

# Selection pressures in *Syngnathus floridae*

Coley Tosto

2023-11-09

```
#This is a cohesive list of all the libraries used in this document
```

```
#MomIDs and embryo counts for each section of the male's brood pouch
em_dat <- read.csv("data/EmbryoParentage.csv")
```

```
#Metadata for males and females from the mesocosm experiments
fem_meso <- read.csv("data/all_fem_meso_floridae.csv")
mal_meso <- read.csv("data/all_mal_meso_floridae.csv")
```

## Calculating mating and reproductive success for individuals who mated

*Syngnathus floridae* (dusky pipefish) were sampled from three distinct seagrass beds around Tampa Bay in Tampa, Florida. Sexually mature females (standard length  $\geq 120$ mm) and pregnant males were collected and brought back to the University of Tampa for mesocosm experiments. In these mesocosms, 8 males and 8 females were housed together in a 140L tank for a period of 14-days and allowed to mate freely. Parentage analysis was done with all of the pregnant males from the trials to figure out how many times each male and female mated, and the number of eggs that were transferred. The results of that are here.

First I had to calculate the mating and reproductive success for each male and female who mated based on the assigned mom for each genotyped embryo.

```
#Row-by-Row analysis of parentage data by male brood pouch section
```

```
#Read in the data
```

```
#em_dat <- read.csv("~/EmbryoParentage.csv")
```

```
#For each row in the dataset(each section of the pouch) apply this function
```

```
mom_counts <- do.call(rbind,apply(em_dat, 1, function(one_section){
```

```
  #Save all of the momIDs into an object
```

```
  mom_ids<-c(one_section[grep("momID",names(one_section))])
```

```
  #Calculate the number of eggs that belongs to each potential mom based on
```

```
  #the proportions and total number of developed and undeveloped embryos
```

```
  mom_props<-c(as.numeric(one_section[grep("prop",names(one_section))]))
```

```
  mom_counts_dev<-mom_props*as.numeric(one_section["num_embryos_dev"])
```

```
  mom_counts_und<-mom_props*as.numeric(one_section["num_embryos_non_dev"])
```

```
  #Create a dataframe that contains the maleID, pouch section number and the
```

```

#number of eggs that belongs to each momID
this_section<-data.frame(
  maleID=one_section["maleID"],
  section_num=one_section["section_num"],
  mom_ids[which((mom_counts_dev + mom_counts_und) > 0)],
  mom_counts_dev[which((mom_counts_dev + mom_counts_und)>0)],
  mom_counts_und[which((mom_counts_dev + mom_counts_und)>0)]
)

#Rename the columns
colnames(this_section)[3:5]<-c("momID", "num_dev", "num_und")

return(this_section)
}))

#Calculate female fitness
fem_fitness<-do.call(rbind,by(mom_counts, mom_counts$momID,function(dat){

  mom_fitness<-data.frame(
    momID=unique(dat$momID),
    MatingSuccess=length(unique(dat$maleID)),
    NumDeveloped=round(sum(dat$num_dev)),
    NumUndeveloped=round(sum(dat$num_und))
  )
  return(mom_fitness)
}))

fem_fitness$totalEggs <- fem_fitness$NumDeveloped + fem_fitness$NumUndeveloped

#Calculate Male Fitness
mal_fitness<-do.call(rbind,by(mom_counts, mom_counts$maleID,function(dat){

  dad_fitness<-data.frame(
    maleID=unique(dat$maleID),
    MatingSuccess=length(unique(dat$momID)),
    NumDeveloped_Calc=round(sum(dat$num_dev)),
    NumUndeveloped_Calc=round(sum(dat$num_und))
  )
  return(dad_fitness)
}))

mal_fitness$totalEggs <- mal_fitness$NumDeveloped_Calc + mal_fitness$NumUndeveloped_Calc

```

After running the above R script we have two datasets, `mal_fitness` and `fem_fitness`. These datasets include information about the mating success (number of mates) and reproductive success (Number of embryos transferred). We can split reproductive success up further later if we want to from the total number of embryos transferred to the number of embryos developed and the number that were undeveloped.

I want to include all of the other metadata that I have for these individuals (traits, collection location, latency to pregnancy, etc.) as well as tack on all of the information for the individuals who did not mate. To do that I am going to need to merge the fitness datasets with `fem_meso` and `mal_meso`.

```

#Make a column in *_meso that contains the full fishID (i.e. FL1M3) to match the
#formatting in the fitness datasets (make sure they have the same name for merging purposes)
fem_meso$momID <- paste0("FL", fem_meso$trial_num, "F", fem_meso$fishID)
mal_meso$maleID <- paste0("FL", mal_meso$trial_num, "M", mal_meso$fishID)

#Merge the datasets based on the columns created above
fem_all <- merge(fem_meso, fem_fitness, by = "momID", all.x = TRUE, all.y = TRUE)
mal_all <- merge(mal_meso, mal_fitness, by = "maleID", all.x = TRUE, all.y = TRUE)

```

There are a few trials that I want to remove from the analysis:

1. All trials where there were no successful matings (7, 9, 10, 11).
2. Trial 1, a male gave birth and the babies were immediately eaten by the adults so the trial was ended early and therefore I was unable to get any parentage information for that trial.

I also want to replace the NAs that were automatically added to the columns from the fitness dataset (MatingSuccess, NumDeveloped, NumUndeveloped, totalEggs) with 0s and add a column to the female dataset that tells me whether or not the female mated (with 1 or 0).

```

#Subset the merged datasets to remove trials without successful matings and Trial 1
fem_succ <- subset(fem_all, !(trial_num %in% c(7, 9, 10, 11, 1)))
mal_succ <- subset(mal_all, !(trial_num %in% c(7, 9, 10, 11, 1)))

#Replace NAs with 0s in the columns related to fitness
mal_succ[,15:18] <- sapply(mal_succ[,15:18],
                           function(x)
                             ifelse(is.na(x), 0, x))

fem_succ[,11:14] <- sapply(fem_succ[,11:14],
                           function(x)
                             ifelse(is.na(x), 0, x))

#Add a column for females to denote mated or unmated
fem_succ$mated <- ifelse(fem_succ$MatingSuccess > 0, paste0(1), paste0(0))

```

## Summary statistics for successfully mated individuals

### Males

Across all 7 trials and 56 total males, there were 24 males that mated at least one time and 6 of those males had two mates.

Looking across all males, including the ones that did not mate, this is what we find as the mean, sd, and se for the number of embryos transferred and how many of those developed versus didn't:

	mean	SD	SE	max	min
Number of Embryos	108.5	155.1902059	20.7381636	500	0
Developed Embryos	94.1964286	139.7347383	18.6728398	463	0
Undeveloped Embryos	14.3035714	27.7907242	3.7136917	131	0

These values will be influenced by the number of 0s coming from males who did not mate. So let's look at the same thing, but this time for only males who had at least one successful mating:

	mean	SD	SE	max	min
Number of Embryos	253.1666667	139.1941611	28.4128892	500	9
Developed Embryos	219.7916667	133.7427704	27.3001287	463	7
Undeveloped Embryos	33.375	34.3901054	7.0198509	131	0

We can see from the bottom table that even when we only include males who mated there is still a wide range in the brood size. I want to see what relationship there is between brood pouch size (in terms of both total area and length) and brood size (total number of embryos).

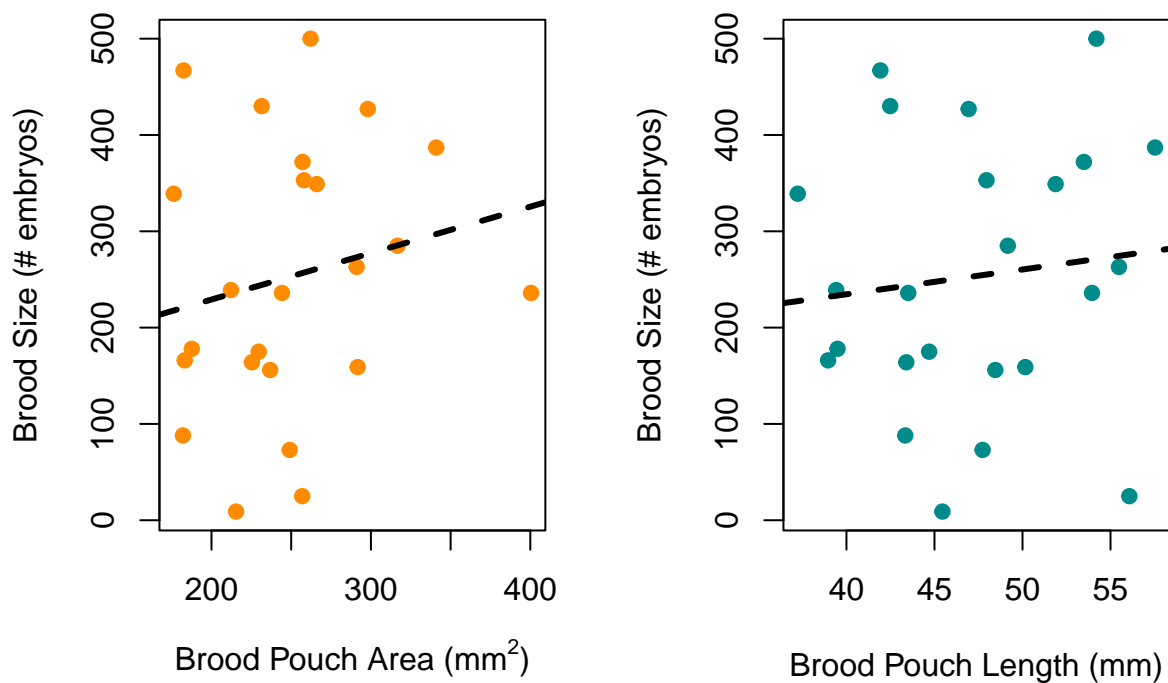


Figure 1: Scatterplot of the relationship between brood pouch size metrics and the number of embryos a male had.

There may be some correlation happening here, but it doesn't look particularly strong (Fig. 1). Let's run some correlations tests to see what they say.

```
##
## Pearson's product-moment correlation
##
## data: as.numeric(mated_mal$bp_area) and mated_mal$totalEggs
## t = 0.90485, df = 22, p-value = 0.3753
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
## -0.2316769 0.5507364
## sample estimates:
##      cor
## 0.1894228

##
## Pearson's product-moment correlation
##
## data: as.numeric(mated_mal$bp_length) and mated_mal$totalEggs
## t = 0.51929, df = 22, p-value = 0.6087
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.3069838 0.4916139
## sample estimates:
##      cor
## 0.1100398
```

There is not a significant correlation between the number of eggs and size of the brood pouch when we look at brood pouch area OR brood pouch length.

## Females

Across all 7 trials and 56 total females, there were 21 females that mated at least one time, 5 females that mated twice, and 2 that mated 3 times.

Looking across all females, including the ones that did not mate, this is what we find as the mean, sd, and se for the total number of embryos transferred from each female (across all of her mates if applicable) and how many of those developed versus didn't:

	mean	SD	SE	max	min
Number of Embryos	108.4821429	195.2328206	26.089083	916	0
Developed Embryos	94.1785714	177.0691808	23.6618646	785	0
Undeveloped Embryos	14.3035714	29.7033329	3.9692748	131	0

These values will be influenced by the number of 0s coming from females who did not mate. So let's look at the same thing, but this time for only females who had at least one successful mating:

	mean	SD	SE	max	min
Number of Embryos	289.2857143	223.3819919	48.745947	916	35
Developed Embryos	251.1428571	211.7324457	46.2038076	785	7
Undeveloped Embryos	38.1428571	38.360508	8.3709491	131	0

We can see from the bottom table that even when we only include females who mated there is still a wide range in the number of eggs transferred. I want to see what relationship there may be between female body size (in terms of standard length, depth, and SVL) and the number of eggs she transferred.

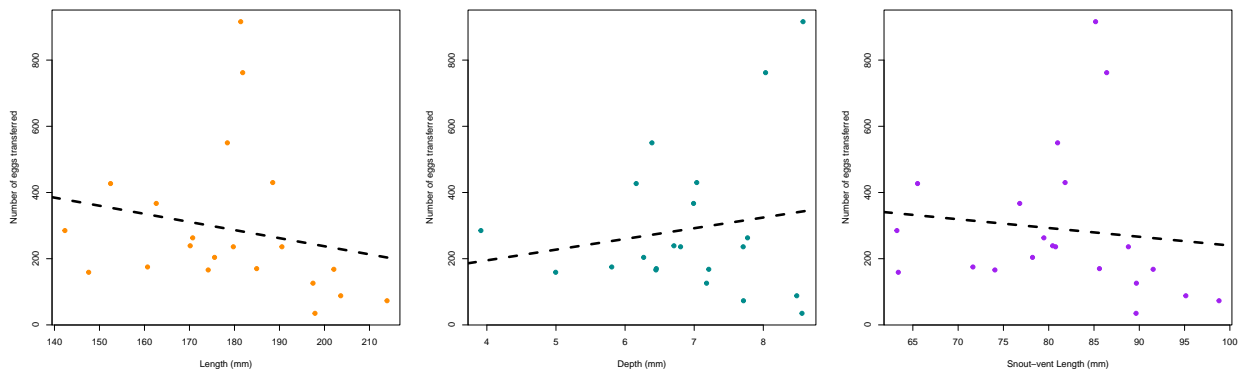


Figure 2: Scatterplot of the relationship between female size metrics and the number of eggs transferred.

There may be some correlation happening here, but it doesn't look particularly strong (Fig. 2). Let's run some correlations tests to see what they say.

```
##
## Pearson's product-moment correlation
##
## data: mated_fem$length and as.numeric(mated_fem$totalEggs)
## t = -0.92304, df = 19, p-value = 0.3676
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.5864091 0.2465727
## sample estimates:
##      cor
## -0.2071652

##
## Pearson's product-moment correlation
##
## data: mated_fem$depth and as.numeric(mated_fem$totalEggs)
## t = 0.74454, df = 19, p-value = 0.4657
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.2839543 0.5593989
## sample estimates:
##      cor
## 0.1683713

##
## Pearson's product-moment correlation
##
## data: mated_fem$svl and as.numeric(mated_fem$totalEggs)
## t = -0.5112, df = 19, p-value = 0.6151
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.5219227 0.3318958
## sample estimates:
##      cor
## -0.1164796
```

There is not a significant correlation between the number of eggs and size of the female in terms of standard length, depth, or snout-vent length. Interestingly, however, there is a negative correlation for length and SVL and a positive correlation for depth (but they are all overall weak).

## Differences between mated individuals and unmated individuals

I want to now see if there are any significant differences in the sizes of individuals who mated vs individuals that didn't mate in males and females.

### Males

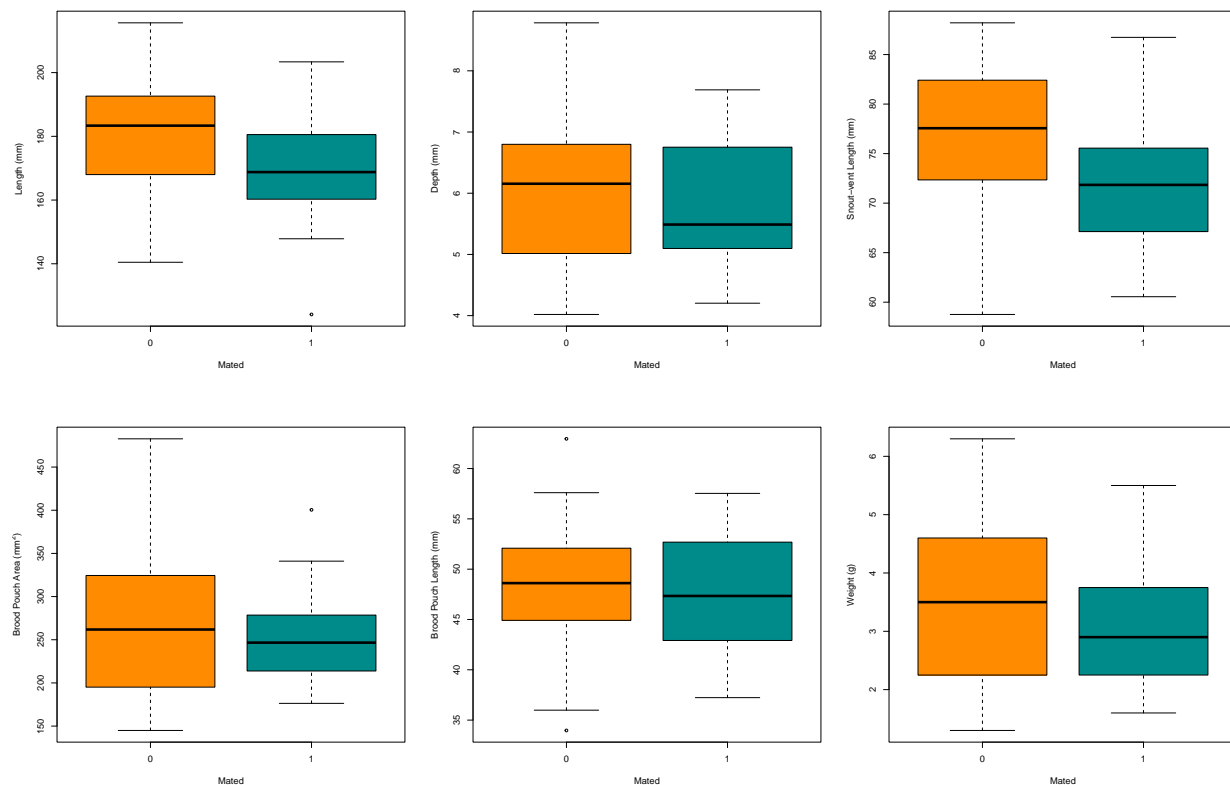


Figure 3: Six different morphometrics compared between males who successfully mated versus those that didn't. Orange represents unmated and blue represents mated males.

```
##
## Wilcoxon rank sum exact test
##
## data: mal_succ$length by mal_succ$preg_status
## W = 503, p-value = 0.02593
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum test with continuity correction
```

```

##
## data:  mal_succ$depth by mal_succ$preg_status
## W = 413, p-value = 0.4919
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum exact test
##
## data:  mal_succ$svl by mal_succ$preg_status
## W = 524, p-value = 0.009344
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum exact test
##
## data:  as.numeric(mal_succ$bp_area) by mal_succ$preg_status
## W = 382, p-value = 0.5525
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum exact test
##
## data:  mal_succ$bp_length by mal_succ$preg_status
## W = 387, p-value = 0.4947
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum test with continuity correction
##
## data:  mal_succ$weight by mal_succ$preg_status
## W = 448.5, p-value = 0.289
## alternative hypothesis: true location shift is not equal to 0

```



## Females

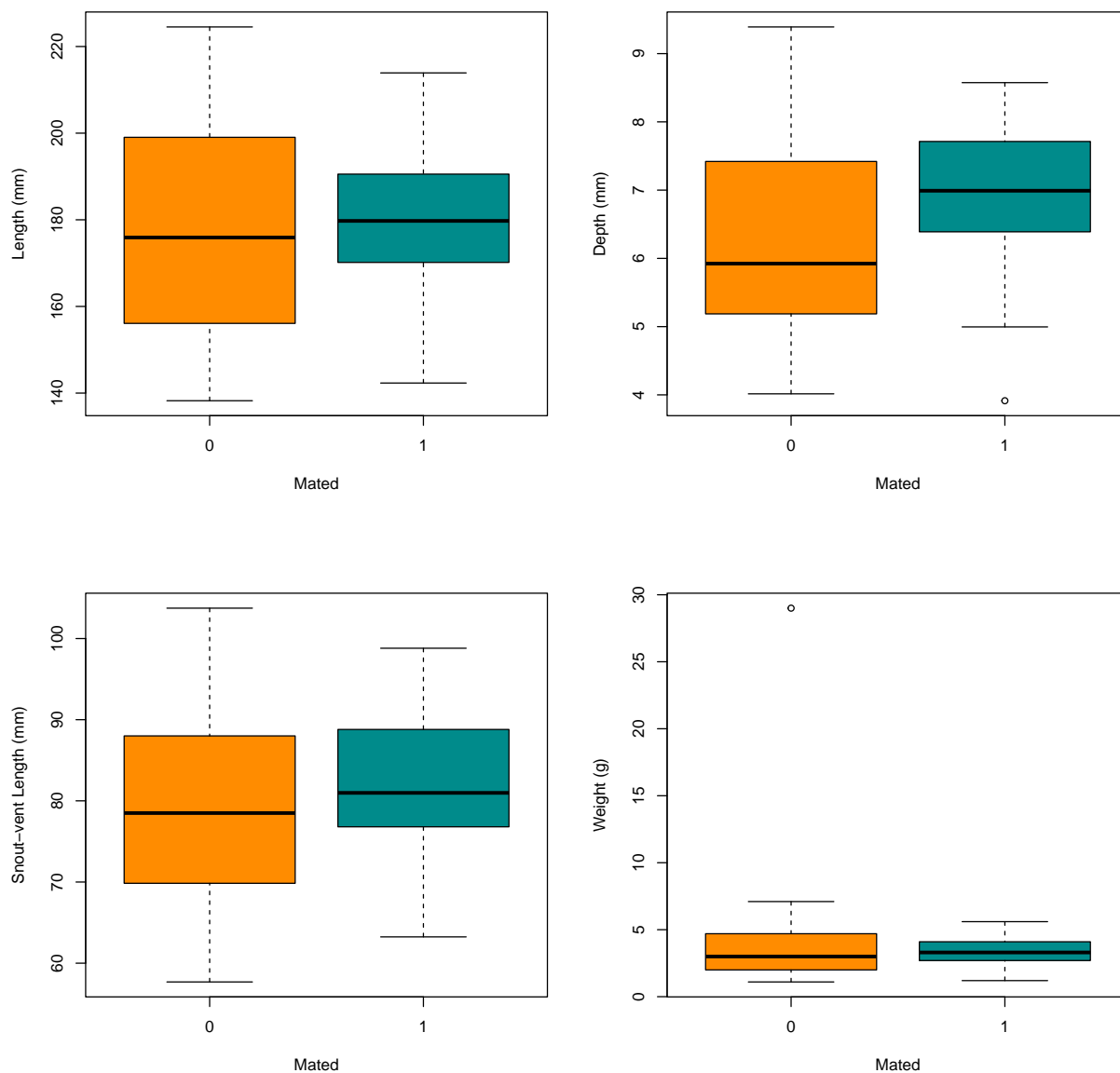


Figure 4: Four different morphometrics compared between females who sucessfully mated versus those that didn't. Orange represents unmated and blue represents mated females.

```
##
## Wilcoxon rank sum exact test
##
## data: fem_succ$length by fem_succ$mated
## W = 353, p-value = 0.8143
## alternative hypothesis: true location shift is not equal to 0
##
```

```
## Wilcoxon rank sum exact test
##
## data: fem_succ$depth by fem_succ$mated
## W = 261, p-value = 0.07263
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum exact test
##
## data: fem_succ$svl by fem_succ$mated
## W = 325, p-value = 0.4805
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum test with continuity correction
##
## data: fem_succ$weight by fem_succ$mated
## W = 338.5, p-value = 0.6292
## alternative hypothesis: true location shift is not equal to 0
```

## Mate success versus Reproductive success

I now want to look at any relationship that may exist between mating success and reproductive success for males and females. The Bateman gradient will be calculated, which is the slope of the weighted least-squares regression of relative reproductive success (number of offspring divided by the mean) on mating success.

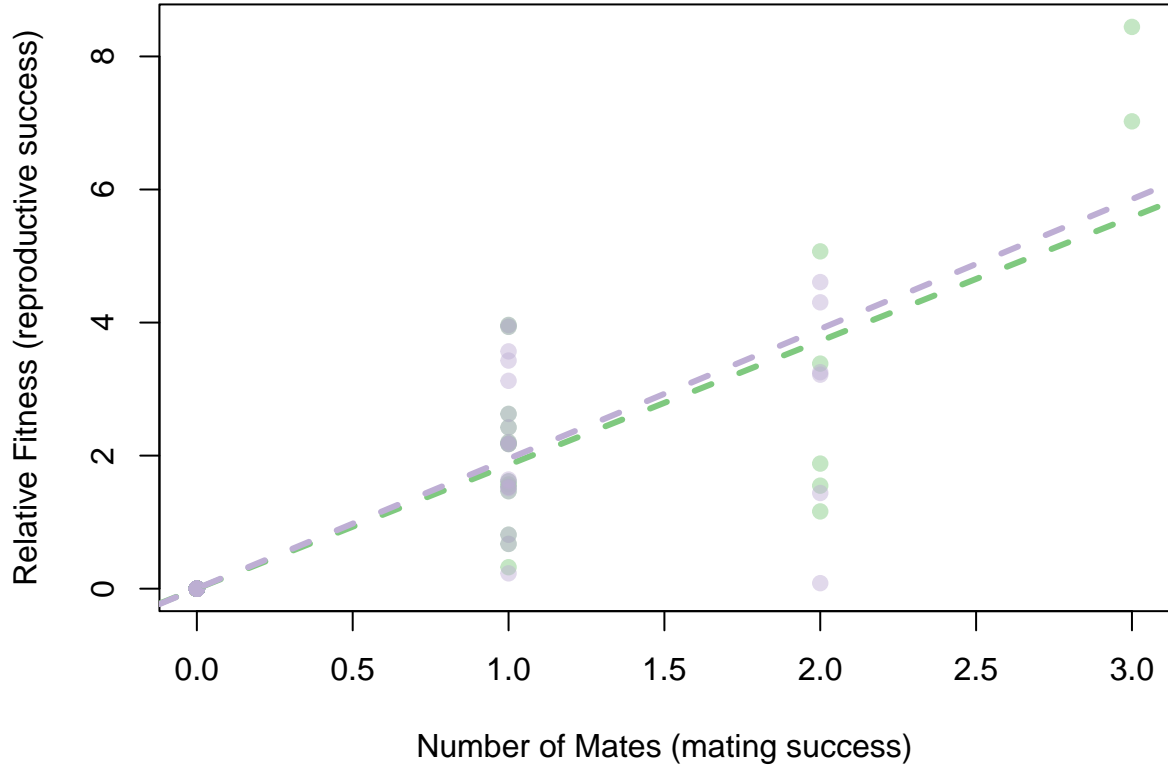


Figure 5: Relationship between reproductive success and mating success for male (purple) and female (green) *Syngnathus floridae*. Reproductive success is shown as relative fitness (i.e. number of offspring produced divided by the mean number of offspring produced). Bateman's gradient is shown as the weighted least-squares regression line (dashed) for males and females.