

Diversity indices

Stijn Schreven

24 April 2022

Contents

Introduction	2
Load packages	2
Input files	2
1. Prepare data	2
2. Calculate diversity indices	3
3. Phylogenetic diversity: LMM regression	3
3.0. Model validation	3
3.1. Substrates	4
3.2. Larvae and substrates	6
4. Phylogenetic diversity: plot	8
4.1. Plot presets	8
4.2. Collect EMM and SE	9
4.3. Plot	9
4.4. Plot (black and white)	10
5. Shannon diversity index	12
5.1. Summarize index	12
5.2. LMM substrates	12
5.3. LMM larvae and substrates	14
5.4. Collect EMM and SE	17
5.4. Plot	17

Introduction

We calculated Faith's phylogenetic diversity for the bacterial communities in our samples, using the decontaminated dataset at genus level. In addition, we calculated Shannon's diversity index for the submission to *Applied and Environmental Microbiology*.

Load packages

```
library(phyloseq)
library(microbiome)
library(picante)
library(plyr)
library(nlme)
library(sciplot)
library(emmeans)
library(knitr)
library(ggplot2)
library(viridis)
```

Input files

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

```
## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 2424 taxa and 284 samples ]
## sample_data() Sample Data: [ 284 samples by 14 sample variables ]
## tax_table() Taxonomy Table: [ 2424 taxa by 6 taxonomic ranks ]
## phy_tree() Phylogenetic Tree: [ 2424 tips and 2423 internal nodes ]
```

1. Prepare data

```
# genus level
pstot.g <- aggregate_taxa(ps1.exp, "Genus")
pstot.g.r <- microbiome::transform(pstot.g, "compositional")

# input for phylog diversity
psg.otu <- as.data.frame(pstot.g.r@otu_table)
psg.tree <- pstot.g.r@phy_tree
pstot.g.r@phy_tree
```

```
##
## Phylogenetic tree with 280 tips and 279 internal nodes.
##
```

```
## Tip labels:
## 3015901058, 301590298, 3015901105, 301590284, 3015901059, 3015901281, ...
##
## Rooted; includes branch lengths.
```

2. Calculate diversity indices

```
# prepare dataframe
pd.g <- data.frame("Description" = colnames(psg.otu),
                  "phylog_div" = NA, "shannon" = NA)

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

# calculate Shannon diversity index
pd.g$shannon <- diversities(psg.otu, index = "shannon")$shannon

# add metadata
pd.g <- merge(meta(pstot.g.r), pd.g, by = "Description")

# summarise
pd.gsum <- ddpoly(pd.g, ~ Diet + Type + Density + Timepoint,
                 summarise, mean = mean(phylog_div),
                           median = median(phylog_div),
                           sd = sd(phylog_div), se = se(phylog_div))

# subset samples for LMM
pd.gS <- subset(pd.g, Type == "substrate" & Timepoint != 0) # 144 samples
pd.gSL <- subset(pd.g, Density != 0 & Timepoint != 0) # 216
```

3. Phylogenetic diversity: LMM regression

Linear mixed model regression.

3.0. Model validation

The R code chunk below was used to validate models by creating QQ plots and plotting residuals against fitted values and independent variables. It is not evaluated in this R markdown file, but can be adapted to evaluate each model in the following R code chunks.

```
# mod = model
# data = dataframe with original data

# check normality of residuals
hist(resid(mod, method = "pearson"), breaks = 30, col = "grey")
qqnorm(mod, ~ ranef (..))
qqnorm(mod)

# check homoskedasticity of residuals
```

```
plot(mod)
plot(resid(mod, method = "pearson") ~ data$Diet); abline(0,0)
plot(resid(mod, method = "pearson") ~ data$Density); abline(0,0)
plot(resid(mod, method = "pearson") ~ data$Timepoint); abline(0,0)
```

3.1. Substrates

Supplementary Table S3 in manuscript Chapter 3 in PhD thesis. Multiple comparisons included in Figure 2 of manuscript submitted to *Applied and Environmental Microbiology*.

```
# mixed model: random term and variance structure
PDs.m0 <- lme(phylog_div ~ Diet * Density * Timepoint, data = pd.gS,
              method = "REML", random = ~ 1 | ContainerID)
PDs.m1 <- update(PDs.m0, random = ~ Timepoint | ContainerID)
## PDs.m2 <- update(PDs.m0, random = ~ Density | ContainerID)
# simple variance structure
PDs.m3 <- update(PDs.m0, weights = varIdent(form = ~ 1 | Diet))
PDs.m4 <- update(PDs.m0, weights = varIdent(form = ~ 1 | Timepoint))
PDs.m5 <- update(PDs.m0, weights = varIdent(form = ~ 1 | Density))
# 2-way interaction structure
PDs.m6 <- update(PDs.m0, weights = varIdent(form = ~ 1 | Diet * Timepoint))
PDs.m7 <- update(PDs.m0, weights = varIdent(form = ~ 1 | Diet * Density))
PDs.m8 <- update(PDs.m0, weights = varIdent(form = ~ 1 | Density * Timepoint))
# compare model AICs:
AIC(PDs.m0, PDs.m1, PDs.m3, PDs.m4, PDs.m5, PDs.m6, PDs.m7, PDs.m8)
```

```
##      df      AIC
## PDs.m0 38 508.6218
## PDs.m1 43 511.2453
## PDs.m3 40 511.3299
## PDs.m4 40 510.9252
## PDs.m5 41 510.4781
## PDs.m6 46 520.5840
## PDs.m7 49 518.1026
## PDs.m8 49 520.3221
```

a random intercept model without variance structure is best (PDs.m0).

```
# model output
aov.pd.s <- anova(PDs.m0)
aov.pd.s[, 3:4] <- round(aov.pd.s[, 3:4], digits = 3)
kable(aov.pd.s)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	72	4853.113	0.000
Diet	2	36	510.573	0.000
Density	3	36	8.101	0.000
Timepoint	2	72	53.259	0.000
Diet:Density	6	36	3.955	0.004
Diet:Timepoint	4	72	17.120	0.000
Density:Timepoint	6	72	4.100	0.001
Diet:Density:Timepoint	12	72	1.832	0.059

```
# posthoc comparisons: EMM
```

```
CLD(emmeans(PDs.m0, pairwise ~ Density + Timepoint | Diet, method = "tukey"),
    Letters = letters)
```

```
## Diet = CF:
```

##	Density	Timepoint	emmean	SE	df	lower.CL	upper.CL	.group
##	0	5	2.04	0.719	47	0.595	3.49	a
##	50	5	2.72	0.719	36	1.259	4.17	ab
##	100	5	3.70	0.719	36	2.238	5.15	ab
##	200	5	4.83	0.719	36	3.377	6.29	abc
##	0	10	5.30	0.719	47	3.855	6.75	bcd
##	0	15	5.55	0.719	47	4.100	6.99	bcd
##	100	10	8.27	0.719	36	6.816	9.73	cde
##	50	10	8.36	0.719	36	6.907	9.82	cde
##	50	15	8.45	0.719	36	6.988	9.90	de
##	200	15	9.80	0.719	36	8.342	11.26	e
##	200	10	10.00	0.719	36	8.540	11.45	e
##	100	15	10.34	0.719	36	8.886	11.80	e

```
##
```

```
## Diet = CS:
```

##	Density	Timepoint	emmean	SE	df	lower.CL	upper.CL	.group
##	0	10	4.64	0.719	36	3.182	6.10	a
##	200	15	4.65	0.719	36	3.198	6.11	a
##	0	5	4.86	0.719	36	3.408	6.32	a
##	100	5	5.24	0.719	36	3.787	6.70	ab
##	100	15	6.00	0.719	36	4.541	7.46	ab
##	50	5	6.07	0.719	36	4.616	7.53	ab
##	200	5	6.20	0.719	36	4.739	7.65	ab
##	0	15	6.56	0.719	36	5.103	8.02	ab
##	200	10	7.08	0.719	36	5.626	8.54	ab
##	50	15	7.16	0.719	36	5.708	8.62	ab
##	100	10	7.34	0.719	36	5.888	8.80	ab
##	50	10	8.43	0.719	36	6.975	9.89	b

```
##
```

```
## Diet = CM:
```

##	Density	Timepoint	emmean	SE	df	lower.CL	upper.CL	.group
##	50	5	13.78	0.719	36	12.325	15.24	a
##	0	10	13.84	0.719	36	12.387	15.30	a
##	100	5	14.07	0.719	36	12.616	15.53	ab
##	200	5	14.58	0.719	36	13.120	16.03	ab
##	200	10	15.11	0.719	36	13.650	16.56	ab
##	0	15	15.71	0.719	36	14.253	17.17	ab
##	50	10	15.92	0.719	36	14.465	17.38	ab
##	0	5	16.28	0.719	36	14.824	17.74	ab
##	100	15	16.76	0.719	36	15.299	18.21	ab
##	200	15	16.81	0.719	36	15.357	18.27	ab
##	100	10	16.88	0.719	36	15.426	18.34	ab
##	50	15	17.54	0.719	36	16.081	19.00	b

```
##
```

```
## Degrees-of-freedom method: containment
```

```
## Confidence level used: 0.95
```

```
## P value adjustment: tukey method for comparing a family of 12 estimates
```

```
## significance level used: alpha = 0.05
```

3.2. Larvae and substrates

Supplementary Table S4 in manuscript Chapter 3 in PhD thesis. Pairwise comparisons included as asterisks in Figure 2 of manuscript submitted to *Applied and Environmental Microbiology*.

```
# mixed model: random term and variance structure
PDsl.m0 <- lme(phylog_div ~ Diet * Density * Timepoint * Type, data = pd.gSL,
              method = "REML", random = ~ 1 | ContainerID)
#PDsl.m1 <- update(PDsl.m0, random = ~ Timepoint | ContainerID)
PDsl.m2 <- update(PDsl.m0, random = ~ Density | ContainerID)
# single variance structure
PDsl.m3 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Diet))
PDsl.m4 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Timepoint))
PDsl.m5 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Density))
PDsl.m6 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Type))
# 2-way interaction structure
PDsl.m7 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Diet * Timepoint))
PDsl.m8 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Diet * Density))
PDsl.m9 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Diet * Type))
PDsl.m10 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Density * Timepoint))
PDsl.m11 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Timepoint * Type))
PDsl.m12 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Density * Type))
# 3-way interaction structure
PDsl.m13 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Diet * Density * Timepoint))
PDsl.m14 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Diet * Timepoint * Type))
PDsl.m15 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Density * Timepoint * Type))
PDsl.m16 <- update(PDsl.m0, weights = varIdent(form = ~ 1 | Diet * Density * Type))
# compare model AICs:
AIC(PDsl.m0, PDsl.m2, PDsl.m3, PDsl.m4, PDsl.m5, PDsl.m6, PDsl.m7, PDsl.m8, PDsl.m9,
    PDsl.m10, PDsl.m11, PDsl.m12, PDsl.m13, PDsl.m14, PDsl.m15, PDsl.m16)
```

##	df	AIC
## PDsl.m0	56	785.3829
## PDsl.m2	61	794.9511
## PDsl.m3	58	786.6372
## PDsl.m4	58	779.1608
## PDsl.m5	58	785.2609
## PDsl.m6	57	785.4254
## PDsl.m7	64	775.3806
## PDsl.m8	64	793.7237
## PDsl.m9	61	785.4662
## PDsl.m10	64	786.3438
## PDsl.m11	61	780.5727
## PDsl.m12	61	784.0501
## PDsl.m13	82	803.4465
## PDsl.m14	73	779.8208
## PDsl.m15	73	790.7805
## PDsl.m16	73	796.5611

```
# model 7 is best: variance structure for Diet * Timepoint.
```

```
# model output
aov.pd.ls <- anova(PDsl.m7)
```

```
aov.pd.ls[, 3:4] <- round(aov.pd.ls[, 3:4], digits = 3)
kable(aov.pd.ls)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	135	5367.633	0.000
Diet	2	27	488.219	0.000
Density	2	27	0.210	0.812
Timepoint	2	135	76.275	0.000
Type	1	135	4.252	0.041
Diet:Density	4	27	3.102	0.032
Diet:Timepoint	4	135	37.629	0.000
Density:Timepoint	4	135	1.742	0.144
Diet:Type	2	135	5.387	0.006
Density:Type	2	135	1.287	0.280
Timepoint:Type	2	135	4.759	0.010
Diet:Density:Timepoint	8	135	1.758	0.091
Diet:Density:Type	4	135	2.064	0.089
Diet:Timepoint:Type	4	135	4.168	0.003
Density:Timepoint:Type	4	135	1.057	0.381
Diet:Density:Timepoint:Type	8	135	0.816	0.590

```
# posthoc comparisons: EMM
CLD(emmeans(PDsl.m7, pairwise ~ Density + Timepoint + Type | Diet, method = "tukey"),
    Letters = letters)
```

```
## Diet = CF:
##   Density Timepoint Type      emmean      SE df lower.CL upper.CL .group
##   50      5          substrate  2.72 0.590 35      1.52      3.91    a
##   100     5          substrate  3.70 0.590 27      2.48      4.91    a
##   100     5          larvae     3.70 0.590 27      2.49      4.91    a
##   50      5          larvae     4.20 0.590 35      3.00      5.40    ab
##   200     5          substrate  4.83 0.590 27      3.62      6.05    abc
##   200     5          larvae     5.54 0.590 27      4.33      6.75    abcd
##   100     10         substrate  8.27 1.020 27      6.18     10.37   bcde
##   50      10         substrate  8.36 1.020 35      6.29     10.43   cde
##   50      15         substrate  8.45 0.841 35      6.74     10.15   cde
##   100     15         larvae     8.71 0.841 27      6.99     10.44   cde
##   50      15         larvae     8.76 0.841 35      7.06     10.47   cde
##   200     15         larvae     9.61 0.841 27      7.89     11.34    e
##   50      10         larvae     9.62 1.020 35      7.55     11.69   de
##   200     15         substrate  9.80 0.841 27      8.07     11.52    e
##   100     10         larvae     9.80 1.020 27      7.71     11.90   de
##   200     10         substrate 10.00 1.020 27      7.90     12.09    e
##   100     15         substrate 10.34 0.841 27      8.62     12.07    e
##   200     10         larvae    11.81 1.020 27      9.71     13.90    e
##
## Diet = CS:
##   Density Timepoint Type      emmean      SE df lower.CL upper.CL .group
##   200     15         substrate  4.65 0.684 27      3.25      6.06    a
##   100      5         substrate  5.24 0.659 27      3.89      6.60   ab
##   100     15         substrate  6.00 0.684 27      4.60      7.40   ab
```

```
## 200      15      larvae      6.04 0.684 27      4.64      7.44 ab
## 50       5      substrate     6.07 0.659 27      4.72      7.43 ab
## 200      5      substrate     6.20 0.659 27      4.84      7.55 ab
## 200      5      larvae      6.61 0.659 27      5.25      7.96 ab
## 100      5      larvae      6.68 0.659 27      5.32      8.03 ab
## 50       5      larvae      6.83 0.659 27      5.48      8.19 ab
## 200     10      substrate     7.08 0.716 27      5.61      8.55 ab
## 50     15      substrate     7.16 0.684 27      5.76      8.57 ab
## 200     10      larvae      7.31 0.716 27      5.84      8.78 ab
## 100     10      substrate     7.34 0.716 27      5.88      8.81 ab
## 100     15      larvae      8.01 0.684 27      6.61      9.41 ab
## 50     15      larvae      8.09 0.684 27      6.69      9.49 ab
## 100     10      larvae      8.26 0.716 27      6.79      9.73 ab
## 50     10      substrate     8.43 0.716 27      6.96      9.90 ab
## 50     10      larvae      8.72 0.716 27      7.25     10.19 b
##
## Diet = CM:
## Density Timepoint Type      emmean      SE df lower.CL upper.CL .group
## 100      15      larvae     11.92 1.191 27      9.48     14.36 a
## 50       5      substrate    13.78 0.585 27     12.58     14.98 a
## 100      5      substrate    14.07 0.585 27     12.87     15.27 a
## 100     10      larvae     14.28 0.538 27     13.17     15.38 a
## 200      5      substrate    14.58 0.585 27     13.38     15.78 ab
## 50       5      larvae     14.94 0.585 27     13.74     16.14 ab
## 200     10      substrate    15.11 0.538 27     14.00     16.21 ab
## 50     10      larvae     15.12 0.538 27     14.01     16.22 ab
## 200     10      larvae     15.14 0.538 27     14.03     16.24 ab
## 200      5      larvae     15.31 0.585 27     14.11     16.51 ab
## 100      5      larvae     15.43 0.585 27     14.23     16.63 ab
## 50     15      larvae     15.70 1.191 27     13.25     18.14 ab
## 50     10      substrate    15.92 0.538 27     14.82     17.03 ab
## 100     15      substrate    16.76 1.191 27     14.31     19.20 ab
## 200     15      substrate    16.81 1.191 27     14.37     19.26 ab
## 100     10      substrate    16.88 0.538 27     15.78     17.99 b
## 50     15      substrate    17.54 1.191 27     15.09     19.98 ab
## 200     15      larvae     17.62 1.191 27     15.17     20.06 ab
##
## Degrees-of-freedom method: containment
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 18 estimates
## significance level used: alpha = 0.05
```

4. Phylogenetic diversity: plot

Figure 3 in manuscript Chapter 3 of PhD thesis; and Figure 2 in manuscript submitted to *Applied and Environmental Microbiology*.

4.1. Plot presets

```
# theme
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))

# facet labels
labs_div <- as_labeller(c("0" = "0 larvae\nper container",
                          "50" = "50 larvae\nper container",
                          "100" = "100 larvae\nper container",
                          "200" = "200 larvae\nper container",
                          "CF" = "chicken feed",
                          "CS" = "camelina",
                          "CM" = "chicken manure"))

```

4.2. Collect EMM and SE

```

# collect and merge data from EMM and mean()
pd.g.emm <- CLD(emmeans(PDs1.m7, pairwise ~Type | Diet + Density + Timepoint,
                       method = "tukey"), Letters = letters)
pd.g.emm.s <- subset(pd.g.emm, select = c(1:6))

# extract EMM of density 0 from model of substrates only
pd.g.emm2 <- CLD(emmeans(PDs.m0, pairwise ~ Diet + Density + Timepoint,
                       method = "tukey"), Letters = letters)
pd.g.emm2.s <- subset(pd.g.emm2, select = c(1:5))
pd.g.emm2.s <- subset(pd.g.emm2.s, Density == 0)
pd.g.emm2.s$Type <- "substrate"

# extract mean and SE for timepoint 0 (not in models)
pd.g.emm3 <- subset(pd.gsum, Timepoint == 0)
pd.g.emm3.s <- subset(pd.g.emm3, select = c(1:5,8))
colnames(pd.g.emm3.s)[5:6] <- c("emmean", "SE")

# combine all EMM and SE
pd.g.emm4 <- rbind(pd.g.emm.s, pd.g.emm2.s, pd.g.emm3.s)

# reorder factor levels
pd.g.emm4$Density <- factor(pd.g.emm4$Density, levels(pd.g.emm4$Density)[c(4, 1:3)])
pd.g.emm4$Timepoint <- factor(pd.g.emm4$Timepoint, levels(pd.g.emm4$Timepoint)[c(4, 1:3)])

```

4.3. Plot

Figure 3 in Chapter 3 of PhD thesis.

```

p.pdg <- ggplot(pd.g.emm4, aes(x = Timepoint, y = emmean,
                              colour = Diet, group = interaction(Diet, Density, Type)))

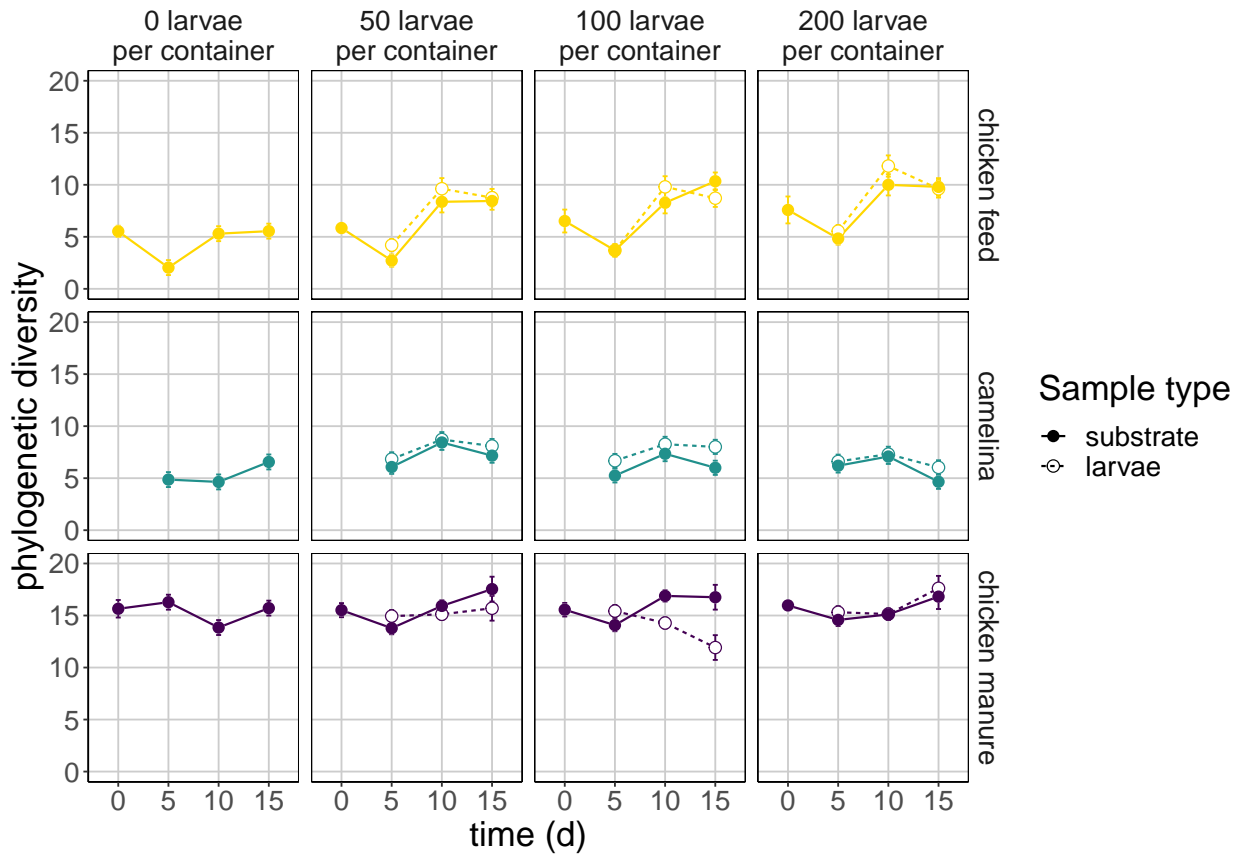
p.pdg <- p.pdg + geom_line(size = .6, aes(linetype = Type)) +
  geom_errorbar(aes(ymin = emmean-SE, ymax = emmean+SE), width = .1) +

```

```

geom_point(shape = 16, size = 3, colour = "white") +
geom_point(aes(shape = Type), size = 3) +
scale_shape_manual(values = c(16,1)) +
scale_color_manual(values = c("gold", "#21908CFF", "#440154FF")) +
labs(x = "time (d)", y = "phylogenetic diversity",
     shape = "Sample type", linetype = "Sample type") +
scale_y_continuous(limits = c(0,20), n.breaks = 6) +
facet_grid(Diet ~ Density, labeller = labs_div) +
theme_div + guides(colour = F)
p.pdg

```



```

ggsave(plot= p.pdg, "./figures/Diversity_Phylogenetic_genus_EMM.png", h = 7, w = 10)
ggsave(plot= p.pdg, "./figures/Diversity_Phylogenetic_genus_EMM.tiff",
       h = 175, w = 250, u = "mm", dpi = 600)

```

4.4. Plot (black and white)

Figure 2 in manuscript submitted to *Applied and Environmental Microbiology*. Same figure as in 4.3, but in black and white.

```

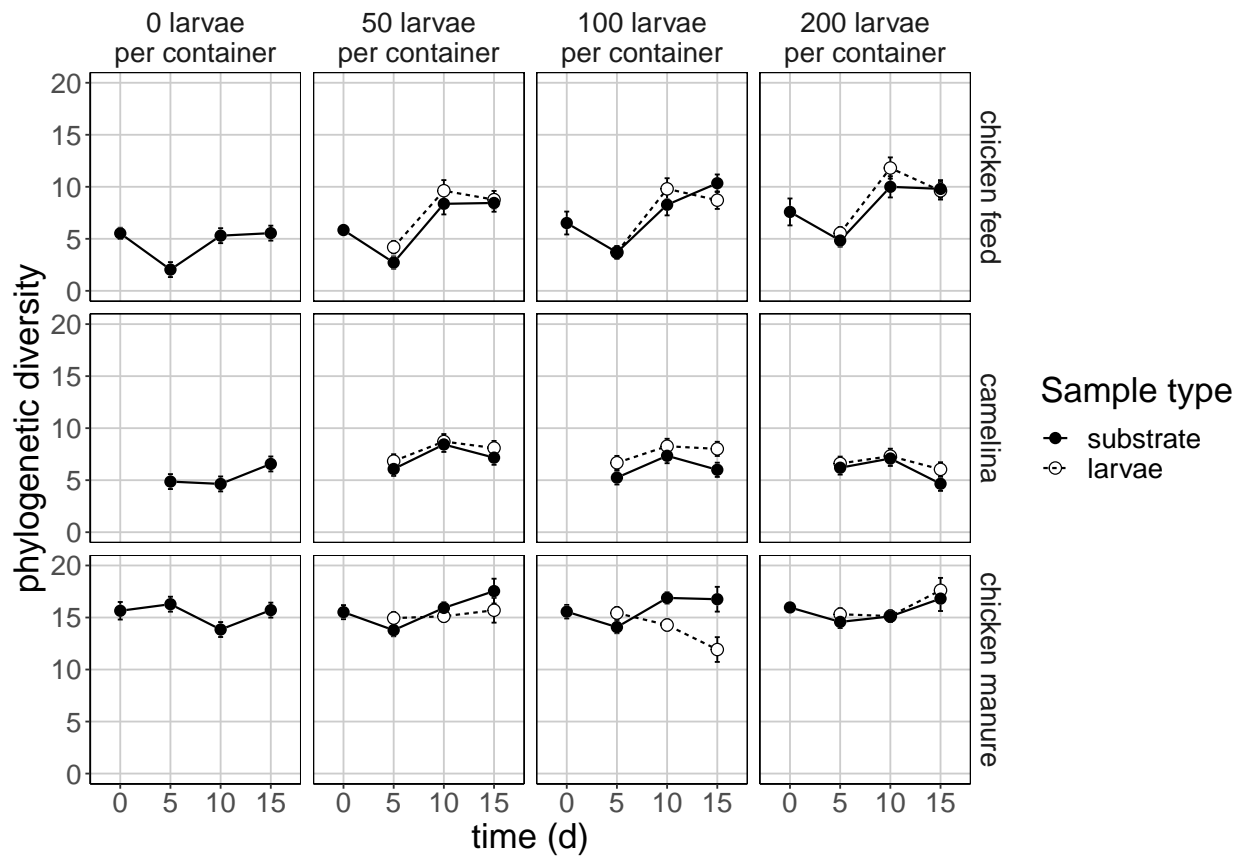
p.pdg.bk <- ggplot(pd.g.emm4, aes(x = Timepoint, y = emmean,
                                group = interaction(Diet, Density, Type)))
p.pdg.bk <- p.pdg.bk + geom_line(size = .6, aes(linetype = Type)) +

```

```

geom_errorbar(aes(ymin = emmean-SE, ymax = emmean+SE), width = .1) +
geom_point(shape = 16, size = 3, colour = "white") +
geom_point(aes(shape = Type), size = 3, colour = "black") +
scale_shape_manual(values = c(16,1)) +
labs(x = "time (d)", y = "phylogenetic diversity",
     shape = "Sample type", linetype = "Sample type") +
scale_y_continuous(limits = c(0,20), n.breaks = 6) +
facet_grid(Diet ~ Density, labeller = labs_div) +
theme_div
p.pdg.bk

```



```

ggsave(plot= p.pdg.bk, "./figures/Diversity_Phylogenetic_genus_EMM_black.png",
       h = 7, w = 10)
ggsave(plot= p.pdg.bk, "./figures/Diversity_Phylogenetic_genus_EMM_black.tiff",
       h = 175, w = 250, u = "mm", dpi = 600)
ggsave(plot= p.pdg.bk, "./figures/Diversity_Phylogenetic_genus_EMM_black.pdf",
       h = 175, w = 250, u = "mm")
# file format used in AEM submission:
ggsave(plot= p.pdg.bk, "./figures/Diversity_Phylogenetic_genus_EMM_black.eps",
       h = 7, w = 10)

```

5. Shannon diversity index

Supplementary Figure S1 in manuscript submitted to *Applied and Environmental Microbiology*. Not included in Chapter 3 of PhD thesis.

5.1. Summarize index

```
# summarise
sh.gsum <- ddpby(pd.g, ~ Diet + Type + Density + Timepoint,
  summarise, mean = mean(shannon),
  median = median(shannon),
  sd = sd(shannon), se = se(shannon))
```

5.2. LMM substrates

```
# mixed model: random term and variance structure
SHs.m0 <- lme(shannon ~ Diet * Density * Timepoint, data = pd.gS,
  method = "REML", random = ~ 1 | ContainerID)
SHs.m1 <- update(SHs.m0, random = ~ Timepoint | ContainerID)
# simple variance structure
SHs.m3 <- update(SHs.m0, weights = varIdent(form = ~ 1 | Diet))
SHs.m4 <- update(SHs.m0, weights = varIdent(form = ~ 1 | Timepoint))
SHs.m5 <- update(SHs.m0, weights = varIdent(form = ~ 1 | Density))
# 2-way interaction structure
SHs.m6 <- update(SHs.m0, weights = varIdent(form = ~ 1 | Diet * Timepoint))
SHs.m7 <- update(SHs.m0, weights = varIdent(form = ~ 1 | Diet * Density))
SHs.m8 <- update(SHs.m0, weights = varIdent(form = ~ 1 | Density * Timepoint))
# compare model AICs:
AIC(SHs.m0, SHs.m1, SHs.m3, SHs.m4, SHs.m5, SHs.m6, SHs.m7, SHs.m8)
```

```
##      df      AIC
## SHs.m0 38 188.4602
## SHs.m1 43 188.1517
## SHs.m3 40 169.7141
## SHs.m4 40 187.8373
## SHs.m5 41 190.7420
## SHs.m6 46 171.8052
## SHs.m7 49 171.7416
## SHs.m8 49 199.9449
```

```
# a random intercept model with variance structure for Diet is best (SHs.m3).
```

```
# model output
aov.sh.s <- anova(SHs.m3)
aov.sh.s[, 3:4] <- round(aov.sh.s[, 3:4], digits = 3)
kable(aov.sh.s)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	72	10384.777	0.000
Diet	2	36	219.696	0.000
Density	3	36	3.065	0.040
Timepoint	2	72	14.196	0.000
Diet:Density	6	36	3.773	0.005
Diet:Timepoint	4	72	15.688	0.000
Density:Timepoint	6	72	9.510	0.000
Diet:Density:Timepoint	12	72	2.113	0.026

```
# posthoc comparisons: EMM
```

```
CLD(emmeans(SHs.m3, pairwise ~ Density + Timepoint | Diet, method = "tukey"),
    Letters = letters)
```

```
## Diet = CF:
```

```
##   Density Timepoint emmean      SE df lower.CL upper.CL .group
##   0         5         0.786 0.1918 47   0.400    1.172    a
##   50        5         1.131 0.1918 36   0.742    1.520   ab
##   100       5         1.266 0.1918 36   0.877    1.655  abc
##   0        15         1.390 0.1918 47   1.004    1.776  abc
##   0        10         1.427 0.1918 47   1.041    1.813  abc
##   200       5         1.442 0.1918 36   1.053    1.831  abc
##   100      10         1.556 0.1918 36   1.167    1.945  abc
##   100      15         1.674 0.1918 36   1.285    2.063  abc
##   50       15         1.757 0.1918 36   1.368    2.146   bc
##   50       10         1.947 0.1918 36   1.558    2.336   bc
##   200      10         2.027 0.1918 36   1.638    2.416   bc
##   200      15         2.134 0.1918 36   1.745    2.523    c
```

```
##
```

```
## Diet = CS:
```

```
##   Density Timepoint emmean      SE df lower.CL upper.CL .group
##   100      15         0.325 0.1847 36  -0.050    0.699    a
##   200      15         0.982 0.1847 36   0.607    1.356   ab
##   50       15         1.121 0.1847 36   0.746    1.495  abc
##   200      10         1.334 0.1847 36   0.959    1.708   bc
##   0        10         1.445 0.1847 36   1.070    1.819   bc
##   100      10         1.447 0.1847 36   1.072    1.821   bc
##   0        15         1.486 0.1847 36   1.111    1.860   bc
##   100       5         1.533 0.1847 36   1.159    1.908   bc
##   0         5         1.613 0.1847 36   1.239    1.988   bc
##   50       10         1.747 0.1847 36   1.372    2.121   bc
##   50        5         1.840 0.1847 36   1.465    2.214   bc
##   200       5         1.916 0.1847 36   1.541    2.290    c
```

```
##
```

```
## Diet = CM:
```

```
##   Density Timepoint emmean      SE df lower.CL upper.CL .group
##   100       5         1.965 0.0869 36   1.789    2.142    a
##   0        10         2.133 0.0869 36   1.957    2.310   ab
##   50        5         2.146 0.0869 36   1.969    2.322   ab
##   0         5         2.295 0.0869 36   2.118    2.471  abc
##   200      15         2.386 0.0869 36   2.210    2.563  abcd
##   100      15         2.427 0.0869 36   2.251    2.603  bcde
```

```
## 50      15      2.492 0.0869 36      2.316      2.668      bcde
## 200      5      2.501 0.0869 36      2.325      2.678      bcde
## 200     10      2.507 0.0869 36      2.331      2.683      bcde
## 50      10      2.707 0.0869 36      2.531      2.883      cde
## 0       15      2.720 0.0869 36      2.544      2.896      de
## 100     10      2.840 0.0869 36      2.664      3.016      e
##
## Degrees-of-freedom method: containment
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 12 estimates
## significance level used: alpha = 0.05
```

5.3. LMM larvae and substrates

```
# mixed model: random term and variance structure
SHsl.m0 <- lme(shannon ~ Diet * Density * Timepoint * Type, data = pd.gSL,
              method = "REML", random = ~ 1 | ContainerID)
#SHsl.m1 <- update(SHsl.m0, random = ~ Timepoint | ContainerID)
SHsl.m2 <- update(SHsl.m0, random = ~ Density | ContainerID)
# single variance structure
SHsl.m3 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Diet))
SHsl.m4 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Timepoint))
SHsl.m5 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Density))
SHsl.m6 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Type))
# 2-way interaction structure
SHsl.m7 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Diet * Timepoint))
SHsl.m8 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Diet * Density))
SHsl.m9 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Diet * Type))
SHsl.m10 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Density * Timepoint))
SHsl.m11 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Timepoint * Type))
SHsl.m12 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Density * Type))
# 3-way interaction structure
SHsl.m13 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Diet * Density * Timepoint))
SHsl.m14 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Diet * Timepoint * Type))
SHsl.m15 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Density * Timepoint * Type))
SHsl.m16 <- update(SHsl.m0, weights = varIdent(form = ~ 1 | Diet * Density * Type))
# compare model AICs:
AIC(SHsl.m0, SHsl.m2, SHsl.m3, SHsl.m4, SHsl.m5, SHsl.m6, SHsl.m7, SHsl.m8, SHsl.m9,
    SHsl.m10, SHsl.m11, SHsl.m12, SHsl.m13, SHsl.m14, SHsl.m15, SHsl.m16)
```

```
##      df      AIC
## SHsl.m0 56 270.6971
## SHsl.m2 61 279.0972
## SHsl.m3 58 272.5045
## SHsl.m4 58 258.6916
## SHsl.m5 58 268.4617
## SHsl.m6 57 271.9909
## SHsl.m7 64 257.6345
## SHsl.m8 64 271.5790
## SHsl.m9 61 250.8429
## SHsl.m10 64 263.3902
## SHsl.m11 61 260.4156
```

```
## SHsl.m12 61 273.1253
## SHsl.m13 82 271.6631
## SHsl.m14 73 251.8113
## SHsl.m15 73 272.9654
## SHsl.m16 73 259.0846
```

```
# model 9 is best: variance structure for Diet * Type
```

```
# model output
```

```
aov.sh.ls <- anova(SHsl.m9)
aov.sh.ls[, 3:4] <- round(aov.sh.ls[, 3:4], digits = 3)
kable(aov.sh.ls)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	135	7465.105	0.000
Diet	2	27	132.664	0.000
Density	2	27	4.745	0.017
Timepoint	2	135	62.518	0.000
Type	1	135	17.944	0.000
Diet:Density	4	27	3.041	0.034
Diet:Timepoint	4	135	19.005	0.000
Density:Timepoint	4	135	3.779	0.006
Diet:Type	2	135	5.435	0.005
Density:Type	2	135	0.378	0.686
Timepoint:Type	2	135	2.950	0.056
Diet:Density:Timepoint	8	135	2.378	0.020
Diet:Density:Type	4	135	2.171	0.076
Diet:Timepoint:Type	4	135	8.730	0.000
Density:Timepoint:Type	4	135	0.670	0.614
Diet:Density:Timepoint:Type	8	135	2.541	0.013

```
# posthoc comparisons: EMM
```

```
CLD(emmeans(SHsl.m9, pairwise ~ Density + Timepoint + Type | Diet, method = "tukey"),
    Letters = letters)
```

```
## Diet = CF:
```

##	Density	Timepoint	Type	emmean	SE	df	lower.CL	upper.CL	.group
##	50	5	substrate	1.131	0.1614	35	0.8033	1.46	a
##	100	5	substrate	1.266	0.1614	27	0.9347	1.60	abcd
##	100	5	larvae	1.292	0.1099	27	1.0667	1.52	abc
##	50	5	larvae	1.305	0.1099	35	1.0824	1.53	ab
##	200	5	substrate	1.442	0.1614	27	1.1105	1.77	abcde
##	100	10	substrate	1.556	0.1614	27	1.2250	1.89	abcde
##	200	5	larvae	1.617	0.1099	27	1.3916	1.84	abcde
##	100	15	substrate	1.674	0.1614	27	1.3429	2.01	abcdef
##	50	15	substrate	1.757	0.1614	35	1.4295	2.08	abcdef
##	50	15	larvae	1.847	0.1099	35	1.6241	2.07	cdef
##	50	10	substrate	1.947	0.1614	35	1.6197	2.27	bcdef
##	100	15	larvae	1.969	0.1099	27	1.7438	2.19	ef
##	200	10	substrate	2.027	0.1614	27	1.6963	2.36	bcdef
##	100	10	larvae	2.046	0.1099	27	1.8206	2.27	ef
##	200	15	substrate	2.134	0.1614	27	1.8027	2.46	def

```

## 50      10      larvae      2.151 0.1099 35      1.9281      2.37      ef
## 200     15      larvae      2.337 0.1099 27      2.1121      2.56      f
## 200     10      larvae      2.367 0.1099 27      2.1417      2.59      f
##
## Diet = CS:
## Density Timepoint Type      emmean      SE df lower.CL upper.CL .group
## 100      15      substrate 0.325 0.1832 27      -0.0514      0.70      a
## 200      15      substrate 0.982 0.1832 27      0.6059      1.36      ab
## 50       15      substrate 1.121 0.1832 27      0.7449      1.50      abc
## 200      10      substrate 1.334 0.1832 27      0.9580      1.71      bc
## 100      10      substrate 1.447 0.1832 27      1.0711      1.82      bc
## 100       5      larvae    1.455 0.1417 27      1.1646      1.75      bc
## 200      15      larvae    1.464 0.1417 27      1.1727      1.75      bc
## 100       5      substrate 1.533 0.1832 27      1.1572      1.91      bc
## 200       5      larvae    1.534 0.1417 27      1.2428      1.82      bc
## 50       5      larvae    1.680 0.1417 27      1.3890      1.97      bc
## 50      15      larvae    1.716 0.1417 27      1.4252      2.01      bc
## 200      10      larvae    1.726 0.1417 27      1.4354      2.02      bc
## 100      10      larvae    1.733 0.1417 27      1.4421      2.02      bc
## 100      15      larvae    1.737 0.1417 27      1.4464      2.03      bc
## 50      10      substrate 1.747 0.1832 27      1.3707      2.12      bc
## 50      10      larvae    1.801 0.1417 27      1.5103      2.09      bc
## 50       5      substrate 1.840 0.1832 27      1.4638      2.22      bc
## 200       5      substrate 1.916 0.1832 27      1.5399      2.29      c
##
## Diet = CM:
## Density Timepoint Type      emmean      SE df lower.CL upper.CL .group
## 100      15      larvae    1.636 0.2223 27      1.1803      2.09      ab
## 100       5      substrate 1.965 0.0917 27      1.7771      2.15      a
## 50       5      substrate 2.146 0.0917 27      1.9575      2.33      ab
## 200       5      larvae    2.380 0.2223 27      1.9235      2.84      abcd
## 200      15      substrate 2.386 0.0917 27      2.1982      2.57      abcd
## 100      15      substrate 2.427 0.0917 27      2.2390      2.62      bc
## 50       5      larvae    2.434 0.2223 27      1.9782      2.89      abcd
## 100       5      larvae    2.444 0.2223 27      1.9878      2.90      abcd
## 100      10      larvae    2.460 0.2223 27      2.0036      2.92      abcd
## 200      10      larvae    2.470 0.2223 27      2.0139      2.93      abcd
## 50      15      substrate 2.492 0.0917 27      2.3037      2.68      bcd
## 200       5      substrate 2.501 0.0917 27      2.3131      2.69      bcd
## 50      15      larvae    2.507 0.2223 27      2.0507      2.96      abcd
## 200      10      substrate 2.507 0.0917 27      2.3189      2.70      bcd
## 50      10      larvae    2.540 0.2223 27      2.0835      3.00      abcd
## 50      10      substrate 2.707 0.0917 27      2.5187      2.90      cd
## 200      15      larvae    2.730 0.2223 27      2.2739      3.19      abcd
## 100      10      substrate 2.840 0.0917 27      2.6519      3.03      d
##
## Degrees-of-freedom method: containment
## Confidence level used: 0.95
## P value adjustment: tukey method for comparing a family of 18 estimates
## significance level used: alpha = 0.05

```


5.4. Collect EMM and SE

```
# collect and merge data from EMM and mean()
sh.g.emm <- CLD(emmeans(SHsl.m9, pairwise ~Type | Diet + Density + Timepoint,
                      method = "tukey"), Letters = letters)
sh.g.emm.s <- subset(sh.g.emm, select = c(1:6))

# extract EMM of density 0 from model of substrates only
sh.g.emm2 <- CLD(emmeans(SHs.m3, pairwise ~ Diet + Density + Timepoint,
                      method = "tukey"), Letters = letters)
sh.g.emm2.s <- subset(sh.g.emm2, select = c(1:5))
sh.g.emm2.s <- subset(sh.g.emm2.s, Density == 0)
sh.g.emm2.s$Type <- "substrate"

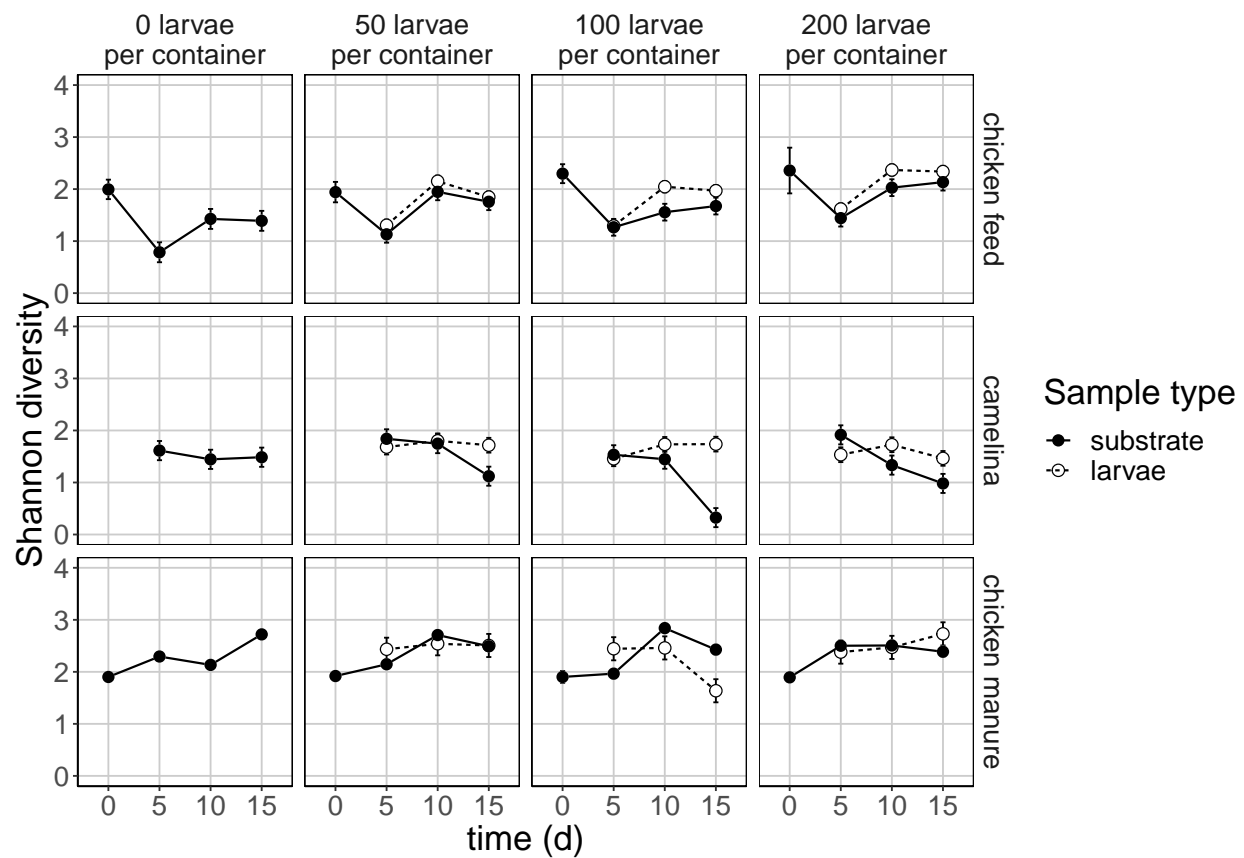
# extract mean and SE for timepoint 0 (not in models)
sh.g.emm3 <- subset(sh.gsum, Timepoint == 0)
sh.g.emm3.s <- subset(sh.g.emm3, select = c(1:5,8))
colnames(sh.g.emm3.s)[5:6] <- c("emmean", "SE")

# combine all EMM and SE
sh.g.emm4 <- rbind(sh.g.emm.s, sh.g.emm2.s, sh.g.emm3.s)

# reorder factor levels
sh.g.emm4$Density <- factor(sh.g.emm4$Density, levels(sh.g.emm4$Density)[c(4, 1:3)])
sh.g.emm4$Timepoint <- factor(sh.g.emm4$Timepoint, levels(sh.g.emm4$Timepoint)[c(4, 1:3)])
```

5.4. Plot

```
p.shg.bk <- ggplot(sh.g.emm4, aes(x = Timepoint, y = emmean,
                                group = interaction(Diet, Density, Type)))
p.shg.bk <- p.shg.bk + geom_line(size = .6, aes(linetype = Type)) +
  geom_errorbar(aes(ymin = emmean-SE, ymax = emmean+SE), width = .1) +
  geom_point(shape = 16, size = 3, colour = "white") +
  geom_point(aes(shape = Type), size = 3, colour = "black") +
  scale_shape_manual(values = c(16,1)) +
  labs(x = "time (d)", y = "Shannon diversity",
       shape = "Sample type", linetype = "Sample type") +
  scale_y_continuous(limits = c(0,4), n.breaks = 5) +
  facet_grid(Diet ~ Density, labeller = labs_div) +
  theme_div
p.shg.bk
```



```
ggsave(plot= p.shg.bk, "./figures/Diversity_Shannon_genus_EMM_black.png",
       h = 7, w = 10)
ggsave(plot= p.shg.bk, "./figures/Diversity_Shannon_genus_EMM_black.tiff",
       h = 175, w = 250, u = "mm", dpi = 600)
ggsave(plot= p.shg.bk, "./figures/Diversity_Shannon_genus_EMM_black.pdf",
       h = 175, w = 250, u = "mm")
```