

Alpha diversity

Stijn Schreven

4 March 2021

Contents

Load packages	1
Input files	1
1. Prepare Data	2
3. Phylogenetic diversity	2
3.1. Calculate metrics	2
3.2. LMM regression	3
4. Errorbar plot	6
4.1. Collect estimates in dataframe	6
4.2. Plot	7

Load packages

```
library(phyloseq)
library(microbiome)
library(picante)
library(plyr)
library(dplyr)
library(emmeans)
library(sciplot)
library(ggplot2)
library(ggpubr)
library(viridis)
```

Input files

```
ps1.work <- readRDS("./phyobjects/ps1.work.rds")
ps1.work
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 2222 taxa and 93 samples ]
## sample_data() Sample Data: [ 93 samples by 16 sample variables ]
## tax_table() Taxonomy Table: [ 2222 taxa by 6 taxonomic ranks ]
## phy_tree() Phylogenetic Tree: [ 2222 tips and 2221 internal nodes ]
```

1. Prepare Data

Prepare data for calculation of Faith's phylogenetic diversity at genus level.

```
ps1.g <- microbiome::aggregate_taxa(ps1.work, "Genus")
ps1.g.r <- microbiome::transform(ps1.g, "compositional")
ps1.g.otu <- as.data.frame(ps1.g.r@otu_table)
ps1.g.tree <- ps1.g.r@phy_tree
# confirm that the tree is rooted
ps1.g.r@phy_tree
```

```
##
## Phylogenetic tree with 220 tips and 219 internal nodes.
##
## Tip labels:
## 216615659, 216615763, 216615306, 216615906, 2166151782, 2166151034, ...
##
## Rooted; includes branch lengths.
```

```
# plot preset
theme_div <- theme_classic() +
  theme(panel.grid.major = element_line(colour = "grey80"),
        panel.spacing = unit(.5, "lines"),
        panel.border = element_rect(color = "black", fill = NA, size = .5),
        strip.background = element_blank(),
        text = element_text(size=20))
```

3. Phylogenetic diversity

3.1. Calculate metrics

```
# create dataframe of metadata
pd.g <- meta(ps1.g.r)

# calculate phylogenetic diversity
pd.g$phylog_div <- pd(t(ps1.g.otu), ps1.g.tree, include.root = T)$PD

# summarise per treatment
```

```
pd.sum <- ddpoly(pd.g, ~ Diet + Treatment + Type + Timepoint,
  summarise, mean = mean(phylog_div),
  sd = sd(phylog_div), se = se(phylog_div))
```

3.2. LMM regression

Need mixed model: repeated measures for Container (timepoints or sample types) - random intercept model.

Analyse diets separately, and test in subsets (balanced datasets):

- substrates including day 0;
- larvae and substrates on day 15.

No fixed term model selection, use full model in all cases.

3.2.0. Model validation

Customize this code chunk for the model and data.frame in question. For each parametric model, check the residuals (“normalized” or “pearson”) using the following code.

```
# mod = model
# data = data.frame with factors

# check normality of residuals and random effects
qqnorm(mod, ~resid(., type = "p"), abline = c(0,1))
hist(resid(mod, method = "pearson"), breaks = 30, col = "grey")
qqnorm(mod, ~ranef(., standard = T))
# check homoskedasticity of residuals
plot(mod)
plot(mod, sqrt(abs(resid(., type = "pearson")) ~ fitted(., type = c("p", "smooth")))
plot(resid(mod, method= "pearson")~data$Treatment); abline(0,0)
plot(resid(mod, method= "pearson")~data$Type); abline(0,0)
```

3.2.1. Subsetting

Subset per day. For CF, day 0 not needed: only 1 treatment with >1 replicate.

```
# CF day 15
pd.cf <- subset(pd.g, Diet == "CF" & Timepoint == 15)
pd.cf$Treatment <- droplevels(pd.cf$Treatment)

# CM day 0
pd.cm0 <- subset(pd.g, Diet == "CM" & Timepoint == 0)
pd.cm0$Treatment <- droplevels(pd.cm0$Treatment)

# CM day 15
pd.cm15 <- subset(pd.g, Diet == "CM" & Timepoint == 15)
```

3.2.2. Chicken feed day 15

```
# mixed model selection - variance structure
m.pd0 <- lme(phylog_div ~ Treatment * Type, data = pd.cf, method = "REML",
            random = ~ 1|ContainerID)
m.pd1 <- update(m.pd0, weights = varIdent(form=~1|Treatment))
m.pd2 <- update(m.pd0, weights = varIdent(form=~1|Type))
m.pd3 <- update(m.pd0, weights = varIdent(form=~1|Treatment * Type))
AIC(m.pd0, m.pd1, m.pd2, m.pd3)
```

```
##          df          AIC
## m.pd0    8 88.37461
## m.pd1   10 83.06470
## m.pd2    9 88.05692
## m.pd3   13 86.09167
```

```
# use model pd1, variance structure for treatment.
```

```
# model output
anova(m.pd1) # effect of Treatment
```

```
##               numDF denDF   F-value p-value
## (Intercept)         1     9 116.22886 <.0001
## Treatment           2     9   6.87218  0.0154
## Type                1     9   3.21038  0.1068
## Treatment:Type      2     9   1.93976  0.1993
```

```
CLD(emmeans(m.pd1, ~ Treatment + Type), Letters = letters, method = "tukey")
```

```
## Treatment Type      emmean    SE df lower.CL upper.CL .group
## Si/E      larvae      2.53 0.570  9      1.24      3.82  a
## Si/Es     larvae      3.21 0.539  9      1.99      4.43  a
## Si/Es     substrate    3.31 0.539  9      2.09      4.53  a
## Si/E      substrate    3.48 0.570  9      2.19      4.77  a
## S/E       substrate    6.51 1.075 11      4.14      8.88  a
## S/E       larvae      6.57 1.075  9      4.14      9.00  a
##
## Degrees-of-freedom method: containment
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 6 estimates
## significance level used: alpha = 0.05
```

```
# but no differences in pairwise comparisons.
```

3.2.3. Chicken manure day 0

```
# mixed model selection
gls.cm0 <- gls(phylog_div ~ Treatment, data = pd.cm0, method = "REML")
gls.cm1 <- update(gls.cm0, weights = varIdent(form=~1|Treatment))
AIC(gls.cm0, gls.cm1)
```

```
##          df          AIC
## gls.cm0  4 67.06889
## gls.cm1  6 67.69183

# use gls.cm0, no variance structure: LM.

lm.cm0 <- lm(phylog_div ~ Treatment, data = pd.cm0)

# model output
anova(lm.cm0) # treatment effect.

## Analysis of Variance Table
##
## Response: phylog_div
##          Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment  2 126.371   63.185   23.004 3.762e-05 ***
## Residuals 14   38.454    2.747
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

CLD(emmeans(lm.cm0, ~ Treatment), Letters = letters, method = "tukey")

## Treatment emmean      SE df lower.CL upper.CL .group
## Si/Es      12.5 0.677 14      11.0      13.9 a
## Si/E       13.6 0.741 14      12.0      15.2 a
## S/E        18.6 0.677 14      17.1      20.0 b
##
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 3 estimates
## significance level used: alpha = 0.05
```

3.2.4. Chicken manure day 15

```
# mixed model selection - variance structure
m.cm0 <- lme(phylog_div ~ Treatment * Type, data = pd.cm15, method = "REML",
            random = ~ 1|ContainerID)
m.cm1 <- update(m.cm0, weights = varIdent(form=~1|Treatment))
m.cm2 <- update(m.cm0, weights = varIdent(form=~1|Type))
m.cm3 <- update(m.cm0, weights = varIdent(form=~1|Treatment*Type))
AIC(m.cm0, m.cm1, m.cm2, m.cm3)

##          df          AIC
## m.cm0 10 171.1630
## m.cm1 13 166.7902
## m.cm2 11 169.9172
## m.cm3 17 169.6193

# use model m.cm1, variance structure for treatment.

# model output
anova(m.cm1)
```

```
##          numDF denDF  F-value p-value
## (Intercept)      1    20 4548.091 <.0001
## Treatment        3    20  505.846 <.0001
## Type             1    19   2.199 0.1545
## Treatment:Type    3    19   9.566 0.0005
```

```
CLD(emmeans(m.cm1, ~ Treatment + Type), Letters = letters, method = "tukey")
```

```
## Treatment Type      emmean    SE df lower.CL upper.CL .group
## Ss/E      larvae      2.38 0.400 19     1.54     3.22    a
## Ss/E      substrate    3.53 0.400 20     2.69     4.36    a
## Si/E      substrate   17.65 0.444 20    16.73    18.58    b
## Si/E      larvae     18.95 0.444 19    18.02    19.88    b
## S/E      substrate   19.64 0.994 23    17.59    21.70   bc
## Si/Es     substrate   19.83 0.483 20    18.82    20.84    b
## S/E      larvae     20.15 0.908 19    18.25    22.05   bc
## Si/Es     larvae     21.98 0.483 19    20.97    22.99    c
##
## Degrees-of-freedom method: containment
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 8 estimates
## significance level used: alpha = 0.05
```

4. Errorbar plot

Figure 4 in manuscript. Phylogenetic diversity, using EMM and SE from models.

4.1. Collect estimates in dataframe

```
# CF larvae + substrates
pd.cf.emm <- CLD(emmeans(m.pd1, ~ Treatment + Type), Letters = letters, method = "tukey")
pd.cf.emm$Diet <- "CF"
pd.cf.emm$Timepoint <- "15"
pd.cf.emm <- subset(pd.cf.emm, select = c(1:4,9,10))
colnames(pd.cf.emm)[3] <- "mean"

# CM larvae + substrates
pd.cm.emm <- CLD(emmeans(m.cm1, ~ Treatment + Type), Letters = letters, method = "tukey")
pd.cm.emm$Diet <- "CM"
pd.cm.emm$Timepoint <- "15"
pd.cm.emm <- subset(pd.cm.emm, select = c(1:4,9,10))
colnames(pd.cm.emm)[3] <- "mean"

# CF substrates day 0
pd.cf.emm0 <- subset(pd.sum, Diet == "CF" & Timepoint == 0)
pd.cf.emm0 <- subset(pd.cf.emm0, select = c(1:5,7))
colnames(pd.cf.emm0)[6] <- "SE"

# CM substrates day 0
pd.cm.emm0 <- CLD(emmeans(lm.cm0, ~ Treatment), Letters = letters, method = "tukey")
```

```

pd.cm.emm0$Diet <- "CM"
pd.cm.emm0$Type <- "substrate"
pd.cm.emm0$Timepoint <- 0
pd.cm.emm0 <- subset(pd.cm.emm0, select = c(1:3,8:10))
colnames(pd.cm.emm0)[2] <- "mean"

# merge data
pd.emm <- rbind(pd.cf.emm0, pd.cm.emm0, pd.cf.emm, pd.cm.emm)

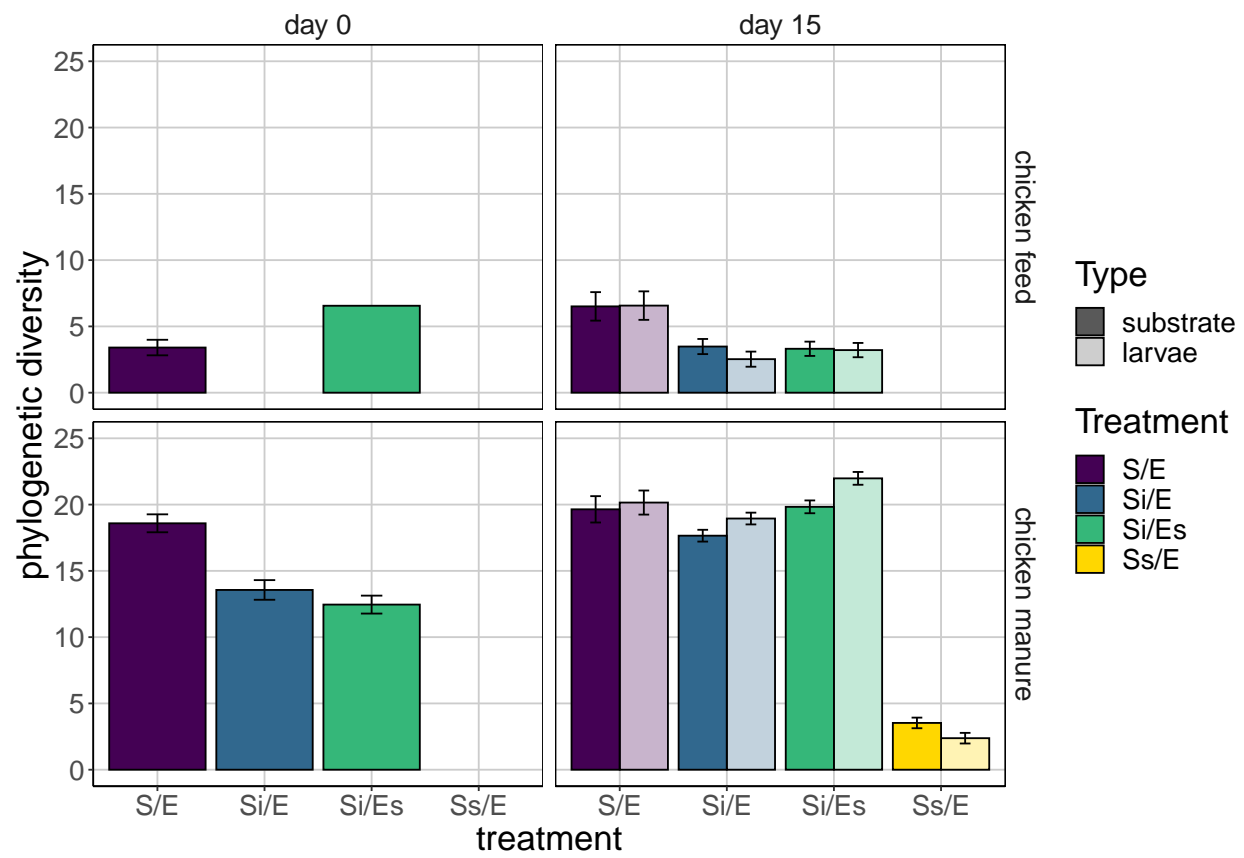
```

4.2. Plot

```

p.pd <- ggplot(pd.emm, aes(x = Treatment, weight = mean,
                           ymin = mean - SE, ymax = mean + SE, group = Type))
p.pd <- p.pd +
  geom_col(position = position_dodge(), aes(y = mean), fill = "white") +
  geom_col(position = position_dodge(), aes(y = mean, fill = Treatment, alpha = Type),
           colour = "black") +
  geom_errorbar(width = .2, position = position_dodge(width = 0.9)) +
  scale_fill_manual(values = c("#440154FF", "#31688EFF", "#35B779FF", "gold")) +
  scale_alpha_manual(values = c(1, .3)) +
  labs(x = "treatment", y = "phylogenetic diversity") +
  scale_y_continuous(limits = c(0, 25), n.breaks = 6) +
  facet_grid(Diet ~ Timepoint, labeller = as_labeller(
    c("0" = "day 0", "15" = "day 15", CF = "chicken feed", CM = "chicken manure"))) +
  theme_div
p.pd

```



```
ggsave(plot = p.pd, "./figures/Fig_4_phylog_div.png", h = 7, w = 10)
ggsave(plot = p.pd, "./figures/Fig_4_phylog_div.pdf", h = 200, w = 320, u = "mm")
```