Combined analyses

Caitlin Nordheim-Maestas

# Compare Bd-inhibitory potency of the AE biofilms to that of the AE microorganisms

# 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   
library(multcomp) # stats  
library(multcompView) # view cld  
library(emmeans) # for pairwise comparisons, especially on mixed effects models and glms  
library(ggpubr) # for making ggqq plot  
library(patchwork) # for combining figures  
  
# load wrangled data from expt2-4 quarto docs in this repo  
eb\_pw\_total\_diff <- read.csv(here("data", "eb\_pw\_total\_diff.csv"))  
pw\_noday0 <- read.csv(here("data", "pw\_noday0.csv"))  
eb\_ae\_bf\_only <- read.csv(here("data", "eb\_ae\_bf\_only.csv"))  
ae\_noday0 <- read.csv(here("data", "ae\_noday0.csv"))  
  
# Colors: these are from Paul Tol's colorblind friendly palette  
with\_microbes\_40\_color <- "#999933"  
no\_microbes\_.22\_color <- "#88ccee"

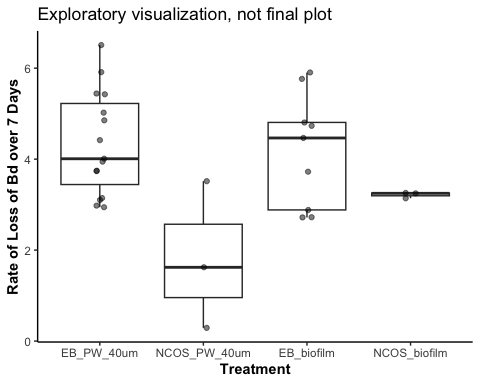
# Data wrangling for all datasets

# data wrangling for comparisons across experiments  
  
## EBpond water  
pw\_WITH\_microbes <- eb\_pw\_total\_diff %>%  
 filter(filter != "0.22um\_filter") %>%   
 mutate(treatment = "EB\_PW\_40um") %>%   
 rename(sample\_ID = site) %>%   
 dplyr::select(sample\_ID, treatment, rate\_loss)  
  
## NCOS PW   
ncos\_pw\_WITH\_microbes <- pw\_noday0 %>%   
 filter(sample\_ID == "PW+microorganism") %>%   
 filter(day != "Day\_3" & day != "Day\_5") %>%  
 dplyr::select(day, sample\_ID, replicate, adh\_plus\_sup) %>%   
 pivot\_wider(names\_from = day, values\_from = adh\_plus\_sup) %>%  
# calculate the rate loss by taking the log of each before subtracting  
 mutate(rate\_loss = log(Day\_1) - log(Day\_7)) %>%   
 mutate(treatment = "NCOS\_PW\_40um") %>%   
 dplyr::select(sample\_ID, treatment, rate\_loss)  
  
## Field bf  
fieldbf\_no\_microbes <- eb\_ae\_bf\_only %>% # no need to filter, the only treatment is no microbe water  
 filter(day != "Day\_0") %>%   
 dplyr::select(day, site, bd\_qty) %>%   
 pivot\_wider(names\_from = day, values\_from = bd\_qty) %>%  
# calculate the rate loss by taking the log of each before subtracting  
 mutate(rate\_loss = log(Day\_1) - log(Day\_7)) %>%   
 rename(sample\_ID = site) %>%   
 mutate(treatment = "EB\_biofilm") %>%   
 dplyr::select(sample\_ID, treatment, rate\_loss)  
  
## NCOS version 3B ae bf with no microbes  
ncos\_ae\_no\_microbes <- ae\_noday0 %>%   
 filter(sample\_ID == "PW+AEBiofilm") %>% # this is a "pond water" treatment of filtered pond water  
 filter(day != "Day\_3" & day != "Day\_5") %>%  
 dplyr::select(day, sample\_ID, replicate, adh) %>%   
 pivot\_wider(names\_from = day, values\_from = adh) %>%  
# calculate the rate loss by taking the log of each before subtracting  
 mutate(treatment = "NCOS\_biofilm") %>%   
 mutate(rate\_loss = log(Day\_1) - log(Day\_7)) %>%   
 dplyr::select(sample\_ID, treatment, rate\_loss)  
   
all\_parts <- bind\_rows(pw\_WITH\_microbes, fieldbf\_no\_microbes, ncos\_pw\_WITH\_microbes, ncos\_ae\_no\_microbes) %>%   
 mutate(location = case\_when(  
 grepl("^EB", treatment) ~ "EB",  
 grepl("^NCOS", treatment) ~ "NCOS"),  
 medium = case\_when(  
 grepl("PW", treatment) ~ "PW",  
 grepl("biofilm", treatment) ~ "biofilm"))  
  
# set PW 40 um as the intercept, comparing back to that one  
all\_parts$treatment <- factor(all\_parts$treatment,  
 levels = c("EB\_PW\_40um", "NCOS\_PW\_40um", "EB\_biofilm", "NCOS\_biofilm"))

# Stats and assumption testing

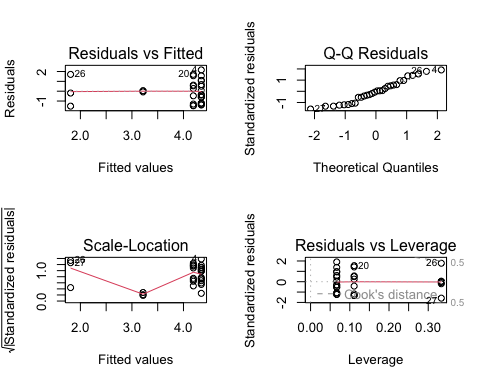
## Exploratory viz:

all\_parts %>%   
 ggplot(aes(y= rate\_loss, x = treatment)) +   
 geom\_boxplot() +  
 geom\_jitter(alpha = 0.5, width = 0.05) +  
 theme\_classic() +  
 theme(legend.position = "bottom",  
 strip.text = element\_text(face="bold"),  
 axis.title = element\_text(face = "bold")) +  
 xlab("Treatment") +  
 ylab("Rate of Loss of Bd over 7 Days") +   
 ggtitle("Exploratory visualization, not final plot")



## Assumptions testing

par(mfrow = c(2,2))  
aov\_5b <- aov(rate\_loss~treatment, data=all\_parts)  
plot(aov\_5b) # diagnostic plot



## interpret

# with microbe depleted  
summary(aov\_5b) # sig effect of treatment

Df Sum Sq Mean Sq F value Pr(>F)   
treatment 3 18.23 6.077 4.424 0.0122 \*  
Residuals 26 35.72 1.374   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## posthoc

TukeyHSD(aov\_5b)

Tukey multiple comparisons of means  
 95% family-wise confidence level  
  
Fit: aov(formula = rate\_loss ~ treatment, data = all\_parts)  
  
$treatment  
 diff lwr upr p adj  
NCOS\_PW\_40um-EB\_PW\_40um -2.5364280 -4.5699791 -0.5028769 0.0104552  
EB\_biofilm-EB\_PW\_40um -0.1544017 -1.5101025 1.2012990 0.9891809  
NCOS\_biofilm-EB\_PW\_40um -1.1302884 -3.1638395 0.9032627 0.4377743  
EB\_biofilm-NCOS\_PW\_40um 2.3820263 0.2384752 4.5255774 0.0252643  
NCOS\_biofilm-NCOS\_PW\_40um 1.4061396 -1.2191635 4.0314428 0.4696326  
NCOS\_biofilm-EB\_biofilm -0.9758867 -3.1194377 1.1676644 0.6023819

# get compact letter display from the tukey using glht (general linear hypotheses test)  
post\_hoc\_rateloss <- glht(aov\_5b,   
 linfct = mcp(treatment = "Tukey")) # run Tukey  
summary(post\_hoc\_rateloss) # show the comparisons

Simultaneous Tests for General Linear Hypotheses  
  
Multiple Comparisons of Means: Tukey Contrasts  
  
  
Fit: aov(formula = rate\_loss ~ treatment, data = all\_parts)  
  
Linear Hypotheses:  
 Estimate Std. Error t value Pr(>|t|)   
NCOS\_PW\_40um - EB\_PW\_40um == 0 -2.5364 0.7413 -3.422 0.00953 \*\*  
EB\_biofilm - EB\_PW\_40um == 0 -0.1544 0.4942 -0.312 0.98865   
NCOS\_biofilm - EB\_PW\_40um == 0 -1.1303 0.7413 -1.525 0.42621   
EB\_biofilm - NCOS\_PW\_40um == 0 2.3820 0.7814 3.049 0.02364 \*   
NCOS\_biofilm - NCOS\_PW\_40um == 0 1.4061 0.9570 1.469 0.45798   
NCOS\_biofilm - EB\_biofilm == 0 -0.9759 0.7814 -1.249 0.59136   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
(Adjusted p values reported -- single-step method)

cld(post\_hoc\_rateloss) # get compact letter display

EB\_PW\_40um NCOS\_PW\_40um EB\_biofilm NCOS\_biofilm   
 "a" "b" "a" "ab"

# export cld results as a dataframe  
cld\_results\_rateloss <- cld(post\_hoc\_rateloss) # get compact letter display (CLD) groupings for treatment comparisons  
labels\_rl <- cld\_results\_rateloss$mcletters$Letters # extract grouping letters from cld object  
treatments\_rl <- names(labels\_rl) # get the treatment names corresponding to the letters  
labels\_rl <- as.character(labels\_rl) # set as character  
  
# df with significance data for the rate loss, will use in plotting  
sig\_data\_rl <- tibble(  
 treatment = treatments\_rl, # treatment names from the CLD results  
 y\_position = c(7,4,6.5,4), # y position (adjusted manually)  
 label = labels\_rl) # CLD group letters for each treatment

# \*PUBLICATION FIGURES

myCustomTheme <- function() {  
 theme\_light() +  
 theme(axis.text = element\_text(size = 7, family = "Helvetica", color = "black"),  
 axis.title.x = element\_text(margin = margin(t = 10), size = 7, face = "plain", family = "Helvetica", color = "black"), # Add space between x-axis label and axis  
 axis.title.y = element\_text(margin = margin(r = 10), size = 7, face = "plain", family = "Helvetica", color = "black"), # Add space between y-axis label and axis  
 title = element\_text(size = 7, face = "bold", family = "Helvetica"),  
 plot.caption = element\_text(size = 7, face = "italic", family = "Helvetica"),  
 legend.text = element\_text(size = 7, family = "Helvetica"), # Increase legend text size  
 panel.grid = element\_blank(), # Remove all grid lines (both major and minor)  
 # axis.line.x = element\_line(color = "grey"), # Keep the x-axis line  
 # axis.line.y = element\_line(color = "grey"), # Keep the y-axis line  
 axis.ticks = element\_line(color = "grey", size = 0.5), # Keep tick markers  
 axis.ticks.x = element\_line(color = "grey", size = 0.5), # ensure bottom axis ticks  
 axis.ticks.y = element\_line(color = "grey", size = 0.5), # <- ensure side axis ticks  
 strip.text = element\_text(size = 7, face = "bold", family = "Helvetica", color = "black"), # Set strip text style  
 strip.background = element\_rect(fill = "white", color = "grey", size = 0.5) # Set strip background to white, outline grey  
 )  
}

# sfeb\_fig <- sfeb %>%   
# ggplot(aes(y= rate\_loss, x = treatment, fill = treatment)) +   
# geom\_boxplot() +  
# geom\_jitter(alpha = 0.3, width = 0.05) +  
# myCustomTheme() +  
# scale\_fill\_manual(values = c("EB\_PW\_40um" = "white",  
# "EB\_biofilm" = "darkgrey")) +  
# theme(legend.position = "none",  
# axis.line.x = element\_line(color = "grey", size = 0.5), # Keep the x-axis line  
# axis.line.y = element\_line(color = "grey", size = 0.5),  
# panel.border = element\_blank()) +  
# xlab("Medium") +  
# ylab("Reduction of Bd over 6 days") +   
# scale\_x\_discrete (labels= c("EB\_PW\_40um" = "microorganisms\n(n = 15)",  
# "EB\_biofilm" = "biofilms\n(n = 9)")) +  
# scale\_y\_continuous(limits = c(0, 7))  
#   
# sfeb\_fig <- sfeb\_fig +  
# labs(tag = "A") +  
# theme(  
# plot.tag = element\_text(family = "Helvetica", size = 20, face = "plain", hjust = -0.1, vjust = 1),  
# plot.tag.position = c(0, 1)  
# )  
#   
# sfeb\_fig  
  
# ncos\_fig <- ncos %>%   
# ggplot(aes(y= rate\_loss, x = treatment, fill = treatment)) +   
# geom\_boxplot() +  
# geom\_jitter(alpha = 0.3, width = 0.05) +  
# myCustomTheme() +  
# scale\_fill\_manual(values = c("NCOS\_PW\_40um" = "white",  
# "NCOS\_biofilm" = "darkgrey")) +  
# theme(legend.position = "none",  
# axis.line.x = element\_line(color = "grey", size = 0.5), # Keep the x-axis line  
# axis.line.y = element\_line(color = "grey", size = 0.5),  
# panel.border = element\_blank()) +  
# xlab("Medium") +  
# ylab("Reduction of Bd over 6 days") +   
# scale\_x\_discrete (labels= c("NCOS\_PW\_40um" = "microorganisms\n(triplicates)",  
# "NCOS\_biofilm" = "biofilms\n(triplicates)")) +  
# scale\_y\_continuous(limits = c(0, 7))  
#   
# ncos\_fig <- ncos\_fig +  
# labs(tag = "B") +  
# theme(  
# plot.tag = element\_text(family = "Helvetica", size = 20, face = "plain", hjust = -0.1, vjust = 1),  
# plot.tag.position = c(0, 1)  
# )  
#   
# ncos\_fig

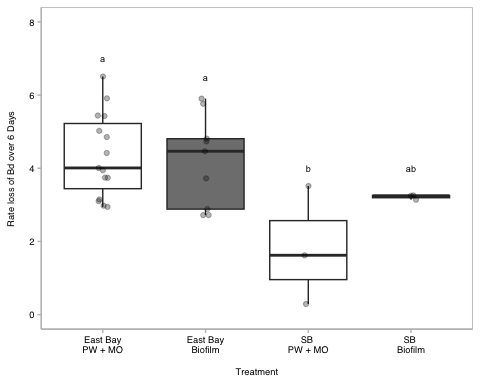
## Fig 6

all\_parts$treatment <- factor(all\_parts$treatment)   
   
si\_alltreat <- all\_parts %>%   
 mutate(site = case\_when(  
 grepl("NCOS", treatment) ~ "Santa Barbara",  
 TRUE ~ "East Bay" # Default to "east Bay"  
 )) %>%   
 ggplot(aes(y= rate\_loss, x = treatment, fill = medium)) +  
 geom\_boxplot() +  
 geom\_jitter(alpha = 0.3, width = 0.05) +  
 myCustomTheme() +  
 # biofilm and pond water colors, both from paul tol colorblind palette  
 scale\_fill\_manual(values = c("PW" = "white",  
 "Biofilm" = "darkgrey")) +  
 #facet\_wrap(~site, scales = "free\_x") +  
 theme(legend.position = "none",  
 axis.line.x = element\_line(color = "grey", size = 0.5), # Keep the x-axis line  
 axis.line.y = element\_line(color = "grey", size = 0.5)) +  
 xlab("Treatment") +  
 ylab("Rate loss of Bd over 6 Days") +   
 scale\_x\_discrete (limits = c("EB\_PW\_40um", "EB\_biofilm",   
 "NCOS\_PW\_40um", "NCOS\_biofilm"), # Specify the order  
 labels= c(  
 "EB\_PW\_40um" = "East Bay\nPW + MO",  
 "NCOS\_PW\_40um" = "SB \nPW + MO",  
 "EB\_biofilm" = "East Bay\nBiofilm",  
 "NCOS\_biofilm" = "SB \nBiofilm")) +  
 scale\_y\_continuous(limits = c(0, 8)) # Set y-axis limits from 0 to 8

Warning: The `size` argument of `element\_line()` is deprecated as of ggplot2 3.4.0.  
ℹ Please use the `linewidth` argument instead.

Warning: The `size` argument of `element\_rect()` is deprecated as of ggplot2 3.4.0.  
ℹ Please use the `linewidth` argument instead.

sig\_data\_rl <- sig\_data\_rl %>%   
 mutate(site = case\_when(  
 grepl("NCOS", treatment) ~ "Santa Barbara",  
 TRUE ~ "East Bay" # Default to "east Bay"  
 ))  
  
si\_alltreat <- si\_alltreat +   
 geom\_text(data = sig\_data\_rl,   
 aes(x = treatment, y = y\_position, label = label, group = site),   
 inherit.aes = FALSE,   
 size = 2.4) # Set fixed text size outside of aes()  
  
si\_alltreat



#ggsave("paper-figures/SI\_combined-expts.pdf", plot = si\_alltreat , width = 3.46, height = 3.46)

# Appendix

### Stats tables

anova table

# anova table  
anova\_output <- tidy(aov\_5b)  
  
aov\_5b\_tbl <- anova\_output %>%  
 dplyr::select(term, df, sumsq, meansq, statistic, p.value) %>%  
 gt() %>%  
 tab\_header(  
 title = "5b 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\_5b\_tbl

Table 1: 5b ANOVA Table

| Term | Df | Sum Sq | Mean Sq | F value | P-value |
| --- | --- | --- | --- | --- | --- |
| treatment | 3 | 18.23 | 6.08 | 4.42 | 0.012 |
| Residuals | 26 | 35.72 | 1.37 | NA | NA |

# prettier, simplified  
anova\_output <- tidy(aov\_5b)  
  
# 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 = 26, 26\*)")  
  
# Create the gt table with selected columns  
aov\_5b\_tbl\_b <- anova\_output %>%  
 dplyr::select(term, statistic, p.value) %>%  
 gt() %>%  
 tab\_header(  
 title = "5b 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\_5b\_tbl\_b

Table 1: 5b ANOVA Table

|  | F value | P-value |
| --- | --- | --- |
| treatment (*df = 3, 26*) | 4.42 | 0.012 |

## Alternate data method: Updated stats May 2025: 2 t-tests, one for each site

New plan: Renwei requests 2 t-tests, with one for each site

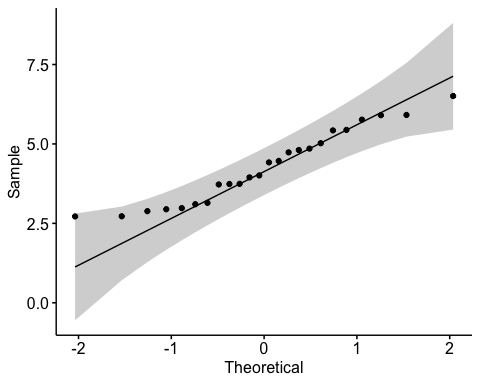
**Results: There is no significant difference in the rate loss of Bd between AE microorganisms and AE biofilms within the SFEB sites (t = 0.31125, df = 22, p-value = 0.7585). There is also no significant difference in the rate loss of Bd between AE microorganisms and AE biofilms within the SBNCOS sites (t = -1.5014, df = 4, p-value = 0.2077)**

## SFEB t-test

sfeb <- all\_parts %>%   
 filter(grepl("eb\_", treatment, ignore.case = TRUE))  
  
# test for homogeneity of variance  
leveneTest(rate\_loss ~ factor(treatment), data = sfeb) # p = 0.9628; var are similar yay!

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 0.0873 0.7704  
 22

# assumptions testing  
sfeb %>%   
 ggqqplot("rate\_loss") # good



shapiro.test(sfeb$rate\_loss) # normal, yay! (p-value = 0.1792)

Shapiro-Wilk normality test  
  
data: sfeb$rate\_loss  
W = 0.94183, p-value = 0.1792

t.test(rate\_loss ~ treatment, data = sfeb, var.equal = TRUE)

Two Sample t-test  
  
data: rate\_loss by treatment  
t = 0.31125, df = 22, p-value = 0.7585  
alternative hypothesis: true difference in means between group EB\_PW\_40um and group EB\_biofilm is not equal to 0  
95 percent confidence interval:  
 -0.8744001 1.1832035  
sample estimates:  
mean in group EB\_PW\_40um mean in group EB\_biofilm   
 4.345978 4.191576

## NCOS t-test

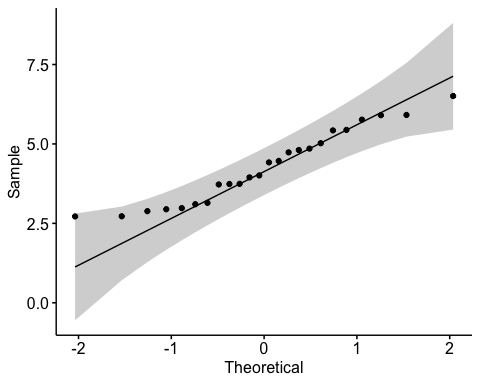
ncos <- all\_parts %>%   
 filter(grepl("ncos\_", treatment, ignore.case = TRUE))  
  
# test for homogeneity of variance  
leveneTest(rate\_loss ~ factor(treatment), data = ncos) # p = 0.1394; var are similar yay!

Levene's Test for Homogeneity of Variance (center = median)  
 Df F value Pr(>F)  
group 1 3.3902 0.1394  
 4

str(ncos)

'data.frame': 6 obs. of 5 variables:  
 $ sample\_ID: chr "PW+microorganism" "PW+microorganism" "PW+microorganism" "PW+AEBiofilm" ...  
 $ treatment: Factor w/ 4 levels "EB\_PW\_40um","NCOS\_PW\_40um",..: 2 2 2 4 4 4  
 $ rate\_loss: num 1.622 3.516 0.291 3.249 3.258 ...  
 $ location : chr "NCOS" "NCOS" "NCOS" "NCOS" ...  
 $ medium : chr "PW" "PW" "PW" "biofilm" ...

# assumptions testing  
sfeb %>%   
 ggqqplot("rate\_loss") # good



shapiro.test(ncos$rate\_loss) # barely did not pass but the qqplot is gorgeous so we will proceed (p-value = 0.04348)...also such a teeny sample size the shapiro isn't the best fit here anyway

Shapiro-Wilk normality test  
  
data: ncos$rate\_loss  
W = 0.78564, p-value = 0.04348

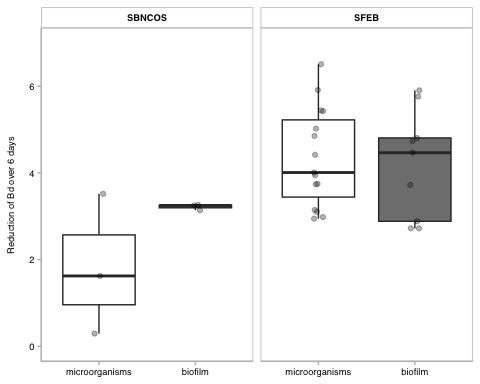
t.test(rate\_loss ~ treatment, data = ncos, var.equal = TRUE)

Two Sample t-test  
  
data: rate\_loss by treatment  
t = -1.5014, df = 4, p-value = 0.2077  
alternative hypothesis: true difference in means between group NCOS\_PW\_40um and group NCOS\_biofilm is not equal to 0  
95 percent confidence interval:  
 -4.006392 1.194113  
sample estimates:  
mean in group NCOS\_PW\_40um mean in group NCOS\_biofilm   
 1.809550 3.215689

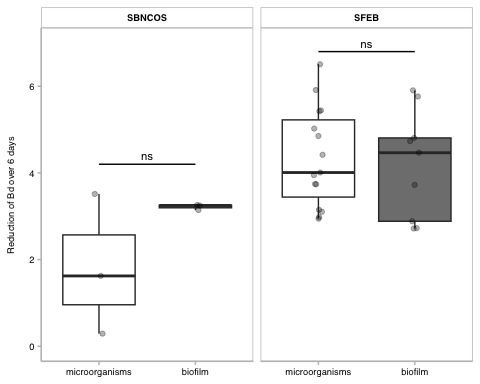
combined <- all\_parts %>%   
 mutate(  
 medium = factor(medium), # convert to factor first  
 medium = fct\_relevel(medium, "PW", "Biofilm"), # pond water first  
 # rename the location so it shows up correct in the facet   
 location = case\_when(  
 location == "EB" ~ "SFEB",  
 location == "NCOS" ~ "SBNCOS",  
 TRUE ~ location)) %>%  
 ggplot(aes(y= rate\_loss, x = medium, fill = medium)) +   
 geom\_boxplot() +  
 geom\_jitter(alpha = 0.3, width = 0.05) +  
 myCustomTheme() +  
 scale\_fill\_manual(values = c("PW" = "white",  
 "Biofilm" = "darkgrey")) +  
 theme(legend.position = "none",  
 axis.line.x = element\_line(color = "grey", size = 0.5), # Keep the x-axis line  
 axis.line.y = element\_line(color = "grey", size = 0.5),  
 panel.border = element\_rect(color = "gray", size = 0.5, fill = NA)) +  
 xlab(NULL) +  
 ylab("Reduction of Bd over 6 days") +   
 scale\_x\_discrete (labels= c("PW" = "microorganisms",  
 "Biofilm" = "biofilms")) +  
 facet\_wrap(~location) +  
 scale\_y\_continuous(limits = c(0, 7))

Warning: There was 1 warning in `mutate()`.  
ℹ In argument: `medium = fct\_relevel(medium, "PW", "Biofilm")`.  
Caused by warning:  
! 1 unknown level in `f`: Biofilm

combined



#ggsave("paper-figures/combined-expts\_fig5\_updated\_NO SIG MARKERS.pdf", plot = combined , width = 3.46, height = 3.46)  
  
ns\_data <- data.frame(  
 location = c("SFEB", "SBNCOS"),  
 x\_start = 1, # corresponds to "PW"  
 x\_end = 2, # corresponds to "Biofilm"  
 y = c(6.8,4.2), # where on Y we want the bar  
 label = "ns"  
)  
  
fig5\_withNS <- combined +  
 geom\_segment(data = ns\_data,  
 aes(x = x\_start, xend = x\_end, y = y, yend = y),  
 inherit.aes = FALSE,  
 linewidth = 0.5) +  
 geom\_text(data = ns\_data,  
 aes(x = 1.5, y = y + 0.2, label = label),  
 inherit.aes = FALSE,  
 size = 3,  
 fontface = "plain")  
  
fig5\_withNS



#ggsave("paper-figures/combined-expts\_fig5\_updated\_SIG MARKERS.pdf", plot = fig5\_withNS, width = 3.46, height = 3.46)