
CENTRO DE CIENCIAS MATEMÁTICAS

CINVESTAV

ENES Análisis exploratorio de datos de microbiomas

Capítulo 7

Andrés Arredondo (email_1@gmail.com)
Adriana Haydé Contreras Peruyero (haydeeperuyero@gmail.com)
David Alberto García Estrada (@gmail.com)

Morelia

Septiembre de 2022

Índice

1. Datos de Ratones y Humanos	3
2. Análisis exploratorio con resumen gráfico	3
2.1. Gráficos de riqueza	3
2.2. Barras de abundancia	5
2.3. Mapas de calor	6
2.4. Redes	6
2.5. Árbol filogenético	6
3. Clusters	6
3.1. Distancias	6
3.2. Diferentes tipos de clústers	6
3.2.1. Single	6
3.2.2. Complete	6
3.2.3. Average	6
3.2.4. Ward	6
4. Ordination	6
4.1. PCA	6
4.2. PCoA	6
4.3. NMDS	6
4.4. CA	6
4.5. RDA	6
4.6. CCA	6
4.7. CAP	6
5. Conclusiones	6

1. Datos de Ratones y Humanos

Las bases de datos que se ocuparon son dos:

- Vdr: es una base de datos de ratones. Contiene datos de microbiomas intestinales y fueron recolectados de heces y muestras de heces cecales. Las que se usan en este capítulo son de heces.
- Troat.otu.tab: estos datos son de fumadores y se encuentran en el paquete **GUniFrac**. Estos datos se usarán para explorar el árbol filogenético.

2. Análisis exploratorio con resumen gráfico

En este capítulo se exploran diferentes gráficos usuales: riqueza, barras de abundancia, mapas de calor, redes y árbol filogenético.

2.1. Gráficos de riqueza

En el capítulo anterior se exploró la diversidad alfa, en este capítulo vamos a explorar un gráfico relacionado a esto. Para esto, se usa la función `plot_richness()` del paquete `phyloseq`. Generalmente, la riqueza se refiere a un gráfico del número total de especies, taxones u OTUs en un ambiente, pero esta función también nos da otras figuras relacionadas a otras diversidades.

Lo primero que debemos hacer es instalar el paquete y leer los datos.

```
#Para instalar el paquete, usamos Bioconductor

#if (!require("BiocManager", quietly = TRUE))
#  install.packages("BiocManager")

#BiocManager::install("phyloseq")
```

En la carpeta `data` se encuentra la base de datos a usar `VdrFecalGenusCounts.csv`.

```
setwd("D:/Users/hayde/Documents/R_sites/Equipo4")

#library(phyloseq)
#library(ggplot2)
abund_table=read.csv("./data/VdrFecalGenusCounts.csv",row.names=1,check.names=FALSE)
abund_table<-t(abund_table)
```

Lo primero que se debe de hacer es contruir nuestro objeto `phyloseq`. Este objeto se contruye tomando en cuenta los siguientes componentes: - Tabla OTU. - Datos muestra. - Tabla de taxonomía. - Árbol filogenético.

Se deben de proporcionar dos objetos de datos pero el orden en que se proporcionan no es importante. Construimos primero nuestra tabla de metadatos con los siguientes comandos.

```
meta_table <- data.frame(row.names=row.names(abund_table),t(as.data.frame(strsplit(row.names(abund_table)
meta_table$Group <- with(meta_table,ifelse(as.factor(X2) %in% c(11,12,13,14,15),c("Vdr-/-"), c("WT"))))
```

Convertimos los datos al formato `phyloseq`.

```
OTU = otu_table(as.matrix(abund_table), taxa_are_rows = FALSE)
SAM = sample_data(meta_table)
physeq <- merge_phyloseq(phyloseq(OTU), SAM)
physeq
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table:      [ 248 taxa and 8 samples ]
## sample_data() Sample Data:  [ 8 samples by 4 sample variables ]
```

Una vez que tenemos nuestro objeto `phyloseq`, podemos usar la función `plot_richness()` para contruir graficar las diversidades alpha observadas y estimadas.

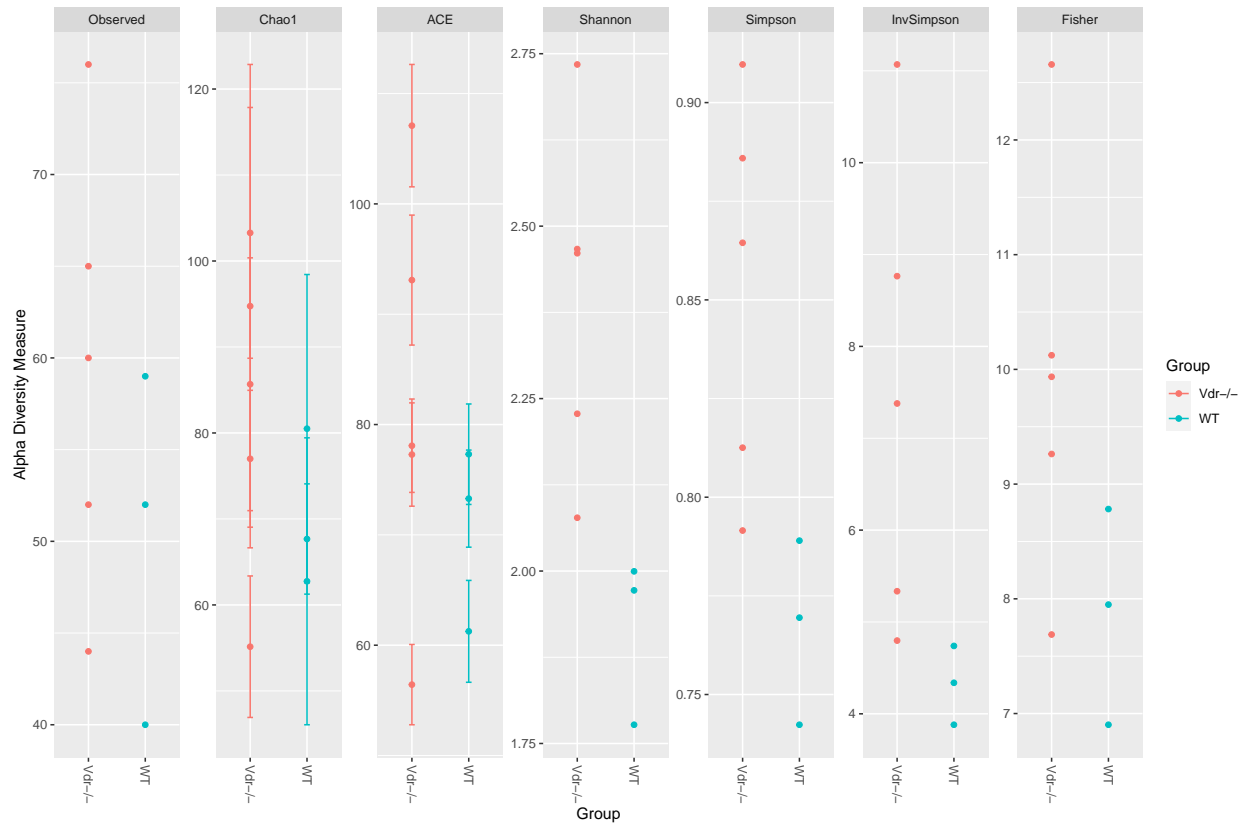


Figura 1: Gráficos de diversidad alpha con Vdr y grupos WT en muestras de heces.

Esta función también nos permite seleccionar solo algunas diversidades. El siguiente es un ejemplo usando solo dos diversidades, la de Chao1 y Shannon.

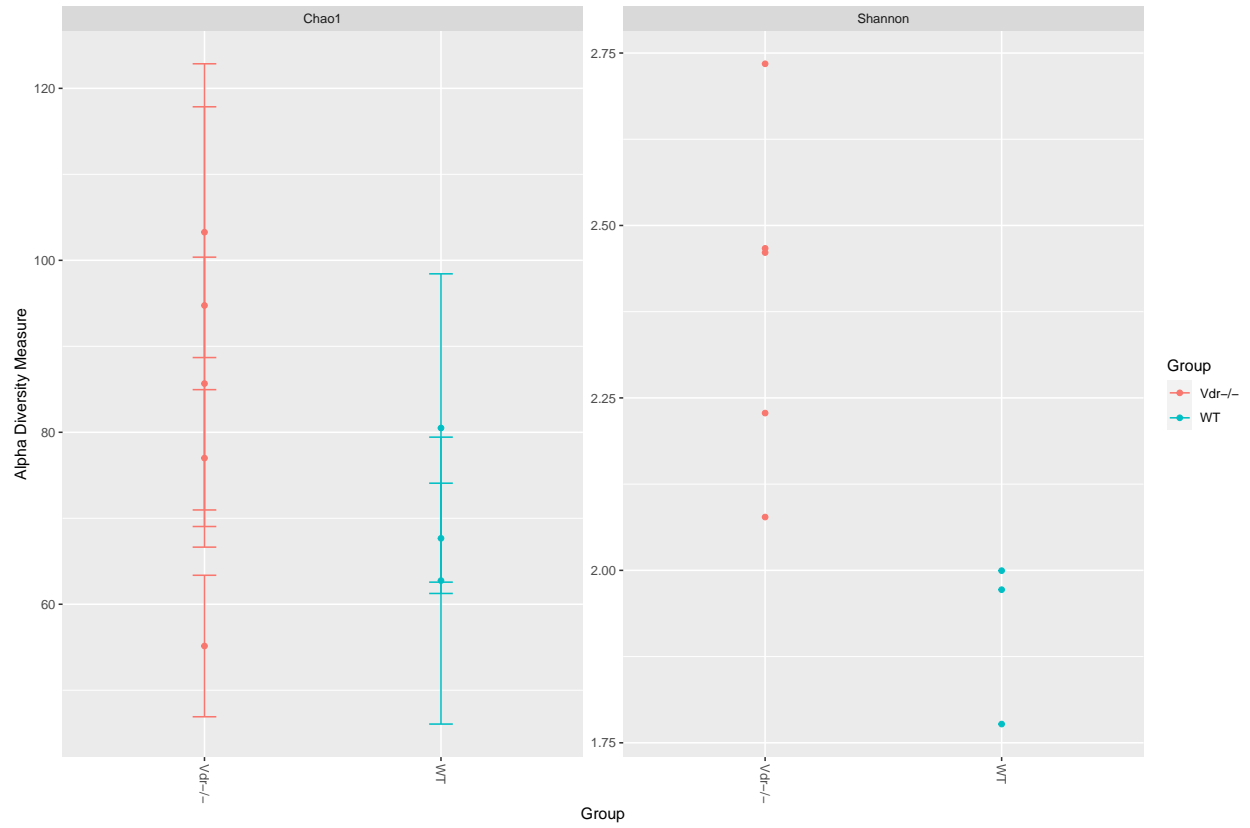


Figura 2: Gráficos de diversidad alpha seleccionando las diversidades de Chao y Shannon.

2.2. Barras de abundancia

La herramienta de `phyloseq` que nos permite graficar las barras de abundancia es `plot_bar()`.

2.3. Mapas de calor

2.4. Redes

2.5. Árbol filogenético

3. Clusters

3.1. Distancias

3.2. Diferentes tipos de clústers

3.2.1. Single

3.2.2. Complete

3.2.3. Average

3.2.4. Ward

4. Ordination

4.1. PCA

4.2. PCoA

4.3. NMDS

4.4. CA

4.5. RDA

4.6. CCA

4.7. CAP

5. Conclusiones