Figures from Plasticity and Artificial Selection for Developmental Mode in a Poecilogonous Sea Slug

This Rmarkdown document contains the figures and code for figures in the MS Plasticity and Artificial Selection for Developmental Mode in a Poecilogonous Sea Slug published in Evolution and Ecology in 2021

Figure 1

Figure one is a schematic of the study design and was made in powerpoint (with lots of copying and pasting).

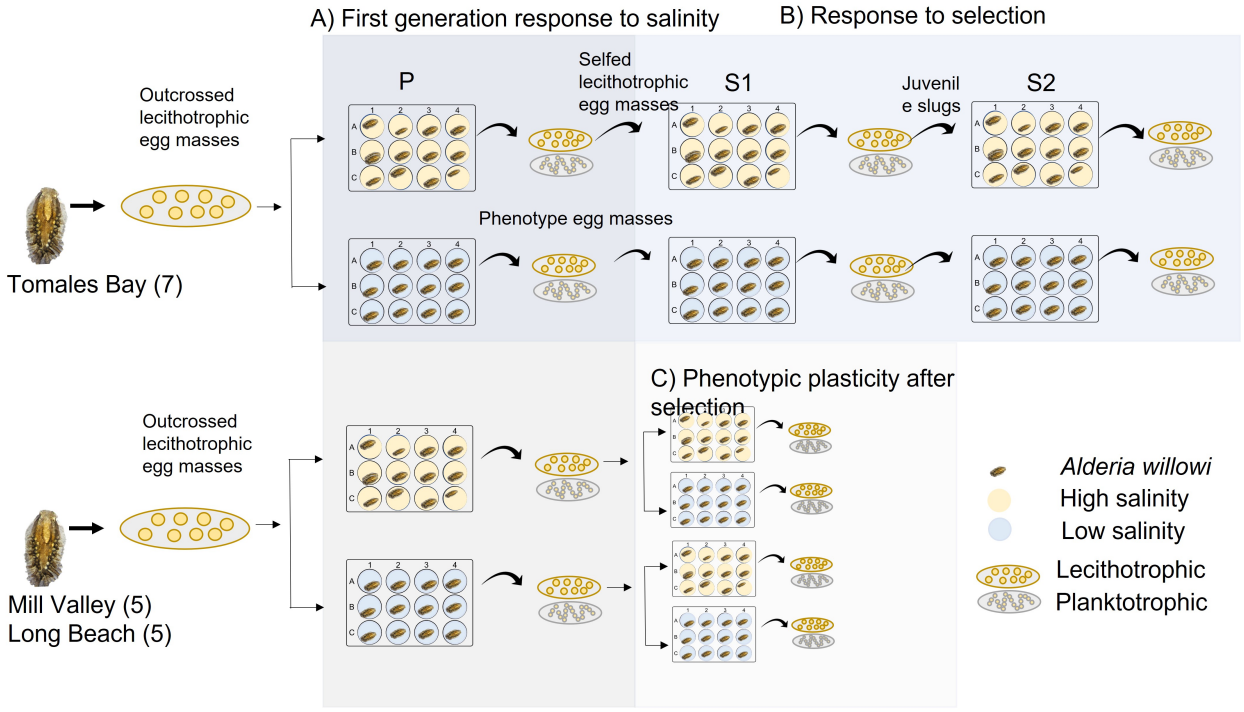

Figure 1: Schematic showing the layout of two split-brood experiments the data for which addresses three main points. The first-generation response to salinity (A) uses pooled data from both experiments. The response to selection (B) was estimated from the first experiment using 7 maternal families collected from Tomales Bay. In the second experiment I evaluate the effect to plasticity following one generation of selection (C) and self-fertilization in low and high salinity using maternal families from Mill Valley and Long Beach.

Figure 2

Figure two combines images of the study species with a scatterplot made in R. While I could combine them in R, I made the final figure in powerpoint. And that's okay.

The code for the scatterplot is included at the end of this document.

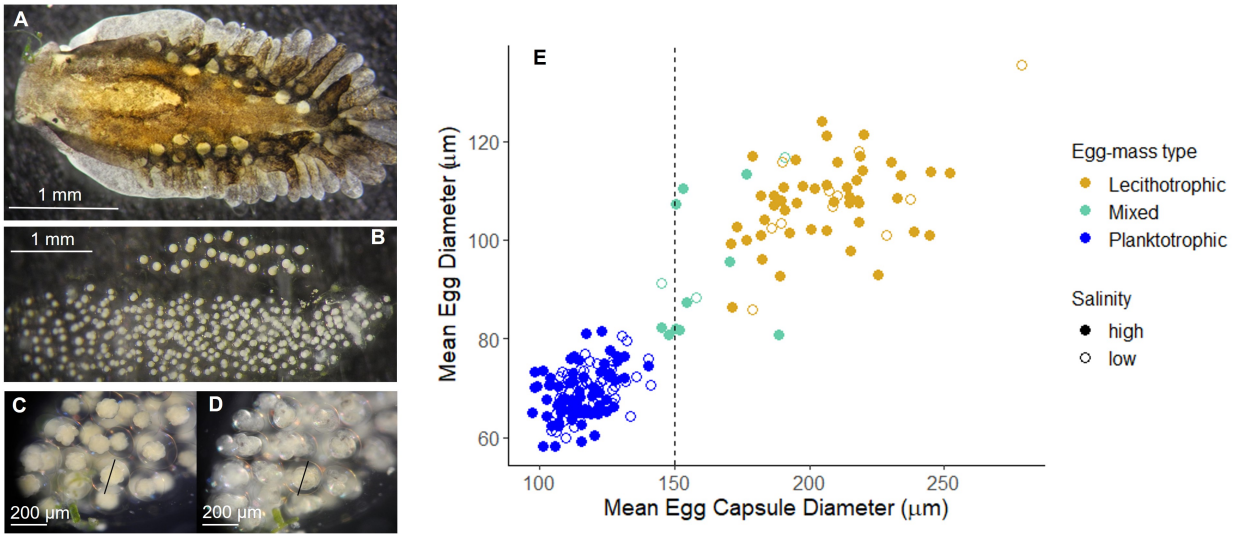

Figure 2: Images showing the study organism, the sacoglossan sea slug Alderia willowi (A). The two types of egg-masses produced by this species with a lecithotrophic egg-mass above a planktotrophic egg-mass (B). Egg capsule size is constant through development. Panel C and D show the same embryo at the 32-64 cell stage (C) and 4 days later at the veliger stage (D). At each time the egg capsule was measured to be 206 µm. Panel (E) shows that the relationship between mean egg diameter and mean egg-capsule diameter for a given egg-mass, is positive (r-squared = 85, p-value < 2.2e-16). Salinity doesn't change the range in egg diameter or egg capsule size but does shift the proportion of egg-mass type. All images were taken by the author.

Figure 3

Shows the family reaction norms for slugs in low versus high salinity

```
#packages
require(ggplot2)
## Loading required package: ggplot2
## Warning: package 'ggplot2' was built under R version 4.0.5


#get the data

map_pop<-read.csv("dataDryad/TomalesBay_capsuledata.csv", header=TRUE, sep=",")
SvN<-read.csv("dataDryad/SoCalvNorCal_Data.csv", header=TRUE, sep=",")

#make svn do the dates right
SvN$date<-as.Date(SvN$date, "%m/%d/%Y")
map_pop$date<-as.Date(map_pop$date, "%m/%d/%Y")

#rbind SuN and maping pops to make a master high versus low (hvl) data set

HvL<-rbind(SvN[,c("Capsule", "original", "date", "Family", "ID")],map_pop[, c("Capsule", "original", "da
HvL$date<-as.Date(HvL$date, "%m/%d/%Y")
```

```
### LETS LOOK AT HvL by individual mean capsule size. this is important
### use the function aggregate, it outputs a dataframe instead of tapply which outputs a vector


ID_mean<-aggregate(HvL$Capsule, list(HvL$ID) , mean)
colnames(ID_mean)<-c("ID", "mean_cap")

#now merge them together with the original data so we can get the Family etc info
ID_mean_tog<-merge(ID_mean, HvL, by=c("ID"))
ID_mean_tog<-subset(ID_mean_tog, select = -c(Capsule, date))
ID_mean_tog_uni<-unique.data.frame(ID_mean_tog)

#now add on the type for different sizes

ID_mean_tog_uni$type_150 <- ifelse(ID_mean_tog_uni$mean_cap>0.150,"l","p")


freq_150<-as.data.frame(table(ID_mean_tog_uni$type_150, as.factor(ID_mean_tog_uni$Family):as.factor(ID_m
#head(freq_150)

#get the dataframe ready to calulate the proportion
freq_150$Var2<-as.vector(freq_150$Var2)
splity<-unlist(strsplit(freq_150$Var2, ":"))

Family<-splity[seq(1, length(splity), by=2)]
ENv<-splity[seq(2, length(splity), by=2)]

freq_150<-cbind(freq_150, Family, ENv)
freq_150<-within(freq_150, rm(Var2))
#head(freq_150)

#this has values for all of the families that we have in the map_pop dataframe.
#we want to remove the 2nd and 3rd gen families - we will tell it which fams to keep

the_fams<-c("SCW46", "SCW47", "SCW48", "SCW49", "SCW50",
            "W128", "W129", "W130", "W131", "W132", "W133", "W134",
            "W143", "W144", "W148","W149", "W150")

freq_l_150_fams<-freq_150[is.element(freq_150$Family, the_fams), ]


#head(freq_l_150_fams)
str(freq_l_150_fams)
## 'data.frame':    68 obs. of  4 variables:
##  $ Var1  : Factor w/ 2 levels "l","p": 1 2 1 2 1 2 1 2 1 2 ...
##  $ Freq  : int  2 9 2 8 1 8 0 10 1 17 ...
##  $ Family: chr  "SCW46" "SCW46" "SCW46" "SCW46" ...
##  $ ENv   : chr  "20C16ppt" "20C16ppt" "20C32ppt" "20C32ppt" ...
length(unique(freq_l_150_fams$Family))
## [1] 17

#write a for loop to calulate the proportion
#create an empty vector to hold the values of prop_l
```

```
prop_l_150<-seq(1:length(freq_l_150_fams$Var1))

for (i in seq(1,length(freq_l_150_fams$Var1),2)){
  calc_prop<-freq_l_150_fams$Freq[i]/(freq_l_150_fams$Freq[i+1]+freq_l_150_fams$Freq[i])
  prop_l_150[i]<-calc_prop

}

#head(prop_l_150)

#bind it to the data frame with the family and env info

prop_l_150<-cbind(prop_l_150, freq_l_150_fams)

#head(prop_l_150)

#pull out just the lecithotrophic props(since those are the only ones that make sense)

prop_l_150_l<-prop_l_150[prop_l_150$Var1 == "l",]

#there are repeating columns in the dataframe pullout the ones we want

prop_l_final<-prop_l_150_l[,c(1,4,5)]

colors <- c("black", "black", "black", "black", "black", #socal black
            "lightblue", "lightblue", "lightblue", "lightblue", "lightblue", "lightblue", "lightblue", 
            "#74C476", "#74C476", "#74C476", "#74C476", "#74C476") #tomales = the greens



prop_l_final$Family<-sort(prop_l_final$Family)

levels(prop_l_final$ENv)<- c("Low", "High")

#define region

prop_l_final$region<-ifelse(grepl("SC*", prop_l_final$Family), "Long Beach", ifelse(grepl("W128|13[0-4]

#now plot that shit

fig2a<-ggplot(data=prop_l_final, aes(x=ENv, y=prop_l_150, group=Family, color=region))+
  geom_point(size=4, alpha=0.65, position = position_dodge(width=0.14))+
  geom_line(size=2, alpha=0.65,  position = position_dodge(width=0.14))+
  scale_color_manual(values=c("#74C476", "black","lightblue"))+
  theme_bw() + theme(panel.grid.major = element_blank(),
                     panel.grid.minor = element_blank(),
                     axis.line = element_line(colour = "black"),
                     plot.title = element_text(size = 16, face = "bold"),
                     text = element_text(size = 16),
                     legend.position="none")+
  labs(x="Salinity", y="Proportion lecithotrophic", title="A")

#now for figure 2b ploting the proportion of low salinity against high salinity
```

```r
#make a table of the freq of l and p

type_150 <- ifelse(ID_mean_tog_uni$mean_cap > 0.150,"l","p")
ID_mean_tog_uni<-cbind(ID_mean_tog_uni, type_150)
#head(ID_mean_tog_uni)

freq_150<-as.data.frame(table(ID_mean_tog_uni$type_150, as.factor(ID_mean_tog_uni$Family):as.factor(ID_m
#head(freq_150)
#get the dataframe ready to calulate the proportion
freq_150$Var2<-as.vector(freq_150$Var2)
splity<-unlist(strsplit(freq_150$Var2, ":"))

Family<-splity[seq(1, length(splity), by=2)]
ENv<-splity[seq(2, length(splity), by=2)]

freq_150<-cbind(freq_150, Family, ENv)
freq_150<-within(freq_150, rm(Var2))
#head(freq_150)

#this has values for all of the families that we have in the map_pop dataframe.
#we want to remove the 2nd and 3rd gen families - we will tell it which fams to keep

the_fams<-c("SCW46", "SCW47", "SCW48", "SCW49", "SCW50",
            "W128", "W129", "W130", "W131", "W132", "W133", "W134",
            "W143", "W144", "W148","W149", "W150")

freq_l_150_fams<-freq_150[is.element(freq_150$Family, the_fams), ]


#head(freq_l_150_fams)
str(freq_l_150_fams)
## 'data.frame':    68 obs. of  4 variables:
##  $ Var1  : Factor w/ 2 levels "l","p": 1 2 1 2 1 2 1 2 1 2 ...
##  $ Freq  : int  2 9 2 8 1 8 0 10 1 17 ...
##  $ Family: chr  "SCW46" "SCW46" "SCW46" "SCW46" ...
##  $ ENv   : chr  "20C16ppt" "20C16ppt" "20C32ppt" "20C32ppt" ...
length(unique(freq_l_150_fams$Family))
## [1] 17

#write a for loop to calulate the proportion
#create an empty vector to hold the values of prop_l

prop_l_150<-seq(1:length(freq_l_150_fams$Var1))

for (i in seq(1,length(freq_l_150_fams$Var1),2)){
  calc_prop<-freq_l_150_fams$Freq[i]/(freq_l_150_fams$Freq[i+1]+freq_l_150_fams$Freq[i])
  prop_l_150[i]<-calc_prop

}

#head(prop_l_150)
```

```
#bind it to the data frame with the family and env info

prop_l_150<-cbind(prop_l_150, freq_l_150_fams)

head(prop_l_150)
##   prop_l_150 Var1 Freq Family      ENv
## 1  0.1818182    l    2  SCW46 20C16ppt
## 2  2.0000000    p    9  SCW46 20C16ppt
## 3  0.2000000    l    2  SCW46 20C32ppt
## 4  4.0000000    p    8  SCW46 20C32ppt
## 5  0.1111111    l    1  SCW47 20C16ppt
## 6  6.0000000    p    8  SCW47 20C16ppt

#pull out just the lecithotrophic props(since those are the only ones that make sense)

prop_l_150_l<-prop_l_150[prop_l_150$Var1 == "l",]

names(colors)<-the_fams

coef(lm(subset(prop_l_150_l, ENv %in% "20C16ppt")[,1]~subset(prop_l_150_l, ENv %in% "20C32ppt")[,1]))
##                                    (Intercept)
##                                    0.009667246
## subset(prop_l_150_l, ENv %in% "20C32ppt")[, 1]
##                                    0.626780764
# (Intercept)
# 0.009200238
# subset(prop_l_150_l, ENv %in% "20C32ppt")[, 1]
# 0.628004209

#we need to sort the families to get colors to work properly

prop_l_150_l$Family<-sort(prop_l_150_l$Family)

#now plot that shit

fig2b<-ggplot()+
  geom_point(aes(y=subset(prop_l_150_l, ENv %in% "20C16ppt")[,1],x=subset(prop_l_150_l, ENv %in% "20C32p
             group=colors,  color=colors, size=4, alpha=0.65,
             position = position_jitter(width=0.04))+
  geom_abline(intercept = 0.009200238 , slope = 0.628004209 , size=2)+
  scale_color_manual(values=colors)+
  theme_bw() + theme(panel.grid.major = element_blank(),
                     panel.grid.minor = element_blank(),
                     axis.line = element_line(colour = "black"),
                     plot.title = element_text(size = 16, face = "bold"),
                     text = element_text(size = 16),
                     legend.position="right")+
  labs(x="Proportion lecithotrophic in high salinity",
       y="Proportion lecithotrophic in low salinity", title = "B")

#now that we've done that we can put them on a single plot

require(gridExtra)
```

```
## Loading required package: gridExtra
## Warning: package 'gridExtra' was built under R version 4.0.5
```

```
lay <- rbind(c(1,2),
             c(1,2))


grid.arrange(fig2a, fig2b, layout_matrix=lay)
```

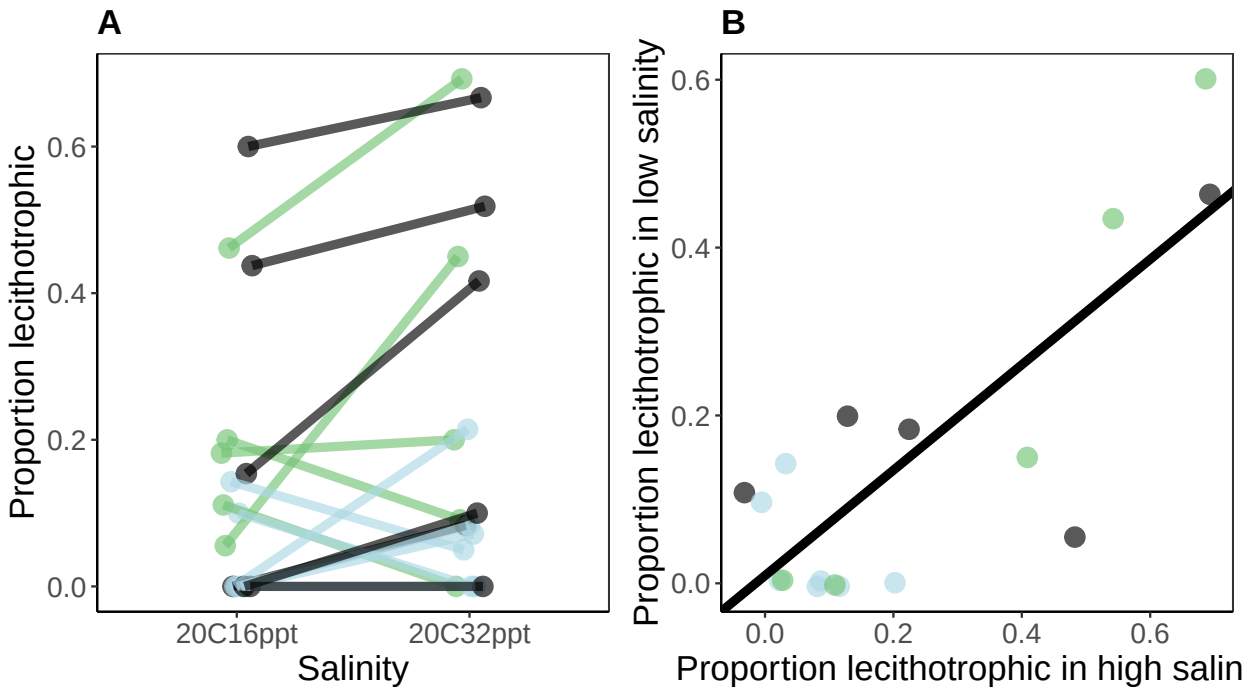

Figure 3. A) Family reaction norm for the proportion of lecithotrophy in low (16 ppt) and high (32 ppt) salinity and (B) the correlation between families for the proportion of lecithotrophy produced in low and high salinity (slope = 0.63, intercept = 0.001). Each line in A and each dot in B are a maternal family. Colors denote sampling sites: Tomales Bay, CA in green, Mill Valley, CA in light blue, and Long Beach, CA in black. Points and lines were jittered slightly to show all the points, as some families laid the same proportion of lecithotrophic egg-masses.

Figure 4

Shows the response to selection for increased lecithotrophy for slugs reared in low and high salinity

```
require(ggplot2)
require(gganimate)
## Loading required package: gganimate
## Warning: package 'gganimate' was built under R version 4.0.5
## No renderer backend detected. gganimate will default to writing frames to separate files
## Consider installing:
## - the `gifski` package for gif output
## - the `av` package for video output
## and restarting the R session
```

```r
#get the data from map_pop
map_pop<-read.csv("dataDryad/TomalesBay_capsuledata.csv", header=TRUE, sep=",")


#look at mean for individual egg masses

map_pop_eggmean<-aggregate(map_pop$Capsule, list(map_pop$date, map_pop$ID, map_pop$eggmass), mean)

colnames(map_pop_eggmean)<- c("date", "ID", "eggmass", "mean_cap")

map_pop_eggmean_tog<-merge(map_pop_eggmean, map_pop, by=c("date", "ID", "eggmass"))

#remove the F2 in low salinity data
map_pop_eggmean_test<-map_pop_eggmean_tog[!(map_pop_eggmean_tog$gen == "F2" & map_pop_eggmean_tog$origin

#set the factor levels for gen to get the plot to plot p first
map_pop_eggmean_tog$gen<-factor(map_pop_eggmean_tog$gen, levels=c("p", "F1", "F2", "F3"))

#change the salinity labels
levels(map_pop_eggmean_tog$original)
## NULL

levels(map_pop_eggmean_tog$original)<-c("Low Salinity 16 ppt", "High Salinity 32 ppt")

#remove F2 from the low salinity treatment
map_pop_eggmean_tog_fin<-map_pop_eggmean_tog[c(map_pop_eggmean_tog$gen == "F2" & map_pop_eggmean_tog$or


#change the levels of gen
levels(map_pop_eggmean_tog_fin$gen)<- c("p","S1","S2","S3")


#plot it
ggplot(data=subset(map_pop_eggmean_tog_fin, gen %in% c("p", "S1", "S2")))+
  #geom_boxplot(aes(y=mean_cap,x=gen, color=gen))+
  geom_histogram(position="identity", bins=40, aes(x=Capsule*1000, y=..density..,  group=gen, color=gen
  geom_density(aes(x=Capsule*1000, y=..density.., group=gen,  color=gen, fill=gen), alpha=0.1, size=1)+
  scale_color_manual(values=c("blue", "aquamarine3","gold"))+
  scale_fill_manual(values=c("blue", "aquamarine3", "gold"))+
  geom_vline(xintercept=150, linetype=2)+
  facet_wrap(~original)+
  theme_bw() + theme(panel.grid.major = element_blank(),
                     panel.grid.minor = element_blank(),
                     axis.line = element_line(colour = "black"),
                     plot.title = element_text(size = 16, face = "bold"),
                     text = element_text(size = 16))+
  guides(fill=guide_legend(title="Generation"),
         color=guide_legend(title="Generation"))+
  labs(y="Density", x=expression(paste("Egg capsule size (", mu,"m)")))
```

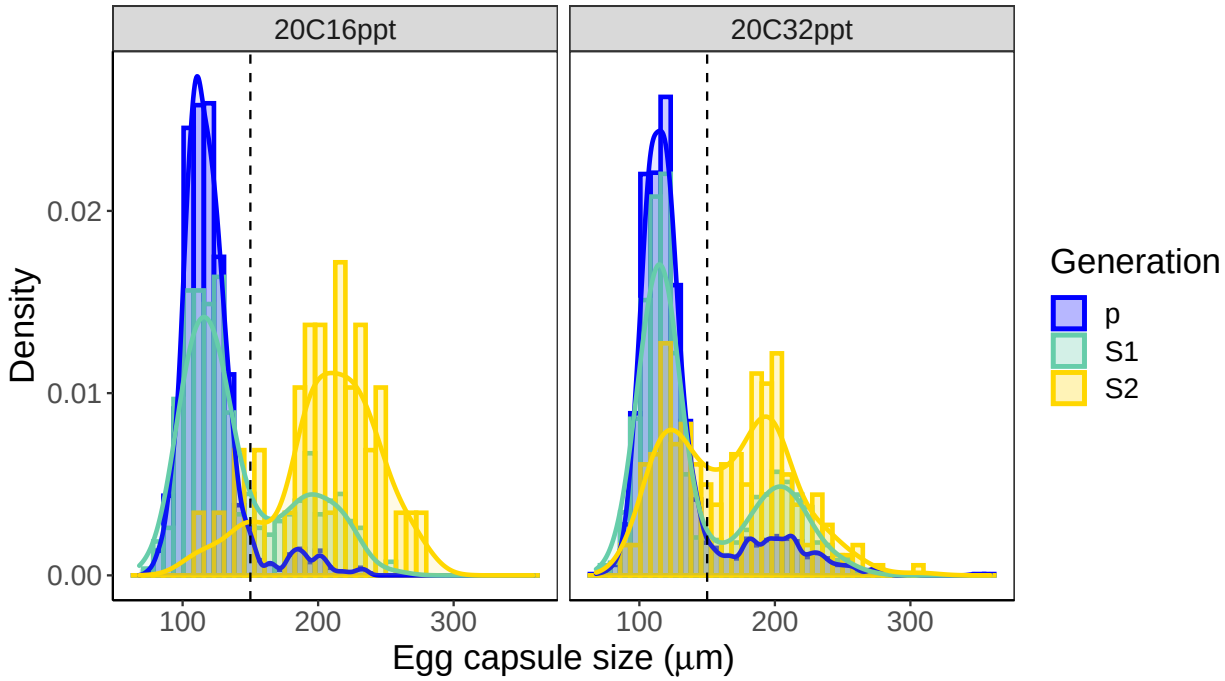

Figure 4. Barplot with density overlay showing the response to selection for lecithotrophy across several generations in low (16 ppt) and high (32 ppt) salinity from slugs collected from Tomales Bay (Figure 1B). The vertical dashed line indicates the cut-off for lecithotrophic or planktotrophic development (egg capsule size $>$ 150 µm). Generations S1 and S2 are 'selfed' (see Methods) while the parental generation is the product of outcrossing in the field. The S2 generation in low salinity is not shown due to small sample size (4 individuals from a single family, all of which laid lecithotrophic egg masses). The response to selection is shown for each generation on the plot where R1 refers to the response from parental to S1, and R2 is the response from S1 to S2.

Figure 5

```
## Figure 5 selection in high and low salinity

library(ggplot2)

#get the data

F1_sal<-read.table("dataDryad/F1_Capsule_data.csv", header=TRUE, sep=",")
SvN<-read.csv("dataDryad/SoCalvNorCal_Data.csv", header=TRUE, sep=",")


#make a plot for before and after selection by family
#data for before selection is in SvN
#head(SvN)
#add a column for before selection
SvN$Generation<-as.vector(strrep("Parental", 1))
#adjust the column names
colnames(SvN)<- c("date","order","Capsule","Family","ID","eggmass","Salinity","region","Generation")
```

```
#add a column to F1_sal for after selection
F1_sal$Generation<-ifelse(F1_sal$mom_env == "20C16ppt","low_selection","high_selection")

#rename off_id to ID and offspring_env to Salinity
colnames(F1_sal)<-c("date", "order", "Capsule", "mom_id", "mom_env", "ID", "eggmass", "Salinity", "Famil

#put svn and f1_sal together
fig_5<-rbind(SvN[,c("date", "Capsule", "ID", "Salinity", "Generation", "Family", "eggmass")],
             F1_sal[,c("date", "Capsule", "ID", "Salinity", "Generation", "Family", "eggmass")])

#by proportion

#get the mean egg capsule size by individual
fig5_ID_mean<-aggregate(fig_5$Capsule, list(fig_5$date, fig_5$ID, fig_5$eggmass), mean)
#head(fig5_ID_mean)
colnames(fig5_ID_mean)<-c("date", "ID", "eggmass", "mean_cap")

#now merge them together with the original data so we can get the Family etc info
fig5_ID_mean_tog<-merge(fig5_ID_mean, fig_5, by=c("ID"))
fig5_ID_mean_tog<-subset(fig5_ID_mean_tog, select = -c(Capsule))
fig5_mean<-unique.data.frame(fig5_ID_mean_tog)

#head(fig5_mean)


#now define L and P

fig5_mean$type_150 <- ifelse(fig5_mean$mean_cap > 0.150,"l","p")

#then define the proportions that we want using table, we must do this for before and after selection s
#before
fig5_data_sel_bs<-fig5_mean[fig5_mean$Generation == "Parental",]
freq_150_bs<-as.data.frame(table(fig5_data_sel_bs$type_150, as.factor(fig5_data_sel_bs$Family):as.facto

#after low
fig5_data_sel_low<-fig5_mean[fig5_mean$Generation == "low_selection",]
freq_150_low<-as.data.frame(table(fig5_data_sel_low$type_150, as.factor(fig5_data_sel_low$Family):as.fa

#after high
fig5_data_sel_high<-fig5_mean[fig5_mean$Generation == "high_selection",]
freq_150_high<-as.data.frame(table(fig5_data_sel_high$type_150, as.factor(fig5_data_sel_high$Family):as
#now stitch them back together

freq_150_bs$Generation<-as.vector(strrep("Parental", 1))
freq_150_low$Generation<-as.vector(strrep("S1 from Low Salinity", 1))
freq_150_high$Generation<-as.vector(strrep("S1 from High Salinity", 1))

freq_150<-rbind(freq_150_bs, freq_150_high, freq_150_low)

#we need to get it back to being a dataframe
freq_150$Var2<-as.vector(freq_150$Var2)
splity<-unlist(strsplit(freq_150$Var2, ":"))
```

```
Family<-splity[seq(1, length(splity), by=2)]
ENv<-splity[seq(2, length(splity), by=2)]

freq_150<-cbind(freq_150, Family, ENv)
freq_150<-within(freq_150, rm(Var2))
#head(freq_150)

#now we can set up a for loop to calculate the proportion

fig5_prop_l_150<-seq(1:length(freq_150$Var1))

for (i in seq(1,length(freq_150$Var1),2)){
  calc_prop<-freq_150$Freq[i]/(freq_150$Freq[i+1]+freq_150$Freq[i])
  fig5_prop_l_150[i]<-calc_prop

}
#this just returns the proportions, see
#head(fig5_prop_l_150)

fig5_prop_l_150<-cbind(fig5_prop_l_150, freq_150)

#head(fig5_prop_l_150)

#yay, remove the p because they don't make sense

fig5_prop_l_150<-fig5_prop_l_150[fig5_prop_l_150$Var1 == "l",]

#head(fig5_prop_l_150)#looks like several families did not have data for after selection,
                      #they are lowering the prop (having zero values)

#to make the plot pretty change env labels
levels(fig5_prop_l_150$ENv)[levels(fig5_prop_l_150$ENv) == "20C16ppt"]<-"Low 16 ppt"
levels(fig5_prop_l_150$ENv)[levels(fig5_prop_l_150$ENv) == "20C32ppt"]<-"High 32 ppt"


#removing families SCW46, SCW47, W143, W149
fig5_prop_l_150<-fig5_prop_l_150[fig5_prop_l_150$Freq != 0,]

#change the order of the legend (Generation)
fig5_prop_l_150$Generation <- factor(fig5_prop_l_150$Generation, levels = c("S1 from High Salinity", "S

#now plot it up

ggplot(data=fig5_prop_l_150, aes(x=ENv, y=fig5_prop_l_150, group=Generation, color=Generation))+
  stat_summary(fun=mean, geom="point", size=2)+
  stat_summary(fun=mean, geom="line", size=2)+
  stat_summary(fun.data = mean_se, geom = "errorbar", aes(width=0.1), size=0.85)+
  #facet_grid(~Generation)+
  scale_color_manual(values=c("black", "grey", "blue"))+
  theme_classic()+
  labs(x="Salinity", y="Proportion Lecithotrophic")
```

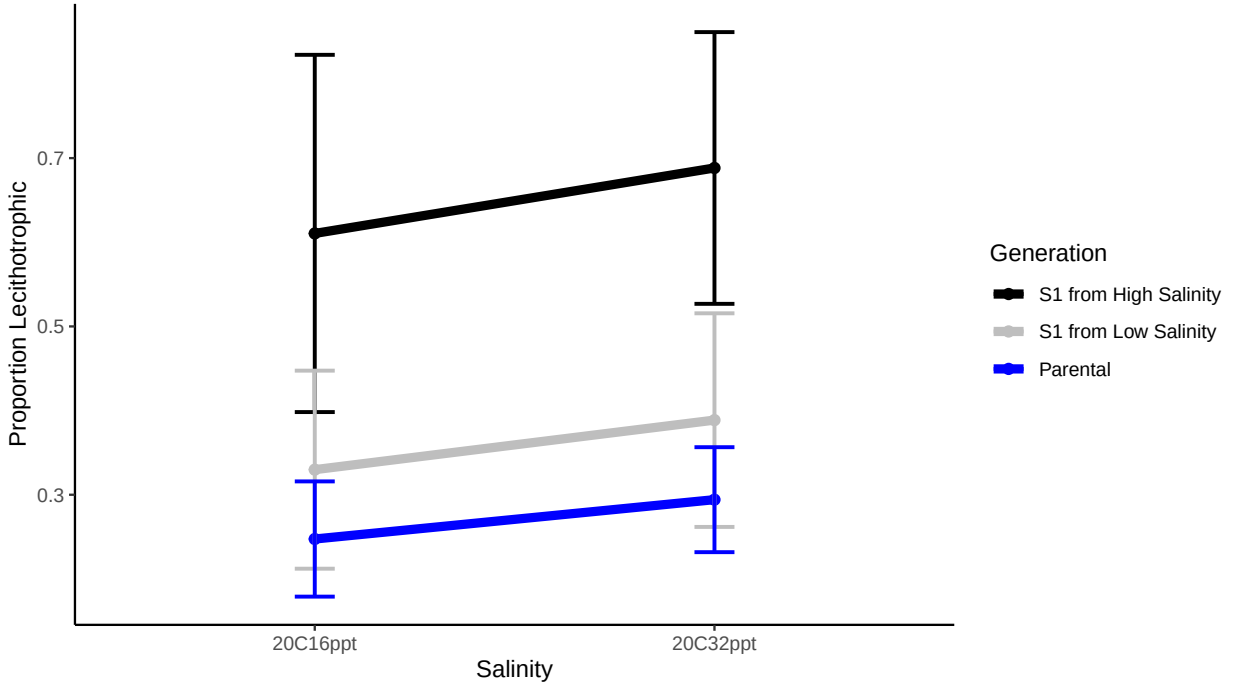

Figure 5. The reaction norm before and after one generation of selection for lecithotrophy in low and high salinity, from slugs collected from Long Beach and Mill Valley. The proportion lecithotrophic is the mean proportion between families and the error bars show standard deviation. The response to selection in high salinity was significant, but was not significant for low salinity, indicating a significant effect of the maternal environment (maternal effects).

Scatterplot from Figure 2

```
# Egg diameter figure for panel 1E
require(ggplot2)

#load the data

map_pop<-read.csv("dataDryad/TomalesBay_capsuledata.csv", header=TRUE, sep=",")
SvN<-read.csv("dataDryad/SoCalvNorCal_Data.csv", header=TRUE, sep=",")
egg_di<-read.csv("dataDryad/EggDiameterMeasurements.csv", header=TRUE, sep=",")

#make svn do the dates right
SvN$date<-as.Date(SvN$date, "%m/%d/%Y")
map_pop$date<-as.Date(map_pop$date, "%m/%d/%Y")

#rbind SuN and maping pops to make a master high versus low (hvl) data set

HvL<-rbind(SvN[,c("Capsule", "original", "date", "Family", "ID", "eggmass")],
           map_pop[, c("Capsule", "original", "date", "Family", "ID", "eggmass")])
HvL$date<-as.Date(HvL$date, "%m/%d/%Y")


#look at mean of egg di and mean of egg capsule
```

```r
#calculate the mean per egg mass for egg di
egg_di_mean<-aggregate(egg_di$EggDi, list(egg_di$date, egg_di$ID, egg_di$eggmass), mean)
colnames(egg_di_mean)<-c("date", "ID", "eggmass", "eggdi")
egg_di_mean$eggmass<-as.numeric(egg_di_mean$eggmass)
egg_di_mean$date<-as.Date(egg_di_mean$date, "%m/%d/%Y")

#again for hvl
HvL_mean<-aggregate(HvL$Capsule, list(HvL$date, HvL$ID, HvL$eggmass), mean)
colnames(HvL_mean)<-c("date", "ID", "eggmass", "Capsule")
HvL$eggmass<-as.numeric(HvL$eggmass)

#merge the two datasets
egg_di_map<-merge(egg_di_mean, HvL_mean, by=c("date", "ID", "eggmass"))
head(egg_di_map)
##         date               ID eggmass      eggdi   Capsule
## 1 2017-07-18 W128.1C320C32ppt       1 0.08100000 0.1172857
## 2 2017-07-18 W128.2A320C32ppt       1 0.11633333 0.1946667
## 3 2017-07-18 W128.2C320C16ppt       1 0.07133333 0.1171667
## 4 2017-07-18 W128.2C320C16ppt       2 0.07366667 0.1152000
## 5 2017-07-18 W131.3A120C16ppt       3 0.07500000 0.1212857
## 6 2017-07-18 W131.3A120C16ppt       4 0.08050000 0.1305714

#change to um
egg_di_map$egg_di_um<-egg_di_map$eggdi*1000
egg_di_map$egg_cp_um<-egg_di_map$Capsule*1000

#get some colors
eggcol<-ifelse(egg_di_map$Capsule > 0.15,"goldenrod", "blue")

#identify(egg_di_map$egg_di_um~egg_di_map$egg_cp_um)
#as ggplot


#some of these points are close together. Are these mixed egg masses? Let's see
#write a function to define eggs by the egg capsule diameter
map_pop_mlp<-aggregate(HvL$Capsule, list(HvL$date, HvL$eggmass, HvL$ID), function(x) ifelse(x > 0.15, "

colnames(map_pop_mlp)<-c("date", "eggmass", "ID", "x")
map_pop_mlp$is_l<-grepl("l", map_pop_mlp$x)
map_pop_mlp$is_p<-grepl("p", map_pop_mlp$x)
map_pop_mlp$is_m<- map_pop_mlp$is_l == T & map_pop_mlp$is_p == T
map_pop_mlp$eggmass_type<-ifelse(map_pop_mlp$is_m == T, "m", ifelse(map_pop_mlp$is_l == T, "l", "p"))
table(map_pop_mlp$eggmass_type)
##
##    l    m    p
##  518  152 1852

head(map_pop_mlp)
##         date eggmass              ID    x  is_l is_p  is_m eggmass_type
## 1 2017-10-24       1 SCW46A120C16ppt p, p FALSE TRUE FALSE           p
## 2 2017-10-07       1 SCW46A120C32ppt p, p FALSE TRUE FALSE           p
## 3 2017-10-07       2 SCW46A120C32ppt p, p FALSE TRUE FALSE           p
## 4 2017-10-07       1 SCW46A220C16ppt p, l  TRUE TRUE  TRUE           m
```

```
## 5 2017-10-10       1 SCW46A220C16ppt p, p FALSE TRUE FALSE            p
## 6 2017-10-13       1 SCW46A220C16ppt p, p FALSE TRUE FALSE            p

egg_id<-merge(map_pop_mlp, egg_di_map, by=c("date", "ID", "eggmass"))

#pull out salinity from ID

egg_id$salinity<-ifelse(grepl("20C16ppt",egg_id$ID),"low", "high")


ggplot()+
  geom_point(data=egg_id, aes(x=egg_cp_um, y=egg_di_um, col=eggmass_type, shape=salinity), cex=3)+
  #facet_wrap(~salinity)+
  scale_color_manual(breaks=c("l", "m",  "p"),
                     labels=c("Lecithotrophic", "Mixed", "Planktotrophic"),
                     name="Egg-mass type",
                     values=c("goldenrod", "aquamarine3",  "blue"))+
  scale_shape_manual(values=c(16, 21), name="Salinity")+
  labs(x=expression(paste("Mean Egg Capsule Diameter (", mu, "m)")),
       y=expression(paste("Mean Egg Diameter (", mu,"m)")))+
  theme_classic()+theme(legend.position = "right",
                        axis.text=element_text(size=12),
                        axis.title=element_text(size=14,face="bold"),
                        legend.title = element_text(size=12),
                        legend.text=element_text(size=12))+
  geom_vline(xintercept = 150, lty=2)
```