

envfit_analysis

Jacob Westaway

Last updated on 2021-07-31

About

EnvFit analysis (Vegan package) was used to explore the effect of covariates on microbial community composition between groups (beta diversity), to determine if probiotic treatment has an effect. The EnvFit function was applied to an ordination using NMDS that was based on a Bray-Curtis dissimilarity matrix calculated from data normalised with Total Sum Scaling. The significance was based on 10,000 permutations and was transformed based on the Benjamini-Hochberg (BH) procedure.

Load packages

```
sapply(c("DESeq2", "phyloseq", "dplyr", "ggplot2", "grid",  
        "gridExtra", "ggpubr", "sjPlot", "pheatmap", "tidyverse", "vegan"),  
       require, character.only = TRUE)
```

Explore clustering of variables with PCoA and Bray-Curtis

```
# define function  
ordination_plots <- function(filtered_ps, variable, vis_method, dist_method){  
# ordinate  
ps_ordination <- ordinate(filtered_ps, method = vis_method, distance = dist_method)  
# get eigenvalues  
evals <- ps_ordination$values$Eigenvalues  
# generate plot  
plot_ordination(filtered_ps, ps_ordination, color = variable,  
  title = "PCoA (Bray-Curtis)" +  
  labs(col = variable) +  
  coord_fixed(sqrt(evals[2] / evals[1])) +  
  geom_point(size = 3) +  
  stat_ellipse(type = "norm", linetype = 2) +  
  scale_color_hue(labels = c("Probiotic-treated", "Non-treated"))  
}  
  
ordination_plots(ps2.TSS, "Primary_Group", "PCoA", "bray")
```

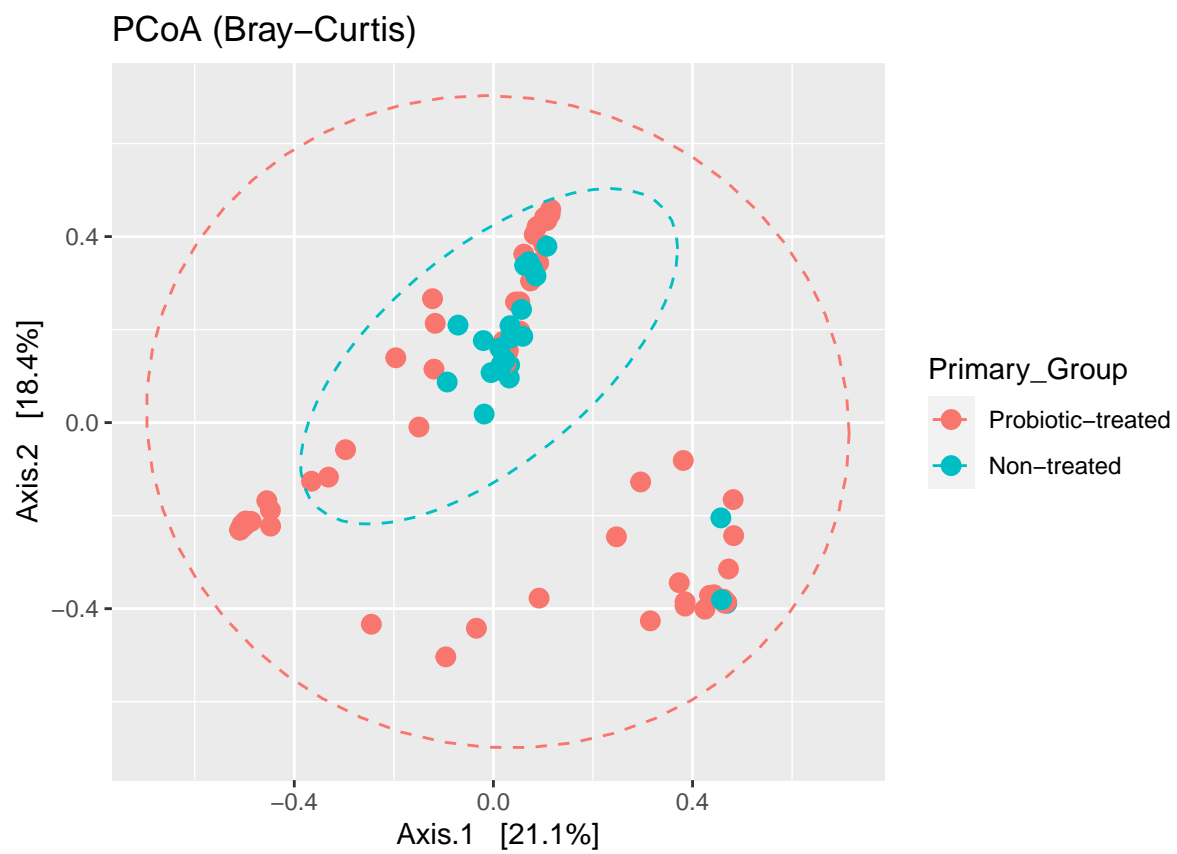


Figure 1: PCoA plot of Bray-Curtis distances coloured by probiotic group.

envfit analysis

```
ordination_distance <- ordinate(
  physeq = ps2.TSS,
  method = "NMDS",
  distance = "bray",
  formula = ~ Primary_Group + Feeding_Type + NEC + Sepsis + Mode_of_Delivery +
    Neonatal_Antibiotics + Chorioamnionitis + Preeclampsia + ROP + Batch +
    Gestational_Age_at_Birth + Diabetes + Antenatal_Antibiotics)

envdat.phy <- ps2.TSS %>%
  sample_data() %>%
  unclass() %>%
  as.data.frame() %>%
  select(Primary_Group, Feeding_Type, NEC, Sepsis, Mode_of_Delivery,
    Neonatal_Antibiotics, Chorioamnionitis, Preeclampsia, ROP, Batch,
    Gestational_Age_at_Birth, Diabetes, Antenatal_Antibiotics) %>%
  centre_and_scale()

set.seed(1)

envfit_object <- envfit(ordination_distance, envdat.phy, permutations = 10000)
```

Adjust p values

```
p.adjust.envfit <- function (x, method = 'bonferroni', n){
  x.new <- x
  if (!is.null (x$vectors)) pval.vectors <- x$vectors$pvals else pval.vectors <- NULL
  if (!is.null (x$factors)) pval.factors <- x$factors$pvals else pval.factors <- NULL
  if (missing (n)) n <- length (pval.vectors) + length (pval.factors)
  if (!is.null (x$vectors)) x.new$vectors$pvals <- p.adjust (x$vectors$pvals,
    method = method, n = n)
  if (!is.null (x$factors)) x.new$factors$pvals <- p.adjust (x$factors$pvals,
    method = method, n = n)
  cat ('Adjustment of significance by', method, 'method')
  return (x.new)
}

set.seed(1)

envfit(ordination_distance, envdat.phy, permutations = 10000) %>%
  p.adjust.envfit(., method = "BH")
```

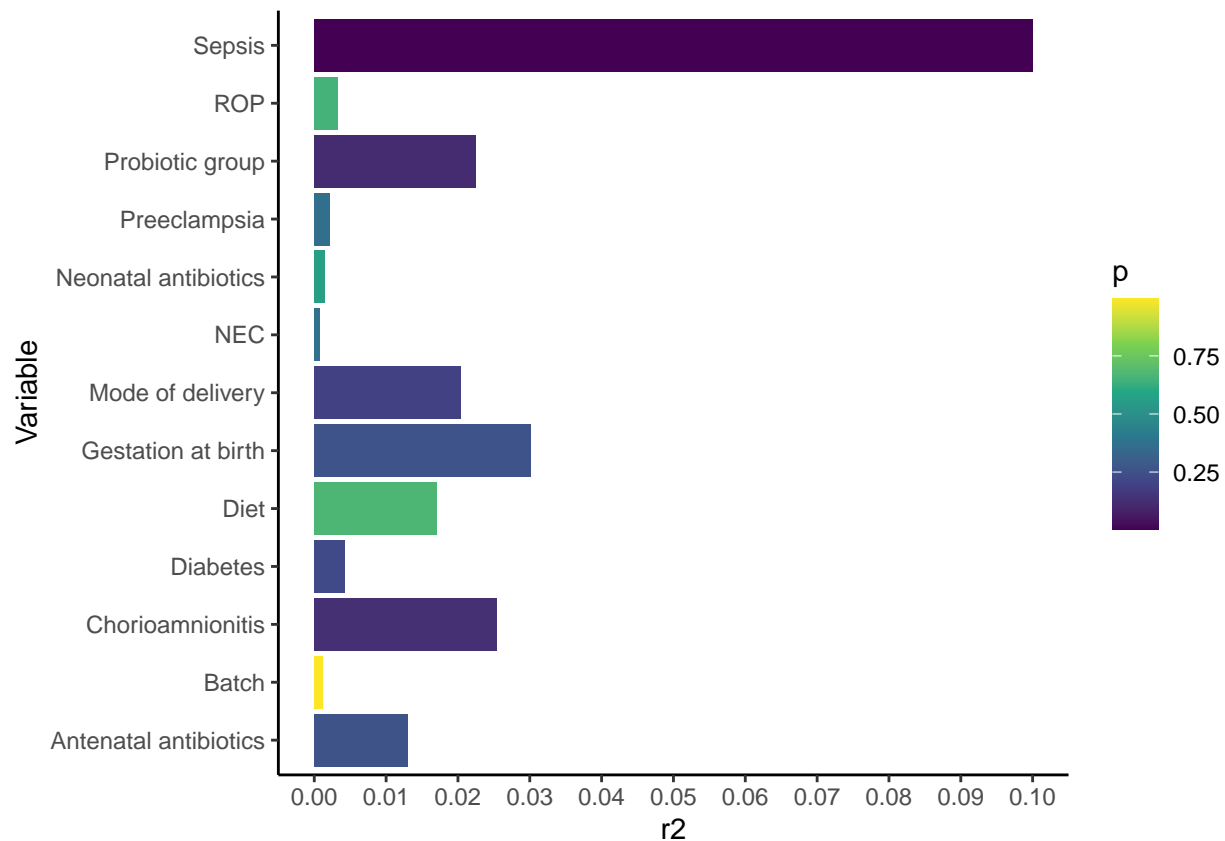
Plot model

```
envfit_object$vectors$r %>%
  as.data.frame() %>%
```

```

rownames_to_column("Variable") %>%
rename("r2" = ".") %>%
left_join(
  (envfit_object$vectors$pvals %>%
as.data.frame() %>%
add_column(Variable = "Gestational_Age_at_Birth") %>%
rename("p" = "."))) %>%
rbind(
envfit_object$factors$r %>%
as.data.frame() %>%
rownames_to_column("Variable") %>%
rename("r2" = ".") %>%
rows_update(., tibble(Variable = "Sepsis", r2 = 0.1)) %>% # changed to 0.1 for plot
left_join(
  envfit_object$factors$pvals %>%
as.data.frame() %>%
rownames_to_column("Variable") %>%
rename("p" = "."))
) %>%
ggplot(aes(x = r2, y = Variable, fill = p)) +
geom_col() +
scale_x_continuous(breaks = scales::pretty_breaks(n = 10), limits = c(0, 0.1)) +
theme(axis.line = element_line(size = .5), panel.background = element_blank()) +
scale_fill_continuous(type = "viridis") +
scale_y_discrete(labels = c("Primary_Group" = "Probiotic group",
  "Neonatal_Antibiotics" = "Neonatal antibiotics",
  "Mode_of_Delivery" = "Mode of delivery",
  "Feeding_Type" = "Diet",
  "Birth_Weight" = "Birthweight",
  "Gestational_Age_at_Birth" = "Gestation at birth",
  "Antenatal_Antibiotics" = "Antenatal antibiotics"))

```



Calculate major contributors (based on adonis)

```
major_contributors <- function(ps2.TSS, variable){
  # perform permanova
  ps_otu <- data.frame(otu_table(ps2.TSS))
  ps_metadata <- data.frame(sample_data(ps2.TSS))
  permanova <- adonis(ps_otu ~Primary_Group + Feeding_Type + NEC + Sepsis +
    Mode_of_Delivery + Neonatal_Antibiotics + Chorioamnionitis +
    Preeclampsia + ROP + Batch + Gestational_Age_at_Birth + Diabetes +
    Antenatal_Antibiotics, data = ps_metadata, method = "bray")

  # coefficients
  coef <- coefficients(permanova)[paste0(variable, "1"),]
  top.coef <- coef[rev(order(abs(coef)))[1:20]]

  genus_contributors <- tax_table(ps2.TSS) %>%
    unclass() %>%
    as.data.frame() %>%
    select("Genus") %>%
    rownames_to_column(var = "ASV") %>%
    right_join((as.data.frame(top.coef) %>%
      rownames_to_column(var = "ASV"))) %>%
    select(!"ASV") %>%
}
```

```

    arrange(top.coef)

return(genus_contributors)
}

major_contributors(ps2.TSS, "Primary_Group") %>%
  filter(top.coef > 0) %>%
  group_by(Genus) %>%
  summarise("top.coef" = mean(top.coef)) %>% # ASVs from same genus
  rbind(
    major_contributors(ps2.TSS, "Primary_Group") %>%
    filter(top.coef < 0) %>%
    group_by(Genus) %>%
    summarise("top.coef" = mean(top.coef))
  ) %>%
  ggplot(aes(x = top.coef, y = Genus, fill = top.coef)) +
  geom_col() +
  theme(axis.line = element_line(size = .5), panel.background = element_blank()) +
  geom_vline(xintercept = 0.0, linetype = 1, size = .5) +
  scale_fill_continuous(type = "viridis", name = "Coefficient") +
  theme(axis.title.x = element_blank())

```

