Evolution at GPI Loci among Mosquitofish Populations in Miami, FL
By: The Fish Grinders

Students:
Lorenzo Cummings
Bo Graves
Jack Lawhorn
Michael Ritchie
Orko Sarkar
Carlee Soeder

Teacher:
Mr. Nexer Palacios
Ms. Euphrasia Requena

Undergraduate Facilitator:
Ms. Griffin Udelson

Graduate Mentor:
Mr. Sean Beckmann

Faculty Research Mentor:
Dr. Michael Gaines

July 26, 2009
Research in Ecology
University of Miami
Abstract

Genetic variation among six populations of mosquito fish, *Gambusia holbrooki*, was analyzed at the Glucose-6-Phosphate Isomerase locus (GPI) using protein electrophoresis. Two GPI loci were identified, one positively charged and the other negative. Two alleles were identified at each locus, F and S. Results were analyzed using homogeneity tests, Nei’s genetic distance, Wright’s F-statistics, and Chi-square test for Hardy Weinberg equilibrium. The results indicate that Lake Osceola is significantly different from all other populations likely due to a lack of gene flow. Additionally, FIU Pond and Athula’s Pond may be experiencing genetic drift, and Kendall Lakes and Tamiami Canal do not appear to be connected genetically. Finally, all populations are in Hardy Weinberg equilibrium indicating that there is currently no evolution. These results may not indicate legitimate trends due to low sample size or human error. They may also indicate common environmental pressures across sites.

Introduction

Genetic variation within and among populations can be affected by geographic distances, barriers to migration, and time. Over time changes in genetic variation between populations can lead to evolution differences between those populations. These differences can be identified using allele frequencies at protein loci. Both gene flow and genetic drift can cause significant changes in allele frequencies at protein loci. Gene flow is the movement of genes from one population to another. This process requires individuals to migrate between populations and to breed in their new populations. Genetic drift is an evolutionary force that can cause alleles to be eliminated over time by random chance.

One organism that can be used to study the effects of gene flow and genetic drift is the mosquitofish, *Gambusia holbrooki*. *G. holbrooki* is a habitat generalist that can be found in
freshwater and brackish lakes, rivers, and canals. They grow from 3-5 cm in length, with females growing larger. Their diet consists primarily of mosquito larvae and eggs. The species was introduced to the United States to reduce and control mosquito populations. Due to their generalist nature they are capable of rapid reproduction and migration into a variety of habitats. As such, they outcompete other fish for resources and have become an invasive species.

*G. holbrooki* contain a protein called Glucose-6-phosphahte isomerase (GPI). This protein is involved in glycolysis and can be used to look at genetic variation in this species. This protein is a dimeric enzyme that is made up of two alleles. These alleles are Fast (F) and Slow (S), and are identified by the speed they move when placed in an electric field. The speed is based on the charge of the protein. The frequency of these alleles can be compared among population throughout Miami to explore genetic variation.

The purpose of this study is to identify genetic variation among six populations of mosquitofish in Miami. The six populations are: Tamiami Canal (TC), Kendall Lakes (KL), FIU Pond (FIU), Athula’s Pond (AP), UM Pond (UM), and Lake Osceola (LO). FIU, UM, and AP are isolated populations incapable of migration with any other population.

Our null hypothesis is there will not be any significant differences in allele frequencies among the six populations studied. Our alternative hypothesis is that there will be significant differences in allele frequencies among the six populations studied. We predict that TC and KL will be the most similar because they are connected by 10 miles of canals. We also predict that AP will be the most different population because it is the most isolated. Further, UM should be similar to LO and FIU should be similar to TC if these are the source of fish for those populations. Finally, LO should be similar to KL and TC if they are part of the same canal system.
Materials and Methods

Twenty to thirty fish were collected from six populations throughout Miami using nets. Fish were kept in jars containing water from their natural habitat until they were returned to the lab. The first site was KL, located 16 km from the University of Miami. The area was clean with little pollution. The water appeared shallow to moderately deep and was clear. Numerous bird species were present that could prey on the fish. Next was TC, located 25.1 km from the university, and connected to KL by 16 km of canals. The area appeared highly polluted and was next to a highway. The water was very murky, and there was oil and large amounts of algae present. Numerous predators were present including other fish, turtles, and alligators. Third, fish were collected from FIU, an isolated man made pond, located 12 km from TC. There were moderate levels of pollution and detritus at this site. The water was cloudy with oil runoff and some algae. The water was shallow and other species of small fish were present, as well as frogs and turtles. Fish were then collected from AP, a small back yard pond in South Miami 2 km from the university. The pond was murky and shallow and no predators were present. However, the site is accessible by birds and frogs. Fish were then collected from two sites on the University of Miami campus. LO was a brackish lake with a film on the water. The water was clear with little algae growth. The site contained many larger fish, birds, and reptiles which may act as predators. Finally, fish were collected from UM a small pond located behind Cox Science Center. The water was cloudy with no pollution. Frogs, crawfish, and birds were also present.

The fish were euthanized using 500g of Tricaine anesthetic per site. After the fish died they were sorted by site. Each fish was ground in a labeled Eppendorf tube containing 50-200 ul of Tris-Glycine buffer. Volume of buffer was determined by the size of the fish. Tubes were then vortexed for ten seconds to thoroughly mix the contents. Samples were then centrifuged for eight
minutes at 13,000 rpm to isolate the proteins from the solids. Next, the supernatant was
transferred to clean labeled tubes and the solid was discarded. The samples were stored in the
freezer to prevent the breakdown of proteins.

Cellulose acetate plates were soaked for 20 minutes in Tris-Glycine buffer. During this
time, 10 ul of each sample was placed in a sample well. The plate was then dried and the samples
were applied to the plate using an applicator. The plate was placed in an electrophoresis chamber
and run at 200 volts for 45 minutes to separate the proteins by charge. Five ml of GPI stain was
then applied to the plate in a petri dish wrapped in aluminum foil. The foil was used to block out
the light because the stain works more rapidly and effectively in the dark. Results were recorded
for all samples. These results were then analyzed using Chi square tests to identify population in
Hardy Weinberg equilibrium. Homogeneity tests, Wright’s F-statistics ($F_{ST}$), and Nei’s Genetic
Distance were also calculated to analyze the results.

**Results**

Two GPI loci were identified in each mosquitofish population studied. One locus was
positively charged while the other was negatively charged. Each locus was composed of two
alleles, F and S, as in previous years. For KL, the frequency of the F allele at the first locus was
0.80 and the second locus was 0.72. At TC, the frequency at the first locus was 0.70 and at the
second locus was 0.58. The frequency of F at the first locus at FIU was 0.85 and at the second
locus was 0.65. For AP, the F allele at the first locus was 0.88 and at the second locus was 0.65. In
LO, the F allele at the first locus was 0.44 and was the same at the second locus. Finally, at UM
the frequency of F at the first locus was 0.64 and at the second locus was 0.53. The test of Hardy
Weinberg equilibrium (HWE) indicated that all populations were in HWE at both loci except
FIU which was not in HWE at the second locus.
A number of population pairs were significantly different based on the results of the homogeneity test. At the first locus LO was significantly different than all other sites. UM was also significantly different from AP, KL, and FIU. Additionally, AP was also significantly different from TC and UM. FIU was also significantly different from UM. At the second locus LO was significantly different from all sites except UM and TC but was approaching significance from TC. Also, KL was significantly different than UM.

Populations were also analyzed using Nei’s genetic distance and F_{ST}. A Nei’s genetic distance of greater than 0.25 was considered significant. No populations were significantly different from each other based on Nei’s genetic distance. However, LO was approaching significance from FIU, KL, and AP. Additionally, the distances between LO and all other sites were at least three times larger than any other population pair. This does not hold true for LO-UM which had similar distances to most population pairs. An F_{ST} value between 0.05 and 0.15 indicated large differences between populations. Values greater than 0.15 indicated that multiple species may be present. LO was different from all other populations except UM. Additionally, AP and UM were also different from each other.

According to the results, AP and FIU were the most similar sites. On the other hand, LO and AP were the most different, with LO being highly different from all sites except UM.

**Discussion**

Our null hypothesis was not supported, indicating that there are statistically significant differences between some of the populations studied (Table 1). The homogeneity test indicates that at both loci LO is significantly different from all other populations with one exception. At the second locus LO is not significantly different from UM. Also, at the first locus LO is most similar to UM even though they are significantly different. Based on these results it is likely that
LO is the source population for the fish in UM. UM is also significantly different from FIU, KL, and AP at this locus. This further supports LO as the source population. Additionally, due to the differences between LO and KL/TC it is highly unlikely that there are any common canals connecting these sites.

In addition to LO, AP is significantly different from both TC and UM. This is not surprising because AP is the most isolated population. In fact, it is expected that AP should be significantly different from more populations due to its level of isolation. Due to its small population size, and degree of isolation the results may indicate that AP is experiencing genetic drift. Further, previous results indicate that AP is significantly different from more populations. In turn, it appears that allele frequencies in AP are randomly fluctuating and that genetic drift is occurring. Finally, KL is significantly different from UM at the second locus, which is expected due to their geographic distance and the fact that UM is isolated.

The two populations that are most similar are AP and FIU. These results are consistent with the previous two years. On the surface these results are surprising because both of these populations are isolated and there is a large geographic distance between them. However, both populations have small populations size and share similar environments. This may indicate similar selective pressures and may explain why they have similar allele frequencies. Since prior to 2007 these populations were different, it is unlikely that they share a common source population (see Figure 1).

It is expected that KL and TC should be the most similar due to the fact that the two sites are connected. In fact, numerous population pairs are more similar than these two. This indicates that these two populations may no longer be connected. This may be the result of high pollution levels noted along the Tamiami Canal in recent years. Additionally, until recently both sites had
low water levels and a greatly reduced number of fish. Both of these factors may contribute to a lack of gene flow. Recently water levels have increased. Also, it was noted this year that there was a large increase in the number of fish at both sites. This may indicate a recent increase in gene flow that is not yet indicated by the allele frequencies. In the future there may be a further increase in population size and a homogenization of allele frequencies.

It is also predicted that KL/TC is the source population for the fish at FIU. If this is true it is expected that the allele frequencies in FIU are similar to those in TC and KL. Given that KL and TC no longer appear to be connected it is difficult to evaluate this prediction. However, other than AP, FIU is most similar to KL at the first locus. Additionally, FIU is very similar to both KL and TC at the second locus. This indicates that these sites may be the likely source of the fish in this population.

Results indicate that all populations are in Hardy-Weinberg equilibrium (HWE) except for FIU at the second locus (see Table 2). This indicates that this population is evolving at this locus. The likely cause of this evolution is due to genetic drift, given the population’s small size and isolated nature. However, this has not been evaluated and this may be due to random chance. The remaining populations are all in HWE, meaning that all five conditions for HWE are met. These are: large population size, random mating, no mutation, no migration, and no natural selection. For the past two years all populations have been in HWE. This indicates that the populations may all be in a static state. Additionally, since all populations are in HWE it is unlikely that there is gene flow between KL and TC.

The results of this study indicate that evolution is not currently occurring in these populations. However, additional confounding factors may be responsible for these results. First, populations were not equally sampled and a small sample size was collected from each
population. This may result in inaccurate allele frequencies being calculated. Additionally, if spatulas were not thoroughly cleaned while grinding samples, then false results may have been identified. Finally, if pregnant female fish were included in the samples it is possible for false heterozygotes to be identified. Since pregnant females were not removed this is a distinct possibility.

Acknowledgments

We would like to thank Dr. Kremples for organizing and running Research in Ecology 2009 and for her interesting lectures. Ms. Pappas for teaching us about statistics and how to use computers in research. Dr. Nobleman for teaching us scientific writing. Dr. Gaines for helping the Fish Grinders in the field and being a great professor. Sean Beckmann for leading the Fish Grinders. Griffin Udelson for helping our group and putting up with us. Ms. Requena and Mr. Palacios for keeping us on task and keeping Sean sane. Christine Hwang for polite suggestions and comic relief. Dr. Athula Wikramanayake for allowing us to collect fish from his house. The University of Miami for providing us with facilities for the program and the Howard Hughes Medical Institute for providing funding for the program to exist.
Literature Cited


White Ibises. 2006. *Genetic differences between Mosquito Fish (Gambusia holbrooki) in connected and isolated populations in southern Florida.* Research in Ecology Program, University of Miami.
Table 1. P-values for comparisons of seven populations of mosquito fish using Chi-square (x²) Homogeneity tests.

<table>
<thead>
<tr>
<th>Locus</th>
<th>TC1</th>
<th>KL1</th>
<th>FIU1</th>
<th>LO1</th>
<th>UM1</th>
<th>AP1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KL1</td>
<td>0.238493</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIU1</td>
<td>0.067686</td>
<td>0.50347593</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LO1</td>
<td>0.0003211</td>
<td>7.4152E-05</td>
<td>6.60989E-06</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UM1</td>
<td>0.473639</td>
<td>0.06376966</td>
<td>0.013339871</td>
<td>0.026323395</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AP1</td>
<td>0.04163</td>
<td>0.31544012</td>
<td>0.693760056</td>
<td>9.04219E-06</td>
<td>0.008985</td>
<td>1</td>
</tr>
</tbody>
</table>

P-values of Chi-Square Distribution at Locus 2

<table>
<thead>
<tr>
<th>Locus</th>
<th>TC2</th>
<th>KL2</th>
<th>FIU2</th>
<th>LO2</th>
<th>UM2</th>
<th>AP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KL2</td>
<td>0.120786</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIU2</td>
<td>0.444131</td>
<td>0.44723</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LO2</td>
<td>0.102458</td>
<td>0.001867</td>
<td>0.0199</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UM2</td>
<td>0.593099</td>
<td>0.040296</td>
<td>0.203651</td>
<td>0.278153</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>AP2</td>
<td>0.503111</td>
<td>0.45352</td>
<td>0.969375</td>
<td>0.034248</td>
<td>0.254538</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2. P-values indicating which populations and loci are in Hardy Weinberg equilibrium.

<table>
<thead>
<tr>
<th>Locus</th>
<th>TC</th>
<th>KL</th>
<th>FIU</th>
<th>LO</th>
<th>UM</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC</td>
<td>0.693</td>
<td>0.322</td>
<td>1</td>
<td>0.627</td>
<td>0.906</td>
<td>0.033</td>
</tr>
<tr>
<td>HW</td>
<td>HW</td>
<td>HW</td>
<td>HW</td>
<td>HW</td>
<td>NO</td>
<td>HW</td>
</tr>
</tbody>
</table>
Figure 1. Frequencies of F allele at locus 1 for four populations over the course of 10 years.