

## Tarea 8.2 “Análisis de datos de Illumina MiSeq desde AMPtk”

Diego Montesinos Valencia 2020

Este script esta basado en la clase de “Introducción a la bioinformática e investigación reproducible para análisis genómicos”

### Metodología y Resultados

Cambiar el working directory a la localización del script

Cargar las librerías que se van a utilizar

```
library("phyloseq")
library("vegan")
```

```
## Loading required package: permute
```

```
## Loading required package: lattice
```

```
## This is vegan 2.5-6
```

```
library("ggplot2")
```

Importar los datos (el archivo taxonomy.boim, ubicado en la carpeta DATA)

```
suelo <- import_biom("../DATA/taxonomy.boim")
suelo
```

```
## phyloseq-class experiment-level object
```

```
## otu_table() OTU Table: [ 1257 taxa and 12 samples ]
```

```
## sample_data() Sample Data: [ 12 samples by 8 sample variables ]
```

```
## tax_table() Taxonomy Table: [ 1257 taxa by 7 taxonomic ranks ]
```

Cambiar las columnas de la tabla de taxonomía

```
colnames(tax_table(suelo)) <- c("Domain", "Phylum", "Class", "Order", "Family", "Genus", "Species")
```

Ver la tabla de taxonomia

```
head(tax_table(suelo))
```

```
## Taxonomy Table: [6 taxa by 7 taxonomic ranks]:
```

	Domain	Phylum	Class	Order
## OTU1	"k__Fungi"	"p__Ascomycota"	"c__Dothideomycetes"	"o__Capnodiales"
## OTU2	"k__Fungi"	"p__Basidiomycota"	"c__Agaricomycetes"	"o__Sebacinales"
## OTU3	"k__Fungi"	"p__Basidiomycota"	"c__Agaricomycetes"	"o__Agaricales"
## OTU5	"k__Fungi"	"p__Basidiomycota"	"c__Agaricomycetes"	"o__Agaricales"
## OTU6	"k__Fungi"	"p__Ascomycota"	"c__Dothideomycetes"	"o__Mytilinidales"
## OTU7	"k__Fungi"	"p__Basidiomycota"	"c__Agaricomycetes"	"o__Sebacinales"
	Family	Genus	Species	
## OTU1	"f__Cladosporiaceae"	"g__Cladosporium"	NA	
## OTU2	"f__Sebacinaceae"	"g__Sebacina"	NA	
## OTU3	"f__Hygrophoraceae"	"g__Cuphophyllus"	"s__Cuphophyllus virgineus"	
## OTU5	"f__Inocybaceae"	"g__Inocybe"	NA	
## OTU6	"f__Gloniaceae"	"g__Cenococcum"	"s__Cenococcum geophilum"	
## OTU7	"f__Sebacinaceae"	"g__Sebacina"	NA	

Crear una tabla binomial a partir del objeto “suelo”

```

otu_table <- as.data.frame(otu_table(suelo)) # Extrae la tabla de OTUs
A <- decostand(otu_table, method="pa") # Cambia a tabla binaria (presencia-ausencia)
otu_table_bin <- phyloseq(otu_table(A, taxa_are_rows= TRUE)) # Transformar la tabla a formato phyloseq
tax_table <- suelo@tax_table # Extraer la tabla de taxa
sam_data <- suelo@sam_data # Extraer la tabla de muestras

```

Crear un objeto con phyloseq

```

suelo_b <- merge_phyloseq(otu_table_bin,tax_table,sam_data)
suelo_b

```

```

## phyloseq-class experiment-level object
## otu_table() OTU Table:          [ 1257 taxa and 12 samples ]
## sample_data() Sample Data:      [ 12 samples by 8 sample variables ]
## tax_table() Taxonomy Table:     [ 1257 taxa by 7 taxonomic ranks ]

```

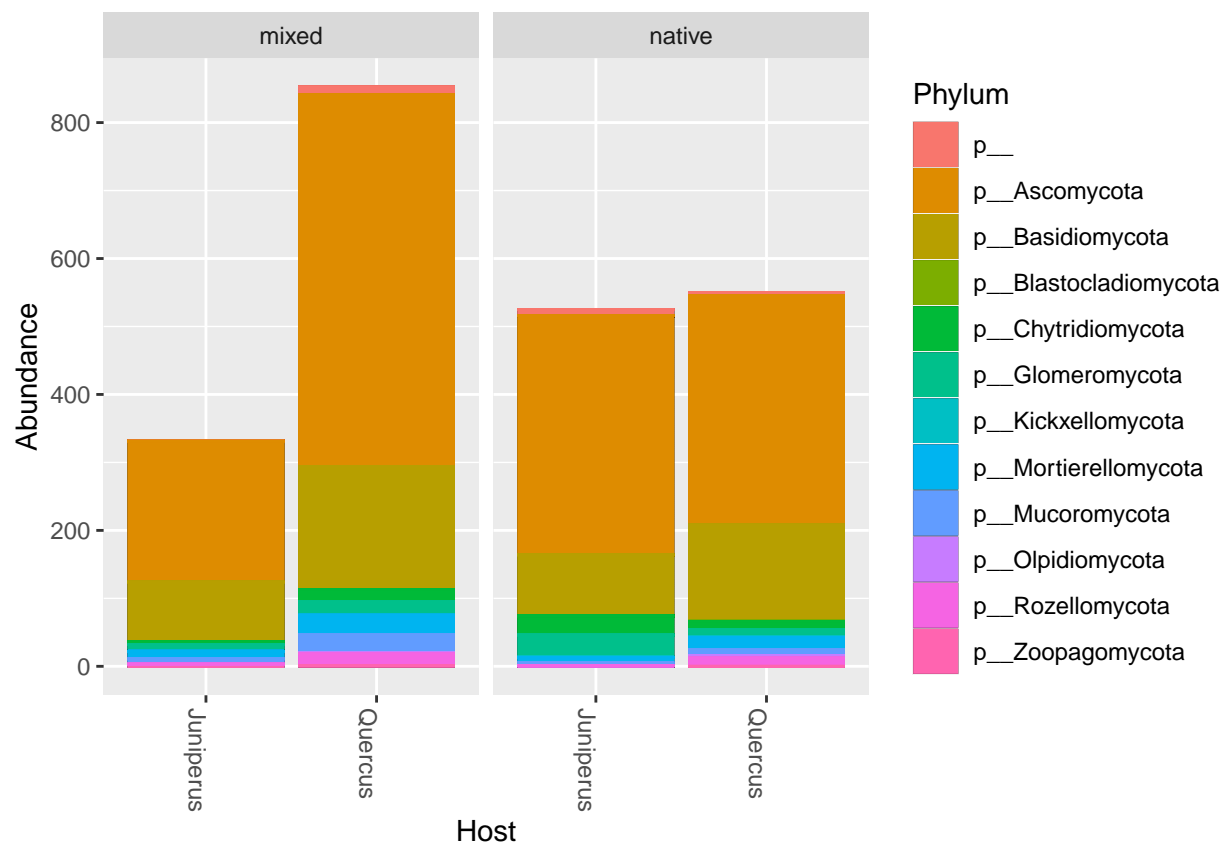
## Diversidad alpha (riqueza en especies)

Visualizar los datos mediante un bar plot pero teniendo tratamiento y hospedero a nivel de phylum

```

plot1 = plot_bar(suelo_b, "Host", fill = "Phylum") +
  geom_bar(aes(color=Phylum, fill = Phylum), stat = "identity", position = "stack")
plot1 + facet_wrap("Treatment")

```



## Hacer un anova de la riqueza observada para host y treatment

Estimar la diversidad observada

```
tab_diversity <- estimate_richness(suelo_b, measures = "Observed")
tab_diversity
```

```
##                Observed
## A1.JM2.2.2621739      80
## A12.QM1.2.2621927    318
## B12.QM2.2.2621928    322
## E3.JN1.2.2621809     90
## F2.QN3.2.2621802    109
## F3.JN2.2.2621810     61
## F9.QM3.2.2621908    212
## G2.QN1.2.2621803     98
## G3.JM1.2.2621811     76
## H2.JM3.2.2621804    176
## H11.JN3.2.2621926   374
## H12.QN2.2.2621934   343
```

Combinar la tabla de diversidad observada con la de las muestras de "suelo\_b"

```
data <- cbind(sample_data(suelo_b), tab_diversity)
data
```

```
##                BarcodeSequence LinkerPrimerSequence RevBarcodeSequence
## A1-JM2-2-2621739             S1  GTGATCATCRARTYTTTG          no_data
## A12-QM1-2-2621927            S12 GTGATCATCRARTYTTTG          no_data
## B12-QM2-2-2621928            S24 GTGATCATCRARTYTTTG          no_data
## E3-JN1-2-2621809             S51 GTGATCATCRARTYTTTG          no_data
## F2-QN3-2-2621802             S62 GTGATCATCRARTYTTTG          no_data
## F3-JN2-2-2621810             S63 GTGATCATCRARTYTTTG          no_data
## F9-QM3-2-2621908             S69 GTGATCATCRARTYTTTG          no_data
## G2-QN1-2-2621803             S74 GTGATCATCRARTYTTTG          no_data
## G3-JM1-2-2621811             S75 GTGATCATCRARTYTTTG          no_data
## H2-JM3-2-2621804             S86 GTGATCATCRARTYTTTG          no_data
## H11-JN3-2-2621926            S95 GTGATCATCRARTYTTTG          no_data
## H12-QN2-2-2621934            S96 GTGATCATCRARTYTTTG          no_data
##                ReversePrimer          phinchID DemuxReads Treatment
## A1-JM2-2-2621739 CCTSCSCTTANTDATATGC A1-JM2-2-2621739    23372   mixed
## A12-QM1-2-2621927 CCTSCSCTTANTDATATGC A12-QM1-2-2621927    24709   mixed
## B12-QM2-2-2621928 CCTSCSCTTANTDATATGC B12-QM2-2-2621928    24694   mixed
## E3-JN1-2-2621809 CCTSCSCTTANTDATATGC E3-JN1-2-2621809    24201  native
## F2-QN3-2-2621802 CCTSCSCTTANTDATATGC F2-QN3-2-2621802    23503  native
## F3-JN2-2-2621810 CCTSCSCTTANTDATATGC F3-JN2-2-2621810    24563  native
## F9-QM3-2-2621908 CCTSCSCTTANTDATATGC F9-QM3-2-2621908    24323   mixed
## G2-QN1-2-2621803 CCTSCSCTTANTDATATGC G2-QN1-2-2621803    23878  native
## G3-JM1-2-2621811 CCTSCSCTTANTDATATGC G3-JM1-2-2621811    24551   mixed
## H2-JM3-2-2621804 CCTSCSCTTANTDATATGC H2-JM3-2-2621804    23365   mixed
## H11-JN3-2-2621926 CCTSCSCTTANTDATATGC H11-JN3-2-2621926    24623  native
## H12-QN2-2-2621934 CCTSCSCTTANTDATATGC H12-QN2-2-2621934    23818  native
##                Host Observed
## A1-JM2-2-2621739 Juniperus      80
## A12-QM1-2-2621927 Quercus      318
## B12-QM2-2-2621928 Quercus      322
```

```
## E3-JN1-2-2621809 Juniperus 90
## F2-QN3-2-2621802 Quercus 109
## F3-JN2-2-2621810 Juniperus 61
## F9-QM3-2-2621908 Quercus 212
## G2-QN1-2-2621803 Quercus 98
## G3-JM1-2-2621811 Juniperus 76
## H2-JM3-2-2621804 Juniperus 176
## H11-JN3-2-2621926 Juniperus 374
## H12-QN2-2-2621934 Quercus 343
```

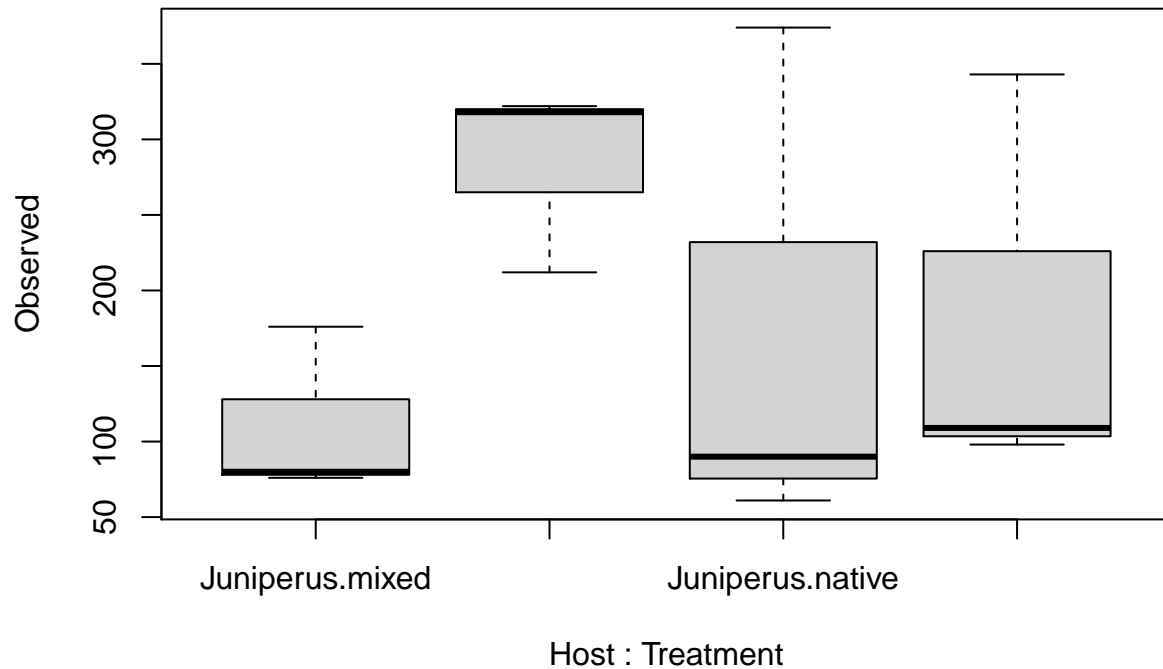
Realizar el ANOVA

```
anova <- aov(Observed ~ Treatment * Host, data = data)
summary(anova)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Treatment    1    990      990   0.071  0.797
## Host          1 24752 24752    1.763  0.221
## Treatment:Host 1 20419 20419    1.454  0.262
## Residuals     8 112317 14040
```

Graficar

```
boxplot(Observed ~ Host * Treatment, data = data)
```



## Diversidad beta

Realizar matriz de disimilitud (Raup-Crick distance)

```
raup <- distance(suelo_b, method = "raup")
raup
```

```
##          A1-JM2-2-2621739 A12-QM1-2-2621927 B12-QM2-2-2621928
## A12-QM1-2-2621927      3.812118e-04
## B12-QM2-2-2621928      1.940208e-02      8.407267e-08
## E3-JN1-2-2621809      2.981085e-04      3.142908e-05      6.091538e-04
## F2-QN3-2-2621802      1.647256e-02      2.194212e-02      6.226159e-01
## F3-JN2-2-2621810      1.154752e-05      1.832723e-03      3.481819e-04
## F9-QM3-2-2621908      3.629450e-02      1.266253e-07      1.532108e-14
## G2-QN1-2-2621803      1.631959e-01      6.168236e-01      5.513875e-01
## G3-JM1-2-2621811      2.127505e-03      1.615661e-03      8.848442e-02
## H2-JM3-2-2621804      8.040961e-02      2.033454e-09      1.139412e-04
## H11-JN3-2-2621926      9.986778e-01      9.992857e-01      9.999252e-01
## H12-QN2-2-2621934      7.233675e-01      3.493853e-03      1.227350e-02
##          E3-JN1-2-2621809 F2-QN3-2-2621802 F3-JN2-2-2621810
## A12-QM1-2-2621927
## B12-QM2-2-2621928
## E3-JN1-2-2621809
## F2-QN3-2-2621802      7.933218e-03
## F3-JN2-2-2621810      9.296648e-10      7.576308e-10
## F9-QM3-2-2621908      6.395760e-02      3.060165e-05      9.862806e-07
## G2-QN1-2-2621803      8.317392e-02      6.731157e-07      1.781737e-03
## G3-JM1-2-2621811      4.075325e-05      5.711590e-02      1.557377e-07
## H2-JM3-2-2621804      8.016741e-10      5.668327e-03      3.255598e-04
## H11-JN3-2-2621926      1.299106e-01      9.772516e-01      5.671620e-01
## H12-QN2-2-2621934      3.124659e-01      6.892796e-01      6.250973e-01
##          F9-QM3-2-2621908 G2-QN1-2-2621803 G3-JM1-2-2621811
## A12-QM1-2-2621927
## B12-QM2-2-2621928
## E3-JN1-2-2621809
## F2-QN3-2-2621802
## F3-JN2-2-2621810
## F9-QM3-2-2621908
## G2-QN1-2-2621803      7.829386e-04
## G3-JM1-2-2621811      1.957659e-01      1.433104e-03
## H2-JM3-2-2621804      1.159980e-04      5.144178e-01      9.663811e-04
## H11-JN3-2-2621926      9.731756e-01      9.889183e-01      9.848998e-01
## H12-QN2-2-2621934      1.694230e-01      5.166819e-01      6.227335e-01
##          H2-JM3-2-2621804 H11-JN3-2-2621926
## A12-QM1-2-2621927
## B12-QM2-2-2621928
## E3-JN1-2-2621809
## F2-QN3-2-2621802
## F3-JN2-2-2621810
## F9-QM3-2-2621908
## G2-QN1-2-2621803
## G3-JM1-2-2621811
## H2-JM3-2-2621804
## H11-JN3-2-2621926      8.059043e-01
## H12-QN2-2-2621934      1.193973e-01      9.999961e-01
```

Realizar NMDS

```
NMDS = ordinate(suelo_b, method = "NMDS", distance = raup)
```

```
## Run 0 stress 0.1419322
## Run 1 stress 0.1322727
## ... New best solution
## ... Procrustes: rmse 0.2084638  max resid 0.4222138
## Run 2 stress 0.1445462
## Run 3 stress 0.1376591
## Run 4 stress 0.144545
## Run 5 stress 0.1466543
## Run 6 stress 0.132276
## ... Procrustes: rmse 0.001738139  max resid 0.0040689
## ... Similar to previous best
## Run 7 stress 0.180785
## Run 8 stress 0.1322743
## ... Procrustes: rmse 0.001148002  max resid 0.002655313
## ... Similar to previous best
## Run 9 stress 0.1445453
## Run 10 stress 0.1322736
## ... Procrustes: rmse 0.0008662604  max resid 0.002003857
## ... Similar to previous best
## Run 11 stress 0.1511893
## Run 12 stress 0.14881
## Run 13 stress 0.1322733
## ... Procrustes: rmse 0.000709403  max resid 0.001632506
## ... Similar to previous best
## Run 14 stress 0.1376593
## Run 15 stress 0.1429636
## Run 16 stress 0.137658
## Run 17 stress 0.137658
## Run 18 stress 0.132275
## ... Procrustes: rmse 0.001351293  max resid 0.003122418
## ... Similar to previous best
## Run 19 stress 0.1322735
## ... Procrustes: rmse 0.0007951687  max resid 0.001835863
## ... Similar to previous best
## Run 20 stress 0.3163909
## *** Solution reached
```

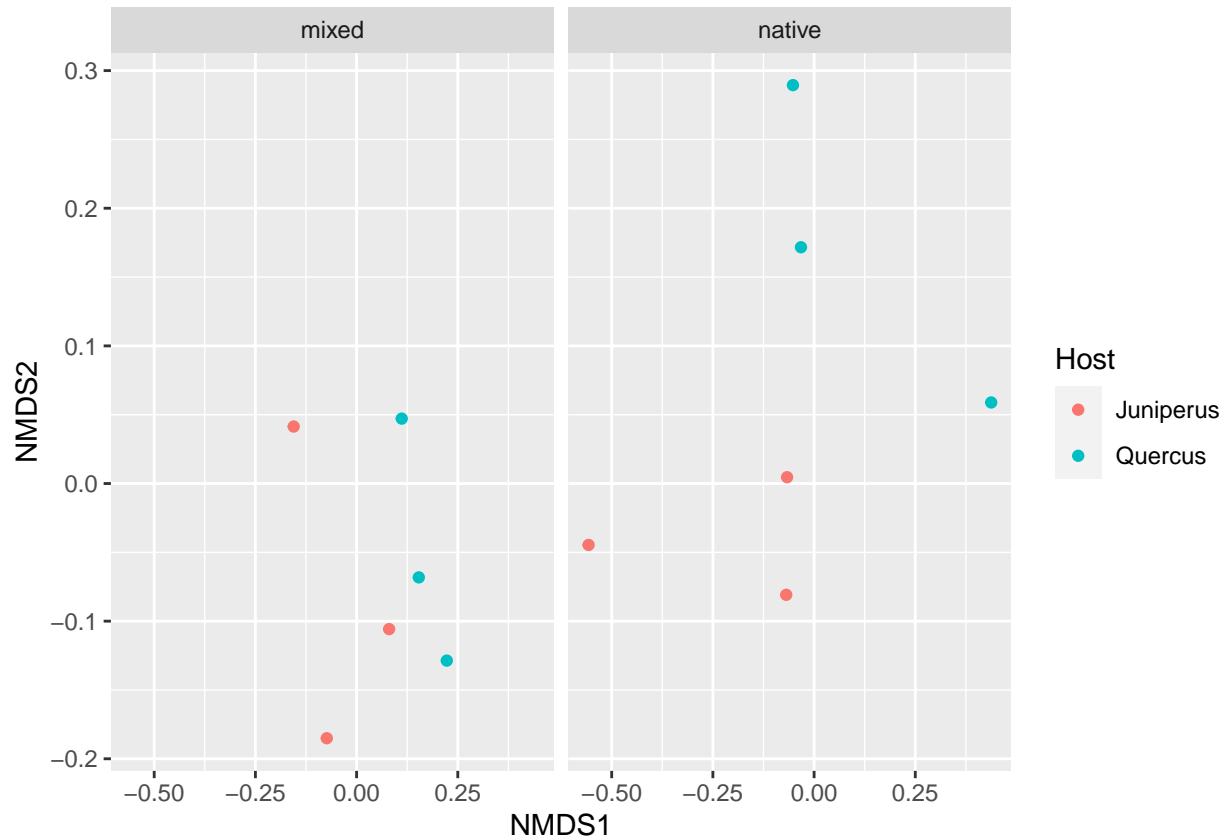
NMDS

```
##
## Call:
## metaMDS(comm = ps.dist)
##
## global Multidimensional Scaling using monoMDS
##
## Data:      ps.dist
## Distance:  raup
##
## Dimensions: 2
## Stress:      0.1322727
## Stress type 1, weak ties
## Two convergent solutions found after 20 tries
```

```
## Scaling: centring, PC rotation
## Species: scores missing
```

Plot de la ordination

```
plot2 = plot_ordination(suelo_b, NMDS, color = "Host") + facet_wrap("Treatment")
plot2
```



Hacer test de adonis de la composicion de la comunidad por host y treatment

```
adonis(raup ~ Host * Treatment, data = data)
```

```
##
## Call:
## adonis(formula = raup ~ Host * Treatment, data = data)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##              Df SumsOfSqs  MeanSqs F.Model    R2 Pr(>F)
## Host           1  0.23671 0.236705  5.2258 0.22812 0.027 *
## Treatment      1  0.27707 0.277068  6.1169 0.26702 0.015 *
## Host:Treatment  1  0.16148 0.161478  3.5650 0.15562 0.089 .
## Residuals      8  0.36236 0.045295          0.34923
```

```
## Total          11    1.03761          1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```