Envr_Genetics_Length_Survival_Hatchery

2025-02-20

This code analyzes the effect of conditions in the environments-of-origin and parent genetics on offspring oyster survival and shell length in the VIMS ABC hatchery and nursery.

##Set working directory

```
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
```

Load 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) #extract significance letters from Tukey-Kramer test
library(grid)#for ggplot arranging
library(cowplot) #arrange ggplots
```

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

```
library(ggpmisc)
```

```
## Loading required package: ggpp
## Registered S3 methods overwritten by 'ggpp':
##   method                from
##   heightDetails.titleGrob ggplot2
##   widthDetails.titleGrob  ggplot2
##
## Attaching package: 'ggpp'
##
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```
library(vegan) #for Mantel tests
```

```
## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.6-8
```

```
library(broom)#save lm results to tibbles
```

##Read in necessary files

```
#this file contains raw lengths of all oysters aged 15-78 days
length <- read.csv(file.path("../../data/performance_H2F/latestage_length.csv"))

#this file has summary statistics on temperature and salinity for the 8 sites (W1-TX, W2-LA, W3-FL, S1-
envr_summary <-  read.csv("../../data/envr_of_origin/envr_summary_stats.csv")

#this file is a matrix of the environmental distances between each site
envr_dist <- read.csv("../../data/envr_of_origin/envr_dist.csv")

#this file contains survival rate data for larval oysters up to day 21 post hatching
survival <-  read.csv("../../data/performance_H2F/CViMVP_larvae_survival.csv")

parent_genetics <- read.csv("../../data/parent_effects_H2F/parent_genetic_effects.csv")
```

Extract survival rates from survival data frame, and reorganize data frame

```
#For survival rates, we only care about survival from day 21 (marked as day_an 15-21), so filter for th
survival_day21 <- survival %>%
  filter(Day_analysis == "15-21") %>%
  filter(Tank_naming != "MVP-LARMIX")

#Exclude any groups with "small" in their label. The length data only considers the eyed oysters.
small <- grep("small", survival_day21$Group_Day_Label, ignore.case = TRUE)

#make new dataframe WITHOUT any indices saved in small, mean remaining data are everything except small
survival_day21_sub <- survival_day21[-small, ]
```

```r
#exclude "MVP" from group names and add in W or S labels
survival_day21_sub["Tank_naming"][survival_day21_sub["Tank_naming"] == "MVP-TX"] <- "W1-TX"
survival_day21_sub["Tank_naming"][survival_day21_sub["Tank_naming"] == "MVP-LA"] <- "W2-LA"
survival_day21_sub["Tank_naming"][survival_day21_sub["Tank_naming"]  == "MVP-FL"] <- "W3-FL"
survival_day21_sub["Tank_naming"][survival_day21_sub["Tank_naming"] == "MVP-LOLA"] <- "S1-LOLA"
survival_day21_sub["Tank_naming"][survival_day21_sub["Tank_naming"] == "MVP-DEBY"] <- "S2-DEBY"
survival_day21_sub["Tank_naming"][survival_day21_sub["Tank_naming"] == "MVP-JR"] <- "W4-VA"
survival_day21_sub["Tank_naming"][survival_day21_sub["Tank_naming"] == "MVP-NH"] <- "W5-NH"
survival_day21_sub["Tank_naming"][survival_day21_sub["Tank_naming"] == "MVP-ME"] <- "W6-ME"

#select only tank naming and survival rate columns
survival_filter <- survival_day21_sub[, c("Tank_naming", "Survival_rate_perc")]

#rename columns
colnames(survival_filter) <- c("site_name", "percent_survival")
```

Length df organizing

```r
#select columns with site name, day, and shell length in mm
length <- length[,c("site_name", "day_an", "shell_length_mm")]

#filter out LARMIX from group
length_filter <- length %>%
  filter(site_name != "LARMIX")

#keep only data from days 15-21
length_21 <- length_filter %>%
  filter(day_an == "15-21")

length_78 <- length_filter %>%
  filter(day_an == "78")

#find mean shell lengths of each group on each day
means_mm_21 <- aggregate(shell_length_mm ~ site_name, FUN = mean, data = length_21)
colnames(means_mm_21) <- c("site_name", "mean_length_mm_21")

means_mm_78 <- aggregate(shell_length_mm ~ site_name, FUN = mean, data = length_78)
colnames(means_mm_78) <- c("site_name", "mean_length_mm_78")
```

Remove extra columns from environmental data frame and add latitude

```r
#remove extra site label and standard deviations from envr_summary
envr <- envr_summary[,-c(1,6,10)]

summary(envr)
```

```
##   site_name          Mean_Annual_Temperature_C Mean_max_temperature_C
##  Length:8           Min.   :15.33              Min.   :23.46
##  Class :character   1st Qu.:16.86              1st Qu.:26.35
##  Mode  :character   Median :17.83              Median :28.65
##                     Mean   :19.09              Mean   :28.54
##                     3rd Qu.:22.33              3rd Qu.:30.98
```

```
##                             Max.    :23.11            Max.    :33.17
##  Mean_min_temperature_C Mean_Annual_Salinity_ppt Mean_max_Salinity_ppt
##  Min.    : 1.758        Min.    :10.46           Min.    :16.88
##  1st Qu.: 3.192         1st Qu.:14.87            1st Qu.:23.69
##  Median : 3.784         Median :18.99            Median :27.90
##  Mean    : 5.121        Mean    :20.02           Mean    :27.25
##  3rd Qu.: 7.481         3rd Qu.:23.01            3rd Qu.:30.54
##  Max.    :10.387        Max.    :32.35           Max.    :37.20
##  Mean_min_Salinity_ppt
##  Min.    : 0.9688
##  1st Qu.: 5.0596
##  Median : 9.3657
##  Mean    : 8.5100
##  3rd Qu.:12.6328
##  Max.    :15.4224
```

```r
#rename cols
colnames(envr) <- c("site_name", "Mean_Annual_Temperature_C", "Mean_max_temperature_C", "Mean_min_tempe

#add sampling site latitudes. Note that these latitudes ARE NOT the same as the latitudes of where env
envr$lat <- c(28.096000, 29.239925, 30.440030, 37.1501163, 37.98030, 37.249107, 43.053746, 44.01330)

#rename column names
colnames(parent_genetics) <- c("site_name", "avg_observed_ho", "avg_expected_hs", "avg_allelic_richness
```

#Join data frames

```r
surv_length_envr_gen <- left_join(envr, means_mm_21, by = "site_name") %>%
  left_join(means_mm_78, by = "site_name") %>%
  left_join(survival_filter, by = "site_name") %>%
  left_join(parent_genetics, by = "site_name")

write.csv(surv_length_envr_gen, "../../data/parent_effects_H2F/Hatchery_Performance_Envr_ParentGenetics
```

The next chunk includes multiple regressions testing the effects of several environmental and genetic parameters on length and survival in the hatchery/nursery. The explanatory variables are 1) mean annual temperature at the environment-of-origin (mean temp), 2) mean annual salinity at the environment of origin (mean sal), 3) mean observed heterozygosity in the parent oysters from each site (Ho), and 4) mean allelic richness (AR) in the parent oysters from each site.

Multiple regressions

```r
#test correlation between temperature and latitude. If they are highly correlated, one must be excluded
temp_lat_corr <- cor.test(surv_length_envr_gen$Mean_Annual_Temperature_C, surv_length_envr_gen$lat)
temp_lat_corr #significant correlation (r = -0.962, p = 0.00014) between temp and lat, cannot include b
```

```
##
##  Pearson's product-moment correlation
##
## data:  surv_length_envr_gen$Mean_Annual_Temperature_C and surv_length_envr_gen$lat
```

```
## t = -8.5844, df = 6, p-value = 0.0001373
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9932434 -0.7970442
## sample estimates:
##        cor
## -0.9616187
```

```r
#multiple regression testing effect of mean annual temp, mean annual salinity, mean Ho, and mean AR on
surv21_mr <- lm(percent_survival ~ Mean_Annual_Temperature_C + Mean_Annual_Salinity_ppt + avg_observed_
summary(surv21_mr)
```

```
##
## Call:
## lm(formula = percent_survival ~ Mean_Annual_Temperature_C + Mean_Annual_Salinity_ppt +
##     avg_observed_ho + avg_allelic_richness, data = surv_length_envr_gen)
##
## Residuals:
##        1        2        3        4        5        6        7        8
##  0.50389 -0.24887 -0.28291 -0.22828 -0.27175 -0.01484  0.56405 -0.02130
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -14.00409   23.47397  -0.597    0.593
## Mean_Annual_Temperature_C   0.26314    0.15173   1.734    0.181
## Mean_Annual_Salinity_ppt   -0.08170    0.05194  -1.573    0.214
## avg_observed_ho           -12.63800   33.97290  -0.372    0.735
## avg_allelic_richness        7.59709    6.49045   1.171    0.326
##
## Residual standard error: 0.5294 on 3 degrees of freedom
## Multiple R-squared:  0.9131, Adjusted R-squared:  0.7971
## F-statistic: 7.876 on 4 and 3 DF,  p-value: 0.06075
```

```r
#model p-value = 0.06075
#model F-stat = 7.876
#adjusted R squared = -.7971
surv21_tidy <- tidy(surv21_mr) #save coefficients as table


#multiple regression testing effect of mean annual temp, mean annual salinity, mean Ho, and mean AR on
length21_mr <- lm(mean_length_mm_21 ~ Mean_Annual_Temperature_C + Mean_Annual_Salinity_ppt + avg_observe
summary(length21_mr)
```

```
##
## Call:
## lm(formula = mean_length_mm_21 ~ Mean_Annual_Temperature_C +
##     Mean_Annual_Salinity_ppt + avg_observed_ho + avg_allelic_richness,
##     data = surv_length_envr_gen)
##
## Residuals:
##         1         2         3         4         5         6         7         8
## -0.011280  0.006474  0.006396  0.010726 -0.003569  0.001885 -0.015313  0.004681
##
```

```
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                0.6846890  0.6259681   1.094    0.354
## Mean_Annual_Temperature_C  0.0011916  0.0040461   0.295    0.788
## Mean_Annual_Salinity_ppt  -0.0001098  0.0013850  -0.079    0.942
## avg_observed_ho           -0.3772591  0.9059375  -0.416    0.705
## avg_allelic_richness      -0.1009795  0.1730774  -0.583    0.601
##
## Residual standard error: 0.01412 on 3 degrees of freedom
## Multiple R-squared:  0.2538, Adjusted R-squared:  -0.7412
## F-statistic: 0.2551 on 4 and 3 DF,  p-value: 0.89
```

```
#model p-value = 0.89
#model F-stat = 0.2551
#Adjusted R squared = -0.74
length21_tidy <- tidy(length21_mr) #save coefficients in table

#multiple regression testing effect of mean annual temp, mean annual salinity, mean Ho, and mean AR on
length78_mr <- lm(mean_length_mm_78 ~ Mean_Annual_Temperature_C + Mean_Annual_Salinity_ppt + avg_observ
summary(length78_mr)
```

```
##
## Call:
## lm(formula = mean_length_mm_78 ~ Mean_Annual_Temperature_C +
##     Mean_Annual_Salinity_ppt + avg_observed_ho + avg_allelic_richness,
##     data = surv_length_envr_gen)
##
## Residuals:
##       1       2       3       4       5       6       7       8
## -1.5701  0.9278  0.7063  1.2637  0.1832  0.8086 -3.2516  0.9321
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -143.1209   107.2568  -1.334    0.274
## Mean_Annual_Temperature_C    -1.2555     0.6933  -1.811    0.168
## Mean_Annual_Salinity_ppt      0.3491     0.2373   1.471    0.238
## avg_observed_ho             274.0808   155.2283   1.766    0.176
## avg_allelic_richness         26.7222    29.6560   0.901    0.434
##
## Residual standard error: 2.419 on 3 degrees of freedom
## Multiple R-squared:  0.5417, Adjusted R-squared:  -0.06934
## F-statistic: 0.8865 on 4 and 3 DF,  p-value: 0.5623
```

```
#model p-value = 0.5623
#model F-stat = 0.8865
#adjusted R squared = -0.06934
length78_tidy <- tidy(length78_mr) #save coefficients as table
```

#adjust p-values

```
model_p_values <- c(0.06075, 0.89, 0.5623)
```

```
model_p_adj <- p.adjust(model_p_values, method = "BH")
```

```
model_p_adj
```

```
## [1] 0.18225 0.89000 0.84345
```

Adjusted model p-values Survival at day 21: 0.18225 Length at day 21: 0.89000 Length at day 78: 0.84345