3A & B: AE Biofilm

Caitlin Nordheim-Maestas

# Load in data and libraries

## read in and clean data  
library(tidyverse) # for cleaning and viewing data  
library(gt) # pretty stats tables  
library(broom) # cleaning for gt  
library(here) # for importing data  
library(car) # stats tests like Levene's  
library(multcompView) # view cld  
library(multcomp) # stats  
library(nlme) # mixed effects models  
library(emmeans) # for pairwise comparisons, especially on mixed effects models and glms  
library(ggpubr) # for making ggqq plot  
library(patchwork) # for combining figures  
  
ns\_biofilm\_bd <- read.csv(here("data", "nine-sites-biofilm-on-Bd - Sheet1.csv"))  
fig\_3b\_raw <- read.csv(here("data", "final\_NCOS\_2024\_reformatted\_for\_R.xlsx - Fig3B.csv"))  
  
# Colors: these are from Paul Tol's colorblind friendly palette  
with\_microbes\_40\_color <- "#999933"  
no\_microbes\_.22\_color <- "#88ccee"  
  
# set up custom theme  
myCustomTheme <- function() {  
 theme\_light() +  
 theme(axis.text = element\_text(size = 12, family = "Times", color = "black"),  
 axis.title.x = element\_text(margin = margin(t = 10), size = 14, face = "bold", family = "Times", color = "black"), # Add space between x-axis label and axis  
 axis.title.y = element\_text(margin = margin(r = 10), size = 14, face = "bold", family = "Times", color = "black"), # Add space between y-axis label and axis  
 title = element\_text(size = 12, face = "bold", family = "Times"),  
 plot.caption = element\_text(size = 10, face = "italic", family = "Times"),  
 legend.text = element\_text(size = 10, family = "Times"), # Increase legend text size  
 panel.grid.major.x = element\_blank(), # Remove major vertical grid lines  
 panel.grid.minor.x = element\_blank(), # Remove minor vertical grid lines  
 panel.grid.major.y = element\_blank(), # Remove major horizontal grid lines  
 panel.grid.minor.y = element\_blank(), # Remove minor horizontal grid lines  
 strip.text = element\_text(size = 12, face = "bold", family = "Times", color = "black"), # Set strip text style  
 strip.background = element\_rect(fill = "white", color = "grey"), # Set strip background to white, # color = "black"  
 axis.ticks = element\_blank() # Remove x and y axis ticks  
 )}

# 3A

Effect of the East Bay aquatic environmental biofilm on Bd growth - “9 sites”

Does the difference in rate loss of Bd in the **adherent** AE biofilm from day 1 to day 7 differ between the two filter types?

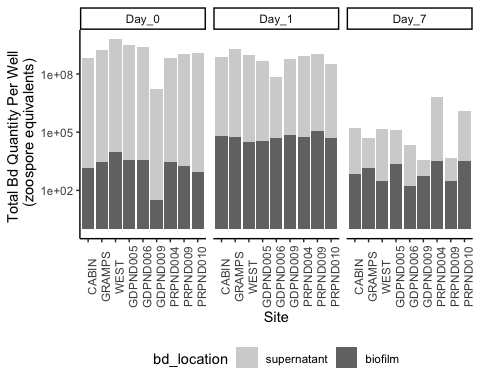
## 3A Data Wrangling

eb\_ae <- ns\_biofilm\_bd  
  
# set as factors with levels so they appear in order in plots later  
eb\_ae$site <- factor(eb\_ae$site,  
 levels = c("CABIN", "GRAMPS", "WEST", "GDPND005", "GDPND006", "GDPND009", "PRPND004", "PRPND009", "PRPND010"))  
eb\_ae$bd\_location <- factor(eb\_ae$bd\_location, levels = c("supernatant", "biofilm"))  
eb\_ae$day <- factor(eb\_ae$day, levels = c("Day\_0", "Day\_1", "Day\_7"))  
  
# Biofilm only, no supernatant, and days 1 and 7 only, no day 0  
eb\_ae\_bf\_only <- eb\_ae %>%   
 filter(bd\_location == "biofilm") %>%   
 filter(day != "Day\_0") %>%   
 #log bd qty  
 mutate(log\_qty = log(bd\_qty))

## EDA

Renwei barplot

eb\_ae %>%  
ggplot(aes(y= bd\_qty, x = site, fill=bd\_location)) +   
 geom\_col() +  
 facet\_grid(.~day)+  
 theme\_classic()+  
 scale\_fill\_manual(values = c("lightgrey", "gray45" )) +  
 theme(axis.text.x = element\_text(angle = 90),  
 legend.position = "bottom") +   
 scale\_y\_log10() +  
 xlab("Site") +  
 ylab("Total Bd Quantity Per Well \n (zoospore equivalents)")



# part2 <- eb\_ae\_bf\_only %>%  
# ggplot(aes(y= bd\_qty, x = day)) +   
# geom\_boxplot(fill = no\_microbes\_.22\_color) +  
# geom\_jitter(alpha = 0.3, width = 0.1) +  
# theme\_classic() +  
# scale\_y\_log10() +  
# theme(legend.position = "none",  
# strip.text = element\_text(face="bold"),  
# axis.title = element\_text(face = "bold")) +   
# xlab("Time (days)") +  
# ylab("Bd Quantity \n (zoospore equivalents)") +  
# scale\_x\_discrete (labels= c("Day\_1" = "Day 1", "Day\_7" = "Day 7"))  
#   
# # Let's add significance letters  
# significance\_data <- tibble(  
# day = factor(c("Day\_1", "Day\_7"), levels = c("Day\_1", "Day\_7")),  
# y\_position = c(1.6e+05, 1.4e+04), # Adjust this depending on your plot's scale  
# label = c("a", "b"))  
#   
# part2 +   
# geom\_text(data = significance\_data, aes(x = day, y = y\_position, label = label),  
# position = position\_dodge(width = 0.75), vjust = 0)

## Stats and assumption testing

Question:

Does the amount of Bd in the biofilm differ between day 1 and day 7?

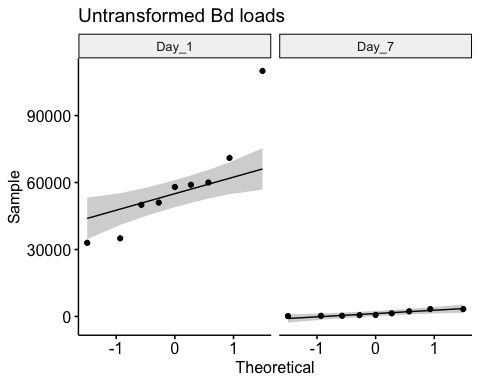
The samples are essentially paired by site, so a paired t-test is most appropriate

Assumptions:

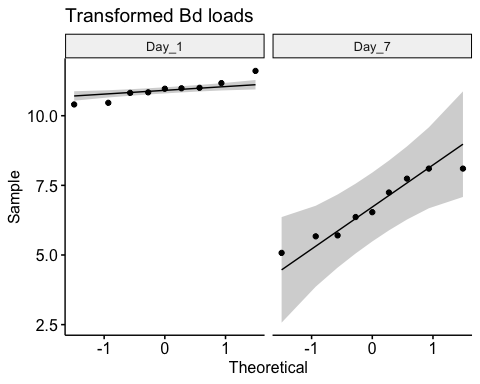
Assumes that the observations from each group represent a random sample from the population. Assumes that the difference of the two observations follow a normal distribution.

Check assumptions

# check normality of the differences across groups let's try untransformed  
eb\_ae\_bf\_only %>%   
 ggqqplot("bd\_qty", title = "Untransformed Bd loads") +  
 facet\_wrap(~day) # gotta transform the data, day 1 is not normal



eb\_ae\_bf\_only %>%   
 ggqqplot("log\_qty", title = "Transformed Bd loads") +  
 facet\_wrap(~day) # not perfect but close, let's try shapiro tests



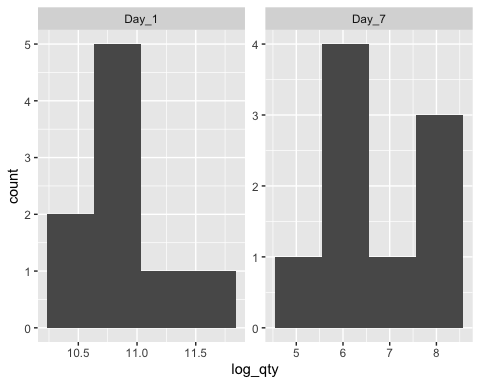
# Shapiro tests  
day\_one <- eb\_ae\_bf\_only %>%   
 filter(day == "Day\_1") %>% # filter to only include day 1  
 pull(log\_qty)  
  
shapiro.test(day\_one) # p >> 0.05, it's normal!

Shapiro-Wilk normality test  
  
data: day\_one  
W = 0.93796, p-value = 0.5606

day\_seven <- eb\_ae\_bf\_only %>%   
 filter(day == "Day\_7") %>% # filter to only include day 1  
 pull(log\_qty)  
  
shapiro.test(day\_seven) # p >> 0.05, it's normal!

Shapiro-Wilk normality test  
  
data: day\_seven  
W = 0.92235, p-value = 0.412

# Histograms for funsies  
eb\_ae\_bf\_only %>%   
ggplot(aes(x = log\_qty)) + # x-axis  
 geom\_histogram(bins = 4) + # make a histogram  
 facet\_wrap(~ day, # make multiple panels by day  
 scales = "free") # let the axes vary between panels



Stats

# Step 3: run the paired t-test  
t.test(day\_one, day\_seven, paired = TRUE)

Paired t-test  
  
data: day\_one and day\_seven  
t = 10.094, df = 8, p-value = 7.915e-06  
alternative hypothesis: true mean difference is not equal to 0  
95 percent confidence interval:  
 3.233993 5.149159  
sample estimates:  
mean difference   
 4.191576

**There is significantly more Bd in the biofilm on Day 1 than on Day 7 (t = 10.094, df = 8, p-value < 0.0001)**

# 3b NCOS AE biofilm

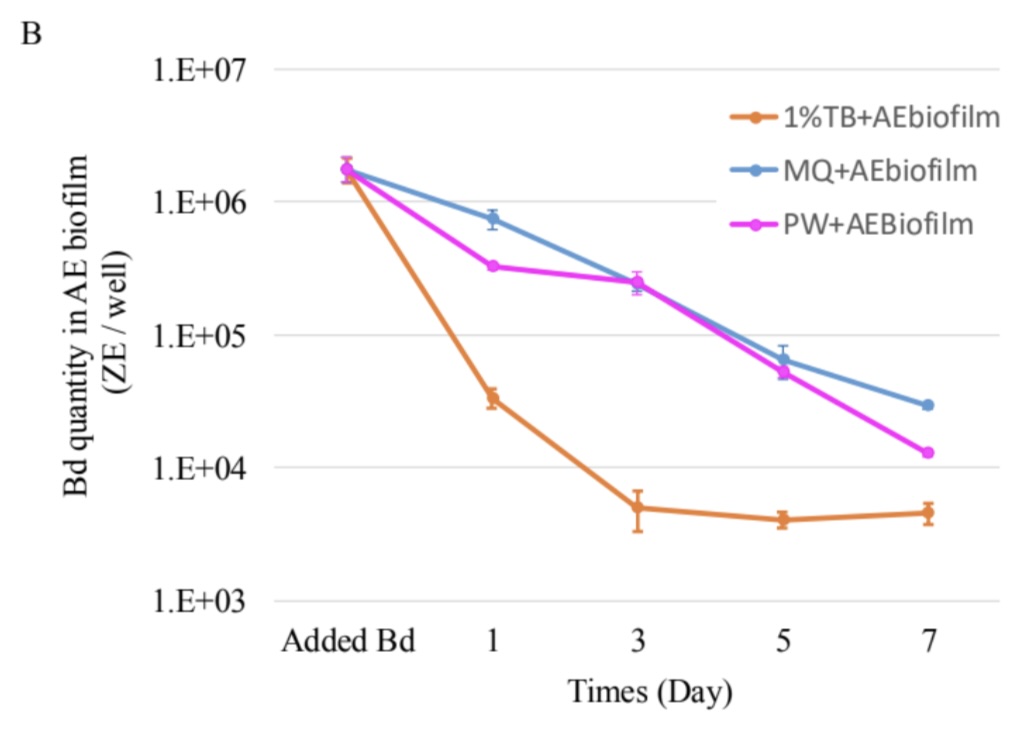
All microbe-depleted, NO pw microbe+ treatment

## 3b Data wrangling

## aquatic environmental biofilm   
# add column for microbes or no  
ae <- fig\_3b\_raw %>%   
 rename(sample\_ID = Adherent.sample.ID) %>%   
   
 # add columns for components y/n  
 # add column for TB or no  
 mutate(TB = case\_when(  
 str\_detect(sample\_ID, "TB") ~ "y",  
 TRUE ~ "n"  
 )) %>%   
 # add column for PW or no  
 mutate(PW = case\_when(  
 str\_detect(sample\_ID, "PW") ~ "y",  
 TRUE ~ "n"  
 )) %>%   
 # update day to have day in it  
 mutate(day = case\_when(  
 day == 1 ~ "Day\_1",  
 day == 3 ~ "Day\_3",  
 day == 5 ~ "Day\_5",  
 day == 7 ~ "Day\_7",  
 day == 0 ~ "Day\_0" # In case you want to include Day\_0 as well  
 )) %>%   
 mutate(day = factor(day)) %>%   
 # add numeric for plotting  
 mutate(day\_numeric = as.numeric(gsub("Day\_", "", as.character(day))))  
  
# control data for ae  
ae\_control\_data <- ae %>%   
 filter(day == "Day\_0") %>%   
 mutate(day\_numeric = as.numeric(gsub("Day\_", "", as.character(day)))) %>%   
 dplyr::select(day, adh, day\_numeric)  
  
ae\_summary <- ae %>%  
 group\_by(day, sample\_ID) %>%  
 reframe(mean = mean(adh), # calculate the mean  
 n = length(adh), # count the number of observations  
 df = n - 1, # calculate the degrees of freedom  
 sd = sd(adh), # calculate the standard deviation  
 se = sd/sqrt(n), # calculate the standard error  
 ) %>%  
 # add column for TB or no  
 mutate(TB = case\_when(str\_detect(sample\_ID, "TB") ~ "y", TRUE ~ "n")) %>%  
 # add column for PW or no  
 mutate(PW = case\_when(str\_detect(sample\_ID, "PW") ~ "y", TRUE ~ "n")) %>%   
 mutate(day\_numeric = as.numeric(gsub("Day\_", "", as.character(day))))  
  
ae\_noday0 <- ae %>%   
 filter(day != "Day\_0") %>%   
 mutate(log\_adh = log(adh)) # note: no zeroes so not log + 1

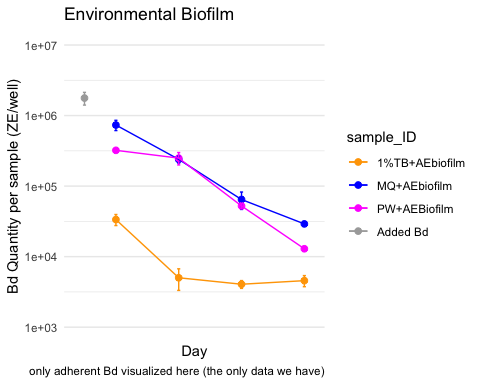
## 3b EDA

Renwei’s figure



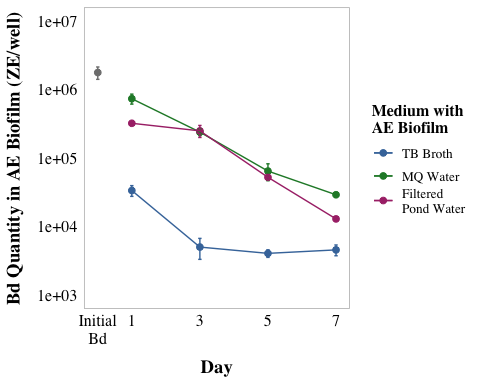
ggplot version: replicating Renwei’s AE

ae\_summary %>%  
 # reorder to match Renwei's plot  
 mutate(sample\_ID = factor(sample\_ID,  
 levels = c("1%TB+AEbiofilm", "MQ+AEbiofilm",  
 "PW+AEBiofilm", "Added Bd" ))) %>%  
  
 ggplot(aes(x = day\_numeric,  
 y = mean,  
 color = sample\_ID)) +  
 geom\_point(size = 2) +  
 geom\_errorbar(aes(ymin = mean - se, # plot the standard error  
 ymax = mean + se),  
 width = 0.1) +  
 geom\_line() +  
 scale\_y\_log10(limits = c(1e3, 1e7),  
 breaks = c(1e3, 1e4, 1e5, 1e6, 1e7)) +  
 # vibes  
 labs(x = "Day",  
 y = "Bd Quantity per sample (ZE/well)",  
 title = "Environmental Biofilm",  
 caption = "only adherent Bd visualized here (the only data we have)") +  
 scale\_color\_manual(values = c("1%TB+AEbiofilm"= "orange",  
 "MQ+AEbiofilm" = "blue",  
 "PW+AEBiofilm" = "magenta",  
 "Added Bd" = "darkgrey")) + # Assign specific colors to match RC's plot  
 theme\_minimal() +  
 theme(  
 panel.grid.major.x = element\_blank(), # Remove major vertical grid lines  
 panel.grid.minor.x = element\_blank()) + # Remove minor vertical grid lines  
 scale\_x\_discrete(labels = c("Day\_0" = "initial Bd", "Day\_1" = "1", "Day\_3" = "3", "Day\_5" = "5", "Day\_7" = "7")) # Use discrete scale for day



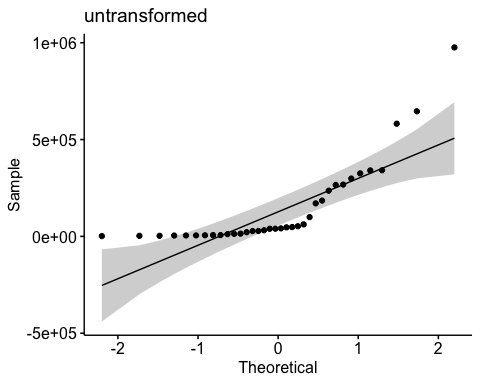
Caitlin’s version AE

ae\_summary %>%  
 # reorder to match Renwei's plot  
 mutate(sample\_ID = factor(sample\_ID,  
 levels = c("1%TB+AEbiofilm", "MQ+AEbiofilm",  
 "PW+AEBiofilm", "Added Bd" ))) %>%  
  
 ggplot(aes(x = day\_numeric,  
 y = mean,  
 color = sample\_ID)) +  
 geom\_point(size = 2) +  
 geom\_errorbar(aes(ymin = mean - se, # plot the standard error  
 ymax = mean + se),  
 width = 0.1) +  
 geom\_line() +  
 scale\_y\_log10(limits = c(1e3, 1e7),  
 breaks = c(1e3, 1e4, 1e5, 1e6, 1e7)) +  
 # vibes  
 labs(x = "Day",  
 y = "Bd Quantity in AE Biofilm (ZE/well)",  
 color = "Medium with\nAE Biofilm")+ # Title for color legend  
  
 scale\_color\_manual(values = c("1%TB+AEbiofilm"= "#4477AA",  
 "MQ+AEbiofilm" = "#228833",  
 #"Added Bd" = "darkgrey",  
 "PW+AEBiofilm" = "#AA3377"),  
 labels = c("1%TB+AEbiofilm" = "TB Broth",  
 "MQ+AEbiofilm" = "MQ Water",  
 "PW+AEBiofilm" = "Filtered\nPond Water",  
 "Added Bd" = "Initial Bd")) + # Custom labels  
  
 myCustomTheme()+  
 scale\_x\_continuous(breaks = c(0, 1, 3, 5, 7),  
 labels = c("Initial\nBd", "1", "3", "5", "7"))



visualize y var: bd load

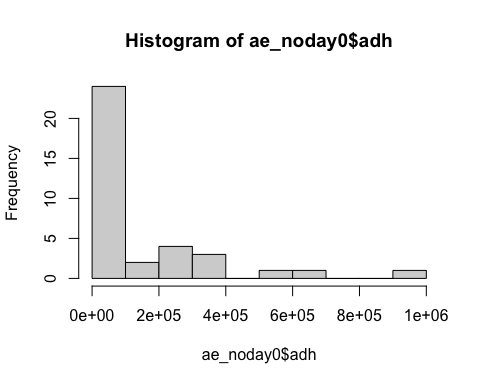
# untransformed  
ggqqplot(ae\_noday0, "adh", title = "untransformed")



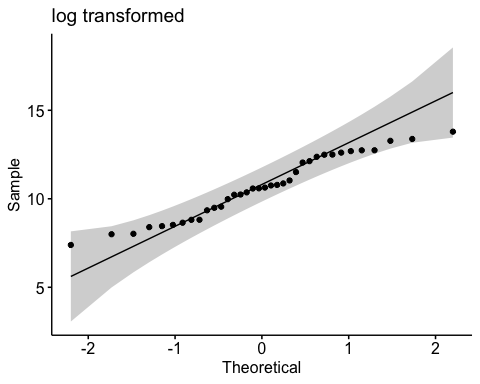
shapiro.test(ae\_noday0$adh) # nope

Shapiro-Wilk normality test  
  
data: ae\_noday0$adh  
W = 0.69154, p-value = 2.089e-07

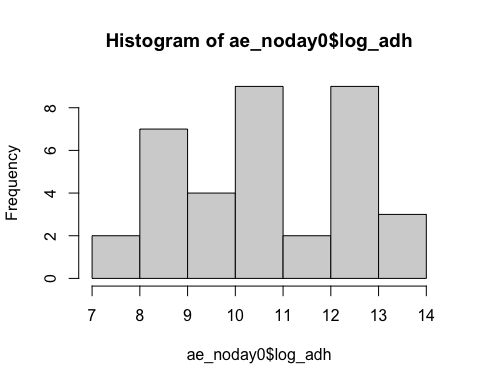
hist(ae\_noday0$adh) # note



# transformed  
ggqqplot(ae\_noday0, "log\_adh", title = "log transformed") # gorgeous



hist(ae\_noday0$log\_adh) # better



shapiro.test(ae\_noday0$log\_adh) # p-value = 0.1699 def normal

Shapiro-Wilk normality test  
  
data: ae\_noday0$log\_adh  
W = 0.95671, p-value = 0.1699

## 3b Stats

Keep interaction, skedasticity isn’t “too bad”, and will show underlying data in plot to be super transparent

y var: amount of Bd in adherent

x vars: day & medium with the AE biofilm and Bd (MQ, TB, PW)

Bd ~ day\*medium

Question: Does the amount of Bd in the aquatic environmental biofilm differ across the media tested and across the days, and do they interact with each other?

Model: 2-way ANOVA

**Summary of results**

Bd qty Day 1 > Day 3 > Day 5 > Day 7 (all p <0.005) in other words, Bd significantly reduced each day

MQ+AEbiofilm > PW+AEBiofilm > 1%TB+AEbiofilm (all p <0.005) In other words, TB plus biofilm has most Bd inhibition power, followed by pond water, then by milliQ with the least inhibition power

null

null <- lm(log\_adh ~ 1,  
 data = ae\_noday0)  
AIC(null) #146.5865

[1] 146.5865

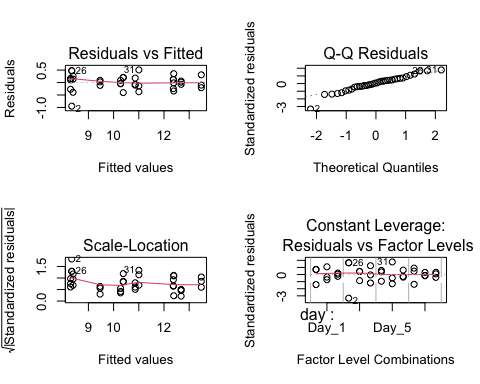
## Bd ~ day\*medium

Note: Not perfectly homoskedstic but “good enough”

# quick check: we want day as a FACTOR  
ae\_noday0 <- ae\_noday0 %>%   
# column for medium  
 mutate(medium = sample\_ID)  
  
# set day 1 as reference  
ae\_noday0$day <- factor(ae\_noday0$day, levels = c("Day\_1", "Day\_3", "Day\_5", "Day\_7"))  
ae\_noday0$day <- relevel(ae\_noday0$day, ref = "Day\_1")  
  
str(ae\_noday0$day)

Factor w/ 4 levels "Day\_1","Day\_3",..: 1 2 3 4 1 2 3 4 1 2 ...

# set MQ as reference  
ae\_noday0$sample\_ID <- factor(ae\_noday0$sample\_ID)  
ae\_noday0$sample\_ID <- relevel(ae\_noday0$sample\_ID, ref = "MQ+AEbiofilm")  
  
# build model  
aov\_3b <- aov(log\_adh ~ day\*medium,  
 data = ae\_noday0)  
  
# diagnostic plot  
par(mfrow = c(2,2))  
plot(aov\_3b) # kinda not homoskedastic



# look at results  
summary(aov\_3b)

Df Sum Sq Mean Sq F value Pr(>F)   
day 3 40.24 13.413 110.10 3.64e-14 \*\*\*  
medium 2 60.07 30.036 246.55 < 2e-16 \*\*\*  
day:medium 6 7.42 1.236 10.14 1.30e-05 \*\*\*  
Residuals 24 2.92 0.122   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(aov\_3b) # all significant and interaction sig

Analysis of Variance Table  
  
Response: log\_adh  
 Df Sum Sq Mean Sq F value Pr(>F)   
day 3 40.239 13.4129 110.100 3.642e-14 \*\*\*  
medium 2 60.073 30.0364 246.553 < 2.2e-16 \*\*\*  
day:medium 6 7.415 1.2359 10.145 1.301e-05 \*\*\*  
Residuals 24 2.924 0.1218   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

AIC(aov\_3b) # 37.78083 better than null

[1] 37.78083

### Posthoc

# posthoc  
#TukeyHSD(aov\_3b)  
  
# posthoc using emmeans  
# use emmeans package to get the t value  
# Perform pairwise comparisons for 'day'  
em\_day <- emmeans(aov\_3b, ~ day)

NOTE: Results may be misleading due to involvement in interactions

tukey\_day <- pairs(em\_day, adjust = "tukey")  
  
# Perform pairwise comparisons for 'medium'  
em\_medium <- emmeans(aov\_3b, ~ medium)

NOTE: Results may be misleading due to involvement in interactions

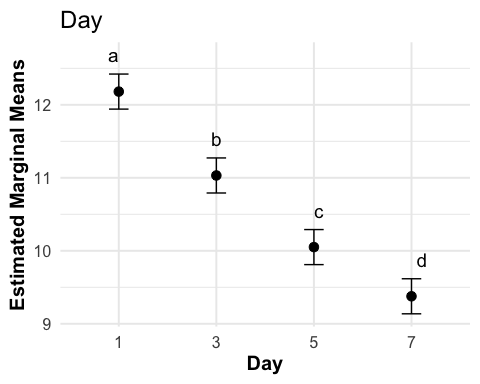
tukey\_medium <- pairs(em\_medium, adjust = "tukey")  
  
# Perform pairwise comparisons for 'day \* medium' (interaction)  
em\_interaction <- emmeans(aov\_3b, ~ day \* medium)  
tukey\_interaction <- pairs(em\_interaction, adjust = "tukey")

#### cld

# Compute CLD letters for 'day'  
cld\_day <- emmeans(aov\_3b, pairwise ~ day, adjust = "tukey") %>%  
 cld(Letters = letters, reverse = TRUE)

NOTE: Results may be misleading due to involvement in interactions

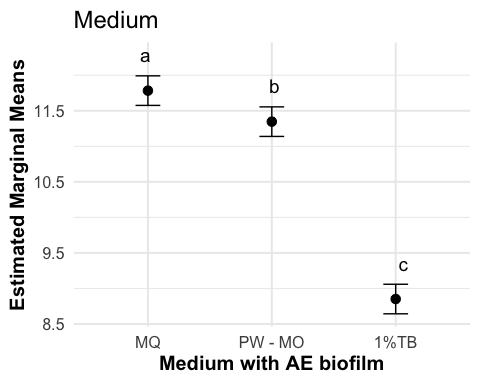
cld\_day\_3b <- ggplot(cld\_day, aes(x = day, y = emmean)) +  
 geom\_point(size = 3) + # Plot the estimated means  
 geom\_errorbar(aes(ymin = lower.CL, ymax = upper.CL), width = 0.2) + # Error bars  
 geom\_text(aes(label = .group), nudge\_y = 0.5, size = 5, color = "black") + # Add CLD letters  
 xlab("Day") +  
 ylab("Estimated Marginal Means") +  
 ggtitle("Day") +  
 scale\_x\_discrete(labels= c("Day\_1" = "1", "Day\_3" = "3", "Day\_5" = "5", "Day\_7" = "7")) +  
 theme\_minimal(base\_size = 15) + # Use minimal theme for clean look  
 theme(axis.text.x = element\_text(size = 12), # Adjust text size for better readability  
 axis.title.x = element\_text(face = "bold"),  
 axis.title.y = element\_text(face = "bold"))  
cld\_day\_3b



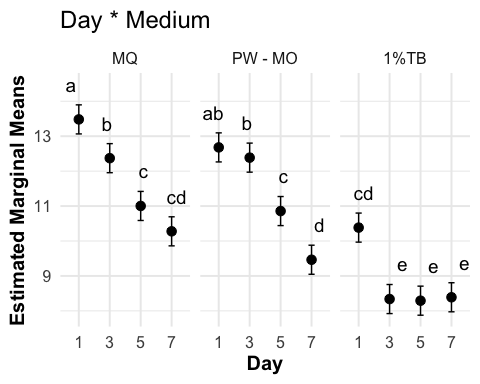
# Compute CLD letters for 'medium'  
cld\_medium <- emmeans(aov\_3b, pairwise ~ medium, adjust = "tukey") %>%  
 cld(Letters = letters, reverse = TRUE)

NOTE: Results may be misleading due to involvement in interactions

cld\_medium\_3b <- ggplot(cld\_medium, aes(x = factor(medium, levels = c("MQ+AEbiofilm", "PW+AEBiofilm", "1%TB+AEbiofilm")), y = emmean)) +  
 geom\_point(size = 3) + # Plot the estimated means  
 geom\_errorbar(aes(ymin = lower.CL, ymax = upper.CL), width = 0.2) + # Error bars  
 geom\_text(aes(label = .group), nudge\_y = 0.5, size = 5, color = "black") + # Add CLD letters  
 xlab("Medium with AE biofilm") +  
 ylab("Estimated Marginal Means") +  
 ggtitle("Medium") +  
 scale\_x\_discrete(labels= c("1%TB+AEbiofilm" = "1%TB", "PW+AEBiofilm" = "PW - MO", "MQ+AEbiofilm" = "MQ")) +  
 theme\_minimal(base\_size = 15) + # Use minimal theme for clean look  
 theme(axis.text.x = element\_text(size = 12), # Adjust text size for better readability  
 axis.title.x = element\_text(face = "bold"),  
 axis.title.y = element\_text(face = "bold"))  
cld\_medium\_3b



# Compute CLD letters for 'day \* medium'  
cld\_day\_medium <- emmeans(aov\_3b, pairwise ~ day \* medium, adjust = "tukey") %>%  
 cld(Letters = letters, reverse = TRUE)  
# Create the plot with custom labels for 'TB'  
cld\_day\_medium\_3b <- ggplot(cld\_day\_medium, aes(x = factor(day), y = emmean)) +  
 geom\_point(size = 3) + # Plot the estimated means  
 geom\_errorbar(aes(ymin = lower.CL, ymax = upper.CL), width = 0.2) + # Error bars  
 geom\_text(aes(label = .group), nudge\_y = 1, size = 5, color = "black") + # Add CLD letters  
facet\_wrap(~ factor(medium, levels = c("MQ+AEbiofilm", "PW+AEBiofilm", "1%TB+AEbiofilm")),   
 labeller = as\_labeller(c("1%TB+AEbiofilm" = "1%TB",   
 "PW+AEBiofilm" = "PW - MO",   
 "MQ+AEbiofilm" = "MQ"))) + # Custom facet labels and order  
 xlab("Day") +  
 ylab("Estimated Marginal Means") +  
 ggtitle("Day \* Medium") +  
 scale\_x\_discrete(labels= c("Day\_1" = "1", "Day\_3" = "3", "Day\_5" = "5", "Day\_7" = "7")) +  
 theme\_minimal(base\_size = 15) + # Use minimal theme for clean look  
 theme(axis.text.x = element\_text(size = 12), # Rotate x-axis labels 45 degrees  
 axis.title.x = element\_text(face = "bold"),  
 axis.title.y = element\_text(face = "bold"))  
cld\_day\_medium\_3b



Results write up

A two-way ANOVA revealed that there was a statistically significant difference in Bd load across days (F(3, 24) = 110.100, p = p <0.0001), across the media (F(2, 24) = 246.55, p <0.0001), and the interaction between the effects of day and medium were also significant (F(6, 24) = 10.145, p = p <0.0001). Bd was significantly lower with each day (Tukey test, p <0.005 for all) and TB plus biofilm has most Bd inhibition power, followed by pond water with no microbes, then by milliQ with the least inhibition power (Tukey test, p <0.05 for all).

# \*PUBLICATION FIGURES

3A: AE Biofilm

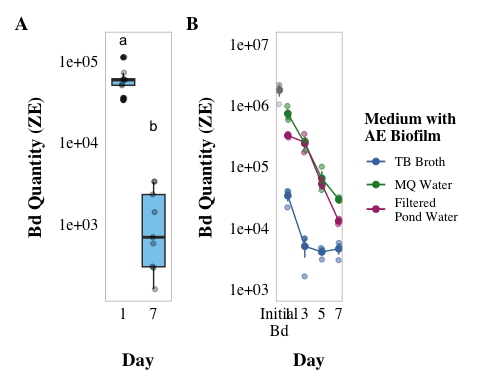
fig\_3a\_boxplot <- eb\_ae\_bf\_only %>%  
ggplot(aes(y= bd\_qty, x = day)) +   
 geom\_boxplot(fill = no\_microbes\_.22\_color) +  
 geom\_jitter(alpha = 0.3, width = 0.05) +  
 myCustomTheme() +  
 scale\_y\_log10() +  
 theme(legend.position = "none",  
 strip.text = element\_text(face="bold"),  
 axis.title = element\_text(face = "bold")) +   
 xlab("Day") +  
 ylab("Bd Quantity (ZE)") +  
 scale\_x\_discrete (labels= c("Day\_1" = "1", "Day\_7" = "7"))  
  
# Let's add significance letters  
significance\_data <- tibble(  
 day = factor(c("Day\_1", "Day\_7"), levels = c("Day\_1", "Day\_7")),  
 y\_position = c(1.6e+05, 1.4e+04), # Adjust this depending on your plot's scale  
 label = c("a", "b"))  
  
fig\_3a <- fig\_3a\_boxplot +   
geom\_text(data = significance\_data, aes(x = day, y = y\_position, label = label),  
 position = position\_dodge(width = 0.75), vjust = 0)

3B: AE Biofilm

# add column for microbes or no  
fig\_3b <- ae\_summary %>%  
 # reorder to match Renwei's plot  
 mutate(sample\_ID = factor(sample\_ID,  
 levels = c("1%TB+AEbiofilm", "MQ+AEbiofilm",  
 "PW+AEBiofilm", "Added Bd" ))) %>%  
  
 ggplot(aes(x = day\_numeric,  
 y = mean,  
 color = sample\_ID)) +  
 geom\_point(size = 2) +  
 geom\_errorbar(aes(ymin = mean - se, # plot the standard error  
 ymax = mean + se),  
 width = 0.1) +  
 geom\_line() +   
   
 # Adding the raw data as a layer with jitter  
 geom\_point(data = ae\_noday0,   
 aes(x = day\_numeric,   
 y = adh,   
 color = sample\_ID), # Raw data points  
 position = position\_jitter(width = 0.1, seed = 1),  
 alpha = 0.3) +  
 # controls raw data  
 geom\_point(data = ae\_control\_data,   
 aes(x = day\_numeric,   
 y = adh,   
 color = "#BBBBBB"), # Raw data points  
 position = position\_jitter(width = 0.1, seed = 1),  
 alpha = 0.3) +  
   
 # Adding the raw data as a layer with jitter  
 geom\_point(data = ae\_noday0,  
 aes(x = day\_numeric,  
 y = adh,  
 color = sample\_ID), # Raw data points  
 position = position\_jitter(width = 0.1, seed = 1),  
 alpha = 0.3) +  
   
 scale\_y\_log10(limits = c(1e3, 1e7),  
 breaks = c(1e3, 1e4, 1e5, 1e6, 1e7)) +  
 # vibes  
 labs(x = "Day",  
 y = "Bd Quantity (ZE)",  
 color = "Medium with\nAE Biofilm")+ # Title for color legend  
   
 scale\_color\_manual(values = c("1%TB+AEbiofilm"= "#4477AA",  
 "MQ+AEbiofilm" = "#228833",  
 #"Added Bd" = "darkgrey",  
 "PW+AEBiofilm" = "#AA3377"),  
 labels = c("1%TB+AEbiofilm" = "TB Broth",  
 "MQ+AEbiofilm" = "MQ Water",  
 "PW+AEBiofilm" = "Filtered\nPond Water",  
 "Added Bd" = "Initial Bd")) + # Custom labels  
  
 myCustomTheme()+  
 scale\_x\_continuous(breaks = c(0, 1, 3, 5, 7),  
 labels = c("Initial\nBd", "1", "3", "5", "7"))

Fig 3 combined

# Combine fig\_2a and fig\_2b side by side  
fig3 <- fig\_3a + fig\_3b +   
 plot\_layout(widths = c(1, 1)) + # Ensure equal widths for both plots  
 plot\_annotation(tag\_levels = 'A') # Adds "A" and "B" to the upper corners  
  
fig3



#ggsave("3a\_3b.png", plot = fig3, width = 14, height = 5, dpi = 1000)

# \*SI figures and tables

## 3B stats 2way anova

### anova table

# anova table  
anova\_output <- tidy(aov\_3b)  
  
aov\_3b\_tbl <- anova\_output %>%  
 dplyr::select(term, df, sumsq, meansq, statistic, p.value) %>%  
 gt() %>%  
 tab\_header(  
 title = "ANOVA Table"  
 ) %>%  
 fmt\_number(  
 columns = c(sumsq, meansq, statistic),  
 decimals = 2  
 ) %>%  
 cols\_label(  
 term = "Term",  
 df = "Df",  
 sumsq = "Sum Sq",  
 meansq = "Mean Sq",  
 statistic = "F value",  
 p.value = "P-value"  
 ) %>% # scientific number format for values <0.001 in p values  
 fmt\_scientific(  
 columns = c(p.value),  
 decimals = 1,  
 rows = p.value < 0.001  
 ) %>%  
 # 3 decimals for p values >=0.001  
 fmt\_number(  
 columns = c(p.value),  
 decimals = 3,  
 rows = p.value >= 0.001  
 )  
aov\_3b\_tbl

Table 1: ANOVA Table

| Term | Df | Sum Sq | Mean Sq | F value | P-value |
| --- | --- | --- | --- | --- | --- |
| day | 3 | 40.24 | 13.41 | 110.10 | 3.6 × 10^-14 |
| medium | 2 | 60.07 | 30.04 | 246.55 | 1.0 × 10^-16 |
| day:medium | 6 | 7.42 | 1.24 | 10.14 | 1.3 × 10^-5 |
| Residuals | 24 | 2.92 | 0.12 | NA | NA |

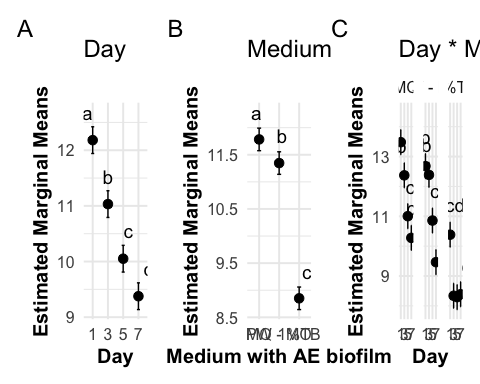
# prettier, simplified  
anova\_output <- tidy(aov\_3b)  
  
# Modify term to include degrees of freedom in \*italics\*  
anova\_output <- anova\_output %>%  
 mutate(term = ifelse(grepl("day:medium", term), "day x medium", term)) %>%   
 mutate (term = paste0(term, " (\*df = ", df, ", ", anova\_output[df == max(df), "df"], "\*)")) %>%   
 filter(term != "Residuals (\*df = 24, 24\*)")  
  
# Create the gt table with selected columns  
aov\_3b\_tbl\_b <- anova\_output %>%  
 dplyr::select(term, statistic, p.value) %>%  
 gt() %>%  
 tab\_header(  
 title = "ANOVA Table"  
 ) %>%  
 fmt\_markdown(  
 columns = c(term)  
 ) %>%  
 fmt\_number(  
 columns = c(statistic),  
 decimals = 2  
 ) %>%  
 cols\_label(  
 term = "",  
 statistic = "F value",  
 p.value = "P-value"  
 ) %>%  
 fmt\_scientific(  
 columns = c(p.value),  
 decimals = 1,  
 rows = p.value < 0.001  
 ) %>%  
 fmt\_number(  
 columns = c(p.value),  
 decimals = 3,  
 rows = p.value >= 0.001  
 )  
  
aov\_3b\_tbl\_b

Table 1: ANOVA Table

|  | F value | P-value |
| --- | --- | --- |
| day (*df = 3, 24*) | 110.10 | 3.6 × 10^-14 |
| medium (*df = 2, 24*) | 246.55 | 1.0 × 10^-16 |
| day x medium (*df = 6, 24*) | 10.14 | 1.3 × 10^-5 |

## cld 3b

pairwise\_cld\_3b <- cld\_day\_3b + cld\_medium\_3b + cld\_day\_medium\_3b +  
 plot\_annotation(tag\_levels = 'A')  
  
# Display the combined plot  
pairwise\_cld\_3b



#ggsave("3b\_pairwise\_cld.png", plot = pairwise\_cld\_3b , width = 14, height = 5, dpi = 1000)

### posthoc table

# post hoc table  
# Convert Tukey emmeans results to data frames  
tukey\_day\_df <- as.data.frame(tukey\_day)  
tukey\_medium\_df <- as.data.frame(tukey\_medium)  
tukey\_interaction\_df <- as.data.frame(tukey\_interaction)  
  
# Add labels to indicate which factor the comparison refers to  
tukey\_day\_df <- tukey\_day\_df %>% mutate(factor = "Day")  
tukey\_medium\_df <- tukey\_medium\_df %>% mutate(factor = "Medium")  
tukey\_interaction\_df <- tukey\_interaction\_df %>% mutate(factor = "Interaction")  
  
all\_tukey\_df <- bind\_rows(tukey\_day\_df, tukey\_medium\_df, tukey\_interaction\_df)  
ph3b\_table <- all\_tukey\_df %>%  
 dplyr::select(factor, contrast, estimate, SE, df, t.ratio, p.value) %>%  
 gt() %>%  
 # change column names  
 cols\_label(  
 factor = "Comparison",  
 contrast = "Group Comparison",  
 estimate = "Estimate",  
 SE = "Standard Error",  
 df = "Degrees of Freedom",  
 t.ratio = "t-Ratio",  
 p.value = "p-value"  
 ) %>%  
 # update header for table  
 tab\_header(  
 title = "Emmeans Post-hoc Test Results"  
 ) %>%  
 # 3 decimal places  
 fmt\_number(  
 columns = c(estimate, SE, t.ratio),  
 decimals = 3  
 ) %>%  
 # scientific number format for values <0.001 in p values  
 fmt\_scientific(  
 columns = c(p.value),  
 decimals = 1,  
 rows = p.value < 0.001  
 ) %>%  
 # 3 decimals for p values >=0.001  
 fmt\_number(  
 columns = c(p.value),  
 decimals = 3,  
 rows = p.value >= 0.001  
 ) %>%  
 #make the headers bold  
 tab\_style(  
 style = list(  
 cell\_text(weight = "bold")  
 ),  
 locations = cells\_column\_labels(everything()))  
ph3b\_table

Table 1: Emmeans Post-hoc Test Results

| Comparison | Group Comparison | Estimate | Standard Error | Degrees of Freedom | t-Ratio | p-value |
| --- | --- | --- | --- | --- | --- | --- |
| Day | Day\_1 - Day\_3 | 1.149 | 0.165 | 24 | 6.985 | 1.8 × 10^-6 |
| Day | Day\_1 - Day\_5 | 2.131 | 0.165 | 24 | 12.949 | 1.5 × 10^-11 |
| Day | Day\_1 - Day\_7 | 2.805 | 0.165 | 24 | 17.046 | 6.2 × 10^-14 |
| Day | Day\_3 - Day\_5 | 0.981 | 0.165 | 24 | 5.964 | 2.1 × 10^-5 |
| Day | Day\_3 - Day\_7 | 1.655 | 0.165 | 24 | 10.060 | 2.5 × 10^-9 |
| Day | Day\_5 - Day\_7 | 0.674 | 0.165 | 24 | 4.097 | 0.002 |
| Medium | (1%TB+AEbiofilm) - (MQ+AEbiofilm) | -2.932 | 0.142 | 24 | -20.579 | 2.1 × 10^-14 |
| Medium | (1%TB+AEbiofilm) - (PW+AEBiofilm) | -2.496 | 0.142 | 24 | -17.515 | 3.2 × 10^-14 |
| Medium | (MQ+AEbiofilm) - (PW+AEBiofilm) | 0.437 | 0.142 | 24 | 3.064 | 0.014 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_3 1%TB+AEbiofilm) | 2.045 | 0.285 | 24 | 7.176 | 1.1 × 10^-5 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_5 1%TB+AEbiofilm) | 2.092 | 0.285 | 24 | 7.340 | 7.8 × 10^-6 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_7 1%TB+AEbiofilm) | 1.994 | 0.285 | 24 | 6.997 | 1.7 × 10^-5 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_1 MQ+AEbiofilm) | -3.098 | 0.285 | 24 | -10.869 | 5.6 × 10^-9 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_3 MQ+AEbiofilm) | -1.988 | 0.285 | 24 | -6.974 | 1.8 × 10^-5 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_5 MQ+AEbiofilm) | -0.620 | 0.285 | 24 | -2.175 | 0.580 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_7 MQ+AEbiofilm) | 0.107 | 0.285 | 24 | 0.374 | 1.000 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_1 PW+AEBiofilm) | -2.296 | 0.285 | 24 | -8.056 | 1.6 × 10^-6 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_3 PW+AEBiofilm) | -2.003 | 0.285 | 24 | -7.028 | 1.6 × 10^-5 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_5 PW+AEBiofilm) | -0.474 | 0.285 | 24 | -1.662 | 0.868 |
| Interaction | (Day\_1 1%TB+AEbiofilm) - (Day\_7 PW+AEBiofilm) | 0.920 | 0.285 | 24 | 3.228 | 0.109 |
| Interaction | (Day\_3 1%TB+AEbiofilm) - (Day\_5 1%TB+AEbiofilm) | 0.047 | 0.285 | 24 | 0.164 | 1.000 |
| Interaction | (Day\_3 1%TB+AEbiofilm) - (Day\_7 1%TB+AEbiofilm) | -0.051 | 0.285 | 24 | -0.178 | 1.000 |
| Interaction | (Day\_3 1%TB+AEbiofilm) - (Day\_1 MQ+AEbiofilm) | -5.143 | 0.285 | 24 | -18.045 | 1.4 × 10^-13 |
| Interaction | (Day\_3 1%TB+AEbiofilm) - (Day\_3 MQ+AEbiofilm) | -4.033 | 0.285 | 24 | -14.150 | 2.3 × 10^-11 |
| Interaction | (Day\_3 1%TB+AEbiofilm) - (Day\_5 MQ+AEbiofilm) | -2.665 | 0.285 | 24 | -9.351 | 1.0 × 10^-7 |
| Interaction | (Day\_3 1%TB+AEbiofilm) - (Day\_7 MQ+AEbiofilm) | -1.938 | 0.285 | 24 | -6.802 | 2.7 × 10^-5 |
| Interaction | (Day\_3 1%TB+AEbiofilm) - (Day\_1 PW+AEBiofilm) | -4.341 | 0.285 | 24 | -15.232 | 4.7 × 10^-12 |
| Interaction | (Day\_3 1%TB+AEbiofilm) - (Day\_3 PW+AEBiofilm) | -4.048 | 0.285 | 24 | -14.203 | 2.1 × 10^-11 |
| Interaction | (Day\_3 1%TB+AEbiofilm) - (Day\_5 PW+AEBiofilm) | -2.518 | 0.285 | 24 | -8.837 | 3.0 × 10^-7 |
| Interaction | (Day\_3 1%TB+AEbiofilm) - (Day\_7 PW+AEBiofilm) | -1.125 | 0.285 | 24 | -3.948 | 0.023 |
| Interaction | (Day\_5 1%TB+AEbiofilm) - (Day\_7 1%TB+AEbiofilm) | -0.098 | 0.285 | 24 | -0.342 | 1.000 |
| Interaction | (Day\_5 1%TB+AEbiofilm) - (Day\_1 MQ+AEbiofilm) | -5.189 | 0.285 | 24 | -18.209 | 1.2 × 10^-13 |
| Interaction | (Day\_5 1%TB+AEbiofilm) - (Day\_3 MQ+AEbiofilm) | -4.079 | 0.285 | 24 | -14.314 | 1.8 × 10^-11 |
| Interaction | (Day\_5 1%TB+AEbiofilm) - (Day\_5 MQ+AEbiofilm) | -2.712 | 0.285 | 24 | -9.515 | 7.5 × 10^-8 |
| Interaction | (Day\_5 1%TB+AEbiofilm) - (Day\_7 MQ+AEbiofilm) | -1.985 | 0.285 | 24 | -6.966 | 1.8 × 10^-5 |
| Interaction | (Day\_5 1%TB+AEbiofilm) - (Day\_1 PW+AEBiofilm) | -4.388 | 0.285 | 24 | -15.396 | 3.7 × 10^-12 |
| Interaction | (Day\_5 1%TB+AEbiofilm) - (Day\_3 PW+AEBiofilm) | -4.095 | 0.285 | 24 | -14.368 | 1.6 × 10^-11 |
| Interaction | (Day\_5 1%TB+AEbiofilm) - (Day\_5 PW+AEBiofilm) | -2.565 | 0.285 | 24 | -9.001 | 2.1 × 10^-7 |
| Interaction | (Day\_5 1%TB+AEbiofilm) - (Day\_7 PW+AEBiofilm) | -1.172 | 0.285 | 24 | -4.112 | 0.016 |
| Interaction | (Day\_7 1%TB+AEbiofilm) - (Day\_1 MQ+AEbiofilm) | -5.092 | 0.285 | 24 | -17.867 | 1.7 × 10^-13 |
| Interaction | (Day\_7 1%TB+AEbiofilm) - (Day\_3 MQ+AEbiofilm) | -3.982 | 0.285 | 24 | -13.972 | 3.0 × 10^-11 |
| Interaction | (Day\_7 1%TB+AEbiofilm) - (Day\_5 MQ+AEbiofilm) | -2.614 | 0.285 | 24 | -9.172 | 1.5 × 10^-7 |
| Interaction | (Day\_7 1%TB+AEbiofilm) - (Day\_7 MQ+AEbiofilm) | -1.888 | 0.285 | 24 | -6.624 | 4.0 × 10^-5 |
| Interaction | (Day\_7 1%TB+AEbiofilm) - (Day\_1 PW+AEBiofilm) | -4.290 | 0.285 | 24 | -15.053 | 6.0 × 10^-12 |
| Interaction | (Day\_7 1%TB+AEbiofilm) - (Day\_3 PW+AEBiofilm) | -3.997 | 0.285 | 24 | -14.025 | 2.8 × 10^-11 |
| Interaction | (Day\_7 1%TB+AEbiofilm) - (Day\_5 PW+AEBiofilm) | -2.468 | 0.285 | 24 | -8.659 | 4.3 × 10^-7 |
| Interaction | (Day\_7 1%TB+AEbiofilm) - (Day\_7 PW+AEBiofilm) | -1.074 | 0.285 | 24 | -3.770 | 0.035 |
| Interaction | (Day\_1 MQ+AEbiofilm) - (Day\_3 MQ+AEbiofilm) | 1.110 | 0.285 | 24 | 3.895 | 0.026 |
| Interaction | (Day\_1 MQ+AEbiofilm) - (Day\_5 MQ+AEbiofilm) | 2.478 | 0.285 | 24 | 8.694 | 4.0 × 10^-7 |
| Interaction | (Day\_1 MQ+AEbiofilm) - (Day\_7 MQ+AEbiofilm) | 3.204 | 0.285 | 24 | 11.243 | 2.8 × 10^-9 |
| Interaction | (Day\_1 MQ+AEbiofilm) - (Day\_1 PW+AEBiofilm) | 0.802 | 0.285 | 24 | 2.813 | 0.235 |
| Interaction | (Day\_1 MQ+AEbiofilm) - (Day\_3 PW+AEBiofilm) | 1.095 | 0.285 | 24 | 3.842 | 0.030 |
| Interaction | (Day\_1 MQ+AEbiofilm) - (Day\_5 PW+AEBiofilm) | 2.624 | 0.285 | 24 | 9.208 | 1.4 × 10^-7 |
| Interaction | (Day\_1 MQ+AEbiofilm) - (Day\_7 PW+AEBiofilm) | 4.017 | 0.285 | 24 | 14.097 | 2.5 × 10^-11 |
| Interaction | (Day\_3 MQ+AEbiofilm) - (Day\_5 MQ+AEbiofilm) | 1.368 | 0.285 | 24 | 4.799 | 0.003 |
| Interaction | (Day\_3 MQ+AEbiofilm) - (Day\_7 MQ+AEbiofilm) | 2.094 | 0.285 | 24 | 7.348 | 7.6 × 10^-6 |
| Interaction | (Day\_3 MQ+AEbiofilm) - (Day\_1 PW+AEBiofilm) | -0.308 | 0.285 | 24 | -1.082 | 0.993 |
| Interaction | (Day\_3 MQ+AEbiofilm) - (Day\_3 PW+AEBiofilm) | -0.015 | 0.285 | 24 | -0.053 | 1.000 |
| Interaction | (Day\_3 MQ+AEbiofilm) - (Day\_5 PW+AEBiofilm) | 1.514 | 0.285 | 24 | 5.313 | 9.2 × 10^-4 |
| Interaction | (Day\_3 MQ+AEbiofilm) - (Day\_7 PW+AEBiofilm) | 2.907 | 0.285 | 24 | 10.202 | 1.9 × 10^-8 |
| Interaction | (Day\_5 MQ+AEbiofilm) - (Day\_7 MQ+AEbiofilm) | 0.726 | 0.285 | 24 | 2.549 | 0.359 |
| Interaction | (Day\_5 MQ+AEbiofilm) - (Day\_1 PW+AEBiofilm) | -1.676 | 0.285 | 24 | -5.881 | 2.3 × 10^-4 |
| Interaction | (Day\_5 MQ+AEbiofilm) - (Day\_3 PW+AEBiofilm) | -1.383 | 0.285 | 24 | -4.853 | 0.003 |
| Interaction | (Day\_5 MQ+AEbiofilm) - (Day\_5 PW+AEBiofilm) | 0.146 | 0.285 | 24 | 0.513 | 1.000 |
| Interaction | (Day\_5 MQ+AEbiofilm) - (Day\_7 PW+AEBiofilm) | 1.540 | 0.285 | 24 | 5.403 | 7.4 × 10^-4 |
| Interaction | (Day\_7 MQ+AEbiofilm) - (Day\_1 PW+AEBiofilm) | -2.402 | 0.285 | 24 | -8.430 | 7.0 × 10^-7 |
| Interaction | (Day\_7 MQ+AEbiofilm) - (Day\_3 PW+AEBiofilm) | -2.109 | 0.285 | 24 | -7.401 | 6.7 × 10^-6 |
| Interaction | (Day\_7 MQ+AEbiofilm) - (Day\_5 PW+AEBiofilm) | -0.580 | 0.285 | 24 | -2.035 | 0.668 |
| Interaction | (Day\_7 MQ+AEbiofilm) - (Day\_7 PW+AEBiofilm) | 0.813 | 0.285 | 24 | 2.854 | 0.219 |
| Interaction | (Day\_1 PW+AEBiofilm) - (Day\_3 PW+AEBiofilm) | 0.293 | 0.285 | 24 | 1.028 | 0.995 |
| Interaction | (Day\_1 PW+AEBiofilm) - (Day\_5 PW+AEBiofilm) | 1.822 | 0.285 | 24 | 6.395 | 6.9 × 10^-5 |
| Interaction | (Day\_1 PW+AEBiofilm) - (Day\_7 PW+AEBiofilm) | 3.216 | 0.285 | 24 | 11.284 | 2.6 × 10^-9 |
| Interaction | (Day\_3 PW+AEBiofilm) - (Day\_5 PW+AEBiofilm) | 1.529 | 0.285 | 24 | 5.366 | 8.1 × 10^-4 |
| Interaction | (Day\_3 PW+AEBiofilm) - (Day\_7 PW+AEBiofilm) | 2.923 | 0.285 | 24 | 10.255 | 1.8 × 10^-8 |
| Interaction | (Day\_5 PW+AEBiofilm) - (Day\_7 PW+AEBiofilm) | 1.393 | 0.285 | 24 | 4.889 | 0.003 |