

Gut microbial enterotypes in COPD patients

Zhaowei Yang

2/18/2020

Reference-based enterotype analysis

We use an enterotype classification model from Paul I Costea et al., Nat Microbiol, 2018. The analysis was ran directly on <http://enterotypes.org/>

Beta-diversity analysis

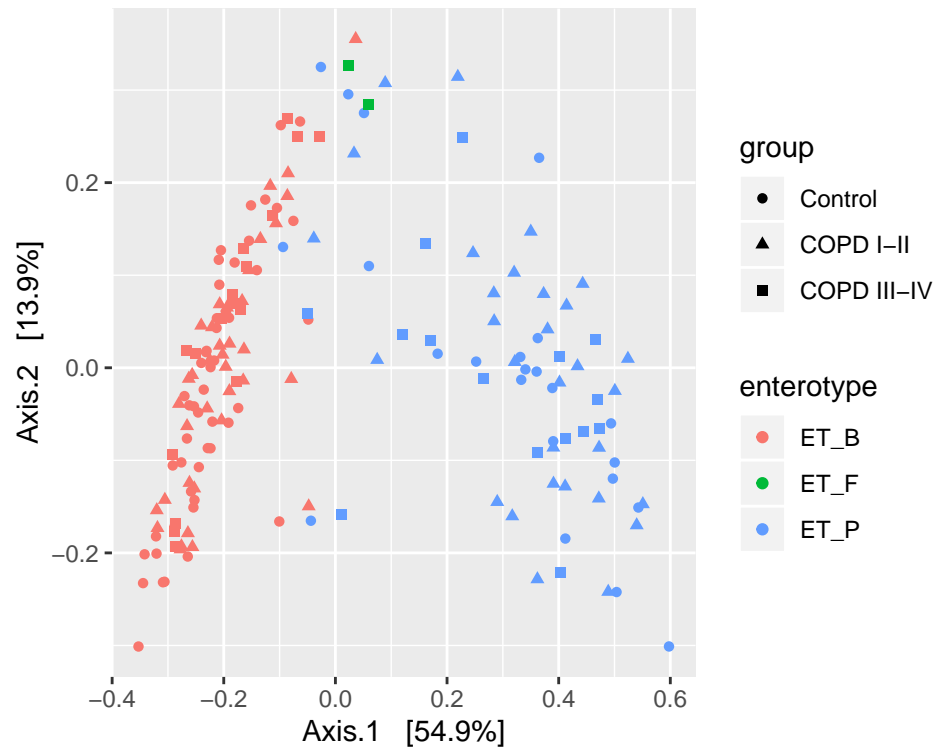
```
## Loading genus level data & reference-based enterotypes
data.genus <- read.csv("data/naijian_genus.csv", row.names = 1)
metadata <- read.csv("data/metadata.csv")
enterotype <- read.csv("enterotypes_org_reanalysis/enterotype_genus_compare.csv", row.names = 1)
metadata$enterotype <- enterotype$Within_ET_space

# rename group
metadata$group <- revalue(metadata$group, c("B"="Control", "C"="COPD I-II", "D"="COPD III-IV"))
metadata$group2 <- as.factor(ifelse(metadata$group=="Control", "Control", "COPD"))

## prepare tax table
taxmat = matrix(NA, nrow = nrow(data.genus), ncol = 7)
rownames(taxmat) <- rownames(data.genus)
colnames(taxmat) <- c("Domain", "Phylum", "Class", "Order", "Family", "Genus", "Species")
taxmat[,7] <- rownames(data.genus)

## create a phyloseq object
OTU = otu_table(data.genus, taxa_are_rows = TRUE)
TAX = tax_table(taxmat)
rownames(metadata) <- metadata$ID
META = sample_data(metadata)
ps = phyloseq(OTU, TAX, META)

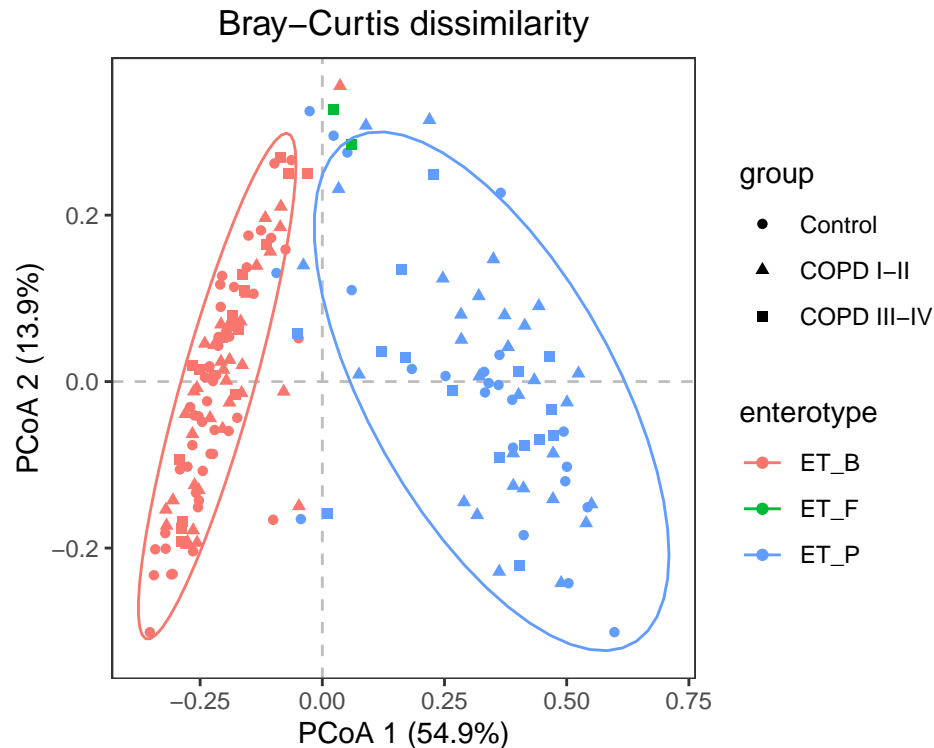
## calculate distance
# Calculate distance matrix
iDist <- distance(ps, method="bray")
# Calculate ordination
iMDS <- ordinate(ps, "PCoA", distance=iDist)
p1 <- plot_ordination(ps, iMDS, color="enterotype", shape="group")
p1
```



reproduce the PCoA plot

```
p2 <- ggplot(p1$data, aes(Axis.1, Axis.2, color=enterotype)) +
  geom_hline(yintercept = 0, color="grey", linetype=2) +
  geom_vline(xintercept = 0, color="grey", linetype=2) +
  geom_point(aes(shape=group)) +
  labs(x="PCoA 1 (54.9%)", y="PCoA 2 (13.9%)", title = "Bray-Curtis dissimilarity") +
  stat_ellipse() + theme_bw() +
  theme(plot.title = element_text(hjust = 0.5),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())
p2
```

Too few points to calculate an ellipse



PERMANOVA test (vegan::Adonis, bray)

We first did a PERMANOVA test on enterotypes and then on two-level disease groups (Control vs. COPD). At last we compared three-level groups (Control vs. COPD I-II vs. COPD III-IV).

```
metadata <- as(sample_data(ps), "data.frame")
a1 <- adonis(distance(ps, method="bray") ~ enterotype, data = metadata)
a1

##
## Call:
## adonis(formula = distance(ps, method = "bray") ~ enterotype,      data = metadata)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##              Df SumsOfSqs MeanSqs F.Model    R2 Pr(>F)
## enterotype    2    11.958   5.9791  75.823 0.47001 0.001 ***
## Residuals   171    13.485   0.0789    0.52999
## Total       173    25.443
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

print(levels(metadata$group2))
```

```
## [1] "Control" "COPD"

a2 <- adonis(distance(ps, method="bray") ~ group2, data = metadata)
a2
```

```
##
## Call:
## adonis(formula = distance(ps, method = "bray") ~ group2, data = metadata)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##              Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## group2        1    0.5984 0.59836  4.1425 0.02352 0.011 *
## Residuals    172    24.8443 0.14444      0.97648
## Total        173    25.4427      1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

print(levels(metadata$group))
```

```
## [1] "Control"      "COPD I-II"    "COPD III-IV"

a3 <- adonis(distance(ps, method="bray") ~ group, data = metadata)
a3
```

```
##
## Call:
## adonis(formula = distance(ps, method = "bray") ~ group, data = metadata)
##
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##              Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## group         2    0.6897 0.34487  2.3824 0.02711 0.034 *
## Residuals    171    24.7530 0.14475      0.97289
## Total        173    25.4427      1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- adonis analysis on enterotype: $p=0.001$
- adonis analysis on group (Control vs. COPD): $p=0.011$
- adonis analysis on group (Control vs. I-II vs. COPD III-IV): $p=0.034$

Enterotype bar plot

```
data.genus <- read.csv("enterotypes_org_reanalysis/enterotype_genus_compare.csv", row.names = 1)
blood_colors <- c("#c10024", "#10497e", "forestgreen")

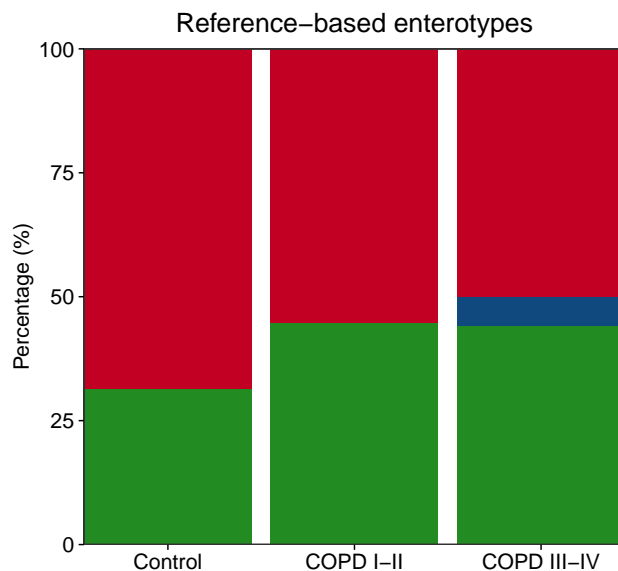
df1 <- table(data.genus[data.genus$subject_group == "B",]$Within_ET_space)/sum(table(data.genus[data.genus$subject_group == "B",]$Within_ET_space))
df2 <- table(data.genus[data.genus$subject_group == "C",]$Within_ET_space)/sum(table(data.genus[data.genus$subject_group == "C",]$Within_ET_space))
df3 <- table(data.genus[data.genus$subject_group == "D",]$Within_ET_space)/sum(table(data.genus[data.genus$subject_group == "D",]$Within_ET_space))

df1 <- data.frame(data=as.numeric(df1), enterotype=names(df1), group="Control")
df2 <- data.frame(data=as.numeric(df2), enterotype=names(df2), group="COPD I-II")
df3 <- data.frame(data=as.numeric(df3), enterotype=names(df3), group="COPD III-IV")
```

```
df <- rbind(df1,df2,df3)

p3 <- ggplot(df, aes(group, data)) +
  geom_col(aes(fill=enterotype)) +
  scale_fill_manual(values=blood_colors) +
  scale_y_continuous(position = "left", expand=c(0,0)) +
  scale_x_discrete(expand=c(0,0)) +
  ylab("Percentage (%)") +
  ggtitle("Reference-based enterotypes") +
  theme_bw(base_size = 8) +
  theme(axis.ticks = element_line(size = 0.75/.pt, colour = "black"),
        axis.text = element_text(size = 8, colour = "black"),
        axis.text.x = element_text(angle = 0, hjust = 0.5),
        axis.title.x = element_blank(),
        title = element_text(size = 8, colour = "black"),
        legend.position = "none",
        panel.grid= element_blank(),
        plot.margin = margin(2,0,0,8, "mm"),
        plot.title = element_text(hjust = 0.5),
        line = element_line(size = 0.75))

p3
```



Save figures

```
ggsave("figure_1.pdf", plot=p2, width = 5, height = 4)
```

```
## Too few points to calculate an ellipse
```

```
ggsave("figure_1.jpg", plot=p2, width = 5, height = 4, dpi = 300)
```

```
## Too few points to calculate an ellipse
```

```
ggsave("figure_2.pdf", plot=p3, width = 3.5, height = 3)
```

```
ggsave("figure_2.jpg", plot=p3, width = 3.5, height = 3, dpi = 300)
```