

*Note: I did all of this in R Markdown but was receiving a number of errors when I tried to Knit to PDF. It seemed like the errors were going to take too long to resolve, so I copied everything over to a document and exported to PDF. Everything is still in the same order as it would appear in R Markdown.

Title: Estimating Dispersal Distance in Maroon Clownfish (*Premnas biaculeatus*)

Author: Kyra Fitz

Introduction:

Obtaining fine-scale estimates of dispersal for a variety of marine species is key to understanding local adaptation, population trends, and to implement conservation actions. Genetic isolation by distance patterns have been observed in many marine species and can be a robust tool for understanding marine dispersal (Selkoe and Toonen 2011). Isolation-by-distance theory describes how with the balance of genetic drift and migration in populations, there will be a positive linear relationship between pairwise geographic distance and pairwise genetic difference between populations or individuals (Wright 1943, Rousset 1997). According to the theory, as the geographic distance between a pair of populations increases, so does the genetic distance between them. The slope of the linear regression between genetic and geographic distance along with the effective density of the population can then be plugged into equations derived by Francois Rousset to calculate dispersal spread (Rousset 1997). Rousset shows that in a 1-dimensional habitat (a habitat that is longer than it is wide), $\sigma = \sqrt{1/4De \cdot m}$ where σ is the spread (standard deviation) of the dispersal kernel, De is the effective population density, and m is the slope of the linear regression between genetic and geographic distance. The dispersal kernel is a probability distribution where height is the probability of a larvae dispersing to a certain distance away from its parents.

Many marine species show isolation-by-distance patterns, but marine larval dispersal may also be impacted by ocean currents, meaning that geographic distance may not be the only factor of interest. In this analysis, I used genotypes of 8 populations of a coral reef fish (*Premnas biaculeatus*) from the islands of Cebu, Bohol, and Leyte in the Philippines at 16 microsatellite loci to generate fine-scale estimates of dispersal. I found that 7 out of the 8 populations follow isolation-by-distance patterns and generated estimates for σ , the spread of the dispersal kernel. For the one population that did not follow isolation-by-distance patterns, I began to analyze if ocean currents may be impacting the dispersal of this population.

Methods and Results:

I used genotypes to generate pairwise F_{st} matrices using the Genepop package. I linearized these F_{st} matrices ($F_{st}/1-F_{st}$) according to Rousset's derivation of an equation to calculate dispersal spread in a 1-dimensional habitat (Rousset 1997). A 1-dimensional habitat is one in which the length of the habitat is greater than the width, and this fits coral reefs well. I then used the geodist package to calculate pairwise geographic distances between populations. I am starting this analysis with these matrices already created. I excluded 4 populations that had

sample sizes less than 5 individuals. Then, I plotted genetic vs. geographic distance in the 8 remaining populations and ran a Mantel test and linear regression.

I was interested in evaluating if there is a significant correlation between genetic and geographic distance, and I used a Mantel test because it can evaluate the significance in the correlation between two matrices that contain pairwise distances. Mantel tests evaluate correlation between entries in the two distance matrices and then use permutations to calculate a p-value (Mantel 1967). A significant p-value is interpreted as there being a correlation between the two distance matrices. The permutation step avoids the issue of pseudo-replication of samples in the distance matrices by essentially randomizing the entries in the first matrix. Mantel tests are frequently used in studies to determine if there is an isolation-by-distance pattern (Selkoe and Toonen 2011). A significant Mantel test run on a matrix of geographic and a matrix of genetic distances is considered confirmation of a isolation-by-distance pattern.

```
##{r}
library(vegan) #package that contains Mantel test function

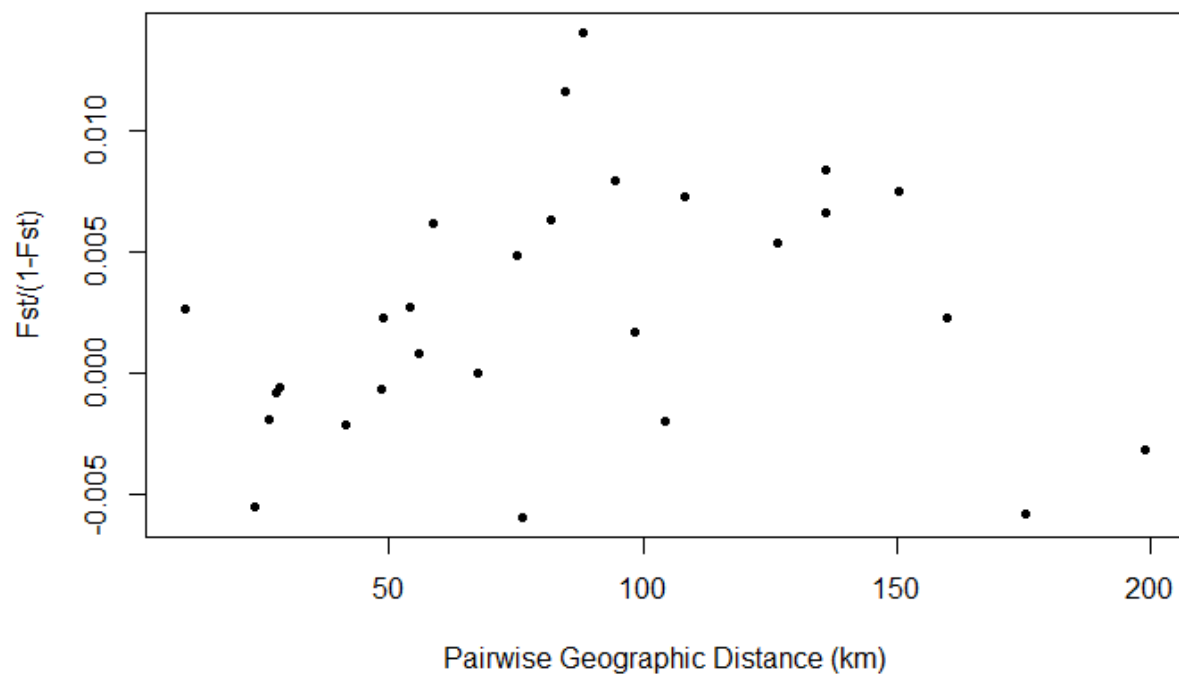
#Initial plot
plot(prbi_EW_geodist_km, prbi_11_12_EW_fstlin,
     xlab="Pairwise Geographic Distance (km)", ylab="Fst/(1-Fst)", pch=20)

#Mantel test
#Using 999 permutations
mantel(prbi_EW_geodist_km, prbi_11_12_EW_fstlin)
#r: 0.1604 p: 0.235

#Linear regression
y19 <- as.numeric(prbi_11_12_EW_fstlin)
x19 <- as.numeric(prbi_EW_geodist_km)

mod_EW <- lm(y19~x19)
summary(mod_EW)

#Plot with regression line
plot(prbi_EW_geodist_km, prbi_11_12_EW_fstlin,
     xlab="Pairwise Geographic Distance (km)", ylab="Fst/(1-Fst)", pch=20)
abline(mod_EW, col="red")
```



Mantel statistic based on Pearson's product-moment correlation

Call:

```
mantel(xdis = prbi_EW_geodist_km, ydis = prbi_11_12_EW_fstlin)
```

Mantel statistic r: 0.1604

Significance: 0.235

Upper quantiles of permutations (null model):

	90%	95%	97.5%	99%
Quantiles	0.281	0.352	0.409	0.513

Permutation: free

Number of permutations: 999

Call:

```
lm(formula = y19 ~ x19)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.0098069	-0.0027813	-0.0003314	0.0038662	0.0115083

Coefficients:

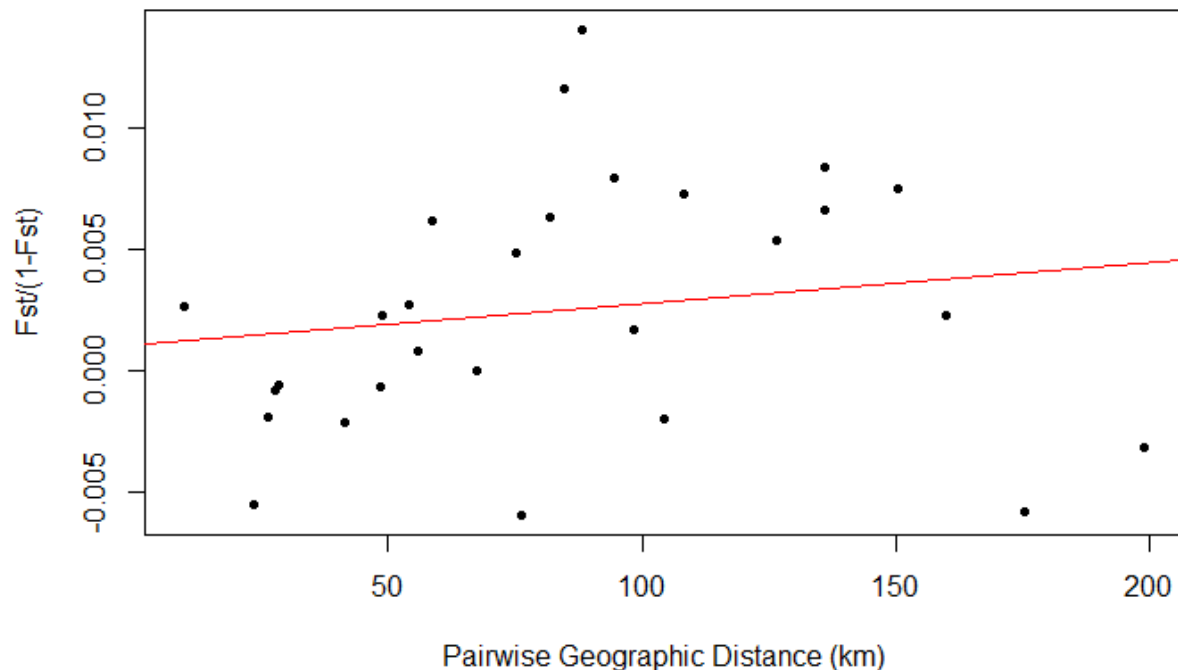
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.062e-03	1.989e-03	0.534	0.598
x19	1.682e-05	2.030e-05	0.829	0.415

Residual standard error: 0.005215 on 26 degrees of freedom

(36 observations deleted due to missingness)

Multiple R-squared: 0.02574, Adjusted R-squared: -0.01174

F-statistic: 0.6868 on 1 and 26 DF, p-value: 0.4148



I found that with all 8 populations included, the Mantel test and linear regression both returned non-significant p-values. I investigated further to find that points on the plot corresponding to one population (labeled as population 19 from the initial assignment of populations from genotypes/allele frequencies) were very spread out and did not seem to be following the same pattern as the other populations. I made the following plot where population 19 is represented in red dots, while the rest of the populations are represented in black dots to visualize this observation:

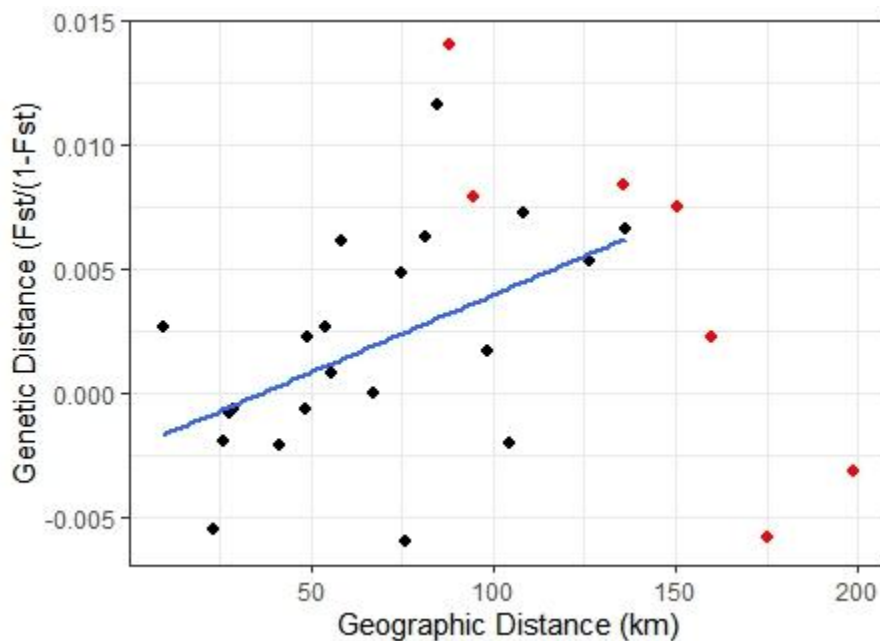
```

{r, echo=FALSE}
library(dplyr)
library(tidyr)
library(ggplot2)

pop19 <- tibble(GeographicDistance = as.vector(prbi_EW_geodist_km[8,]),
                GeneticDistance = as.vector(prbi_11_12_EW_fstlin[8,]))
pop19 <- drop_na(pop19)

ggplot(data=augmented_mod_distanceframe, aes(x=distanceframe$GeographicDistance,
                                              y=distanceframe$GeneticDistance)) +
  geom_point(size=1.75) +
  geom_point(data=pop19, aes(x=GeographicDistance, y=GeneticDistance), color="red", size=1.75) +
  geom_smooth(method="lm", se=FALSE) +
  theme_bw() +
  xlab("Geographic Distance (km)") +
  ylab("Genetic Distance (Fst/(1-Fst))")

```



I then excluded population 19 from the genetic and geographic distance matrices and plotted the data, ran a Mantel test, and ran a linear regression.

```

...{r}
#Taking population 19 out of the matrix
prbi_11_12_EW_no19_fstlin <- prbi_11_12_EW_fstlin[-8, -8]

prbi_EW_no19_geodist_km <- prbi_EW_geodist_km[-8, -8]

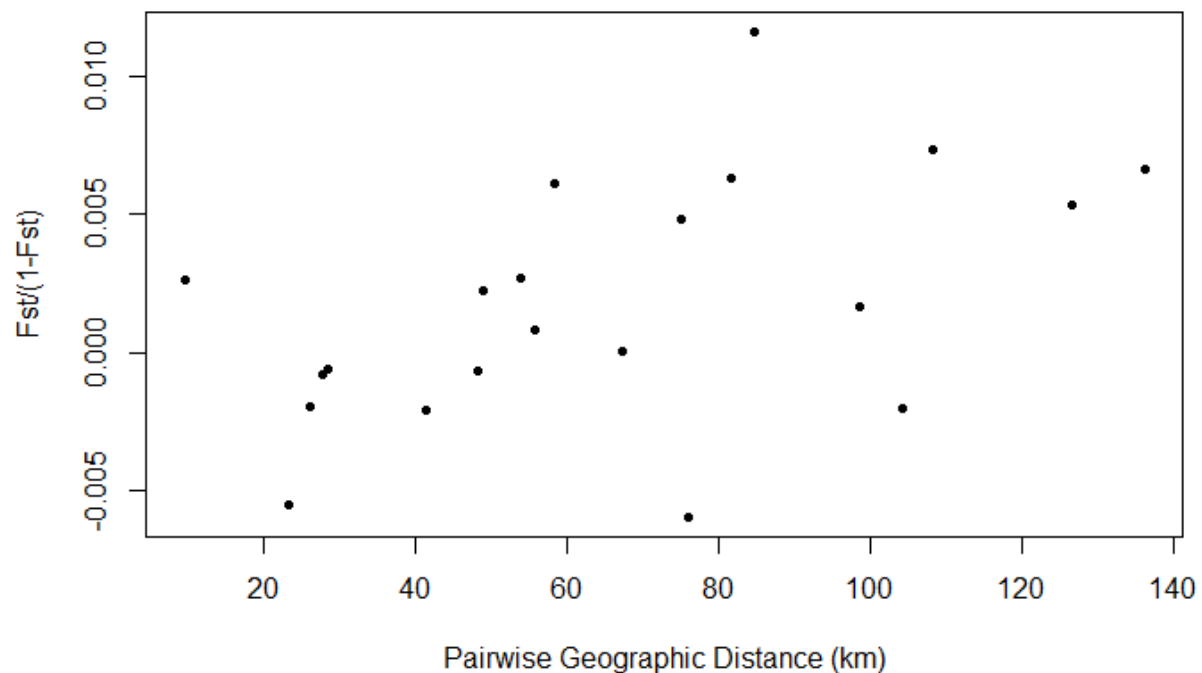
#Plotting data
plot(prbi_EW_no19_geodist_km, prbi_11_12_EW_no19_fstlin,
      xlab="Pairwise Geographic Distance (km)", ylab="Fst/(1-Fst)", pch=20)

#Run linear regression
y <- as.numeric(prbi_11_12_EW_no19_fstlin)
x <- as.numeric(prbi_EW_no19_geodist_km)
mod_EW_no19 <- lm(y~x)
summary(mod_EW_no19)

#Add linear regression to plot
plot(prbi_EW_no19_geodist_km, prbi_11_12_EW_no19_fstlin,
      xlab="Pairwise Geographic Distance (km)", ylab="Fst/(1-Fst)", pch=20)
abline(mod_EW_no19, col="red")

#Mantel test
mantel(prbi_EW_no19_geodist_km, prbi_11_12_EW_no19_fstlin)
...

```



Call:

```
lm(formula = y ~ x)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.0084305	-0.0019328	-0.0002583	0.0024181	0.0085919

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.272e-03	1.866e-03	-1.218	0.2383
x	6.253e-05	2.518e-05	2.484	0.0225 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.003948 on 19 degrees of freedom
(28 observations deleted due to missingness)

Multiple R-squared: 0.2451, Adjusted R-squared: 0.2054

F-statistic: 6.168 on 1 and 19 DF, p-value: 0.0225

Mantel statistic based on Pearson's product-moment correlation

Call:

```
mantel(xdis = prbi_EW_no19_geodist_km, ydis = prbi_11_12_EW_no19_fstlin)
```

Mantel statistic r: 0.4951

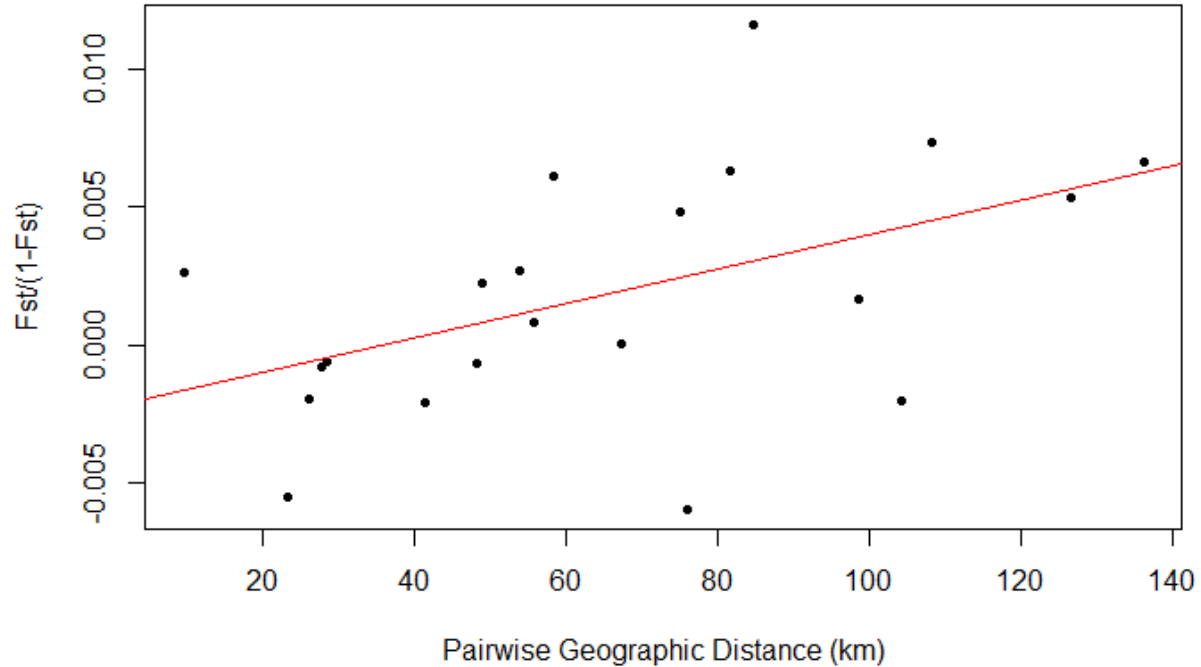
Significance: 0.015

Upper quantiles of permutations (null model):

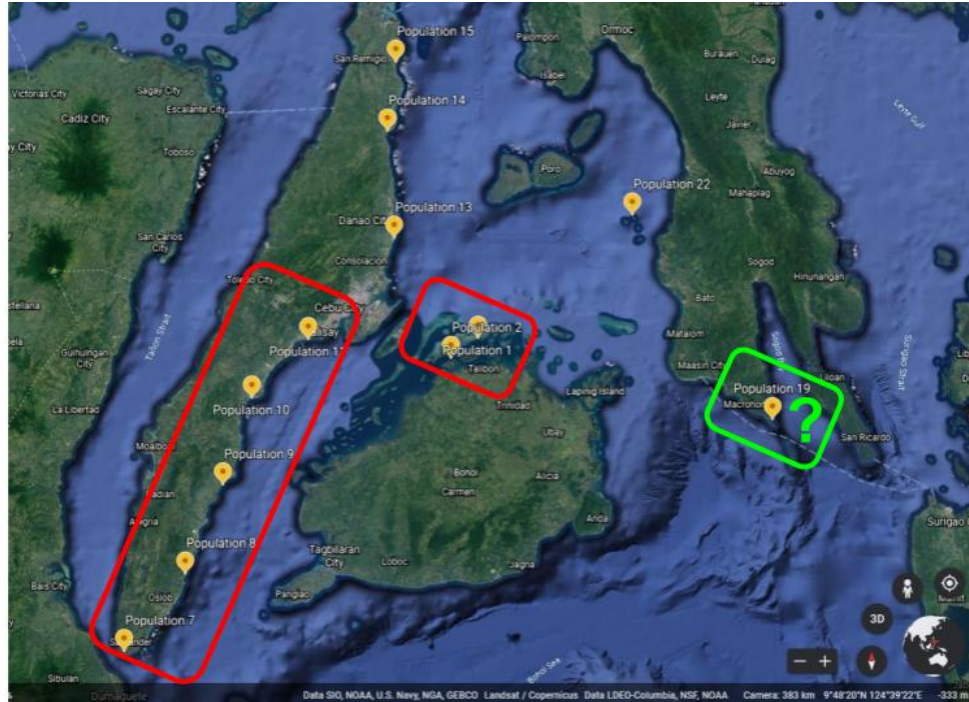
90%	95%	97.5%	99%
0.303	0.400	0.456	0.509

Permutation: free

Number of permutations: 5039



With population 19 excluded, the Mantel test and the linear regressions both yield significant p-values. The Mantel statistic r and the adjusted R-squared show a weak positive relationship. This suggests that as the geographic distance between a pair of populations increases, the genetic distance between the pair also increases. Based on the significant Mantel test and the general positive relationship between geographic and genetic distance, I conclude that 7 out of 8 populations follow an Isolation-by-Distance pattern. Now, I can use the slope from this linear regression and an effective population density estimate for these populations in Rousset's 1D habitat equation to generate an estimate for dispersal spread. I also show a map of the populations and their locations in the Philippines. Populations in red boxes follow Isolation-by-Distance patterns, population 19 is in a green box and does not follow Isolation-by-Distance patterns, and the populations without a box were not included in the analysis due to very low sample size (less than 5 individuals in a population).



```

##{r}
#Estimate of effective population size is 6942 fish. This was generated using the program Ne estimator and linkage disequilibrium methods.

#Effective density (De) is obtained by dividing effective population size by the length of the reef (130m- measured in ArcGIS)
6942/130 #De=53.4 fish/km

#According to Rousset, sigma = sqrt(1/4*De*m) where De is effective density of the population and m is the slope of the linear regression between geographic distance and linearized Est (Est/1-Est)
sqrt(1/(4*de_length*m)) #8.653 km

[1] 53.4
[1] 8.65277

```

This generates a dispersal spread estimate of about 8.7 km. I then worked on further analysis to start to understand what may be happening with population 19, the one population that does not follow Isolation-by-Distance patterns. I began to investigate if ocean currents may explain this difference and did so by obtaining a connectivity matrix from work done by Thompson et al. (2018). The connectivity matrix generated in this study shows larval dispersal probabilities assuming passive dispersal based on ocean currents between geographic coordinates in the Coral Triangle region. I found and extracted the probabilities that matched the coordinates of the 8 *Premnas biaculeatus* populations into a new matrix. This time I included population 19 with the 7 populations that followed Isolation-by-Distance patterns in the matrices. I then conducted Mantel tests between this new larval dispersal probability matrix and the genetic distance matrix to assess if there is a correlation between ocean current driven dispersal probability and genetic distance. I also plotted pairwise probability of larval dispersal vs. pairwise genetic distance.

```

#Perform Mantel test to assess correlation between connectivity matrix
#(gives probability of dispersal between populations, named ocean_matrix) and
#genetic distance (linearized Fst)
mantel(ocean_matrix, prbi_11_12_EW_full_fstlin)

#Plotting the data
plot(ocean_matrix, prbi_11_12_EW_full_fstlin,
      xlab="Probability of Larval Dispersal between Populations", ylab="Fst/(1-Fst)",
      pch=20)

```

Mantel statistic based on Pearson's product-moment correlation

Call:

```
mantel(xdis = ocean_matrix, ydis = prbi_11_12_EW_full_fstlin)
```

Mantel statistic r: -0.4342

Significance: 0.984

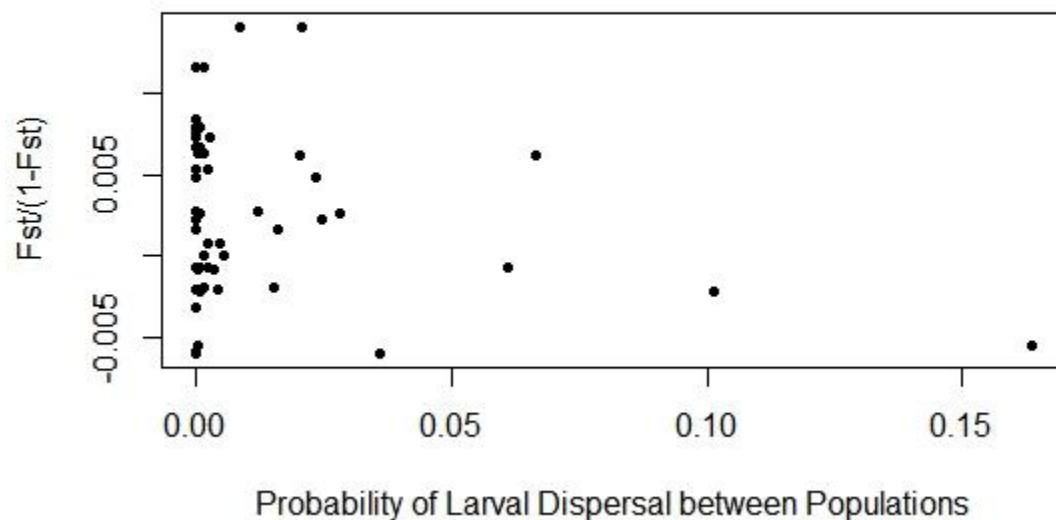
Upper quantiles of permutations (null model):

90% 95% 97.5% 99%

0.282 0.358 0.423 0.504

Permutation: free

Number of permutations: 999



The Mantel test yielded a non-significant p-value. This indicates that there is not a significant correlation between the matrix of larval dispersal probability based on ocean currents and pairwise genetic distance. The plot demonstrates that the points don't show a clear pattern and are generally clustered at low probabilities. The Mantel statistic of $r=-0.43$ indicates a weak negative relationship between larval dispersal probability and genetic distance, suggesting that as the dispersal probability between two populations increases, the genetic distance between them decreases. This makes sense as two populations that are connected by ocean currents would likely be more genetically similar than two populations that are not connected.

I then wanted to analyze what would happen when all three matrices were evaluated at once. To do this, I used partial Mantel tests to analyze the correlation between larval dispersal probability based on ocean currents and genetic distance while controlling for over-water geographic distance. I used a partial Mantel test because it allows for the inclusion of a third distance matrix (Smouse et al. 1986). It yields a Mantel statistic r that is the partial correlation coefficient of the first two matrices given the third matrix (for my data this is the correlation of larval dispersal probability based on ocean currents and genetic distance given the over-water geographic distance). Partial Mantel tests use permutation to obtain a p-value like the full Mantel test. I calculated over-water geographic distance by measuring the shortest route across the ocean between populations using the measurement tool in Google Earth. I used over-water distance rather than using the geodist package here because ocean currents can only push larvae through the ocean, not over land, and the geodist package may calculate a route that goes over land.

```
##Run partial Mantel test of oceanographic distance vs. genetic distance, while controlling for geographic distance
library(gdata) #package with function to make a symmetrical matrix

#Load in matrix with over water distances
water_distance <- read.csv("OverWater_Distance.csv")
water_distance <- as.matrix(water_distance)
water_distance <- water_distance[, -1]
colnames(water_distance) <- c(1, 2, 7, 8, 9, 10, 11, 19)
rownames(water_distance) <- c(1, 2, 7, 8, 9, 10, 11, 19)

#Make matrix symmetrical
upperTriangle(water_distance) <- lowerTriangle(water_distance, byrow=TRUE)

#Partial Mantel test controlling for over water geographic distance
mantel.partial(ocean_matrix, prbi_11_12_EW_full_fstlin, water_distance)

...
```

Partial Mantel statistic based on Pearson's product-moment correlation

Call:
mantel.partial(xdis = ocean_matrix, ydis = prbi_11_12_EW_full_fstlin, zdis = water_distance)

Mantel statistic r : -0.4157
Significance: 0.972

Upper quantiles of permutations (null model):
90% 95% 97.5% 99%
0.293 0.393 0.458 0.526
Permutation: free
Number of permutations: 999

The partial Mantel test results are similar to the full Mantel test results. There is not a significant correlation found between larval dispersal probability based on ocean currents and genetic distance when controlling for over-water geographic distance. The Mantel statistic r is similar to the full Mantel test and suggests that populations with a higher larval dispersal probability based on ocean currents will be more genetically similar. Overall, this indicates that ocean currents are not the main driver of dispersal for all 8 populations together. However, these results do not elucidate what exactly may be happening with population 19, and I would like to isolate population 19 in future analyses.

Conclusion and Discussion:

All populations except for one follow isolation-by-distance patterns. Using effective density and isolation-by-distance patterns, larval dispersal spread is estimated at 8.7 km. This is slightly lower than past dispersal estimates for other anemonefish species, but in the range of other coral reef fishes. There appears to be a variety of dispersal lengths in coral reef fish with anemonefish ranging from 5-20 km, but other coral reef fish such as butterfly fish having larval dispersal lengths upwards of 50 km (Almany et al. 2017). Overall isolation-by-distance patterns allowed us to generate fine-scale estimates of dispersal, which is key to better understanding adaptation and population trends along with informing successful conservation strategies for the species.

There does not appear to be a significant correlation between larval dispersal probabilities based on ocean currents and genetic distance. This suggests that ocean currents are not the main driver for the dispersal patterns found in these 8 populations of *Premnas biaculeatus* near the islands of Cebu, Bohol, and Leyte in the Philippines. This may be due to the locations of the populations and the location, direction, and strength of the currents in the area, especially the Bohol current. There is a current that travels around the south end of Bohol that may be taking larvae from population 19 and dispersing them in locations other than the locations that were sampled for this study. Based on the direction of this current, it is likely that only a few of the populations would be affected by it, which could account for why there were no significant correlations when all populations were evaluated together. I plan to conduct further analysis on population 19 by isolating it and evaluating the way ocean current and other environmental factors may be affecting its dispersal on a fine-scale.

References:

- Almany, G. R., S. Planes, S. R. Thorrold, M. L. Berumen, M. Bode, P. Saenz-Agudelo, M. C. Bonin, A. J. Frisch, H. B. Harrison, V. Messmer, G. B. Nanninga, M. A. Priest, M. Srinivasan, T. Sinclair-Taylor, D. H. Williamson, and G. P. Jones. 2017. Larval fish dispersal in a coral-reef seascape. *Nature Ecology & Evolution* 1:1–7.
- Mantel, N. 1967. The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research* 27:209–220.

Rousset, F. 1997. Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. *Genetics* 145:1219–1228.

Selkoe, K., and R. Toonen. 2011. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series* 436:291–305.

Smouse, P. E., J. C. Long, and R. R. Sokal. 1986. Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence. *Systematic Zoology* 35:627–632.

Thompson, D. M., J. Kleypas, F. Castruccio, E. N. Curchitser, M. L. Pinsky, B. Jönsson, and J. R. Watson. 2018. Variability in oceanographic barriers to coral larval dispersal: Do currents shape biodiversity? *Progress in Oceanography* 165:110–122.

Wright, S. 1943. Isolation by Distance. *Genetics* 28:114–138.