

LU3SV530 - Metagenomique du Sol
Analyse bioinformatique de données gène de l'ARNr 16S
TD 2 - Phyloseq

Djelika TRAORE & Martin LARSEN

www.immulab.fr

INSERM U1135, CHU Pitié-Salpêtrière, Paris, France

1

La diversité alpha



2

Description de l'écologie microbienne

- La diversité alpha
 - Biomasse
 - Richesse
 - Homogénéité de composition

3

La diversité alpha

- La **biomasse**, en microbiologie (microbiomasse) :
 - Valeur quantitative
 - Masse totale de l'ensemble des micro-organismes vivant dans un milieu, qu'il soit naturel ou artificiel, à un moment donné
- La **richesse** fait référence au nombre total d'espèces observées dans un échantillon.

4

Exercice : Déterminer la diversité pour différentes compositions de microbiote

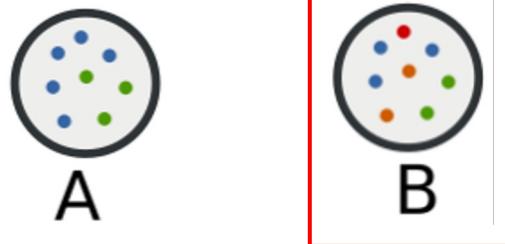


Quel est l'échantillon avec la diversité la plus grande ?
Pourquoi ?

9

Exercice : Déterminer la diversité pour différentes compositions de microbiote

1. La diversité alpha : Richesse observée



2 espèces < 4 espèces

- L'alpha diversité peut être défini par le **nombre d'espèces**.
- Cette mesure se nomme **richesse observée**.

10

Exercice : Déterminer la diversité pour différentes compositions de microbiote



A



B

Quel est l'échantillon avec la diversité la plus grande ?
Pourquoi ?

11

Exercice : Déterminer la diversité pour différentes compositions de microbiote

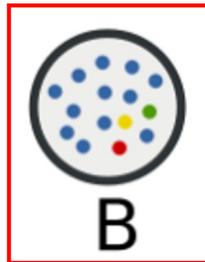
1. La diversité alpha : Richesse observée



A

3 espèces

<



B

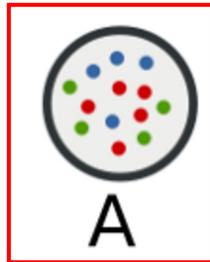
4 espèces

- Basé sur le nombre d'espèces : Echantillon B à une richesse observée plus importante !

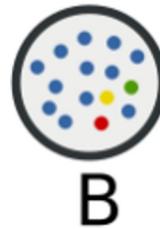
12

Exercice : Déterminer la diversité pour différentes compositions de microbiote

2. La diversité alpha : L' indice de Shannon



3 espèces



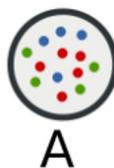
4 espèces

- Dans certain cas c'est l'échantillon A qui à la plus grande alpha diversité !
- Il existe une autre mesure, appelé l'indice de Shannon, qui permet de prendre en compte le **nombre d'espèces** ET leur **abondance**.

13

Exercice : Déterminer la diversité pour différentes compositions de microbiote

2. La diversité alpha : L' indice de Shannon



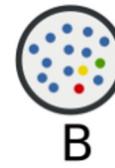
A

3 espèces dont : 4 vertes, 5 rouges, 4 bleues

$$H = -\left(\frac{4}{13} \log\left(\frac{4}{13}\right) + \frac{5}{13} \log\left(\frac{5}{13}\right) + \frac{4}{13} \log\left(\frac{4}{13}\right)\right) \approx 1,09$$

$$E_H = H/\log(k) = 1,09/\log(3) = 0,99$$

- $H = -\sum_{i=1}^k p_i \ln(p_i)$; p=abundance
- $E_H = \frac{H}{\ln(k)}$; [Normalized]



B

4 espèces dont : 1 verte, 1 rouge, 1 jaune, 12 bleues

$$H = -\left(\frac{1}{15} \log\left(\frac{1}{15}\right) + \frac{1}{15} \log\left(\frac{1}{15}\right) + \frac{1}{15} \log\left(\frac{1}{15}\right) + \frac{12}{15} \log\left(\frac{12}{15}\right)\right) \approx 0,72$$

$$E_H = H/\log(k) = 0,72/\log(4) = 0,51$$

14

La diversité bêta



15

Description de l'écologie microbienne

- La diversité alpha
 - Biomasse
 - Richesse
 - Homogénéité de composition
- La diversité bêta
 - *Absolute species turnover*
 - Basé sur l'abondance

16

Description de l'écologie microbienne

- La diversité alpha
 - Biomasse
 - Richesse
 - Homogénéité de composition
- La diversité bêta
 - *Absolute species turnover*
 - Basé sur l'abondance
- Note: La diversité alpha et bêta peuvent être évalués à différent niveau phylogénétique

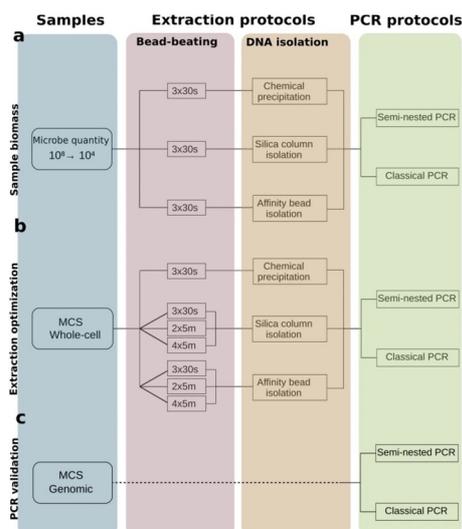
17

Décrire l'écologie microbienne dans un contexte précis

Article | [Open access](#) | Published: 24 May 2021

Refinement of 16S rRNA gene analysis for low biomass biospecimens

Remy Villette, Gaëlle Autaa, Sophie Hind, Johanna B. Holm, Alicia Moreno-Sabater & Martin Larsen



Différentes variables sont présentes :

- Le type d'échantillons
 - **microbiote humain (a),**
 - **communautés microbiennes (b) (c)** (contrôle positif pour : l'extraction d'ADN, l'amplification du gène et le séquençage du gène bactérien)
- La quantité de microbes
- Le type de PCR
- Le type d'extraction d'ADN bactérien

But : La biomasse a t-elle un impact sur le séquençage du gène de l'ARNr 16S ?

18

La biomasse a-t-elle un impact sur le séquençage du gène de l'ARNr 16S ?

1. Mesurer la diversité alpha pour différentes biomasses de microbiote

1. Comparer la composition phylogénétique

- Au niveau *Phylum*
- Au niveau *Class*
- Au niveau *Genus*

3. Comparer la diversité microbienne des échantillons entre eux

19

1. Mesurer la diversité alpha pour différentes biomasses de microbiote

A. Importation de l'objet phyloseq

```
pth = file.choose()
ps = readRDS(pth)
```

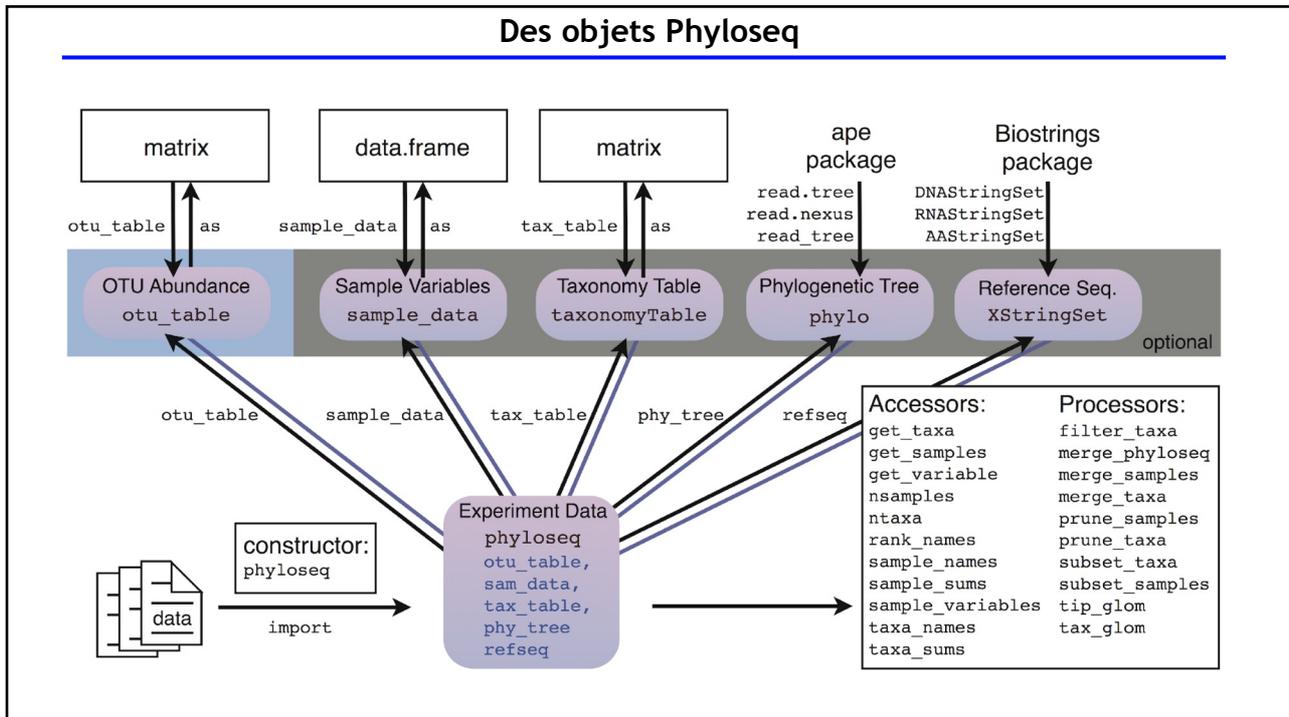


```
`` {r}
ps = readRDS("/Users/djelika/Documents/2019-11 Sci Rep Villette Scarcity Paper/Metadata/2021-05 Sci Rep
Villette et al 16S refinement/2021-05 VILLETTE et al 16S Refinement ps.rds")

ps
````

phyloseq-class experiment-level object
otu_table() OTU Table: [515 taxa and 37 samples]
sample_data() Sample Data: [37 samples by 9 sample variables]
tax_table() Taxonomy Table: [515 taxa by 7 taxonomic ranks]
phy_tree() Phylogenetic Tree: [515 tips and 513 internal nodes]
```

20



21

## 1. Mesurer la diversité alpha pour différente biomasse de microbiote

B. Calculer la richesse observée et l'indice de Shannon pour chaque échantillon

```
Représenter la diversité alpha
```{r}
richness = as.data.frame(estimate_richness(human, measures = c("Shannon", "Observed")))
head(richness)
```
```

Description: df [6 × 2]

|                                   | Observed<br><dbl> | Shannon<br><dbl> |
|-----------------------------------|-------------------|------------------|
| RAW_HD1.104miniprep_16S_TAGGCT    | 37                | 2.724880         |
| RAW_HD1.105miniprep_16S_TGCCTT    | 16                | 2.666398         |
| RAW_HD1.105miniprepbis_16S_TTCTTG | 28                | 2.809098         |
| RAW_HD1.106miniprep_16S_ATAAGA    | 23                | 2.494274         |
| RAW_HD1.106miniprepbis_16S_CCGACC | 118               | 3.913311         |
| RAW_HD1.107miniprep_16S_CACACT    | 74                | 3.654130         |

6 rows

22

## 1. Mesurer la diversité alpha pour différente biomasse de microbiote

### C. Représenter la diversité alpha

```
#Mesurer l'alpha-diversité pour différente concentration de microbiote
```{r}
# Rajouter des variables importantes pour la suite de l'analyse

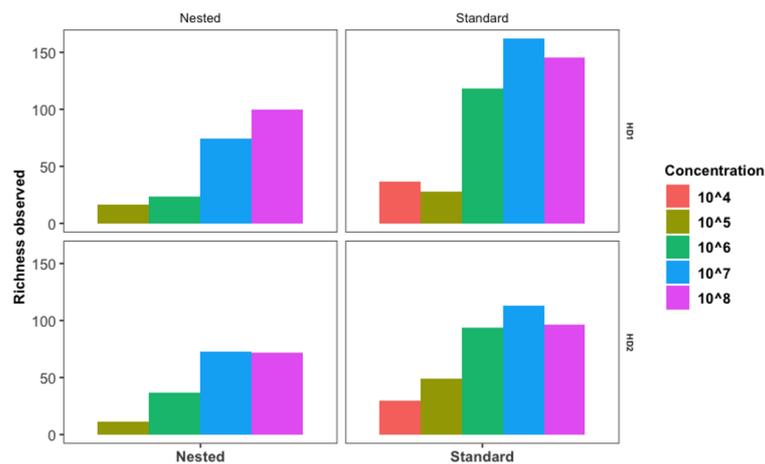
richness$SampleOrigin<- human@sam_data$SampleOrigin #paste sample info
richness$PCR= human@sam_data$PCR
richness$Concentration = human@sam_data$Concentration

ggplot(richness, aes(x=PCR, y=Observed, fill=Concentration))+
  geom_bar(stat='identity', position = "dodge")+
  facet_grid(SampleOrigin~PCR, scales = "free_x", space = "free")+
  ylab("Richness observed")+
  theme_bw()+
  theme(strip.text.y = element_text(size = 6, face = "bold"), strip.background = element_blank(),
        legend.text = element_text(size = 10, face="bold"),
        axis.text.x = element_text(size = 10, face = "bold"),
        legend.title = element_text(size = 10, face="bold"),
        axis.title.x = element_blank(),
        axis.title.y=element_text(size = 10, face="bold"),
        axis.text.y = element_text(size = 10),
        panel.grid = element_blank())
```
```

23

## 1. Mesurer la diversité alpha pour différente biomasse de microbiote

### C. Représenter la diversité alpha



24

## 2. Comparer la composition phylogénétique

### A. Composition au niveau Phylum



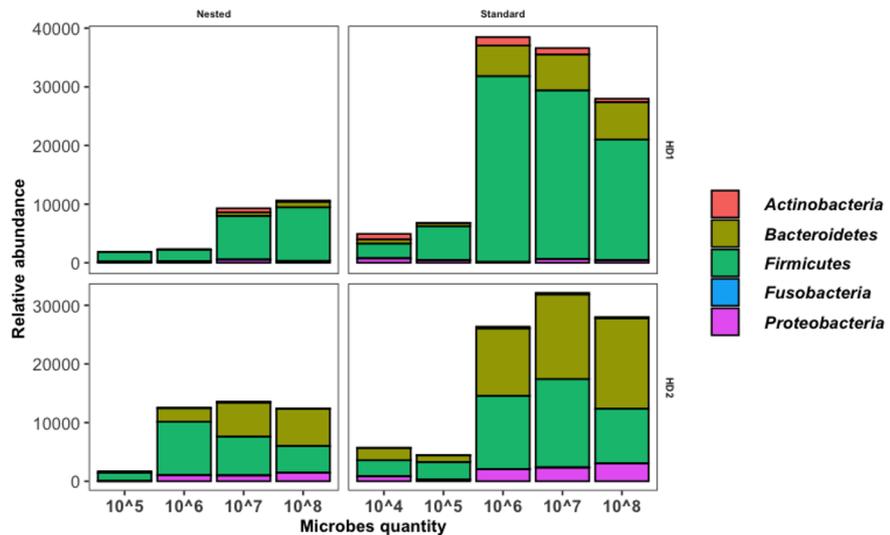
```
#Composition microbienne au niveau Phylum
```{r}
Phylum <- tax_glom(human, "Phylum")

plot_bar(Phylum, x = "Concentration", fill = "Phylum")+
  facet_grid(SampleOrigin-PCR, space = "free", scales = "free")+
  labs(x="Microbes quantity", y="Relative abundance")+
  theme_bw()+
  theme(strip.background = element_blank(),
        axis.title.x = element_text(size=10, face="bold"),
        strip.text = element_text(size = 6, face = "bold"),
        axis.text.x = element_text(size = 10, face = "bold"),
        axis.text.y = element_text(size = 10),
        axis.title.y = element_text(size = 10, face="bold"),
        legend.title = element_blank(),
        legend.text = element_text(size = 10, face="bold.italic"),
        legend.spacing.x = unit(0.5,"cm"),
        panel.grid = element_blank())
```
```

25

## 2. Comparer la composition phylogénétique

### A. Composition au niveau Phylum



26

## 2. Comparer la composition phylogénétique

### B. Composition au niveau Class



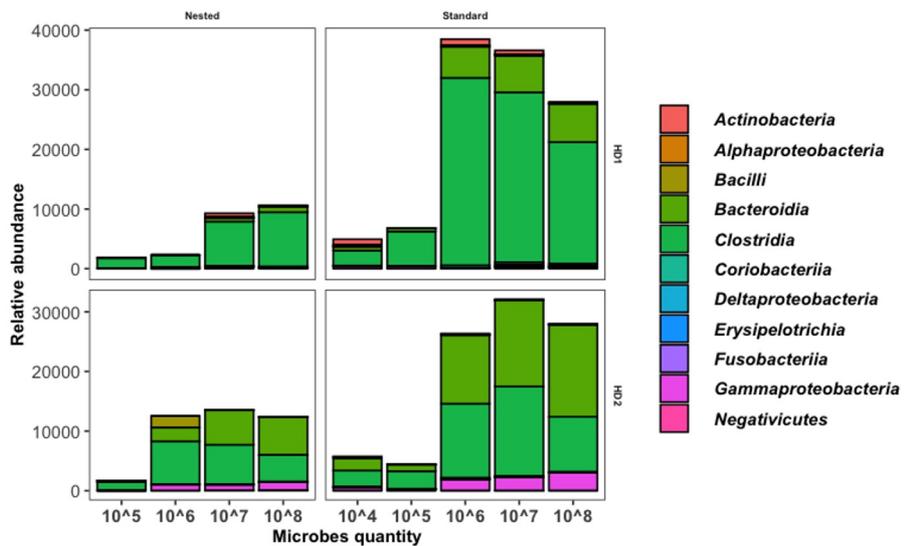
```
#Composition microbienne au niveau Class
```{r}
Class<- tax_glom(human, "Class")

plot_bar(Class, x = "Concentration", fill = "Class")+
  facet_grid(SampleOrigin~PCR, space = "free", scales = "free")+
  labs(x="Microbes quantity", y="Relative abundance")+
  theme_bw()+
  theme_bw()+
  theme(strip.background = element_blank(),
        axis.title.x = element_text(size=10, face="bold"),
        strip.text = element_text(size = 6, face = "bold"),
        axis.text.x = element_text(size = 10, face = "bold"),
        axis.text.y = element_text(size = 10),
        axis.title.y = element_text(size = 10, face="bold"),
        legend.title = element_blank(),
        legend.text = element_text(size = 10, face="bold.italic"),
        legend.spacing.x = unit(0.5,"cm"),
        panel.grid = element_blank())
```
```

27

## 2. Comparer la composition phylogénétique

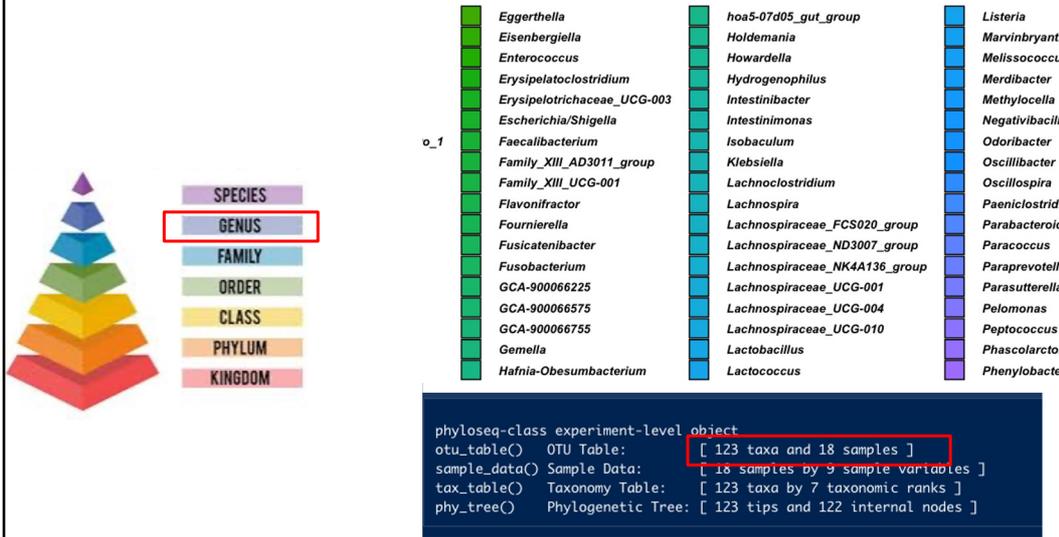
### B. Composition au niveau Class



28

## 2. Comparer la composition phylogénétique

### C. Composition au niveau Genus

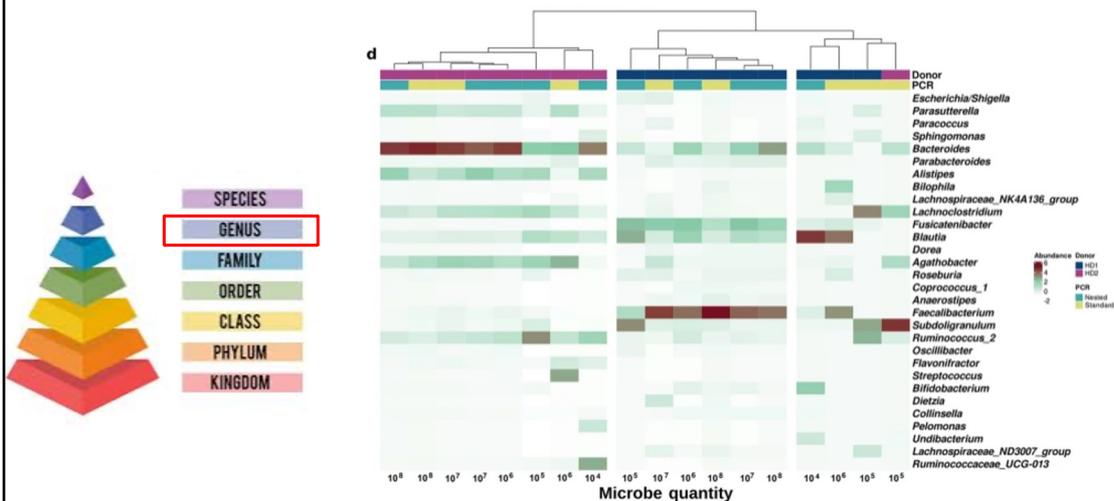


Representation graphique illisible (trop grand nombre de Genus) ⇒ Changer de représentation !

29

## 2. Comparer la composition phylogénétique

### C. Composition au niveau Genus - Heatmap



Exemple de représentation possible : la heatmap !

30

## 2. Comparer la composition phylogénétique

### C. Composition au niveau **Genus** - Heatmap

```
Préparer les annotations
```{r}
Heat<- tax_glom(ps, "Genus")
Heat<- prune_taxa(names(sort(taxa_sums(Heat),TRUE)[1:50]), Heat)

x= t(as.matrix(Heat@otu_table@.Data))

dend= hclust(distance(ps, method="bray"), "ward.D2")
dend= hclust(vegdist(ps@otu_table, method="bray"), "ward.D2")

dend= dendextend:: click_rotate(as.dendrogram(dend))

group= HeatmapAnnotation(Donor=Heat@sam_data$SampleName,
  PCR=Heat@sam_data$PCR ,
  col=list(Donor= c("HD1"="#114477", "HD2"="#AA4488"),
  PCR= c("Nested"="#44AAAA", "Standard"="#DDDD77")),
  show_annotation_name = T,
  annotation_name_gp = gpar(fontsize=15, fontface="bold"))

name= HeatmapAnnotation(Concentration= anno_text(Heat@sam_data$Concentration,
  bp = gpar(fontsize=14, fontface="bold")))
```
```

```
##Heatmap
```{r}

Heatmap(as.matrix(scale(x)),
  top_annotation = group,
  bottom_annotation = name,
  cluster_rows = F,
  cluster_columns = as.dendrogram(dend),
  row_labels = Heat@tax_table[, "Genus"],
  row_names_gp = gpar(fontsize=15, fontface="bold.italic"),
  show_column_names = F,
  col = c("white", "#88CCAA", "#771122"),
  column_dend_height = unit(3, "cm"),
  name = "Abundance",
  column_split = 3,
  column_gap = unit(5, "mm"),
  column_title = NULL)
```
```

31

## 2. Comparer la composition phylogénétique

### C. Composition au niveau **Genus** - Heatmap

- Méthode facilitant la visualisation de larges données à l'aide d'un nuancier de couleur (exemple : bleu → rouge)
- Dans notre cas : Comparer l'abondance relative des *Genus* retrouvés dans les différents échantillons

32

### 3. Comparer la diversité microbienne des échantillons entre eux

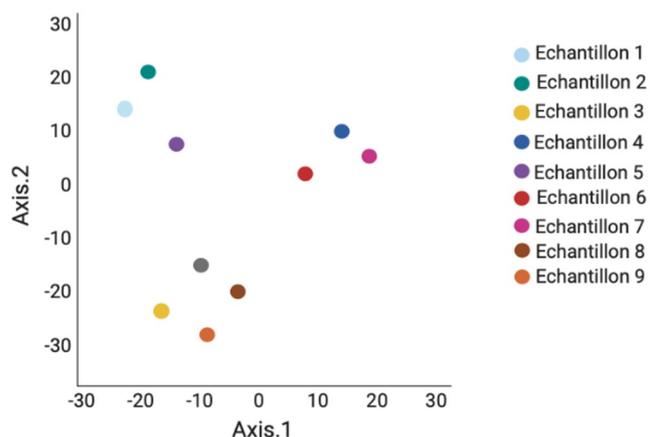
#### A. La diversité bêta

- Comparer la composition taxonomique entre différents échantillons.
- Distance inter-échantillon
- Il existe plusieurs types de matrice de distances possibles basées sur :
  - Occurrence (Jaccard)
  - Abondance (Bray-curtis)
  - Phylogénie (Unifrac)

33

### 3. Comparer la diversité microbienne des échantillons entre eux

#### A. La diversité bêta



- Chaque échantillon est représenté par un point et la distance entre les points représente la similarité entre les échantillons.
- Plus la distance entre les échantillons est courte, plus ils sont similaires.

34

### 3. Comparer la diversité microbienne des échantillon entre eux

#### A. La diversité bêta

```
#Dsitance entre échantillon

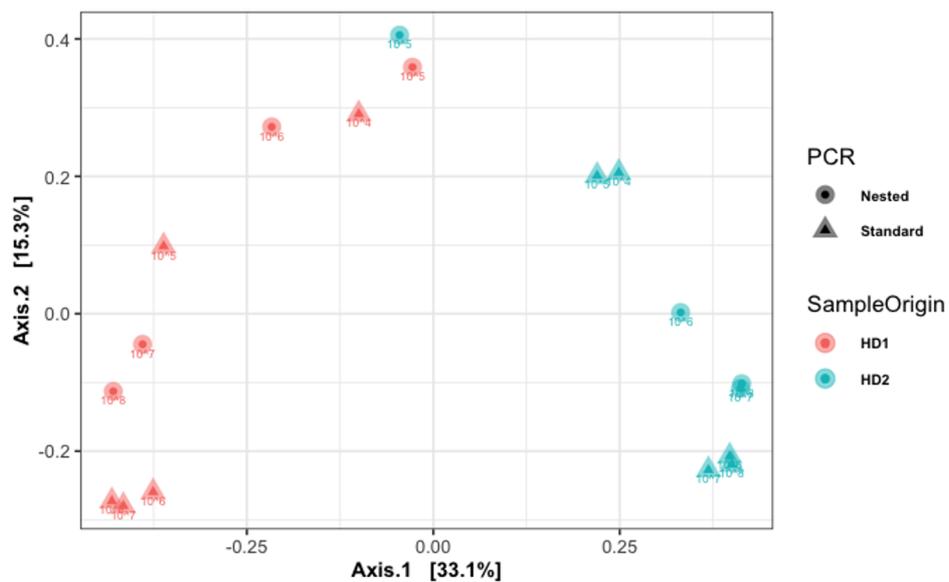
```{r}
ord <- ordinate(human, method = "PCoA", distance = "bray")
dist <- phyloseq::distance(human, method = "bray")

plot_ordination(, ord, color = "Concentration", shape="PCR",label = "SampleOrigin")+
  geom_point(size=4, alpha=0.5)+
  theme_bw()+
  theme(legend.text = element_text(size = 7, face = "bold"),
        axis.title = element_text(size = 10, face = "bold"),
        legend.spacing.x = unit(0.3, "cm"))
```
```

35

### 3. Comparer la diversité microbienne des échantillon entre eux

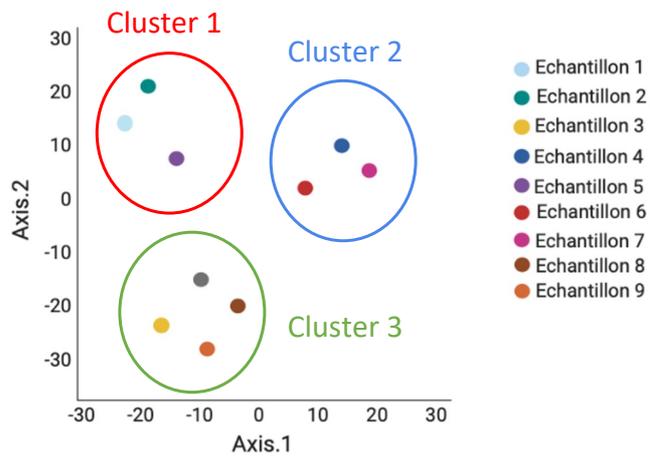
#### A. La diversité bêta



36

### 3. Comparer la diversité microbienne des échantillons entre eux

#### A. La clusterisation

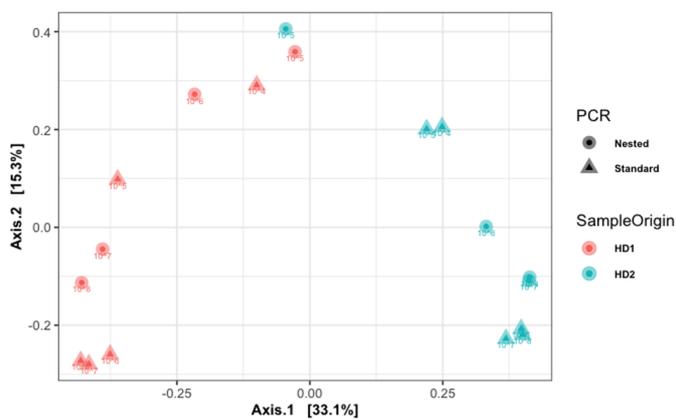


- Chaque échantillon est représenté par un point et la distance entre les points représente la similarité entre les échantillons.
- Plus la distance entre les échantillons est courte, plus ils sont similaires.

37

### 3. Comparer la diversité microbienne des échantillons entre eux

#### A. La clusterisation



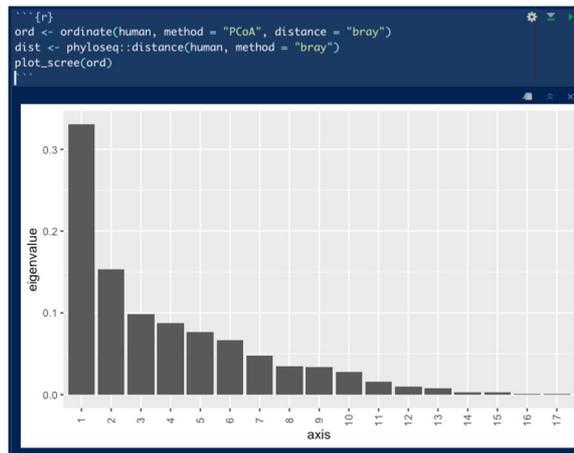
→ Combien de clusters en tout ?

38

### 3. Comparer la diversité microbienne des échantillons entre eux

#### B. La clusterisation

Certaines méthodes de calculs mathématiques permettent de déterminer le nombre exact de clusters à définir en fonction du jeu de donnée.



40

### 3. Comparer la diversité microbienne des échantillons entre eux

#### B. La clusterisation

```
##2 clusters
```{r}
clust <- hclust(dist, method = "ward.D2")
clust$labels = with(sample_data(human), paste(SampleOrigin, PCR, Concentration))
sample_data(human)$cluster = as.factor(cutree(clust, k=2))

tmp2 = ord$vectors[,1:2]
tmp2 = as.data.frame(merge(tmp2, sample_data(human), by="row.names"))

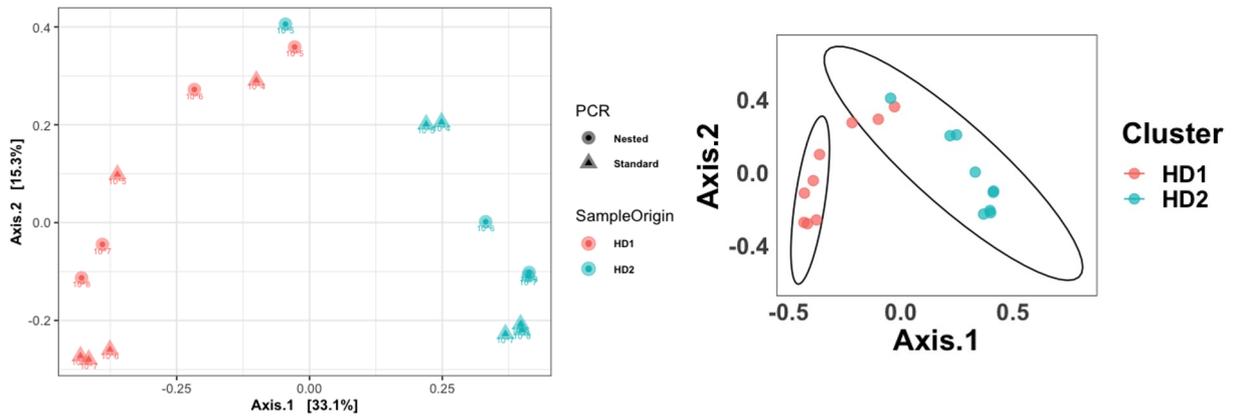
tmp2 %>%
  ggplot(aes(x=Axis.1, y=Axis.2, color=SampleOrigin, group=cluster))+
  stat_ellipse()+
  geom_point(alpha=0.7, size=3)+
  labs(x="Axis.1", y="Axis.2", color="Cluster")+
  theme_bw(base_line_size = 0)+
  theme(legend.position = "right",
        text = element_text(size = 20, face = "bold"))
```

41

3. Comparer la diversité microbienne des échantillons entre eux

B. La clusterisation

2 clusters?

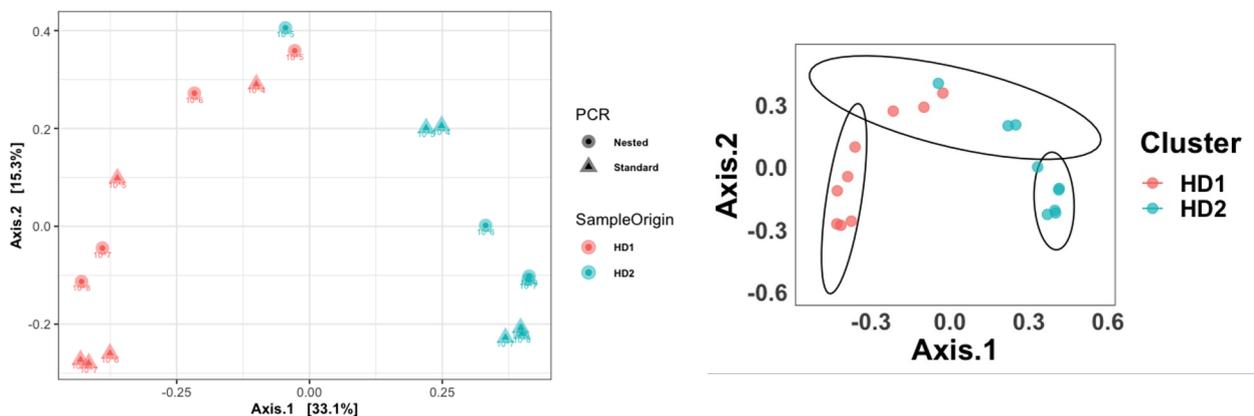


42

3. Comparer la diversité microbienne des échantillons entre eux

B. La clusterisation

3 clusters ?



43

3. Comparer la diversité microbienne des échantillons entre eux

C. La HEATMAP

```
##Heatmap
```{r}
Heatmap(as.matrix(scale(x)),
 top_annotation = group,
 bottom_annotation = name,
 cluster_rows = F,
 cluster_columns = as.dendrogram(dend),
 row_labels = Heat@tax_table[, "Genus"],
 row_names_gp = gpar(fontsize=15, fontface="bold.italic"),
 show_column_names = F,
 col = c("white", "#88CCAA", "#771122"),
 column_dend_height = unit(3, "cm"),
 name = "Abundance",
 column_split = 3,
 column_gap = unit(5, "mm"),
 column_title = NULL)
```
```

46