

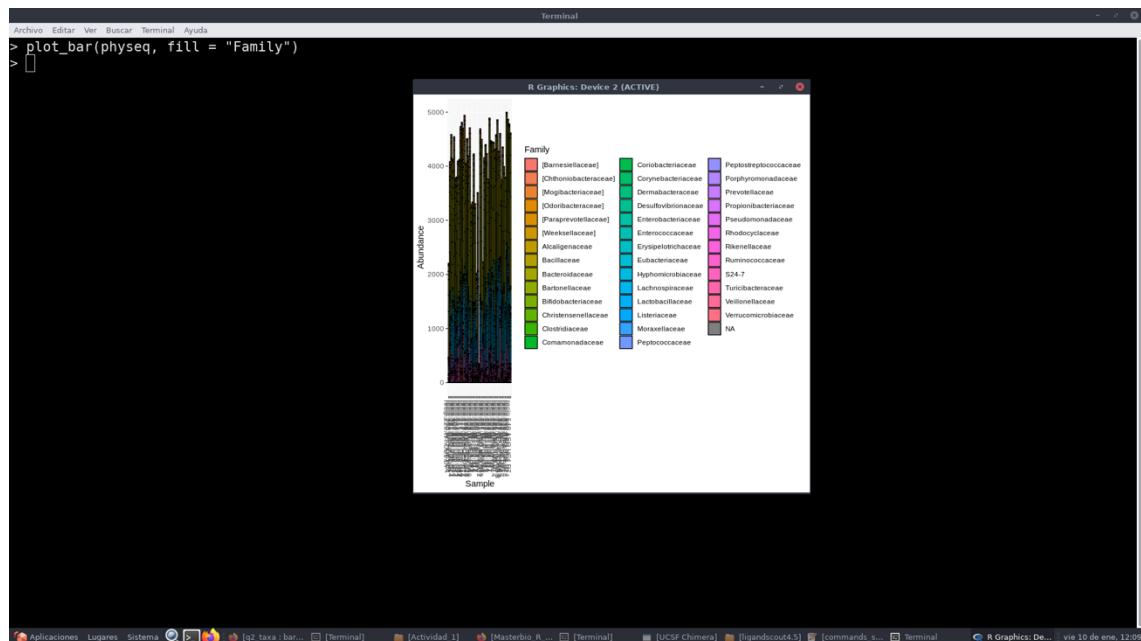
1. Haz un subset de la orden Clostridiales, ¿Cuántos taxones había y cuántos taxones hay ahora?

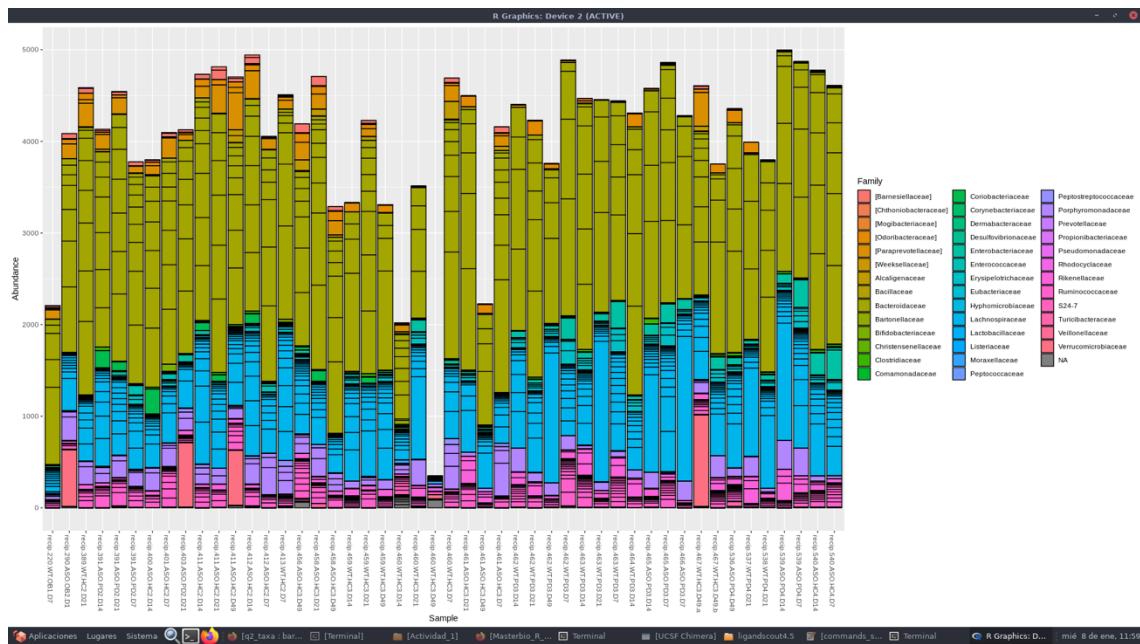
```
Archivo Editar Ver Buscar Terminal Ayuda
> physeq.clos = subset_taxa(physeq, Order == "Clostridiales")
> physeq.clos
phyloseq-class experiment-level object
otu_table() OTU Table: [ 156 taxa and 48 samples ]
sample_data() Sample Data: [ 48 samples by 9 sample variables ]
tax_table() Taxonomy Table: [ 156 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 156 tips and 155 internal nodes ]
> physeq
phyloseq-class experiment-level object
otu_table() OTU Table: [ 287 taxa and 48 samples ]
sample_data() Sample Data: [ 48 samples by 9 sample variables ]
tax_table() Taxonomy Table: [ 287 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 287 tips and 286 internal nodes ]
>
```

Al realizar el subset del orden Clostridiales con el código: physeq.clos = subset_taxa(physeq, Order == "Clostridiales"), antes había 287 taxones, mientras que ahora quedan 156 taxones, al filtrar por el orden Clostridiales.

2. Representa para cada muestra su abundancia para cada una de las familias.

Con la función `plot_bar(physeq, fill = "Family")`, filtrando por el campo `family` obtenemos para cada muestra su abundancia para cada una de las familias.





3. Calcula el valor medio de diversidad alfa (Índice de Shannon) en las muestras según la variable `donor_status`, que diferencia a las muestras según tengan un trasplante de microbioma de personas sanas o con la enfermedad de Parkinson, haciendo uso del paquete `microbiome`. ¿Hay diferencias significativas entre los grupos? ¿Qué grupo de muestras tiene una mayor diversidad alfa según este índice? Incluye el código utilizado para dar tu respuesta.

El valor medio de la diversidad alfa para el grupo de donantes sanos (“Healthy”) es: 2.90387
El valor medio de la diversidad alfa para el grupo de donantes con enfermedad de Parkinson (“PD”) es: 2.541922

El p-valor es de 0.0001040143, por lo que sí que hay diferencias significativas entre el grupo de donantes sanos y donantes con la enfermedad de Parkinson, tanto si se fija un umbral de p-valor de 0.5 que suele ser el más habitual, y es menor de 0.5, como si se fija un umbral de p-valor de 0.1 o 0.001 si se quiere ser más estricto, ya que el p-valor obtenido es sumamente pequeño y menor a todos estos umbrales. Además, el p-valor ajustado tiene el mismo valor que el p-valor por lo que confirma que sea significativo. Esto quiere decir que el microbioma de los organismos con trasplante de microbioma de donantes sanos tiene una diversidad microbiana diferente (y en este caso mayor) que los organismos con trasplante de microbioma de donantes con enfermedad de Parkinson.

El grupo de donantes sanos es el que tiene una mayor diversidad alfa.

El código empleado para obtener estos valores:

- Primero se obtienen todos los índices de Shannon:
`d = meta(physeq)`
`d$diversity <- microbiome::diversity(physeq, "shannon")$shannon`
 - Despues obtenemos en concreto los valores de diversidad de la variable `donor_status`:
`spl = split(d$diversity, d$donor_status)`

Y calculamos el valor medio del índice de Shannon para cada grupo:

```
mean(spl$Healthy)
mean(spl$PD)
```

- Después se realiza el test de Kolmogorov-Smirnov para obtener el p-valor, y se realiza un ajuste del p-valor obtenido.

```
pv <- ks.test(spl$Healthy, spl$PD)$p.value
padj <- p.adjust(pv)
```

```

Archivo Editar Ver Buscar Terminal Ayuda
> d = meta(physeq)
> d$diversity <- microbiome::diversity(physeq, "shannon")$shannon
> d$diversity
[1] 2.369679 2.689382 2.863240 3.041673 3.000449 2.972395 2.842200 3.046244
[9] 2.642987 3.057670 2.916675 3.117208 3.114286 2.715667 2.892660 3.380230
[17] 2.965027 2.679207 2.581345 2.876582 2.561989 3.472265 2.563211 2.268194
[25] 2.882049 3.050586 2.517621 2.703146 2.634748 2.622611 2.027248 2.708270
[33] 2.944224 1.996112 3.037790 2.578882 2.375542 2.897508 1.972140 2.870625
[41] 2.597007 2.790595 2.235320 2.453876 2.404599 2.863774 2.603077 2.821195
> spl = split(d$diversity, d$donor_status)
> spl
$Healthy
[1] 2.369679 2.689382 2.863240 3.041673 3.000449 2.972395 2.842200 3.046244
[9] 2.642987 3.057670 2.916675 3.117208 3.114286 2.715667 2.892660 3.380230
[17] 2.965027 2.679207 2.581345 2.876582 2.561989 3.472265 2.563211 2.870625
$PD
[1] 2.561989 2.563211 2.268194 2.530586 2.517621 2.634748 2.622611 2.027248
[9] 2.708270 2.944224 1.996112 3.037790 2.578882 2.375542 2.897508 1.972140
[17] 2.597007 2.790595 2.235320 2.453876 2.404599 2.863774 2.603077 2.821195

> mean(spl$Healthy)
[1] 2.90387
> mean(spl$PD)
[1] 2.541922
> pv <- ks.test(spl$Healthy, spl$PD)$p.value
> pv
[1] 0.0001040143
> padj <- p.adjust(pv)
> padj
[1] 0.0001040143
> 
```

4. Utiliza el paquete microbiome para obtener los índices de diversidad alfa, ¿Cuál es la muestra con menor índice de Shannon y qué valor tiene?

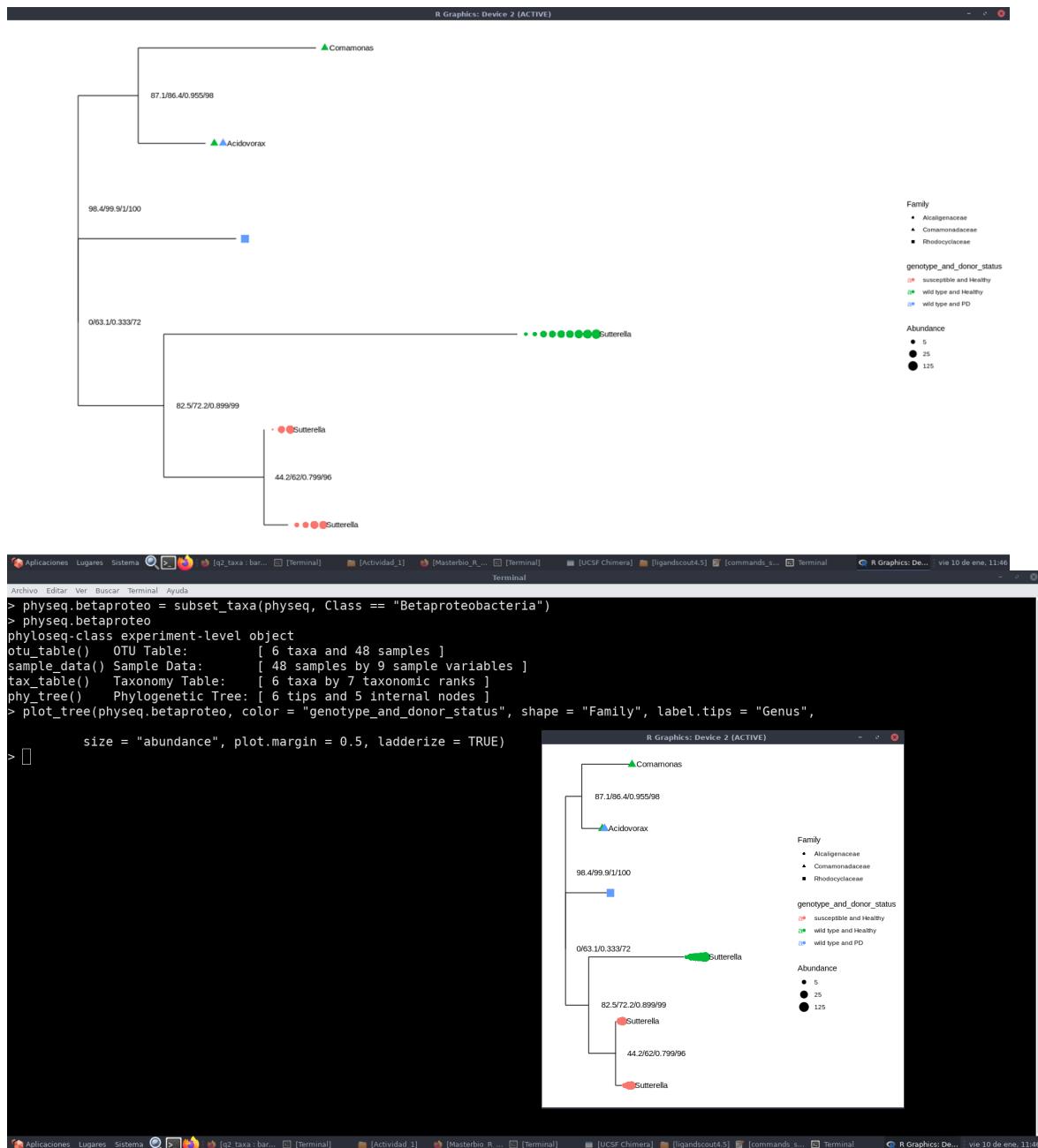
La muestra con menor índice de Shannon es: recip.466.ASO.PD3.D7, con un índice de Shannon de 1.972140.

```

Archivo Editar Ver Buscar Terminal Ayuda
> orden_shannon <- microbiome::diversity(ASV)[order(as.numeric(as.character(microbiome::diversity(ASV)$shannon)), decreasing = FALSE),]
> View(orden_shannon)
> head(orden_shannon)
  inverse_simpson gini_simpson shannon fisher coverage
recip.466.ASO.PD3.D7 4.703581 0.7873960 1.972140 4.015996 2
recip.463.WT.PD3.D21 4.773133 0.7904944 1.996112 4.323769 2
recip.462.WT.PD3.D49 4.490889 0.7773272 2.027248 4.626499 2
recip.537.WT.PD4.D21 6.513110 0.8464635 2.235320 4.405411 3
recip.460.WT.HC3.D49 7.876562 0.8730411 2.268194 2.665856 3
recip.220.WT.OB1.D7 5.265387 0.8100804 2.369679 6.109143 2

Data: orden_shannon
   row.names  inverse_simpson gini_simpson shannon fisher coverage
 1 recip.466.ASO.PD3.D7 4.703581 0.7873960 1.972140 4.015996 2
 2 recip.465.WT.PD3.D14 6.796492 0.8519424 2.379542 5.056213 3
 3 recip.465.WT.PD3.D14 6.960102 0.8695242 2.380124 5.056213 3
 4 recip.539.WT.PD3.D21 7.512519 0.8686265 2.268194 2.480370 3
 5 recip.460.WT.HC3.D49 7.07862 0.9704411 2.369679 6.489956 3
 6 recip.220.WT.OB1.D8 5.263937 0.8100094 2.369679 6.109143 2
 7 recip.465.WT.PD3.D14 6.796492 0.8519424 2.379542 5.056213 3
 8 recip.465.WT.PD3.D14 6.960102 0.8695242 2.380124 5.056213 3
 9 recip.539.WT.PD3.D21 7.512519 0.8686265 2.268194 2.480370 3
10 recip.461.WT.HC3.T21 7.512548 0.9704407 2.351762 5.409957 3
11 recip.461.WT.HC3.T21 7.962438 0.9741015 2.351762 5.409957 3
12 recip.459.WT.HC3.D21 6.895740 0.8540831 2.561989 6.210812 3
13 recip.459.WT.HC3.D21 6.952105 0.8695242 2.561989 6.210812 3
14 recip.454.WT.PD3.D4 9.291000 0.9704407 2.351762 5.409957 3
15 recip.459.WT.HC3.D4 7.078674 0.9582474 2.561989 6.210812 3
16 recip.457.WT.HC3.D4 6.798896 0.8526594 2.561989 6.210812 3
17 recip.540.WT.HC4.D14 6.854695 0.8844479 2.683097 6.585690 4
18 recip.462.WT.PD3.D4 9.047469 0.9804236 2.561989 6.210812 3
19 recip.462.WT.PD3.D4 9.047469 0.9804236 2.561989 6.210812 3
20 recip.465.WT.PD2.T21 9.876595 0.9897883 2.542897 5.269874 4
21 recip.459.WT.HC3.D4 8.944940 0.8880950 2.679007 5.646959 3
22 recip.290.WT.OB1.D11 10.371238 0.9035795 2.688830 5.453139 4
23 recip.462.WT.PD3.D7 9.188539 0.9716531 2.561989 6.210812 3
24 recip.462.WT.PD3.D7 9.188539 0.9716531 2.561989 6.210812 3
25 recip.412.WT.HC2.D7 9.876595 0.9897883 2.542897 5.269874 4
26 recip.538.WT.PD4.149 9.027310 0.8895251 2.796595 7.754638 4
27 recip.540.WT.HC4.D14 10.426836 0.9441044 2.821195 7.284882 4
28 recip.400.WT.HC2.D14 12.216895 0.9351447 2.642201 6.870706 5
29 recip.399.WT.HC7.O7 12.06789 0.9711302 2.683246 5.277016 5
  
```

5. Haz un subset de la clase Betaproteobacteria y representa en un árbol filogenético cada uno de los géneros contenidos en esta clase, distinguiéndolos con un marcador que indique la familia en el árbol y coloreando las muestras según la variable "genotype_and_donor_status". Adjunta el código que utilizas para crearlo, así como el gráfico generado y qué conclusiones extraes de la observación de este árbol.



Código empleado:

Se hace un subset de la clase Betaproteobacteria

`physeq.betaproteo = subset_taxa(physeq, Class == "Betaproteobacteria")`

Y se hace el árbol filogenético coloreando por la variable “genotype_and_donor_status”, usando una forma para distinguir cada familia y etiquetando al final de las líneas los géneros representados con el código:

```
plot_tree(physeq.betaproteo, color = "genotype_and_donor_status", shape = "Family",
label.tips = "Genus", size = "abundance", plot.margin = 0.5, ladderize = TRUE)
```

Hay 3 géneros representados y 3 familias. La familia Alcaligenaceae con el género Sutterella representado, la familia Comamonadaceae con el género Acidovorax y Comamonas representado, y la familia Rhodocyclaceae, con ningún género asignado.

Los géneros están agrupados por su proximidad filogenética, por lo que se representan más juntos los que pertenecen a la misma familia.

En los organismos con el genotipo salvaje y estado saludable el género más abundante que tienen de la clase Betaproteobacteria parece ser el género Sutturella. En cuanto al genotipo susceptible y estado saludable (algo menos abundante que el anterior), pero que el género también mayor representado es Sutturella, seguido de Comamonas o Acidovorax, igual representados. El genotipo salvaje y con enfermedad de Parkinson parece el menos abundante, representado por el género Acidovorax y género no asignado de la familia Rhodocyclaceae.