwise, a bootstrap test.
- We also calculate a 95% confidence interval for the ratio of ratios of proportions of each genus using the technique employed.
- Next, we adjust the p-values for each pair of samples using the Benjamini-Yekutieli method. We round the adjusted p-values to 4 decimal places.
- For each pair of samples, we consider that a genus represents significantly different ratios of proportions from the samples when its adjusted p-value is <0.05.

  Significantly different ratios of proportions yields statistical evidence of intercation between the pair of factors that define the pair of pairs of samples. For instance, an adjusted p-value <0.05 for a genus in the comparison of the quotients p_{i20}/p_{c20} and p_{i28}/p_{i20} yields statistical evidence that the change of proportions between the condition “c” and the condition “i” is affected by the change of temperature from 20º to 28º.
- We generate a CSV file for each pair of pairs of samples that contains, for each genus, the ratio of ratios of proportions (for example, (p_i28/p_c28)/(p_c28/p_oculina)), the adjusted p-value, and the 95% confidence interval.

  We show in this report only the genera of interest.

```
RR.RR=function(E1,E2,conf.level=0.95,B=10000){
  E=rbind(E1[2:1,],E2[2:1,])
  e=E[,2]
  n=rowSums(E)
  e[e==0]=10^(-16)
  p=e/n
  RR1=p[1]/p[2]
  RR2=p[3]/p[4]
  RRR=RR1/RR2
#
  if (any(n==0)){Res=rbind(c(NA,NA,NA,NA,1))}
  if (!any(n==0)){
  if (any(E < 5)|any(n<50)){
  E[n==0,]=1
    Sim=function(E){
e.sim=c(rbinom(1,n[1],p[1]),rbinom(1,n[2],p[2]),
        rbinom(1,n[3],p[3]),rbinom(1,n[4],p[4]))
  p.sim=e.sim/n
  if(p[1]>10^(-6)){
e1.sim=c(rbinom(1,n[1],p[1]), rbinom(1,n[2],p[2]),
        rbinom(1,n[3],p[3]), rbinom(1,n[4],(p[2]*p[3])/p[1]))
  } else {
  e1.sim=c(rbinom(1,n[1],p[1]), rbinom(1,n[2],p[2]),
        rbinom(1,n[3],(p[1]*p[4])/p[2]), rbinom(1,n[4],p[4]))
}
```

```
  p1.sim=e1.sim/n
  c((p.sim[1]*p.sim[4])/(p.sim[2]*p.sim[3]),
    (p1.sim[1]*p1.sim[4])/(p1.sim[2]*p1.sim[3]))
}

X0=replicate(B,Sim(E))
IC.1=quantile(X0[1,],(1-conf.level)/2,na.rm=TRUE)
IC.2=quantile(X0[1,],1-(1-conf.level)/2,na.rm=TRUE)
p.val=length(X0[2,][X0[2,]>=max(RRR,1/RRR)|X0[2,]<=min(RRR,1/RRR)])/B
  }  else {
  se.lnRR1=sqrt(1/e[1]+1/e[2]-1/n[1]-1/n[2])
  se.lnRR2=sqrt(1/e[3]+1/e[4]-1/n[3]-1/n[4])
  se.lnRRR=sqrt(se.lnRR1^2+se.lnRR2^2)
  IC.1=exp(log(RRR)-se.lnRRR*qnorm((1+conf.level)/2))
  IC.2=exp(log(RRR)+se.lnRRR*qnorm((1+conf.level)/2))
  p.val=2-2*pnorm(abs(log(RRR))/se.lnRRR)
  }

  Res=rbind(c(RRR,p.val,IC.1,IC.2))
  if (!is.na(Res[1]) & Res[1]>10^8){Res[1]=Inf}
    }
  return(Res)
}


RR.RR.sig=function(x,y,z,t,etiqueta){
DF=DFG[c(x,y,z,t),]
n=rowSums(DF)
Res=c()
for (i in 1:dim(DF)[2]){
E1=matrix(c(n[1]-DF[1,i],DF[1,i],n[2]-DF[2,i],DF[2,i]),nrow=2,byrow=TRUE)
E2=matrix(c(n[3]-DF[3,i],DF[3,i],n[4]-DF[4,i],DF[4,i]),nrow=2,byrow=TRUE)
Res=rbind(Res,RR.RR(E1,E2))
}
Resultado=data.frame(Género=Generos,RRRR=Res[,1],p.val=Res[,2],IC1=Res[,3],IC2=Re
s[,4])
Resultado[,c(2,4,5)]=round(Resultado[,c(2,4,5)],3)
Resultado[,6]=round(p.adjust(Resultado[,3],method="BY"),4)

Resultado=Resultado[,c(1,2,6,4,5)]
names(Resultado)=c("Genus", etiqueta, "adjusted p-value", "95% CI lower end", "95%
CI upper end")
return(Resultado)
}
```

c28/oculina vs c20/oculina

```
Resultado.sig=RR.RR.sig(1,2,1,3,"(P_c28/P_oculina)/(P_c20/P_oculina)")
write.csv2(Resultado.sig,"Generos_Base.c20_vs_Base.c28.csv",row.names=FALSE)
```

There are 1372 with significantly different ratios of proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	(P_c28/P_oculina)/(P_c20/P_oculina)	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	1.201	0.0000	1.143	1.261
Fusarium	1.368	0.0000	1.265	1.480
Saccharomycodes	0.885	1.0000	0.681	1.149
Rhizophagus	2.124	0.0000	1.975	2.285
Symbiodinium	1.766	0.0000	1.733	1.800
Nitrosopumilus	1.015	1.0000	0.961	1.072
Streptomyces	0.728	0.0000	0.711	0.746
Acinetobacter	1.850	0.0000	1.781	1.923
Bordetella	0.793	0.0002	0.718	0.876
Salmonella	1.759	0.0000	1.573	1.967
Neisseria	2.376	0.0000	2.126	2.654
Klebsiella	0.684	0.0000	0.645	0.726
Desulfovibrio	0.216	0.0000	0.201	0.232
Halodesulfovibrio	0.128	0.0000	0.101	0.161
Loktanella	0.010	0.0000	0.008	0.012
Marinifilum	0.452	0.0000	0.409	0.500
Marivita	0.014	0.0000	0.011	0.019
Roseobacter	0.018	0.0000	0.015	0.022
Ruegeria	0.135	0.0000	0.126	0.146

Yoonia	0.011	0.0000	0.008	0.013
Vibrio	0.076	0.0000	0.073	0.080

I20/c20 vs C20/oculina

```
Resultado.sig=RR.RR.sig(1,2,2,4,"(P_i20/P_c20)/(P_c20/P_oculina)")
write.csv2(Resultado.sig,"Generos_i20.c20_vs_c20.base.csv",row.names=FALSE)
```

There are 1714 with significantly different ratios of proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	(P_i20/P_c20)/(P_c20/P_oculina)	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	0.395	0.0000	0.375	0.416
Fusarium	0.632	0.0000	0.582	0.687
Saccharomycodes	271.118	0.0000	225.124	326.508
Rhizophagus	0.716	0.0000	0.667	0.769
Symbiodinium	1.826	0.0000	1.781	1.873
Nitrosopumilus	1.113	0.0029	1.054	1.174
Streptomyces	0.446	0.0000	0.434	0.457
Acinetobacter	5.122	0.0000	4.980	5.269
Bordetella	32.005	0.0000	29.712	34.475
Salmonella	24.596	0.0000	22.575	26.798
Neisseria	63.391	0.0000	58.347	68.872
Klebsiella	1.352	0.0000	1.290	1.417
Desulfovibrio	0.385	0.0000	0.360	0.413
Halodesulfovibrio	10.740	0.0000	9.078	12.706

Loktanella	1.372	0.1805	1.083	1.738
Marinifilum	9.471	0.0000	8.793	10.201
Marivita	403.869	0.0000	322.302	506.079
Roseobacter	1.234	0.6270	1.014	1.502
Ruegeria	8.776	0.0000	8.222	9.368
Yoonia	2.282	0.0000	1.751	2.976
Vibrio	0.106	0.0000	0.101	0.111

I28/c28 vs c28/oculina

```
Resultado.sig=RR.RR.sig(1,3,3,5,"(P_i28/P_c28)/(P_c28/P_oculina)")
write.csv2(Resultado.sig,"Generos_i28.c28_vs_c28.base.csv",row.names=FALSE)
```

There are 1665 with significantly different ratios of proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	(P_i28/P_c28)/(P_c28/P_oculina)	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	0.438	0	0.413	0.465
Fusarium	0.580	0	0.526	0.641
Saccharomycodes	484.363	0	402.139	583.400
Rhizophagus	0.429	0	0.387	0.477
Symbiodinium	0.190	0	0.186	0.194
Nitrosopumilus	11.658	0	10.569	12.859
Streptomyces	1.116	0	1.089	1.144
Acinetobacter	14.675	0	14.189	15.177
Bordetella	99.165	0	91.930	106.968

Salmonella	7.200	0	6.586	7.871
Neisseria	6.142	0	5.640	6.688
Klebsiella	13.943	0	13.278	14.643
Desulfovibrio	3.225	0	3.066	3.392
Halodesulfovibrio	286.301	0	242.508	338.004
Loktanella	703.113	0	615.805	802.799
Marinifilum	4.881	0	4.543	5.244
Marivita	2310.042	0	1889.082	2824.809
Roseobacter	243.176	0	216.360	273.315
Ruegeria	149.604	0	141.800	157.838
Yoonia	950.195	0	812.653	1111.016
Vibrio	12.614	0	12.283	12.954

I28/c28 vs i20/c20

```
Resultado.sig=RR.RR.sig(2,4,3,5,"(P_i28/P_c28)/(P_i20/P_c20)")
write.csv2(Resultado.sig,"Generos_i28.c28_vs_i20.c20.csv",row.names=FALSE)
```

There are 1436 with significantly different ratios of proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	(P_i28/P_c28)/(P_i20/P_c20)	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	1.332	0.0000	1.253	1.416
Fusarium	1.256	0.0005	1.135	1.390
Saccharomycodes	1.581	0.0000	1.535	1.628
Rhizophagus	1.273	0.0002	1.148	1.411

Symbiodinium	0.184	0.0000	0.178	0.189
Nitrosopumilus	10.638	0.0000	9.650	11.727
Streptomyces	1.825	0.0000	1.776	1.874
Acinetobacter	5.301	0.0000	5.189	5.416
Bordetella	2.456	0.0000	2.368	2.548
Salmonella	0.515	0.0000	0.488	0.543
Neisseria	0.230	0.0000	0.221	0.240
Klebsiella	7.058	0.0000	6.822	7.302
Desulfovibrio	1.806	0.0000	1.729	1.887
Halodesulfovibrio	3.402	0.0000	3.325	3.479
Loktanella	5.126	0.0000	4.346	6.046
Marinifilum	0.233	0.0000	0.228	0.238
Marivita	0.081	0.0000	0.074	0.090
Roseobacter	3.533	0.0000	3.098	4.029
Ruegeria	2.307	0.0000	2.217	2.401
Yoonia	4.376	0.0000	3.638	5.263
Vibrio	9.078	0.0000	8.797	9.367

ms28/c28 vs ms20/c20

```
Resultado.sig=RR.RR.sig(2,6,3,7,"(P_ms28/P_c28)/(P_ms20/P_c20)")
write.csv2(Resultado.sig,"Generos_ms28.c28_vs_ms20.c20.csv",row.names=FALSE)
```

There are 515 with significantly different ratios of proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	(P_ms28/P_c28)/(P_ms20/P_c20)	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	1.689	0.0000	1.587	1.798
Fusarium	1.287	0.0001	1.165	1.422
Saccharomycodes	1.216	0.0000	1.184	1.250
Rhizophagus	2.440	0.0000	2.183	2.728
Symbiodinium	1.293	0.0000	1.260	1.328
Nitrosopumilus	2.272	0.0000	2.133	2.419
Streptomyces	1.319	0.0000	1.283	1.356
Acinetobacter	1.468	0.0000	1.440	1.496
Bordetella	1.402	0.0000	1.357	1.448
Salmonella	1.118	0.0002	1.068	1.170
Neisseria	0.400	0.0000	0.385	0.415
Klebsiella	1.902	0.0000	1.834	1.972
Desulfovibrio	1.344	0.0000	1.269	1.423
Halodesulfovibrio	2.666	0.0000	2.589	2.746
Loktanella	1.465	0.0088	1.210	1.774
Marinifilum	0.379	0.0000	0.370	0.388
Marivita	0.031	0.0000	0.028	0.034
Roseobacter	1.393	0.0004	1.211	1.601
Ruegeria	0.340	0.0000	0.326	0.355
Yoonia	1.127	1.0000	0.919	1.383
Vibrio	1.378	0.0000	1.326	1.431

mx28/c28 vs mx20/c20

```
Resultado.sig=RR.RR.sig(2,8,3,9,"(P_mx28/P_c28)/(P_mx20/P_c20)")
write.csv2(Resultado.sig,"Generos_mx28.c28_vs_mx20.c20.csv",row.names=FALSE)
```

There are 555 with significantly different ratios of proportions (at a 5% global signification level). As far as the genera of interest goes:

```
rownames(Resultado.sig)=Resultado.sig$Genus
Resultado.sig.Interesting=Resultado.sig[Interesting,]
rownames(Resultado.sig.Interesting)=NULL

Resultado.sig.Interesting %>%
  kbl() %>%
  kable_styling()
```

Genus	(P_mx28/P_c28)/(P_mx20/P_c20)	adjusted p-value	95% CI lower end	95% CI upper end
Aspergillus	1.177	0.0001	1.103	1.257
Fusarium	0.893	1.0000	0.808	0.987
Saccharomycodes	3.716	0.0000	3.602	3.835
Rhizophagus	1.189	0.0414	1.078	1.312
Symbiodinium	0.200	0.0000	0.195	0.204
Nitrosopumilus	0.323	0.0000	0.288	0.362
Streptomyces	1.266	0.0000	1.229	1.303
Acinetobacter	0.683	0.0000	0.672	0.695
Bordetella	2.867	0.0000	2.744	2.995
Salmonella	0.984	1.0000	0.929	1.043
Neisseria	0.723	0.0000	0.688	0.760
Klebsiella	5.235	0.0000	5.061	5.415
Desulfovibrio	0.543	0.0000	0.508	0.581
Halodesulfovibrio	0.016	0.0000	0.013	0.019
Loktanella	0.490	0.0000	0.406	0.591
Marinifilum	0.035	0.0000	0.034	0.037
Marivita	9.553	0.0000	8.338	10.944
Roseobacter	0.523	0.0000	0.449	0.610
Ruegeria	1.389	0.0000	1.320	1.461

Yoonia	0.359	0.0000	0.295	0.437
Vibrio	1.514	0.0000	1.448	1.582

Who produces the antimicrobian genes?

Let's call p_E the proportion of antimicrobial genes that are expressed by eukaryotes and p_P the proportion of antimicrobial genes expressed by prokaryotes.

- In the natural sample, we perform a one-sided test with the alternative hypothesis $p_E>p_P$; since the samples are large and the numbers allow it, we use the usual test based on the normal approximation.
- In the aquarium samples, we perform a one-sided test with the alternative hypothesis $p_P>p_E$. For the individual aquariums, we adjust the p-values to control the risk of false positives. In each case, we also give the 95% confidence interval of the value of p_E or p_P corresponding to the one-sided test performed.
- We round all p-values to 4 decimal places.

```
Datos.EE=data.frame(read_excel("Genes.xlsx",sheet="Euk"))
Datos.PP=data.frame(read_excel("Genes.xlsx",sheet="Prok"))

Samples=names(Datos.EE)[-1]
Samples[1]="natural"

Eucariotas=Datos.EE[,1]
DFE=t(Datos.EE[,-1])
colnames(DFE)=Eucariotas
row.names(DFE)=Samples

Procariotas=Datos.PP[,1]
DFP=t(Datos.PP[,-1])
colnames(DFP)=Procariotas
row.names(DFP)=Samples

N.E=rowSums(DFE)
N.P=rowSums(DFP)
```

Natural

```
e=N.E[1]
n=N.E[1]+N.P[1]
PT=prop.test(e,n,alternative="greater")

res=data.frame("Natural",PT$p.value,PT$conf.int[1])

res[,2]=round(res[,2],6)
res[,3]=round(res[,3],4)

names(res)=c("Muestra","p-value p_E>p_P?","p_E a partir de...")




res %>%
  kbl() %>%
  kable_styling()
```

Muestra	p-value p_E>p_P?	p_E a partir de…
Natural	0	0.7142

Aquariums

```
res=c()
for (i in 2:9){
e=N.P[i]
n=N.E[i]+N.P[i]
PT=prop.test(e,n,alternative="greater")
res=rbind(res,c(PT$p.value,PT$conf.int[1]))
}
e=sum(N.P[2:9])
n=sum(N.E[2:9])+sum(N.P[2:9])
PTG=prop.test(e,n,alternative="greater")

res=rbind(res,c(PTG$p.value,PTG$conf.int[1]))
res=data.frame(c(Samples[-1],"Global Acuarios"),res)

res[,4]=NA
res[1:8,4]=round(p.adjust(res[1:8,2],method="BY"),6)
res[,2]=round(res[,2],4)
names(res)=c("Muestra","p-value p_P>p_E?","p_P a partir de...","adjusted p-value")




res[,c(1,2,4,3)] %>%
  kbl() %>%
  kable_styling()
```

Muestra	p-value p_P>p_E?	adjusted p-value	p_P a partir de…

c20	1	1	0.4727977
c28	0	0	0.7830996
i20	0	0	0.7499165
i28	0	0	0.8447125
ms20	0	0	0.5534443
ms28	0	0	0.8214144
mx20	1	1	0.3906610
MX28	0	0	0.7780377
Global Acuarios	0		0.7668473