

Nursery_Performance

2024-02-22

```
setwd("~/Desktop/GitHub/MVP-H2F-HatcheryField/src/performance_H2F")
```

Packages

```
#load necessary packages  
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.4      v readr      2.1.5  
## v forcats    1.0.0      v stringr   1.5.1  
## v ggplot2    3.5.1      v tibble    3.2.1  
## v lubridate  1.9.3      v tidyr     1.3.1  
## v purrr      1.0.2  
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)  
library(gridExtra)
```

```
##  
## Attaching package: 'gridExtra'  
##  
## The following object is masked from 'package:dplyr':  
##  
##      combine
```

```
library(multcompView)  
library(car) #upload package with Anova function
```

```
## Loading required package: carData  
##  
## Attaching package: 'car'  
##  
## The following object is masked from 'package:dplyr':  
##  
##      recode  
##  
## The following object is masked from 'package:purrr':  
##  
##      some
```

```
library(cowplot)
```

```
##  
## Attaching package: 'cowplot'  
##  
## The following object is masked from 'package:lubridate':  
##  
##     stamp
```

Load data

```
#download spawn treatment data  
spawn <- read.csv("../data/performance_H2F/CviMVP_Larvae_Growth-spawn_trt.csv")  
#download group data  
group <- read.csv("../data/performance_H2F/CviMVP_Larvae_Growth-group.csv")  
#download length data  
larvae_img <- read.csv("../data/performance_H2F/CviMVP_Larvae_Growth-larvae_img.csv")
```

Data preparation

```
#combine larvae_img and group dfs  
len_df <- group %>%  
  left_join(larvae_img, by = "Group_Day_Key")  
  
#remove some columns  
len_analysis_df <- subset(len_df, select = -c(Ind_ID, Group_Day_Label.y, Image_Key, Img_Notes, SpawnTrt_L))
```

Merge dataframes

```
#subset data for oysters aged 15-78 days (later stage larvae)  
latestage <- subset(len_analysis_df, Day_analysis == "78" | Day_analysis == "15-21")  
  
#a sample marked WV is erroneously in this data set- remove  
latestage <- subset(latestage, Image_Label != "WV_06_200L_6")  
  
#rename cols  
colnames(latestage) <- c("site_name", "site_day_key", "site_day_label", "day_trt", "day_an", "day", "day_an")  
  
#make column with measurements in millimeters instead of microns  
latestage$shell_length_mm <- (latestage$shell_length_um)/1000
```

Subset data

```

#prep data for ANOVA

#use grep function to make a vector of all indices in latestage_len with the phrase "small" in them. Save them in small
small <- grep("small", latestage$site_day_label, ignore.case = TRUE)

small_img <- grep("small", latestage$image_label, ignore.case = TRUE)

#make new dataframe WITHOUT any indices saved in small, mean remaining data are everything except small
no_small_df <- latestage[-small, ]
no_small_df <- latestage[-small_img, ]

#rename rows
colnames(no_small_df) <- c("site_name", "site_day_key", "site_day_label", "day_trt", "day_an", "day", "day_len")

#replace JR with updated group name "VA"
no_small_df["site_name"][no_small_df["site_name"] == "MVP-JR"] <- "MVP-VA"

#exclude "MVP" from group names and add in W or S labels
groups <- c(unique(no_small_df$site_name))
wild <- substring(groups[c(2,1,6,5,4,3)], 5)
w_label <- paste(paste("W", 1:6, sep = ""), wild, sep = "-")

for(x in wild){
  no_small_df["site_name"][no_small_df["site_name"] == paste("MVP", x, sep = "-")] <- w_label[grep(x, w_label)]
}

#rename selection line sites to include S1-S2 labels
selection <- substring(groups[7:8], 5)
s_label <- paste(paste("S", 1:2, sep = ""), selection, sep = "-")

for(x in selection){
  no_small_df["site_name"][no_small_df["site_name"] == paste("MVP", x, sep = "-")] <- s_label[grep(x, s_label)]
}

#add H1 label to LARMIX (hybrid mix)
no_small_df["site_name"][no_small_df["site_name"] == "MVP-LARMIX"] <- "H1-LARMIX"

no_small_df <- no_small_df %>%
  mutate(site_name = site_name %>%
    fct_relevel("W1-TX", "W2-LA", "W3-FL", "W4-VA", "S1-LOLA", "S2-DEBY", "W5-NH", "W6-ME", "H1-LARMIX"))

#save latestage_df as a csv file for future analyses
write.csv(no_small_df, "../data/performance_H2F/latestage_length.csv")

```

Data product: latestage_length.csv

```

#calculate mean oyster length per site_name per day using aggregate

#make standard error equation

```

```

se <- function(x){
  sd(x)/sqrt(length(x))
}

#find mean lengths of each site_name on each day
means_mm <- aggregate(shell_length_mm ~ site_name + day_an, FUN = mean, data = no_small_df)
colnames(means_mm) <- c("site_name", "days", "mean_length_mm")

#find standard error of length of each site_name on each day
se_mm <- aggregate(shell_length_mm ~ site_name + day_an, FUN = se, data = no_small_df)
colnames(se_mm) <- c("site_name", "days", "se_length_mm")

```

Calculate mean length and standard error

Prepare pre-deployment (day 78) data

```

#convert site_name and day from chr to factor
no_small_df$site_name <- as.factor(no_small_df$site_name)
no_small_df$day_an <- as.factor(no_small_df$day_an)
str(no_small_df)

## 'data.frame':    1773 obs. of  10 variables:
## $ site_name      : Factor w/ 9 levels "W1-TX","W2-LA",...: 2 2 2 2 2 2 2 2 2 ...
## $ site_day_key   : chr  "7E8E995C" "7E8E995C" "7E8E995C" "7E8E995C" ...
## $ site_day_label : chr  "LA_19_Eyed" "LA_19_Eyed" "LA_19_Eyed" "LA_19_Eyed" ...
## $ day_trt        : chr  "19_Eyed" "19_Eyed" "19_Eyed" "19_Eyed" ...
## $ day_an         : Factor w/ 2 levels "15-21","78": 1 1 1 1 1 1 1 1 1 ...
## $ day            : int   19 19 19 19 19 19 19 19 19 ...
## $ date           : chr   "2023-05-28" "2023-05-28" "2023-05-28" "2023-05-28" ...
## $ image_label    : chr   "LA_19_eyed_1" "LA_19_eyed_2" "LA_19_eyed_3" "LA_19_eyed_4" ...
## $ shell_length_um: num   316 320 333 324 332 ...
## $ shell_length_mm: num    0.316 0.32 0.333 0.324 0.332 ...

#subsetting based on oyster age (days) for oysters on day 78 (end of hatchery stage, pre field deployment)
predep <- subset(no_small_df, day_an == "78")

```

Run ANOVA for end of hatchery (day 78) shell lengths

```

predep.lm <- lm(shell_length_mm ~ site_name, data = predep) #fit to lm
predep.lm.resid <- predep.lm$residuals #save residuals
shapiro.test(predep.lm.resid) #use Shapiro Wilk test to assess normality

##
## Shapiro-Wilk normality test
##
## data:  predep.lm.resid
## W = 0.9974, p-value = 0.08176

```

```

# p = 0.082, fail to reject H0, data are normal
#plot(predep.lm)
#this data looks good, normal, homoscedastic, no outliers

#Tukey-Kramer labels only allow one "-" per comparison label, so I will replace the "-" with "_"
predep$site_name <- gsub("-", "_", predep$site_name)

predep.anova3 <- Anova(lm(shell_length_mm ~ site_name, data = predep), type = 3)
predep.anova3

## Anova Table (Type III tests)
##
## Response: shell_length_mm
##           Sum Sq   Df F value    Pr(>F)
## (Intercept) 20630.6    1 1867.501 < 2.2e-16 ***
## site_name    4509.8    8   51.029 < 2.2e-16 ***
## Residuals   11765.3 1065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

predep.anova <- aov(lm(shell_length_mm ~ site_name, data = predep))
summary(predep.anova)

```

Run Tukey Kramer post-hoc test

```

##           Df Sum Sq Mean Sq F value Pr(>F)
## site_name    8   4510   563.7   51.03 <2e-16 ***
## Residuals  1065  11765    11.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

#extract p-value
predep_pval <- summary(predep.anova)[[1]][["Pr(>F)"]][1]
print(predep_pval)

```

```
## [1] 4.998853e-70
```

```

#post-hoc comparison- Tukey-Kramer test
predep.tuk <- TukeyHSD(x=predep.anova,conf.level = 0.95)

#extract significant letters from TK test and store
require(multcompView)
predep_cld <- multcompLetters4(predep.anova, predep.tuk)
predep_dt <- group_by(predep, site_name) %>%
  summarise(m=mean(shell_length_mm)) %>%
  arrange(desc(m))
predep_cld <- as.data.frame.list(predep_cld$site_name)
predep_dt$predep_cld <- predep_cld$Letters

#make shell length numeric in predep
predep$shell_length_mm <- as.numeric(predep$shell_length_mm)

```

Figure product: pre-deployment length boxplot

```
predep <- predep %>%
  mutate(site_name = site_name %>%
    fct_relevel("W1_TX", "W2_LA", "W3_FL", "W4_VA", "S1_LOLA", "S2_DEBY", "W5_NH", "W6_ME", "H1_

#plot. fatten = NULL will remove the median lines in each box, allowing me to display the mean instead.
predep_box <- ggplot(data = predep, aes(site_name, shell_length_mm))+
  geom_boxplot(color = c("#3e4989", "#31688e", "#26828e", "#1f9e89", "#440154", "#482878", "#35b779", "#
  scale_x_discrete(name = "Broodstock Group", limits = c("W1_TX", "W2_LA", "W3_FL", "W4_VA", "S1_LOLA"
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.2, linewidth = .3, color = c("white", "
  stat_summary(fun = mean, geom = "errorbar", aes(ymax = ..y.., ymin = ..y..), width = 0.75, linewidth =
  ylab("Day 78 Shell Length (mm)") +
  xlab("Broodstock Group") +
  labs(subtitle = "p < 0.001") +
  ylim(c(0, 32)) +
  geom_text(data = predep_dt, size = 3, aes(label = predep_cld, y = 29, x = site_name), vjust = -0.5) +
  ggtitle("Pre-Field-Deployment (Day 78) Shell Lengths") +
  theme_bw() +
  theme(plot.subtitle = element_text(size = 9), axis.title.y = element_text(size = 11), axis.title.x =

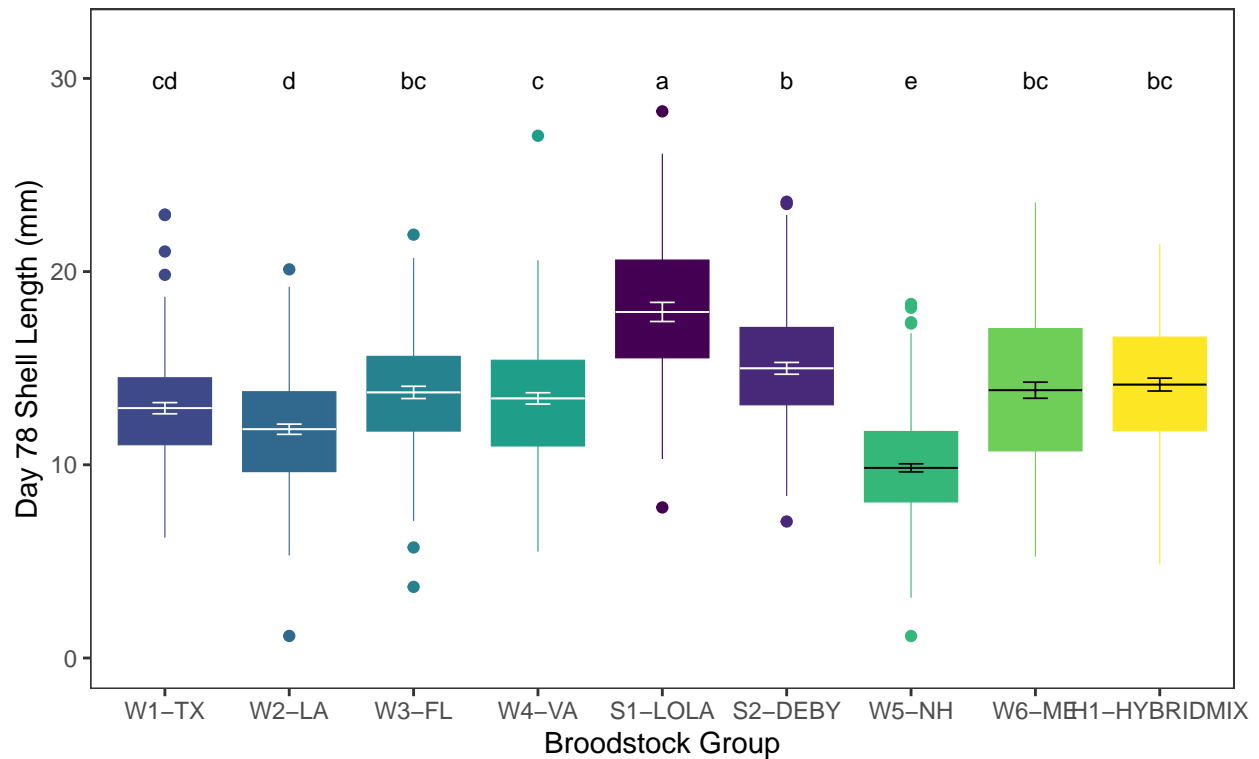
predep_box
```

```
## Warning: The dot-dot notation ('..y..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(y)' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
```

Pre-Field-Deployment (Day 78) Shell Lengths

$p < 0.001$



```
ggsave("predep_length.pdf",
  plot = predep_box,
  device = "pdf",
  path = "../figures/performance_H2F")
```

```
## Saving 6.5 x 4.5 in image
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
## Removed 1 row containing missing values or values outside the scale range
## ('geom_segment()').
```