Does site location and skin microbial diversity impact ectoparasite load in painted turtles?

Morgan Clark and Brandon Kristy

Introduction

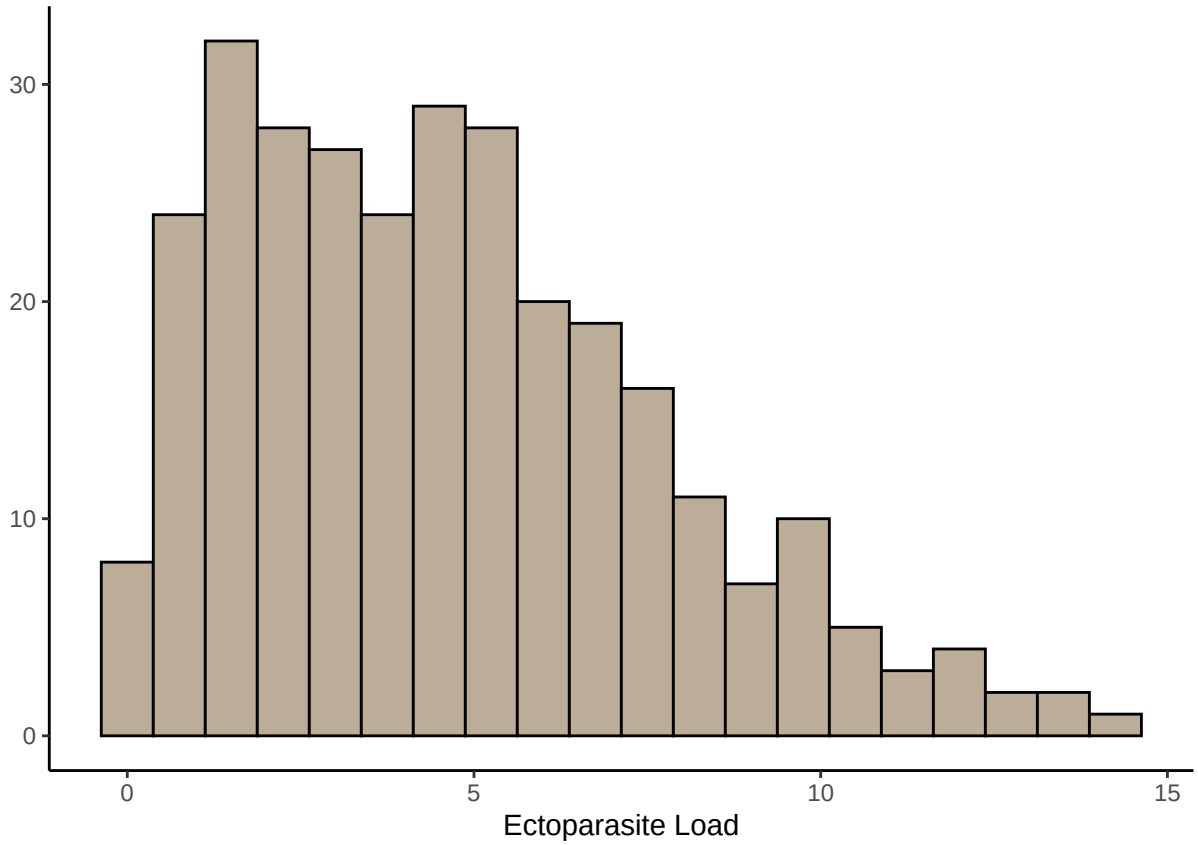
The following scenario represents our hypothetical experimental design and hypothesis-testing for our simulated dataset:

Painted turtles (*Chrysemys picta*) are a common freshwater turtle species in North America. As they inhabit rivers, lakes, and ponds they are often subject to aquatic ectoparasites, such as leeches, which negatively impact host health by absorbing nutrients and serve as vectors for more harmful blood parasites. Recently there has been an uptick in research investigating the microbiome of painted turtles, focusing largely on quantifying bacterial diversity and function in the gut. However, little work has been done to quantify and survey the microbial community dwelling on the turtle's shell, or how this microbial community interacts with stressors in their environment, such as ectoparasites. The skin is the turtle's primary barrier against the outside world; it is possible that microbes living on, in, and around the turtle's skin tissue may functionally impact host fitness. In this study, we survey three populations of painted turtles, quantifying ectoparasite load and shell-microbiota diversity across each population. Because endophytic microbial communities functionally impact host health, we predict that turtles with higher microbial diversity may be less susceptible to ectoparasite predation, and consequently have fewer attached leeches. Further, because food availability, prey diversity, and environmental stressors vary across each of the locally adapated populations, we predict ectoparasite load will vary across each site.

We used a standardized sampling protocol across three sites - each with locally adapted red-painted turtle populations. Over the course of the summer season, we used sardine-baited cathedral traps to mark and capture 100 turtle individuals across three sites. We transported turtles individually back to Kellogg Biological Station (KBS) in damp cloth bags to keep the leeches from desiccating. To quantify ectoparasite load, we counted leeches on the shell, anterior body, and posterior body, normalizing ectoparasite counts by total body area.

To evaluate turtle shell microbial diversity, we collected skin swabs in three regions of the shell: the carapace, the plastron, and the bridge. We collected an additional skin swab on the scales around the head, neck, and appendages. For each turtle, these swabs were pooled together; we extracted DNA from these swab samples from 100 turtles in each of the selected populations, resulting in 300 total samples. To profile the bacterial community, we performed a 2-step PCR reaction of the 16S rRNA gene marker with the 515F-806R primer set. We sequenced clean PCR products on the Illumina Miseq platform. In short, filtered sequences with a high degree of similarity (97%) were clustered into Operational Taxonomic Units (OTUs). In this study, we define microbial diversity as the number of OTUs within each sample, which is a proxy for within-sample species richness: alpha diversity. We returned all turtles back to their capture locations within 24 hours of their initial capture!

Before generating the models, we first inspect the distribution of our response variable: ectoparasite load.



The response variable, ectoparasite load, appears to be normally distributed, so we will generate our models using the lm() function. Below, we create 5 models for model comparison.

Model Comparison

```
# NULL MODEL
m1 <- lm(ecto.load ~ 1, data = dat)

# ecto ~ diversity
m2 <- lm(ecto.load ~ microbe_div.scaled, data = dat)

# ecto ~ site
m3 <- lm(ecto.load ~ site, data = dat)

# ecto ~ diversity + site
m4 <- lm(ecto.load ~ microbe_div.scaled + site, data =dat)

# ecto ~ diversity * site
m5 <- lm(ecto.load ~ microbe_div.scaled * site, data = dat)
```

To compare the models we used AIC model comparison. This allows us to select the best-fitting model based on AIC values.

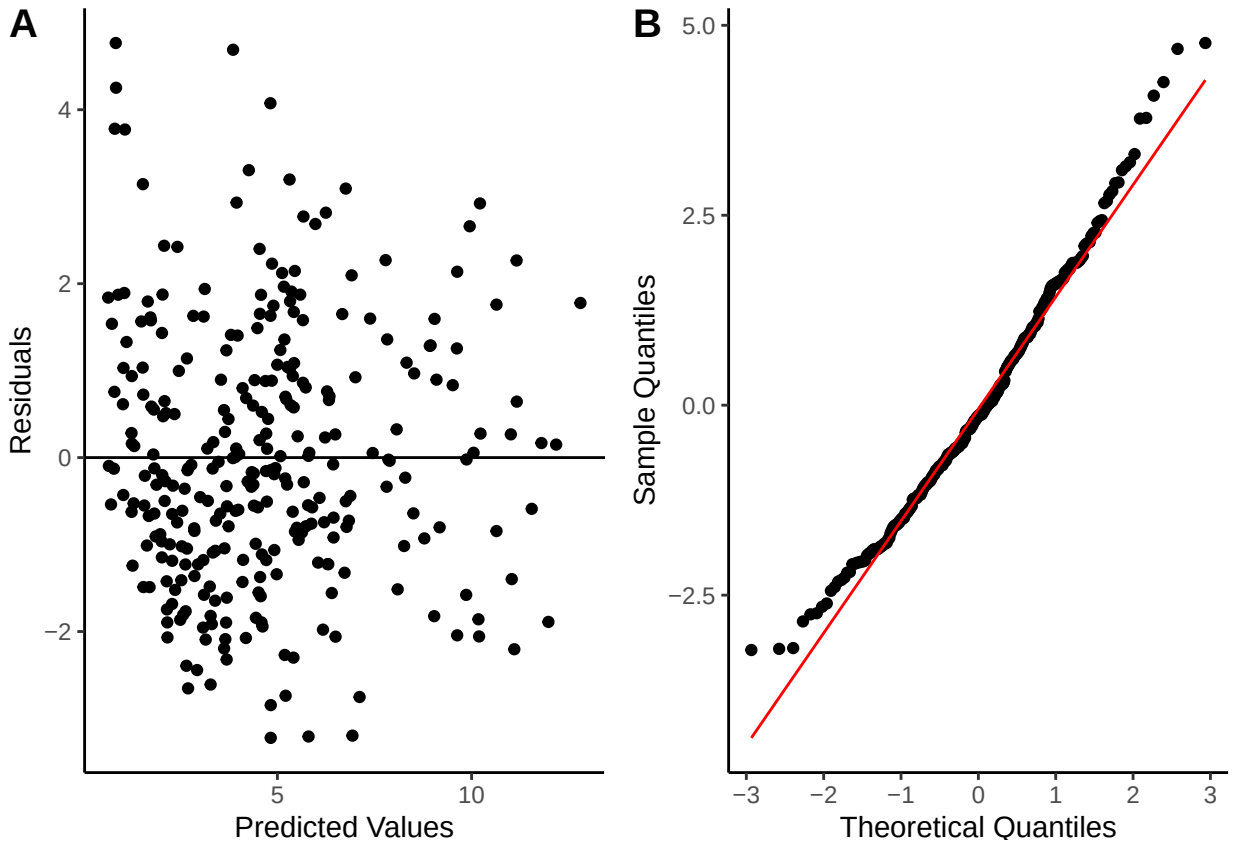
```
##    logLik AICc   dLogLik dAICc  df weight
## m5 -546.4 1107.1  212.5     0.0 7  1
## m4 -594.4 1199.1  164.5    92.0 5  <0.001
## m2 -672.8 1351.7   86.1   244.6 3  <0.001
```

```
## m3 -724.7 1457.6   34.2   350.5 4  <0.001
## m1 -758.9 1521.9    0.0   414.8 2  <0.001
```

According to the AIC model comparison, m5 is the best-fitting model, explaining nearly 100% of the weight and the lowest AIC value compared to the other models. The difference in the log likelihood between m5 and m4, the additive model, was 212.5. Because of this, we we will move forward with the analysis with model 5, the interactive model, in our model evaluation. We performed model testing by plotting the residuals of model 5 with its predicted values. In addition, we also created a Normal Q-Q plot. The summary of the interactive model, the residuals plot, and the Q-Q plot are now shown below:

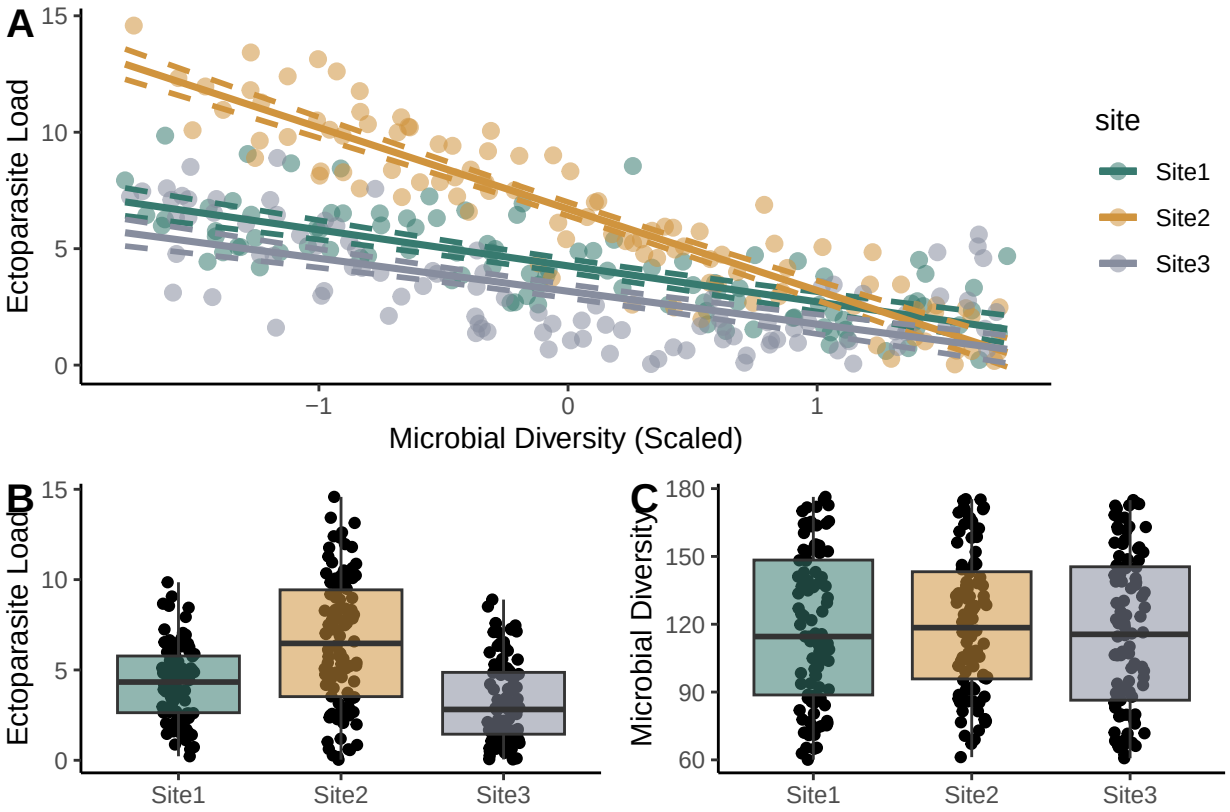
Model Evaluation

```
##
## Call:
## lm(formula = ecto.load ~ microbe_div.scaled * site, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2223 -1.0442 -0.1431  0.9453  4.7670
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    4.2618     0.1510  28.216  < 2e-16 ***
## microbe_div.scaled            -1.5453     0.1495 -10.338  < 2e-16 ***
## siteSite2                      2.4424     0.2138  11.425  < 2e-16 ***
## siteSite3                     -1.0903     0.2137  -5.102 6.05e-07 ***
## microbe_div.scaled:siteSite2  -1.9531     0.2201  -8.876  < 2e-16 ***
## microbe_div.scaled:siteSite3   0.1312     0.2081   0.631    0.529
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.51 on 294 degrees of freedom
## Multiple R-squared:  0.7576, Adjusted R-squared:  0.7534
## F-statistic: 183.7 on 5 and 294 DF,  p-value: < 2.2e-16
```



We performed model evaluation using two methods, first by plotting the predicted model values against the residuals in panel A, and then generating a Q-Q plot in panel B. By plotting the predicted values against the residuals for model 5, we are able to see that the values are randomly scattered around zero without any clear outliers, indicating that model 5 is a good fit for the data. The Q-Q plot allows us to visualize the distribution of the residuals. Here our Q-Q plot shows that the residuals are normally distributed because the points fall along a straight line. Because the interactive model does a sufficient job explaining the variability in our response variable, we can now plot our model predictions over the raw dataset to identify significant trends worth evaluating:

Model Analysis



Here we visualized the raw data for the relationship between microbial diversity (scaled) and ectoparasite load in panel A. The model fit is displayed across each site; the 95% confidence intervals are presented as the dashed lines. We can visualize general trends in the data, evaluate model fit, and identify any outliers that may not be included in the model predictions. From this figure, we determined that model 5 continues to be the most appropriate model for our data as it fits the overall trend and variability of our data sufficiently. Ultimately, increasing microbial diversity negatively impacts Ectoparasite load across all sites. In addition, Ectoparasite load across each site is plotted in panel B. Lastly, microbial diversity across each site is plotted in panel C. There is no significant difference in microbial diversity across each site.

Evaluating the effect sizes from the terms in model five allows us to compare the relative strength of relationships between predictor and response variables. In this case, this will include microbial diversity (scaled) across and between sites. The interactive model explained 75.76% of the variability in our response variable: Ectoparasite load (R-squared = 0.7576). Across all three sites, there is a significant, negative effect of microbial diversity on ectoparasite load ($p < 0.01$). A 1-unit change in microbial diversity decreases ectoparasite load by up to 101% (95% CI: 41 % - 160%). Ectoparasite load was variable across all three sites. When microbial diversity is zero, Site 2's mean ectoparasite load was 1.57 times higher than Site 1's ectoparasite load (95% CI: 1.53 - 1.61) and 2.11 times higher than Site 3's ectoparasite load (95% CI: 2.02 - 2.23). Further, Site 3's ectoparasite load was 1.34 times lower than Site 1's ectoparasite load (95% CI: 1.31 - 1.38) when microbial diversity is zero. In addition to their individual effects on ectoparasite load, both site and microbial diversity interacted to influence ectoparasite load ($p < 0.01$). Specifically, a 1-unit change in microbial diversity decreases ectoparasite load by 54.53% (95% CI: 25.10% - 83.94%) in site 1 and 95% in site 2 (95% CI: 52% - 134%). Interestingly, there was no significant relationship between ectoparasite load and microbial diversity in site 3 (95% CI: -28% - 54%; $p = 0.529$).

Overall, our results identify a potential relationship between skin microbial diversity and site on ectoparasite load for painted turtles. It is likely that site-specific influences on food availability, prey diversity, and

environmental stressors results in variable ectoparasite load. These site-specific differences may result in variable skin microbial community composition. Overall, additional work is needed to evaluate site-specific impacts on microbial community composition and functional diversity within the turtle's skin microbiome. This work furthers our understanding of how host-associated microbial communities across locally adapted populations play a role on turtle host fitness and health.

The code used to simulate the data, compare models, evaluate the best-fit model, and analyze the best-fit model are shown below:

```
## MODEL SIMULATION
set.seed(123)
# Define the number of groups and the number of individuals sampled within each group
n.groups <- 3
n.sample <- 100

n <- 300

# Create a numerical indicator describing the region where each individual was recovered
x <- gl(n = n.groups, k = n.sample, length = n)
site <- factor(x, labels = c("Site1", "Site2", "Site3"))

turtle_id <- seq(1:300)
# Microbial Diversity
microbe_div <- runif(n, 60, 177)
microbe_div.scaled <- scale(microbe_div, center=TRUE, scale = TRUE)

# Build the design matrix. Put the site-specific effects
# first, followed by turtle size and microbe diversity effects, followed by their interactions.
X.mat <-  model.matrix(~ site * microbe_div.scaled)
print(X.mat, dig = 2)

# Select the parameter values for each of the parameters that you defined
beta.vec <- c(4.24, 2.23, -2.34, -1.5, -2.2, -1.3)


# Here's the recipe for assembling the parasite counts in three steps:
# 1. Add up all components of the linear model to get the linear predictor, which is the #    expected
#    Obtain the value of the linear predictor by matrix multiplication of the
#    design matrix (Xmat) and the parameter vector (beta.vec).

lin.pred <- X.mat %*% beta.vec

# 3. Generate the data from a normal distribution with linear predictor as the mean.
ecto.load <- abs(rnorm(n =n, mean = lin.pred, sd = 1.45))

# Package the data ecto.load, site, and microbial diversity into
# a data frame.
dat <- data.frame(turtle_id, ecto.load, site, microbe_div, microbe_div.scaled)

## MODEL COMPARISON
# AIC
tab = AICctab(m1, m2, m3, m4, m5, base=TRUE, delta=TRUE, weights=TRUE, logLik=TRUE)
tab
```

```
## MODEL EVALUATION
summary(m5)
# plot residuals of best-fit model here:
predicted <- predict(m5)
residuals <- resid(m5)
d <- data.frame(predicted, residuals)

d %>%
  ggplot(aes(x=predicted, y = residuals)) +
  geom_point() +
  geom_abline(intercept = 0, slope = 0) +
  ylab("Residuals") +
  xlab("Predicted Values") +
  theme_classic() -> p1

d %>%
  ggplot(aes(sample = residuals)) +
  stat_qq() +
  stat_qq_line(col = "red") +
  ylab("Sample Quantiles") +
  xlab("Theoretical Quantiles") +
  theme_classic() -> p2



ggdraw() +
  draw_plot(p1, x = 0.0, y = 0.0, width = 0.5, height = 1) +
  draw_plot(p2, x = 0.5, y= 0.0, width = 0.5, height =1) +
  draw_plot_label(label = c("A", "B"), size = 15,
                  x = c (0, 0.5), y = c(1, 1))
## MODEL ANALYSIS

## SITE 1
newdata.site1 <- data.frame(ecto.load = seq(min(dat$ecto.load), max(dat$ecto.load), length.out =100),
                            microbe_div.scaled = seq(min(dat$microbe_div.scaled), max(dat$microbe_div.sca
                            site = "Site1",
                            turtle_id = 19)



## SITE 2
newdata.site2 <- data.frame(ecto.load = seq(min(dat$ecto.load), max(dat$ecto.load), length.out =100),
                            microbe_div.scaled = seq(min(dat$microbe_div.scaled), max(dat$microbe_div.sca
                            site = "Site2",
                            turtle_id = 19)



## SITE 3
newdata.site3 <- data.frame(ecto.load = seq(min(dat$ecto.load), max(dat$ecto.load), length.out =100),
                            microbe_div.scaled = seq(min(dat$microbe_div.scaled), max(dat$microbe_div.sca
                            site = "Site3",
                            turtle_id = 19)
```

```r
# Then use predict() to get predictions and SE of predictions for ctrl and excl
# SITE 1
m5.pred.site1 <- predict(m5, newdata = newdata.site1, allow.new.levels=TRUE, se = TRUE)
m5.pred.site1 <- as.data.frame(m5.pred.site1)

m5.pred.site1$upperCI <- (m5.pred.site1$fit + 1.96*(m5.pred.site1$se.fit))
m5.pred.site1$lowerCI <- (m5.pred.site1$fit - 1.96*(m5.pred.site1$se.fit))

# SITE 2
m5.pred.site2 <- predict(m5, newdata = newdata.site2, allow.new.levels=TRUE, se = TRUE)
m5.pred.site2 <- as.data.frame(m5.pred.site2)

m5.pred.site2$upperCI <- (m5.pred.site2$fit + 1.96*(m5.pred.site2$se.fit))
m5.pred.site2$lowerCI <- (m5.pred.site2$fit - 1.96*(m5.pred.site2$se.fit))

# SITE 3
m5.pred.site3 <- predict(m5, newdata = newdata.site3, allow.new.levels=TRUE, se = TRUE)
m5.pred.site3 <- as.data.frame(m5.pred.site3)

m5.pred.site3$upperCI <- (m5.pred.site3$fit + 1.96*(m5.pred.site3$se.fit))
m5.pred.site3$lowerCI <- (m5.pred.site3$fit - 1.96*(m5.pred.site3$se.fit))

# Plot model predictions with raw data:
  ggplot() +
    geom_point(data =dat, aes(x = microbe_div.scaled, y = ecto.load, col = site), alpha = 0.55, size = 1
    geom_line(aes(x = newdata.site1$microbe_div.scaled, y = m5.pred.site1$fit, col = newdata.site1$site)
    scale_color_manual(values = c('#377A6F', '#D0943E', '#878D9E' )) +
    geom_line(aes(x = newdata.site1$microbe_div.scaled, y = m5.pred.site1$upperCI, col = newdata.site1$s
    geom_line(aes(x = newdata.site1$microbe_div.scaled, y = m5.pred.site1$lowerCI, col = newdata.site1$s
    geom_line(aes(x = newdata.site2$microbe_div.scaled, y = m5.pred.site2$fit, col = newdata.site2$site)
    geom_line(aes(x = newdata.site2$microbe_div.scaled, y = m5.pred.site2$upperCI, col = newdata.site2$s
    geom_line(aes(x = newdata.site2$microbe_div.scaled, y = m5.pred.site2$lowerCI, col = newdata.site2$s
    geom_line(aes(x = newdata.site3$microbe_div.scaled, y = m5.pred.site3$fit, col = newdata.site3$site)
    geom_line(aes(x = newdata.site3$microbe_div.scaled, y = m5.pred.site3$upperCI, col = newdata.site3$s
    geom_line(aes(x = newdata.site3$microbe_div.scaled, y = m5.pred.site3$lowerCI, col = newdata.site3$s
    theme_classic() +
    ylab("Ectoparasite Load") +
    xlab("Microbial Diversity (Scaled)") -> p1

# Plot Boxplot of ectoparasite load across the sites
dat %>%
  ggplot(aes(x=site, y = ecto.load, fill = site)) +
  scale_fill_manual(values = c('#377A6F', '#D0943E', '#878D9E' )) +
  geom_jitter(width = 0.1) +
  geom_boxplot( alpha = 0.55) +
  xlab("Site") +
  ylab("Ectoparasite Load") +
  theme_classic() -> p2

# Plot Boxplot of microbial diversity across the sites
dat %>%
  ggplot(aes(x=site, y = microbe_div, fill = site)) +
```

```
  scale_fill_manual(values = c('#377A6F', '#D0943E', '#878D9E' )) +
  geom_jitter(width = 0.1) +
  geom_boxplot( alpha = 0.55) +
  xlab("Site") +
  ylab("Microbial Diversity") +
  theme_classic() -> p3

## Arrange all plots
## Arrange all plots
ggdraw() +
  draw_plot(p1, x = 0.0, y = 0.45, width = 1.0, height = 0.55) +
  draw_plot(p2, x = 0.0, y= 0.0, width = 0.5, height =0.45) +
  draw_plot(p3, x=0.5, y = 0.0, width = 0.5, height = 0.45) +
  draw_plot_label(label = c("A", "B", "C"), size = 15,
                  x = c (0, 0,0.5 ), y = c(1, 0.45,0.45 ))

# Effect size analysis
## R - squared value
# 0.7576
## Effect of microbial diversity on ectoparasite load on across ALL sites
summary(m5) # -1.5468 is the effect (slope) between microbial diversity and ectoparasite load
confint(m5, level = 0.95)

## Means parameterization
m5.means <- lm(ecto.load ~ microbe_div.scaled * site - 1, data = dat)
summary(m5.means)
confint(m5.means, level = 0.95)

# How many times greater is Site 2's ectoparasite load relative to Site 1's?
6.7042 / 4.2618 #1.57 effect
6.4065 / 3.9645 # 1.61 upper CI
7.0019022 / 4.5590521 # 1.535 lower CI

# How many times greater is Site 2's ectoparasite load relative to Site 3's?
6.7042/3.1715 # 2.11 effect
6.4065 / 2.8739 # 2.23 upper CI
7.0019022 / 3.4690 # 2.018 lower CI

# How much lower is Site 3's ectoparasite load relatie to Site 1's?
4.2618 / 3.1715 # 1.343 effect
3.9645 / 2.8739 # 1.38 upper CI
4.5590521 / 3.4690495 # 1.314 lower CI

# Microbial diversity effect on ectoparasite load (using m2)
m2.means <- lm(ecto.load ~ microbe_div.scaled - 1, data =dat)
summary(m2.means)
confint(m2.means)

# Interactive effects
summary(m5.means)
```