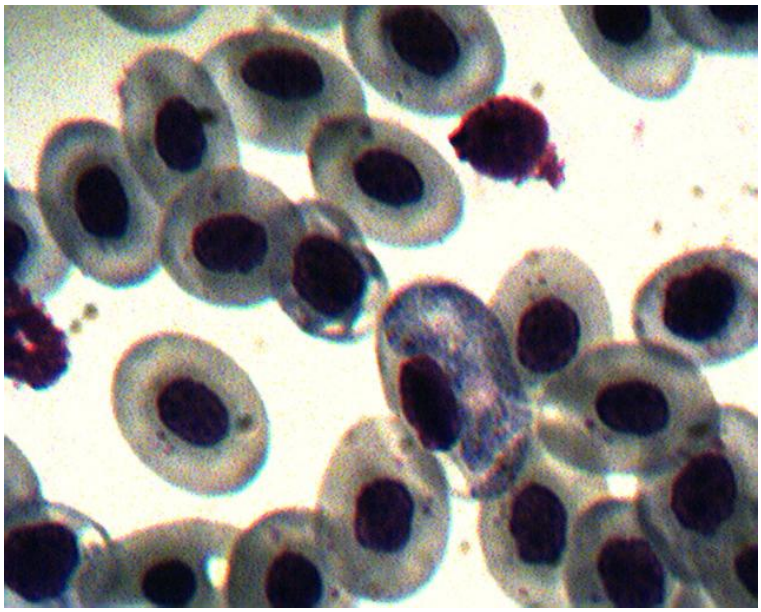
Microsatellite Allele Associations in Mixed-Clone Infections of a Lizard Malaria Parasite, *Plasmodium mexicanum*



Megan Carole Lind

Department of Biology

University of Vermont

Honors Thesis

Defended May 2, 2013

Thesis Committee:

Dr. Joseph Schall- Biology Department

Dr. Nicholas Gotelli- Biology Department

Dr. Yolanda Chen- Plant & Soil Science Department

Table of Contents

Abstract

The ecology of how organisms of different genotypes interact within an environment is an exciting, and relatively new, area of research. Generally, we think of individuals competing to better their own individual fitness. However, there are examples of cooperative individuals such as the yeast, *Candida albicans*, that forms biofilms within indwelling medical devices to increase their resistance to antifungals (Kumamoto 2002). Not much is definitively known about how multiclonal malarial infections interact within a host. There is some support that the clones compete within infections for dominance (De Roode et al., 2005). Other studies show that there is variation in the way clones interact (Vardo-Zalik and Schall, 2009). Here, my goal is to add to the body of knowledge on how malaria parasites interact within infections. I do this by examining *Plasmodium mexicanum* infections of the western fence lizard, *Scleroporus occidentalis*, obtained over a long-term study from 1996 through 2012 from a natural parasite-host system located in northern California at the Hopland Research and Extension Center. The data from four microsatellite loci, Pmx306, Pmx732, Pmx747, and Pmx839, for the presence of alleles were examined for a total of 547 infected lizards. A novel approach of using Ulrich's Pairs analysis (2008), which was originally designed to be applied to species associations, was utilized to determine significant positive and negative associations between alleles at each locus. For the four loci combined, there was a total of 84 negative significant negative relationships and 18 positive significant relationships. A trend of more negative significant results in comparison to positive significant results was seen across all four loci. When looking at the total number of possible allele combinations, the neutral interactions account for most relationships. However, due to the conservative nature of the Pairs analysis and its tendency towards type II errors, there may be a higher proportion of negative and positive associations that have been missed. There was no relationship seen between the lengths (repeat length) of alleles within a significant pair and their involvement in a positive or negative significant pair. Therefore, I cannot conclude that alleles which are more similar in size and perhaps more closely related are interacting positively with one another. I found a significant positive correlation between the allele's abundance over the time period sampled and the number of significant results it was involved in. I suggest that this is a factor of the high allele abundances for *P.mexicanum* being similar to the intermediate abundances of species. The Pairs analysis is best at identifying significant pairs for species that have intermediate abundances and therefore *P. mexicanum* alleles which are higher in abundance tend to have larger numbers of significant results found. Finally, I conclude that there is variation in the types of interactions between clones seen in *P. mexicanum* infections where neutral interactions are the most numerous, negative interactions are the next in abundance, and positive interactions are the least numerous.

Chapter 1: Importance of the Genetics of Parasites

Genetics is pivotal to the theory of evolution. Genes are what natural selection acts on over time to cause species to adapt and evolve to their surroundings. Parasites have adapted to their environments and hosts over their association together and genetic differentiation is a visible monument to that process. Parasites are important for us due to the numerous infections that occur worldwide every year and the mortality rate associated with these infections. For example, the World Health Organization reports an estimate of 219 million cases of malaria worldwide in 2010 and approximately 660,000 deaths from malaria which is caused by parasites of the genus, *Plasmodium*, in humans (World Health Organization, 2012). Malaria not only takes a toll on the health of the world but also on the economy. There is a strong correlation seen between malaria and poverty. Countries with regions where *Plasmodium falciparum*, the deadliest of the human malaria parasites, is prevalent the average per-capita growth of GDP is 0.4% per year while other countries experienced an average of 2.3% per year (Sachs and Malaney, 2002).

Due to the process of natural selection, parasites are sometimes able to evolve resistance to the medical treatments we have devised to treat parasitic infections. When a drug becomes ineffective it is important that we understand the mechanism the parasite has evolved to become resistant as well as other general mechanisms for how the parasite becomes virulent in the host so that we can develop new treatments to treat people affected by the disease. For example, *Plasmodium falciparum* has increasingly become more resistant to the antimalarial drug, chloroquine, which works by disrupting the parasite's mechanism to detoxify the heme

molecules resulting from the parasite's digestion of host hemoglobin. Researchers are working on discovering the specific mechanism that now allows *P. falciparum* to avoid the remedial effects of chloroquine. Molecular genetic techniques play a significant role in this research. Fidock *et al.* used PCR amplification with primers to locate the genetic sequences in the cDNA library of the *P. falciparum* genome. They found mutations in the *pfcrt* gene that codes for a new transmembrane protein which seems to confer chloroquine resistance to the individuals that have it. It is research like this that will help us understand why some drugs will be more effective in treating disease than others (Fidock *et al*., 2000).

Not only can we learn about ways to combat resistance to antimalarial drugs through the use of molecular genetics techniques, but we can also learn how to treat malaria at the level of transmission. This way, we can take preventative measures and stop people from getting the disease rather than treating new cases of the disease as they occur. Researchers are working on transmission-blocking vaccines for human malaria that target the Pfs25H antigen that resides on the surface of *P. falciparum* zygotes and ookinetes as well as the homologue, Pvs25, in *Plasmodium vivax*. Knowing the genetic sequence of the antigen allows researchers to amplify the antigen within a yeast host and then use the product as a vaccine (Kongkasuriyachai *et al.*, 2004).

These examples demonstrate that genetic variation in any parasite population will determine the ability of the parasite to evolve responses to host changes, including medical interventions. But, genetic variation in the parasite also has its costs if different genotypes interact negatively – competition – when in mixed infections. This leads to the goals of my research. I would like to determine if malaria parasites are interacting with one another when

they are found in multiclonal infections (infections with more than one genotype). It is unclear if different genotypes of malaria parasites within a single host interact negatively (competition), positively (cooperation), or have no effect on one another. To find out which of these relationships occur, I will examine *Plasmodium mexicanum* infections of western fence lizards from 1996 through 2012 using four known microsatellite loci.

Chapter 2: Allele Association Study of *Plasmodium mexicanum*

Introduction

Richard Dawkins has coined the phrase 'Paradox of the Organism'. This phrase represents how what we define an organism by the genetic information contained within it. With the multitude of genes in a single space, why isn't there conflict between these genes? From what we observe, genes within an organism most often cooperate to form a cohesive organism that acts as a single unit. Why doesn't one gene compete in order to ensure its own transmission to the next generation? Dawkins' answer to this is that all of the cells within an organism share the same means of being passed on to the next generation. This would be the gametes in terms of sexually reproducing organisms. Therefore, the organism is not torn apart by conflict within its genome because all of these genes are passed on to offspring through the gametes. Dawkins uses parasites as an example of how in order to have the exact same 'desiderata lists' there must be a single genotype of cell within a multicellular species. Parasites can have shared desiderata lists with the host when parts of their desiderata lists overlap, but because they are not the same genotype they will differ in some of their goals. This is what tears them apart; that is, what drives an antagonistic interaction. Dawkins proposes that since the parasite and the host ultimately have different desires or genotypes, they cannot completely cooperate together as a unit. By this definition, even organisms within the same species do not have identical desires because they both want their own genotypes to be passed on to the next generation (Dawkins, 1990).

Is it really true that two genotypes cannot have a positive relationship with one another? What about organisms that participate in social behavior or altruism? For example, the social amoeba, *Dictyostelium discoideum*, can be a solitary unicellular organism but the solitary

amoebae can also come together under harsh conditions to form multicellular aggregates called fruiting bodies. Together, these fruiting bodies have higher mobility, better protection from predators, and better dispersal than the solitary amoebae (Li and Purugganan, 2011). Another example would be the biofilms formed by *Candida albicans*, a fungus that can inhabit medical devices such as prosthetic heart valves and central venous catheters. Biofilms are communities of microorganisms that exhibit increased expression of genes that lead to the production of exopolymeric matrix to anchor the biofilm to a surface as well as genes that allow the biofilm to develop antifungal resistance that are not seen in planktonic (floating or drifting) microorganisms. For instance, *Candida albicans* biofilms upregulate the expression of genes, such as *MDR1,* that produce drug efflux transporters (these pumps expel drugs that make it to the interior of the cell back outside of the cell) (Kumamoto, 2002).

Now consider parasitic infections that contain more than one genotype, called multiclonal infections. These different genotypes (commonly referred to as clones) are contained within a single host and both have the goal of being transmitted to the next stages in their life histories. They appear to share some common desires. How do they interact with each other within this host? By Dawkins' definition, they would be in conflict with one another because they are different genotypes. But couldn't they also interact in a positive way with one another and work together against the immune system of the host? There is also the third possibility that the genotypes would have a neutral interaction and have no effect on one another. Little is known about how different parasite genotypes interact with one another while sharing the same space (Taylor and Read, 1998).

The parasite-host-vector system that is utilized for this study involves the parasite, *Plasmodium mexicanum*, the vertebrate host, *Scleroporus occidentalis*, and the insect vectors, *Lutzomyia vexator* and *Lutzomyia stewarti*. These sand fly vectors take blood meals from the infected lizard host and, in doing so, transmit the malaria parasite to the next lizard they bite. This model system is located in Northern California and has been studied extensively at the Hopland Research and Extension Center (HREC). There are many benefits to this model system including the fact that it is a natural system rather than an experimental system and that data have been collected as far back as 1978 so longitudinal studies can be performed (Schall *et al*., 1982).

Genetic clones, identified by scoring specific alleles, of a malaria parasite may reveal three kinds of association in mixed-clone infections: neutral in which clones do not interact, negative (an antagonistic interaction), or positive. The best experimental studies of interaction of malaria parasites come from the study of *Plasmodium chabaudi* in laboratory mice. For example, de Roode *et al*. found that clones that infected mice at the same time usually ended up with one clone being dominant to the other after some time. However, they also found that if a clone which was normally suppressed by a more dominant clone was established in the mouse first, it could be the dominant clone in the infection. (De Roode *et al*., 2005). This previous study suggests that there is a competitive or negative interaction between genotypes within infections. What could lead to a positive association of genotypes for a malaria parasite among infections? One possibility is that because the vector takes a blood meal from a single vertebrate host and transmits it to another host, that clones in an infection often travel together. A study done by Vardo-Zalik on the clonal diversity of *P. mexicanum* showed that the transmission of multiple clones from *S. occidentalis* to the sand fly vectors is high. In experimental infections with two clones, 89% of the infections matched exactly from host to vector (Vardo-Zalik, 2009).

The benchmark studies of de Roode *et al* in 2005 with *P. chabaudi*, however, may not be what is occurring in natural systems and only looks at experimental infections (Ford *et al*., 2010). By observing a natural system, such as the *P. mexicanum* system of northern California, we can further the knowledge about clonal interactions in multiclonal malaria infections.

To examine the nature of how clones are associated among infections of a malaria parasite, this study examines specific microsatellite loci of *P. mexicanum* in multiclonal infections to determine if specific alleles end up with other alleles located at the same loci more often than by chance. For example, if we examine how often allele "A" is found with allele "B" in infections and compare it to the odds that allele "A" has just been randomly placed with allele "B" in a host, we can determine if allele "A" and allele "B" are found together in an infection more than simply expected by chance. This would indicate that there is some kind of positive association between these genotypes. These clones may be cooperating to some degree or are simply being transmitted as a group by the vector and remain over multiple passages. On the other hand, we can explore if certain pairs of alleles are *less often* found together than by random chance. In this example, allele "C" and allele "D" would not occur within a single infection more than is expected by chance. This would imply that those genotypes are interacting negatively when they happen to occur in the same infection and therefore competition tends towards them not being in the same host.

The broad question that is addressed by this thesis is: Do genotypes of the malaria parasites found in an infection have any sort of effect on one another? The null hypothesis for this study is that alleles will be found randomly together within infections. The alternative hypothesis is that alleles will be found together more often than by chance (positive interaction)

or alleles will be found together less often than by chance (negative interaction). Also, I sought to determine if more closely related alleles (alleles that have diverged more recently) will have positive relationships and more distantly related alleles will have negative relationships. Another goal of the research is to explore a new method to examine such genetic data; this is a method of null model analysis recently devised for studies in community ecology.

Materials and Methods

Sample collection

Lizards were caught at nearby sites at the Hopland Research and Extension Center (HREC) near the town of Hopland in Mendocino County, CA. Nine sites on the field station were sampled that have been named Greenhouse, Water Tank, Middle Lower Horse, Gorge, Lamb Barn, Parson's Creek, Horse Coral, Joy's Site, and Chicken Coop. These are all within less than a kilometer distance from one another. The method of capture involves using fishing poles with string nooses tied at the end of them. The noose is carefully slipped over the head of the lizard and the fishing pole is pulled upwards to tighten the noose around the neck of the lizard. The captured lizards are transported to a lab setting in pillowcases with breathable fabric and put into plastic bins with holes for aeration in the lids once reaching the lab. In the lab, the lizards are sexed, measured in length, checked for ticks/mites, checked for tail condition, and given a number. A toe clip is then performed to obtain enough blood for a blood smear and a blood dot dried on filter paper. The blood dots are then stored in sealable bags with silica beads to keep the moisture to a minimum. The number of lizards caught over the course of the summer as well as the techniques for handling them are in accordance with IACUC protocol which has been approved and given to Dr. Joseph Schall. Blood smears are stained with Giemsa. The blood smears are used to identify infected lizards by microscopic scans using a light microscope at

1000x magnification using oil immersion for approximately 6 minutes for each slide. The lizards have the last two digits of their personal number painted on their dorsal surface and they are released back at the sites from which they were initially taken within 24 hours of the time they were captured.

Molecular techniques

I extracted DNA samples from blood dots of infected lizards collected from the 2011 and 2012 field seasons at HREC. Extraction was done using a DNeasy kit (Quiagen, Valencia, CA) and was performed according to the protocol typically used in the lab (Vardo and Schall, 2007).

The extracted DNA was amplified by polymerase chain reaction (PCR) through the use of primers for *P. mexicanum* specific microsatellites Pmx306, Pmx732, Pmx747 and Pmx839 (Vardo and Schall, 2007). Ford *et al*. has shown that the microsatellite markers Pmx747, Pmx306, and Pmx710s are able to show the relative densities of clones in multiclonal infections of *P. mexicanum* in an accurate and reproducible manner. This experiment confirms that the microsatellites used in this study are reliable to detect clones in a *P. mexicanum* infection (2010). A single PCR reaction consists of a tube containing one Ready-to-Go PCR bead (GE Healthcare, Piscataway NJ), 1 µL of forward primer for a specific locus, 1 µL of reverse primer for the same locus, 4 µL of the extracted DNA, and 19 µL of PCR-quality water for a total of 25 µL in each reaction. In addition to the extracted DNA samples of infected lizards, one control where water replaces the 4 µL of DNA is included in each batch to make sure there is no contamination of samples.

The PCR products were visualized by gel electrophoresis using a 1% agarose gel. According to the relative brightness of the bands visualized during gel electrophoresis, 5µL of

PCR product was diluted with water properly and put into one well of a 96-well dilution plate (45µL for no band, 67µL for weak band, and 90µl for strong band). 15µL from a mixture of 20µL of size standard LIZ500 (ABI) and 1.5mL of Hi-Di formamide (ABI, Foster City, CA, USA) was inserted into each well of a separate 96-well plate. Then, 1µL of the diluted PCR product is transferred into the corresponding wells of the separate 96-well plate. The plates were sent to Cornell University Core Lab for genotyping using an ABI prism genetic analyzer. Data from the Cornell University Core Lab was extracted using Peak Scanner Software v1.0 (Life Technologies, Grand Island, NY, USA). Each peak on the returned pherograms indicates an individual allele of a single clone because the parasites are haploid during this stage, as they are for the majority of their life cycle. The height and size of the peaks for all clones were recorded for each infection.

Statistical analysis

The microsatellite data for infected lizards was analyzed using the Pairs program (Ulrich, 2008). This program uses a null model approach to detect non-random associations of the alleles in a matrix of the form *allele* x *lizard*. The program first tests if the overall matrix is significantly different from a random pattern, and then examines all the possible pairs to detect non-random associations (Gotelli and Ulrich, 2010). All available data from 1996 to 2012 for the four microsatellite loci Pmx306, Pmx732, Pmx747, and Pmx839 were organized in an Excel spreadsheet along with lizard numbers, years, and site location. There was a total of 547 infected lizards with available data used for this study. Every allele was binned into standard allele lengths according to the binning used in Schall and St. Denis's paper on prevalence of alleles in *Plasmodium mexicanum* (2013). Each locus was then separated into another spreadsheet of its

own and converted into a comma separated values (.csv) format. R programming (R Core Team, 2012) was used to design a program to convert these comma separated values formats into text files (.txt) that could be read and analyzed correctly by the Pairs program (Ulrich, 2008), as well as separate the files into groupings by year. The year groups were formed so that each group had around 50-80 infections. Also, the microsatellite alleles are less likely to be linked to specific coding loci over longer periods of time, so data are analyzed only for periods of a few years. The resulting year groupings were 1996-1997, 1998-2002, 2003-2005, 2006, 2007-2008, 2009, 2010, 2011-2012.

These text files for each of the year groupings were then run through the Pairs program twice, once using the null model of randomization with fixed rows and columns (input=s) and once using the null model of randomization with fixed rows and equiprobable columns (input=f) (Ulrich, 2008). Having the rows fixed allows the typical frequency of each allele to be preserved during the analysis. The randomization which has the fixed columns is the more conservative of the two tests since it preserves the uniqueness of each infection (that is, number of clones) in addition to preserving the frequency of each allele. This biases the results toward Type II errors. The randomization that does not keep the columns fixed treats every column equally and therefore implies that each infection has the same likelihood of occurring. For both analyses, the pairwise co-occurrence measure used was the C-score (input=c) since it is a commonly utilized method. This score looks at the checkerboards found within the matrices and checks for a pattern of segregation rather than a random checkerboard (Gotelli and Ulrich, 2010). The rest of the options for the analysis were kept at the default for the program (see Appendix C for an example of a Pairs analysis). Further analysis of the resulting data such as linear regressions and Fisher's Exact Test was performed in JMP 9 (SAS institute, Cary, NC, USA).

Expectations

Overall, the methods that I am using will bias the results towards missing real associations, both positive and negative. This is due to the biological and statistical aspects of the methods. In terms of biology, I am using microsatellite markers to look at associations between genotypes. Microsatellite markers are, in general, assumed to be neutral, non-coding regions of the DNA. However, they can follow loci which code for important things if they are linked to those loci. There is the chance that the four microsatellite loci used in this experiment are not linked to coding loci. But, a study done by Vardo-Zalik and Schall showed variation in life history traits and virulence of mixed-clone infections compared to single-clone infections using these microsatellite markers. This suggests that the microsatellites are linked to important coding loci (Vardo-Zalik and Schall, 2009). Another reason it will be difficult to find significant results is that mating of the parasite could break this linkage of the microsatellites to coding regions due to recombination. There are many opportunities for recombination across the time period that I am studying so a relationship that might have been present could be lost. This is one of the reasons that I divided the time span into smaller year groupings to reduce the chance of a linkage break hiding a significant relationship. The smaller year groupings also served another purpose. The possibility for mutations in alleles means that the same length microsatellite alleles may not have all arisen from the same mutational origin. For example, one microsatellite allele may come to be a certain size through deletion of repeats while another may come to be the same size by an insertion of repeats. Therefore, alleles in infections across many years may not be homologous. By reducing the number of years examined in each analysis, the chance of alleles not being homologous is lessened.

The other way that the methods are biasing the results towards missing significant results is through the choice of the analysis program. The Pairs program (Ulrich, 2008) is very conservative in the way that it analyzes the data, especially with the two options that I have chosen to use to generate the null randomization models. The Pairs program is designed to reduce the problem of Type I errors that occur when looking at multiple comparisons. By constraining the rows and columns or just the rows, the program is limiting the number of possible random matrices it can generate to compare to the actual matrix in order to determine the data's significance. Therefore, there is a higher probability of the data matrix to be considered insignificant than if there had been more random matrices generated.

In summary, if significant results are found through the Pairs analysis, then there is a strong probability that these results are indeed significant due to both the biological and statistical aspects explained above. If the clones within an infection are cooperating, I would expect to see a significant positive association (clones seen together more often than expected by chance). If the clones are in conflict, I would expect to see negative associations (clones seen together less often than expected by chance). If the clones are not interacting with each other, I would expect to see no significant results (clones are seen together and apart the same as expected by chance).

Results

Overall, the fixed row and column randomization Pairs analysis had a lower number of significant matrices when compared to the fixed row and equiprobable column randomization Pairs analysis (Table 1). This is the case for each microsatellite locus. These results were expected because the fixed row and column randomization is a more conservative test that would tend towards a type II error where we would miss a real association. However, the number of significant pairs for both randomizations is not very different. I decided to use the fixed row and equiprobable column test for the rest of the comparisons since I feel that it would be the more biologically relevant. The reason this test would more likely mimic what is found in nature is that the columns represent infected lizards. We are assuming, in this case, that each lizard is equally likely to be infected by *P. mexicanum* clones. A critique of this approach would be that each lizard has a unique immune system that may be more or less likely to fight off infection by *P. mexicanum*. For each locus, combining results for all years, I found the number of significant pairs of alleles to be greater than expected by chance. Overall, the findings are striking because the data fill sparse matrices with many cells having low counts of alleles (many rare alleles) which makes detection of real associations difficult.

Table 1- Comparing significant results of the two types of randomization used in the Pairs analyses (Ulrich, 2008), the fixed row and column randomization (fixed-fixed) and the fixed row and equiprobable column randomization (row fixed). Both randomizations were run for each microsatellite locus (see Appendix B). Simulations to examine the behavior of the Pairs program by Gotelli and Ulrich (2010) predict a 0.03% and 0.97% remaining Type I errors, so these values are used to calculate the number of "significant" pairs expected as Type I errors.

	Fixed-Fixed				**Row fixed**			
Microsatellite Locus	**# significant matrices**	**# significant pairs**	**# possible pairs**	**# significant results expected by chance**	**# significant matrices**	**# significant pairs**	**# possible pairs**	**# significant results expected by chance**
Pmx306	2	17	780	0.234 to 7.566	6	13	780	0.234 to 7.566
Pmx732	0	16	817	0.2451 to 7.9249	4	19	817	0.2451 to 7.9249
Pmx747	0	11	682	0.2046 to 6.6154	8	10	682	0.2046 to 6.6154
Pmx839	1	13	363	0.1089 to 3.5211	6	9	363	0.1089 to 3.5211

Looking at the number of negative interactions that alleles have with one another as opposed to the number of positive interactions, every microsatellite locus shows a higher number of significant negative interactions than significant positive interactions. For example, with Pmx306 there are a total of 8 significant positive interactions from 1996-2012 while there are 18 significant negative interactions over the course of 1996-2012 (Appendix A: Table 1). The largest difference between significant negative interactions and significant positive interactions is seen in Pmx732 where there are 6 significant positive interactions and 32 significant negative

interactions (Appendix A: Table 2). In total, there were 18 significant positive interactions found and 84 significant negative interactions found (Table 2).

Table 2- Totals for the number of positive and negative relationships within each microsatellite locus over the years of 1996 to 2012 (See Appendix A: Tables 1-4). Data obtained from the Pairs analysis (Ulrich, 2008) using the fixed rows and equiprobable columns for the null model for randomization (see Appendix B).

Summary of relationships for *P. mexicanum* microsatellite alleles		
	Positive Relationships (+)	**Negative Relationships (-)**
Pmx306	8	18
Pmx732	6	32
Pmx747	4	16
Pmx839	0	18
Total for all microsatellites	**18**	**84**

I then tested if there was a relationship between the difference in base pairs between the two alleles within a significant pairing and evaluation of the association as positive or negative interaction. Basically, I wanted to know if smaller differences in allele size, and therefore possibly more closely related alleles by descent, would be associated with positive or negative interactions. The same could also be examined for larger differences in allele size. Table 3 shows the data for each microsatellite locus in full. The data were examined for significance by Fisher's Exact test because the sample size was small. None were found to be significant (Table 3 and 4-7).

Table 3- The distance in base pairs between the two alleles in a significant pair and the relationship to whether it had a positive or negative interaction. This is looked at for significant pairs of each of the four microsatellite loci from 1996 to 2012 (see Appendix B for significant pairs).

Distance between significant pairs (in bp) and corresponding relationship for 1996-2012							
Pmx306		**Pmx732**		**Pmx747**		**Pmx839**	
# bp apart	**+ or -**	**# bp apart**	**+ or -**	**# bp apart**	**+ or -**	**# bp apart**	**+ or -**
3	-	2	-	3	-	3	-
3	-	2	-	3	-	3	-
3	+	5	-	3	-	3	-
6	-	5	-	3	+	3	-
9	-	5	+	9	-	3	-
9	-	9	-	15	-	6	-
12	-	10	-	21	-	9	-
12	-	11	-	21	-	9	-
12	+	11	+	21	+	9	-
21	-	14	-	39	-		
27	+	14	-				
36	-	16	-				
36	+	18	-				
		21	-				
		25	-				
		27	+				
		37	-				
		63	-				
		66	-				

Table 4- The number of smaller distances in base pairs apart (below median) or larger distances in base pairs apart (above median) for the distances between significant pairs for microsatellite locus, Pmx306 (see Table 3 for complete list of distances between pairs). Fisher's Exact Test shows there is not a significant relationship.

Pmx306 (median= 12)	**Below median**	**Above median**
Positive relationship (+)	1	2
Negative relationship (-)	5	2
Fisher's Exact Test	0.5000	

Table 5- The number of smaller distances in base pairs apart (below median) or larger distances in base pairs apart (above median) for the distances between significant pairs for microsatellite locus, Pmx732 (see Table 3 for complete list of distances between pairs). Fisher's Exact Test shows there is not a significant relationship.

Pmx732 (median= 14)	**Below median**	**Above median**
Positive relationship (+)	2	1
Negative relationship (-)	7	7
Fisher's Exact Test	1.0000	

Table 6- The number of smaller distances in base pairs apart (below median) or larger distances in base pairs apart (above median) for the distances between significant pairs for microsatellite locus, Pmx747 (see Table 3 for complete list of distances between pairs). Fisher's Exact Test shows there is not a significant relationship.

Pmx747 (median= 12)	**Below median**	**Above median**
Positive relationship (+)	1	1
Negative relationship (-)	4	4
Fisher's Exact Test	1.0000	

Table 7- The number of smaller distances in base pairs apart (below median) or larger distances in base pairs apart (above median) for the distances between significant pairs for microsatellite locus, Pmx839 (see Table 3 for complete list of distances between pairs). Median values were included in this case because they were equal to the minimum value. Fisher's Exact Test shows there is not a significant relationship.

Pmx839 (median= 3)	**Below median**	**Above median**
Positive relationship (+)	0	0
Negative relationship (-)	5	4
Fisher's Exact Test	1.0000	

To determine if the number of significant results for an allele was related to its frequency in the lizard population, I asked if there was a relationship between the allele's abundance over the years of 1996 to 2012 and the number of significant results it was involved in. The linear regressions for the significant results in relationship to abundance for all four loci, Pmx306, Pmx732, Pmx747, and Pmx839, show a positive relationship between the two variables. All four of the linear regressions are highly significant with p-values less than 0.0001 and have high R-squared values (Figures 1-4).

To determine if this result was an artifact, I then looked at the positive and negative significant results separately for all of the four loci combined. Both of these linear regressions also showed highly significant positive relationships with p-values less than 0.0001 (Figures 5-6). Since I see the same relationship for both positive and negative significant results, this suggests that the Pairs analysis is more readily able to detect significant relationships for alleles of intermediate abundance. In a conversation with Dr. Nicholas Gotelli, he confirmed that the Pairs analysis works best at finding significant results for the species of intermediate abundance.

Since our allele abundances are never above 50%, this suggests that our most common alleles are equivalent to an intermediate abundance that you would see when using the Pairs analysis for species associations. Therefore, the Pairs analysis is better able to find significant results, be them positive or negative, for the allele associations that are the most abundant or common.

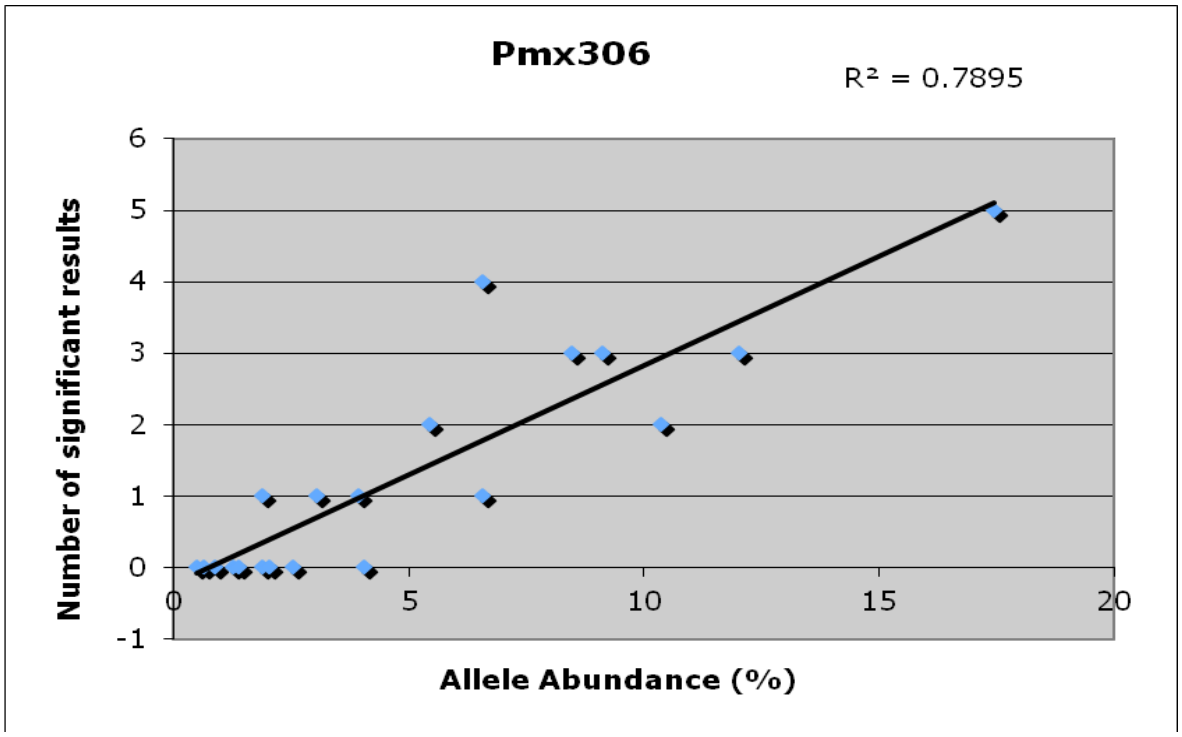


Figure 1- Pmx306 alleles with each of their corresponding abundances for 1996-2012 versus the number of significant pairs the alleles appeared in. The p-value of the linear regression is significant at <0.0001 R-squared value is 0.7895.

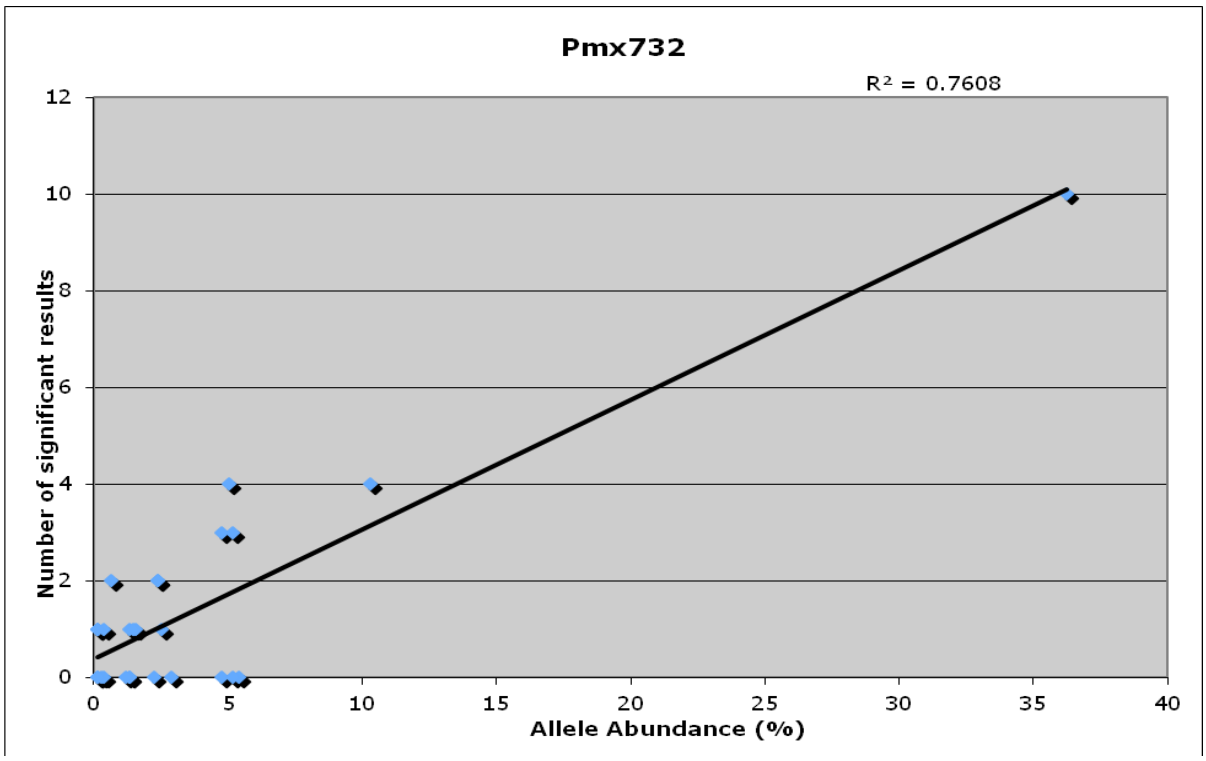


Figure 2- Pmx732 alleles with each of their corresponding abundances for 1996-2012 versus the number of significant pairs the alleles appeared in. The p-value of the linear regression is significant at <0.0001 and the R-squared value is 0.7608.

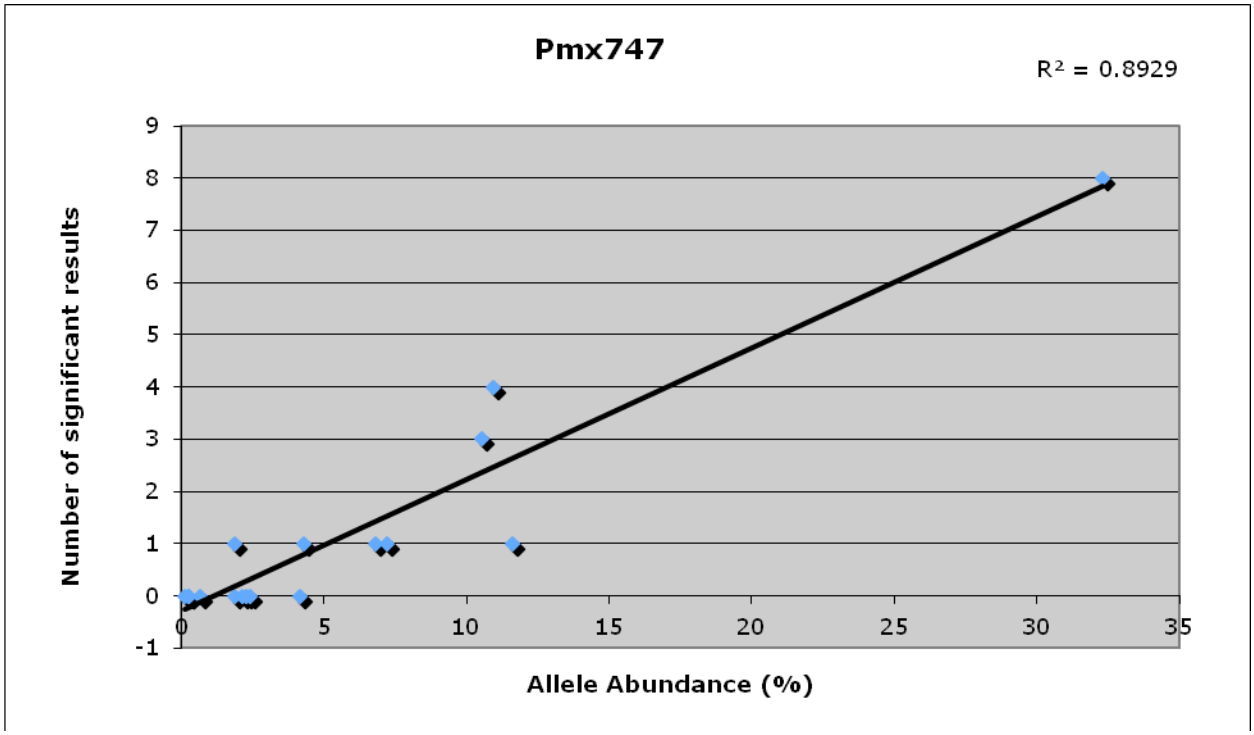


Figure 3- Pmx747 alleles with each of their corresponding abundances for 1996-2012 versus the number of significant pairs the alleles appeared in. The p-value of the linear regression is significant at <0.0001 and the R-squared value is 0.8929.

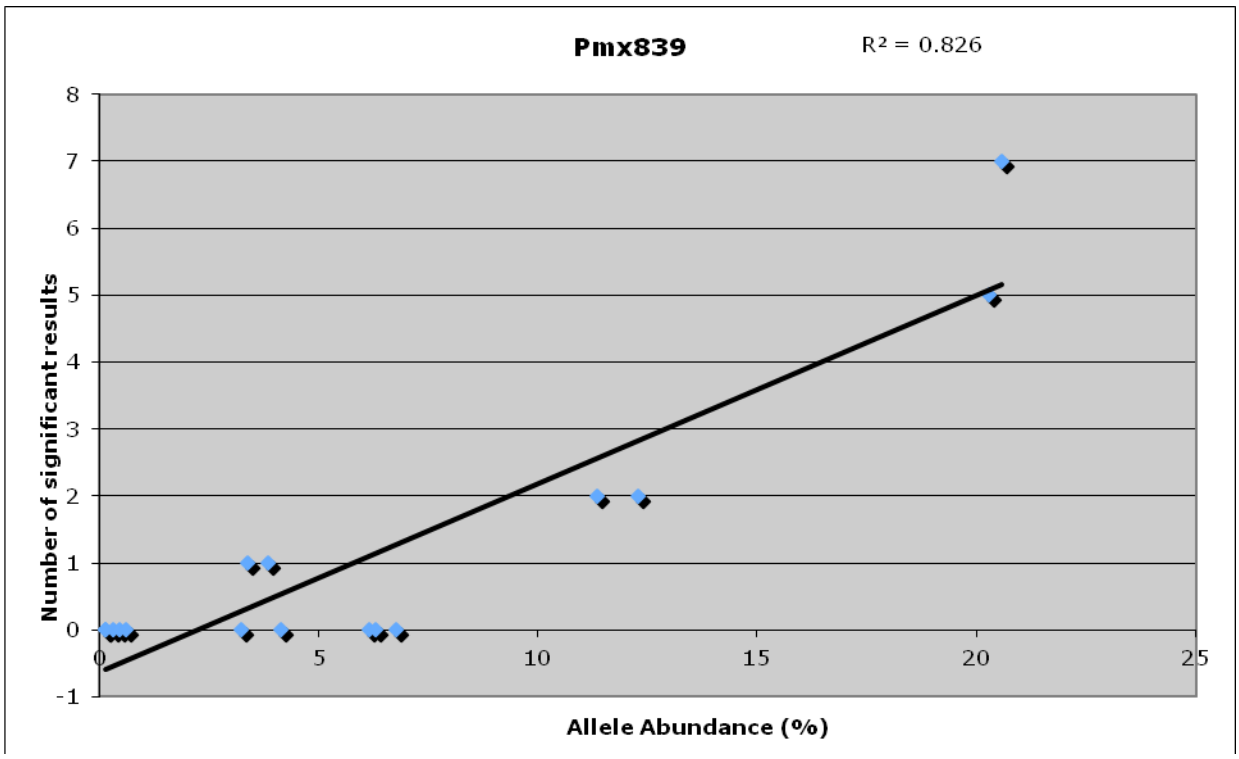


Figure 4- Pmx839 alleles with each of their corresponding abundances for 1996-2012 versus the number of significant pairs the alleles appeared in. The p-value is significant at <0.0001 and the R-squared value is 0.826.

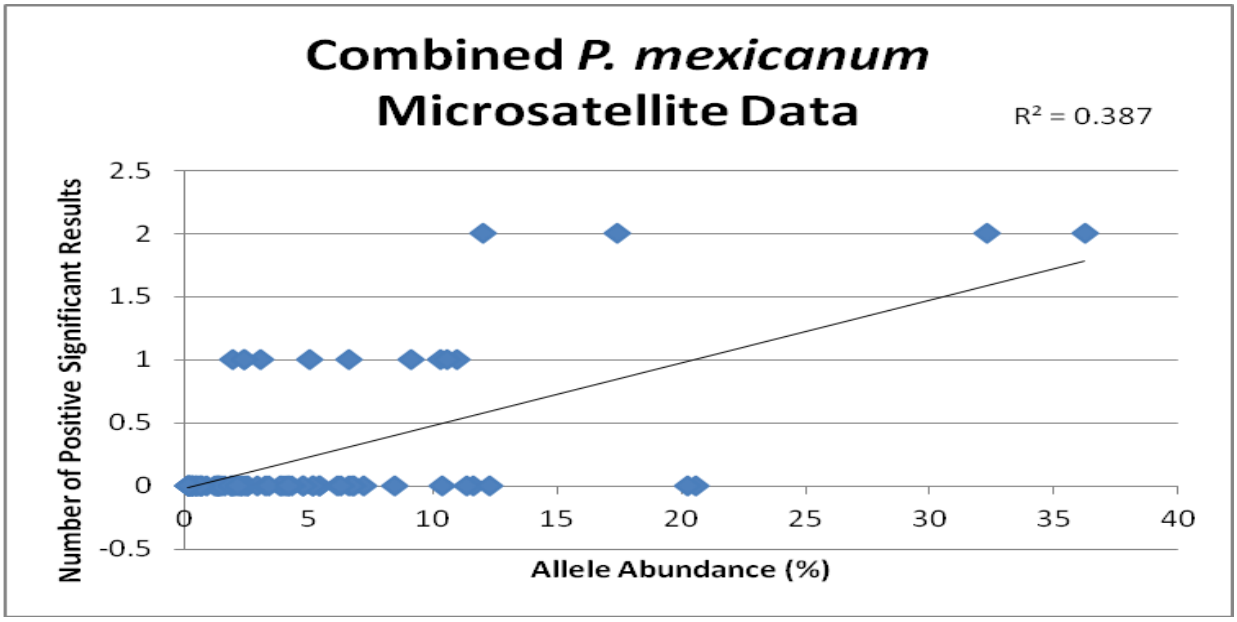


Figure 5- Combined data for all four microsatellites (Pmx306, Pmx732, Pmx747, and Pmx839) looking at the allele abundance over the years of 1996-2012 in relationship to the number of positive significant results the allele showed. The p-value of the linear regression is significant at <0.0001 and the R-squared value is 0.387.

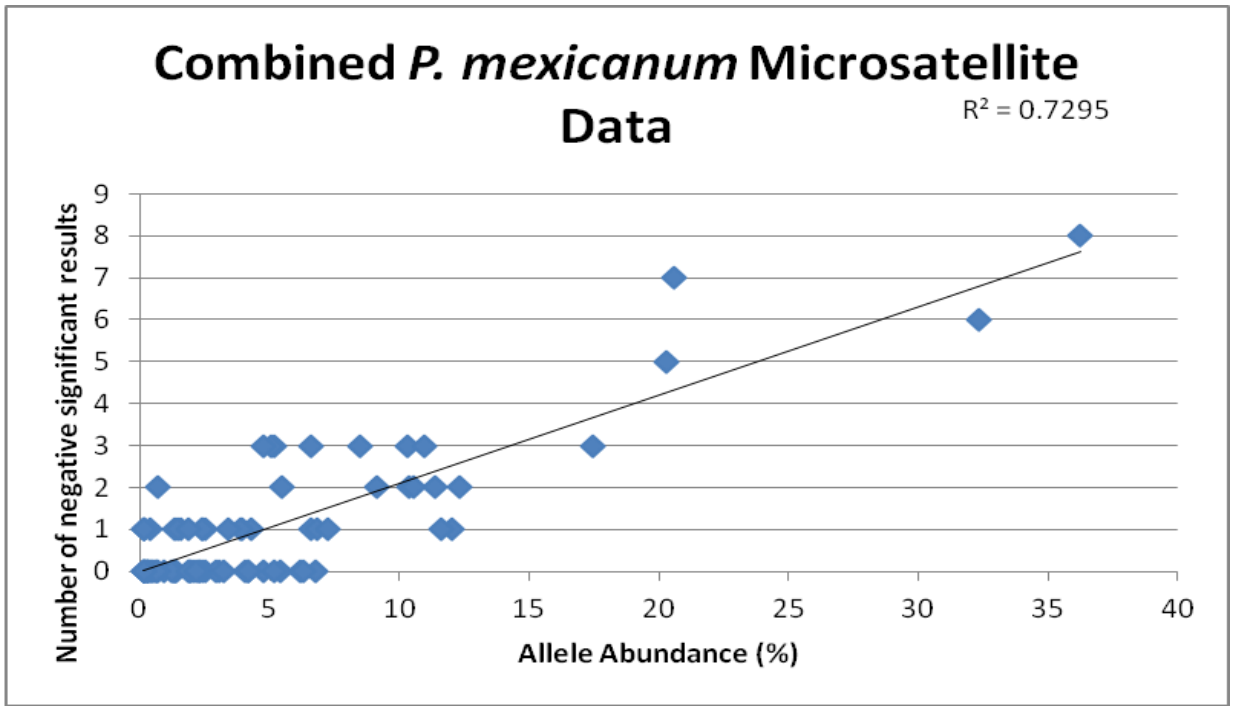


Figure 6- Combined data for all four microsatellites (Pmx306, Pmx732, Pmx747, and Pmx839) looking at the allele abundance over the years of 1996-2012 in relationship to the number of negative significant results the allele showed. The p-value of the linear regression is <0.0001 and the R-squared value is 0.7295.

Discussion

This study of microsatellite allele associations for *Plasmodium mexicanum* in its lizard host revealed significant overall associations (significant matrix) and significant pairs of alleles for all microsatellite loci. This is the first such study for any malaria parasite. Further, negative associations were more common than positive associations. No pattern was found for a measure of assumed degree of relation between alleles (length of the repeat region in the microsatellite) and a positive or negative association of those alleles.

Due to the nature of the Pairs analysis which is conservative to eliminate Type I errors for multiple comparisons, the significant results found would have to be a result of a very strong

biological effect. The tendency to have Type II errors for this analysis leads me to conclude that the significant results found are true positive and negative associations between alleles. In addition to the analysis tending towards the side of caution, the generation time of the parasite is small enough that between the year groupings you have lots of chances for recombination. Recombination of alleles may move the alleles around and break their previous linkages to coding loci, making them less likely to participate in the same relationships from year to year. With our results, there are not only significant matrices but significant pairs. This means that the pairs had a strong enough relationship to be singled out as significant within the already significant matrix. For these reasons, I have confidence in the significance of the pairs detected by the Pairs analysis.

For each of the four microsatellite loci of *P. mexicanum*, there were more negative significant associations found between allele pairs than positive significant associations. In general, looking at the number of possible allele pairs there are more non-significant neutral relationships where the alleles do not have an effect on one another. However, since the analysis is so conservative it is possible that there are more significant pairs missed due to a Type II error. I can say that for the relationships where the alleles do interact with one another there are more conflicting, negative interactions rather than positive interactions. These results support the hypothesis that there are either positive or negative interactions between alleles.

I found a significant correlation between allele frequency and the number of significant results that an allele was involved in. This relationship was also seen when separating the positive and negative significant results. This suggests that this is due to the tendency of the Pairs analysis to better identify significant pairs that have intermediate abundance. In my samples, the

alleles never rose above 50% abundance and therefore would be equivalent to the species with intermediate abundance. This is a novel way of using the Pairs analysis and it will be interesting to see if other researchers apply the analysis differently from the species associations it was designed for as well.

I did not find a relationship in terms of how far apart the alleles within a significant pair were from one another. The smaller distances between alleles or larger distances between allele pairs showed no relationship to the whether the interaction was positive or negative. Therefore, we cannot support the hypothesis that more closely related (smaller distance) alleles are in positive relationships with each other. However, the sample size for that analysis was small, hence the use of the Fisher's Exact Test, and it would be interesting to test this idea with an even larger sample size.

In conclusion, I find that clones vary in the types of interactions they have with other clones within an infection with the majority of them being neutral towards one another, a portion of them conflicting with one another, and a smaller portion of them positively interacting with one another. These conclusions support Vardo-Zalik and Schall's data which show that the rate of increase for meronts and gametocytes were variable in mixed-clone infections of *P. mexicanum*. Some of the multiclonal infections they followed had higher rates of parasite replication and parasitemia whereas other multiclonal infections did not vary from the single clone infections in their rates of replication or parasitemia (2009). It seems as though there is not a single type of interaction occurring that accounts for multiclonal infection dynamics. Instead, genotypes vary in their interactions.

Chapter 3: Conclusions and Future Directions

The process of writing a thesis that will hopefully someday become a published scientific paper has been an extremely rewarding experience. Through the mentorship of Dr. Joseph Schall and the rest of the Schall lab I have learned the diligence and thoroughness it requires to conduct a proper scientific experiment. There are so many details that go into completing a scientific study that you don't realize when you are reading the final product in a journal. Not to mention the incredible amount of work ethic it takes. It's hard to work on a project for so long and be awaiting the results. When everything works out you feel so accomplished. Thankfully, I got to experience that feeling while writing this thesis on my allele association study. On the other hand, I also learned the frustrations that researchers go through when a project isn't working out the way you had hoped. We had a whole other project on the genetics of the vectors for the *P. mexicanum* system that I worked on for what felt like endless hours that hit a roadblock. No matter how much troubleshooting we did the project remained at a standstill. Recently, Dr. Schall has been testing new microsatellite primers for *Lutzomyia vexator* and has found a few that are amplifying well. This seems promising and I am excited to see the conclusion of the vector geographic genetic differentiation study. Hopefully, someone will be able to pick up where I left off and use the many samples of *Lutzomyia vexator* I collected last summer. Finishing the vector geographic genetic differentiation study would help complete the overall picture since similar studies have been done by Fricke *et al*. on the parasite, *Plasmodium mexicanum*, and the vertebrate host, *Scleroporus occidentalis*.

I hope that the findings of our research will spark new questions for other researchers, including future students in the lab. I think it would be very interesting to look at the genetics of

the social amoeba or another more evidently cooperative organism with a similar analysis technique to this study. Would you see more positive interactions when compared to the *P. mexicanum* study? Another idea that would be intriguing to look at would be to look at the relative abundance of each allele within an infection in relation to which allele pairs are interacting positively or negatively. We would expect to see infections with clones of equal abundance in positive relationships or neutral relationships and clones where one dominates in abundance to have negative relationships. In addition to this, it would also be interesting to follow the relative clone abundance for significant pairs over time. However, it would take a lot of time and resources to set up a longitudinal study that looks at this question.

Acknowledgements

I would like to first thank Dr. Joseph Schall for his continuous guidance over the course of my career as an undergraduate, and more specifically for his help during the execution of this thesis. I would also like to thank Allison Neal for her invaluable input for this thesis and her help with the R programming needed to get the analysis done. In addition, I would like to thank Dr. Anne Vardo-Zalik for her mentorship during my fieldwork at HREC. I would like to thank Krystina Katterman for her help with my molecular work during the fall semester. Also, I would like to thank Dr. Nicholas Gotelli for helping us understand the Pairs program for the analysis. Finally, I would like to give my sincere appreciation to Dr. Joseph Schall, Dr. Nicholas Gotelli, and Dr. Yolanda Chen for taking the time to give me feedback and be a part of my thesis committee.

Literature Cited

Dawkins R. 1990. Parasites, desiderata lists and the paradox of the organism. *Parasitology* **100**: 563-573.

De Roode J.C., Helinski M. E. H., Anwar M.A., and Read A.F. 2005. Dynamics of multiple infection and within-host competition in genetically diverse malaria infections. *The American Naturalist* **166**(5): 531-542.

Fidock D.A., Nomura T., Talley A.K., Cooper R.A., Dzekunov S.M., Ferdig M.T., Ursos L.M.B., Sidhu A.B.S., Naudé B., Deitsch K.W., Su X., Wootton J.C., Roepe P.D., and Wellems T.E. 2000. Mutations in the P. falciparum Digestive Vacuole Transmembrane Protein PfCRT and Evidence for Their Role in Chloroquine Resistance. *Molecular Cell* **6**: 861-871.

Ford A.F., Vardo-Zalik A.M., and Schall, J.J. 2010. Relative clonal density of malaria parasites in mixed-genotype infections: validation of a technique using microsatellite markers for *Plasmodium falciparum* and *P. mexicanum*. *Journal of Parasitology* **96**(5): 908-913.

Gotelli N.J. and Ulrich W. 2010. The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia* **162**: 463-477.

Kongkasuriyachai D., Bartels-Andrews L., Stowers A., Collins W.E., Sullivan J., Sattabongkot J., Torii M., Tsuboi T., and Kumar N. 2004. Potent immunogenicity of DNA vaccines encoding *Plasmodium vivax* transmission-blocking vaccine candidates Pvs25 and Pvs28- evaluation of homologous and heterologous antigen-delivery prime-boost strategy. *Vaccine* **22**: 3205-3213.

Kumamoto C.A. 2002. *Candida* biofilms. Current Opinion in Microbiology **5**: 608-611.

Li S. I. and Purugganan M.D. 2011. The cooperative amoeba: *Dictyostelium* as a model for social evolution. *Trends in Genetics* **27** (2): 48-54.

R Core Team. 2012. R: A Language and Environment for Statistical Computing. http://www.R-project.org

Read L.H. and Read A.F. 1998. Determinants of transmission success of individual clones from mixed-clone infections of the rodent malaria parasite, *Plasmodium chabaudi*. *International Journal for Parasitology* **28**: 719-725.

Sachs J. and Malaney P. 2002. The economic and social burden of malaria. *Nature* **415**: 680-685.

Schall J.J, Bennett A.F., and Putnam R.W. 1982. Lizards Infected with Malaria: Physiological and Behavioral Consequences. *Science* **217**: 1057-1059.

Schall J.J. and St. Denis K.M. 2013. Microsatellite loci over a thirty-three year period for a malaria parasite (*Plasmodium mexicanum*): bottleneck in effective population size and effect on allele frequencies. Parasitology **140**: 21-28.

Schall J.J. and Vardo A.M. 2007. Identification of microsatellite markers in *Plasmodium mexicanum*, a lizard malaria parasite that infects nucleated erythrocytes. *Molecular Ecology Notes* **7**: 227-229.

Ulrich W. 2008. Pairs- a FORTRAN program for studying pair wise species associations in ecological matrices. www.uni.torun.pl/~ulrichw

Vardo-Zalik A.M. 2009. Clonal diversity of a malaria parasite, *Plasmodium mexicanum*, and its transmission success from its vertebrate-to-insect host. *International Journal for Parasitology* **39**: 1573-1579.

Vardo-Zalik A.M. and Schall J.J. 2009. Clonal diversity alters infection dynamics of a malaria parasite (*Plasmodium mexicanum*) and its vertebrate hosts. *Ecology* **90**(2): 529-536.

World Health Organization. 2012. World Malaria Report: 2012.

Appendix A

Table 1- The number of relationships, either positive or negative, for each allele of the microsatellite locus Pmx306 for *P. mexicanum* for the years 1996 to 2012. Data obtained from the Pairs analysis (Ulrich, 2008) using the fixed row and equiprobable column null model for randomization (see Appendix B).

	Number of Relationships (using row fixed data)	
Allele for Pmx306	**Positive Relationships (+)**	**Negative Relationships (-)**
166	0	0
169	**1**	**2**
172	**2**	**3**
175	0	**2**
178	0	0
181	**1**	**3**
184	**2**	**1**
187	0	**3**
190	0	0
193	0	**2**
196	0	**1**
199	**1**	0
202	0	0
205	**1**	0
208	0	0
211	0	0
214	0	0
217	0	**1**
220	0	0
223	0	0
Total for Pmx306	**8**	**18**

Table 2- The number of relationships, either positive or negative, for each allele of the microsatellite locus Pmx732 for *P. mexicanum* for the years 1996 to 2012. Data obtained from the Pairs analysis (Ulrich, 2008) using the fixed row and equiprobable column for the null model for randomization (see Appendix B).

	Number of Relationships (using row fixed data)	
Allele for Pmx732	**Positive Relationships (+)**	**Negative Relationships (-)**
186	0	0
189	0	0
191	0	0
193	0	0
196	0	**1**
206	0	**1**
213	0	**1**
216	0	**1**
220	0	0
222	0	**2**
225	0	0
228	0	**1**
240	0	**1**
250	0	**1**
256	0	0
258	**1**	**1**
260	**1**	**1**
263	0	0
264	0	0
265	0	0

268	0	0
271	**2**	**8**
273	0	**3**
276	**1**	**3**
279	0	0
282	0	**3**
285	**1**	**3**
289	0	0
292	0	**1**
Total for Pmx732	**6**	**32**

Table 3- The number of relationships, either positive or negative, for each allele of the microsatellite locus Pmx747 for *P. mexicanum* for the years 1996 to 2012. Data obtained from the Pairs analysis (Ulrich, 2008) using the fixed row and equiprobable column for the null model for randomization (see Appendix B).

	Number of Relationships (using row fixed data)	
Allele for Pmx747	**Positive Relationships (+)**	**Negative Relationships (-)**
160	0	**1**
163	0	0
166	0	**1**
169	0	0
172	**1**	**3**
175	**2**	**6**
178	0	**1**
181	0	0
184	0	0
187	0	0
190	0	0
193	0	**1**
196	**1**	**2**
199	0	**1**
202	0	0
205	0	0
208	0	0
211	0	0
Total for Pmx747	**4**	**16**

Table 4- The number of relationships, either positive or negative, for each allele of the microsatellite locus Pmx839 for *P. mexicanum* for the years 1996 to 2012. Data obtained from the Pairs analysis (Ulrich, 2008) using the fixed row and equiprobable column for the null model for randomization (see Appendix B).

	Number of Relationships (using row fixed data)	
Allele for Pmx839	**Positive Relationships (+)**	**Negative Relationships (-)**
247	0	0
250	0	0
253	0	0
256	0	0
259	0	**1**
262	0	**7**
265	0	**5**
268	0	**2**
271	0	**2**
274	0	**1**
277	0	0
285	0	0
290	0	0
293	0	0
296	0	0
298	0	0
Total for Pmx839	**0**	**18**

Appendix B

Locus ____306____ Randomization = Fixed-Fixed

Years	Fixed-Fixed
1996-1997	N alleles =17 N lizards=66 N possible pairs = 136 Sig of matrix = no Sig pairs Na/Nb/Nboth, +/-, Sig **172/169 31/22/2 - 0.00000458** **181/193 19/16/8 + 1.00000000**
1998-2002	N alleles =17 N lizards= 63 N possible pairs = 136 Sig of matrix = yes Sig pairs Na/Nb/Nboth, +/-, Sig **172/193 19/9/5 + 1.00000000** **172/199 19/8/6 - 0.01468766** **169/181 15/9/6 - 0.00347280** **169/187 15/4/3 - 0.00266882** **181/187 9/4/3 - 0.00047709**
2003-2005	N alleles =13 N lizards=65 N possible pairs = 78 Sig of matrix = no Sig pairs Na/Nb/Nboth, +/-, Sig **172/181 22/12/6 - 0.03251427**
2006	N alleles =15 N lizards=62 N possible pairs = 105 Sig of matrix = yes Sig pairs Na/Nb/Nboth, +/-, Sig **169/205 9/6/3 - 0.00664201** **181/217 6/4/2 - 0.00138409**
2007-2008	N alleles =13 N lizards=59 N possible pairs = 78 Sig of matrix = no Sig pairs Na/Nb/Nboth, +/-, Sig **172/175 16/4/4 - 0.00003764**

2009	N alleles =14 N lizards=73 N possible pairs = 91 Sig of matrix = no Sig pairs Na/Nb/Nboth, +/-, Sig **196/223 15/5/3 - 0.00211563** **193/169 5/1/1 - 0.00000000** **178/166 4/1/1 - 0.00000000**
2010	N alleles =13 N lizards=72 N possible pairs = 78 Sig of matrix = no Sig pairs Na/Nb/Nboth, +/-, Sig **187/223 19/2/2 - 0.00064875** **193/178 8/7/3 - 0.00051553**
2011-2012	N alleles =13 N lizards= 61 N possible pairs = 78 Sig of matrix = no Sig pairs Na/Nb/Nboth, +/-, Sig **169/199 2/2/1 - 0.00000000**

Locus ____306____ Randomization= Row Fixed

Years	Row Fixed
1996-1997	N alleles =17 N lizards=66 N possible pairs = 136 Sig of matrix = no Sig pairs Na/Nb/Nboth, +/-, Sig **172/169 31/22/2 - 0.00000196**
1998-2002	N alleles =17 N lizards=63 N possible pairs = 136 Sig of matrix = yes Sig pairs Na/Nb/Nboth, +/-, Sig **172/184 19/8/0 + 1.00000000** **172/199 19/8/6 + 1.00000000** **169/181 15/9/6 - 0.03389327** **181/187 9/4/3 - 0.00424864**
2003-2005	N alleles =13 N lizards=65 N possible pairs = 78 Sig of matrix = yes Sig pairs Na/Nb/Nboth, +/-, Sig **172/193 22/9/0 - 0.02681080**
2006	N alleles =15 N lizards=62 N possible pairs = 105 Sig of matrix = yes Sig pairs Na/Nb/Nboth, +/-, Sig **184/181 23/6/0 + 1.00000000** **169/205 9/6/3 + 1.00000000** **181/217 6/4/2 - 0.02582812**
2007-2008	N alleles =13 N lizards=59 N possible pairs =78 Sig of matrix = no Sig pairs Na/Nb/Nboth, +/-, Sig **184/193 20/17/1 - 0.00666484** **172/175 16/4/4 - 0.00346461**

2009	N alleles =14 N lizards=73 N possible pairs = 91 Sig of matrix = yes Sig pairs Na/Nb/Nboth, +/-, Sig **187/175 23/11/0 - 0.00387218**
2010	N alleles =13 N lizards=72 N possible pairs =78 Sig of matrix = yes Sig pairs Na/Nb/Nboth, +/-, Sig **187/196 19/17/1 - 0.04356450**
2011-2012	N alleles =13 N lizards=61 N possible pairs =78 Sig of matrix = yes Sig pairs Na/Nb/Nboth, +/-, Sig

Locus ___732___ Randomization= Fixed-Fixed

Years	Fixed-Fixed
1996-1997	N alleles = 14 N lizards=47 N possible pairs =91 Sig of matrix = no Sig pairs Na/Nb/Nboth, +/-, Sig **196/282 2/1/1 + 1.00000000** **196/216 2/1/1 - 0.00000000** **282/216 1/1/1 - 0.00000000**
1998-2002	N alleles =19 N lizards=66 N possible pairs =171 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **271/273 41/16/1 - 0.00000273** **271/282 41/7/1 + 1.00000000** **193/282 18/7/5 - 0.01165951** **273/282 16/7/5 - 0.02099086** **273/279 16/5/4 - 0.02264307** **285/222 7/2/2 - 0.00007772** **260/276 6/3/2 - 0.00221975**
2003-2005	N alleles =17 N lizards=66 N possible pairs =136 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **189/260 2/2/1 - 0.00000000** **213/250 1/1/1 - 0.00000000**
2006	N alleles =14 N lizards=61 N possible pairs =91 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **271/276 50/5/1 - 0.00323471** **222/240 1/1/1 - 0.00000000**
2007-2008	N alleles =14 N lizards=58 N possible pairs =91 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **258/285 7/6/3 - 0.00275857**

2009	N alleles =12 N lizards=71 N possible pairs =66 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig
2010	N alleles =12 N lizards=67 N possible pairs =66 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig
2011-2012	N alleles =15 N lizards=61 N possible pairs =105 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **265/289 3/1/1 - 0.00000000**

Locus ____732____ Randomization= Row Fixed

Years	Row Fixed
1996-1997	N alleles =14 N lizards=47 N possible pairs =91 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **271/260 17/7/0 + 1.00000000** **282/216 1/1/1 - 0.00000000**
1998-2002	N alleles =19 N lizards=66 N possible pairs =171 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **271/273 41/16/1 - 0.00000000** **271/282 41/7/1 - 0.02368102** **271/292 41/4/0 - 0.01379507** **273/282 16/7/5 - 0.03537530** **285/222 7/2/2 - 0.00096910** **260/276 6/3/2 - 0.00250950** **196/206 1/1/1 - 0.00000000**
2003-2005	N alleles =17 N lizards=66 N possible pairs =136 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **271/285 32/12/2 - 0.03718250** **271/273 32/11/1 - 0.00768545** **213/250 1/1/1 - 0.00000000**
2006	N alleles =14 N lizards=61 N possible pairs =91 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **271/276 50/5/1 - 0.00182297** **228/258 2/1/1 - 0.00000000** **222/240 1/1/1 - 0.00000000**

2007-2008	N alleles =14 N lizards=58 N possible pairs =91 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **271/276 35/5/0 - 0.01226525** **258/285 7/6/3 + 1.00000000**
2009	N alleles =12 N lizards=71 N possible pairs =66 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **271/276 41/8/1 + 1.00000000**
2010	N alleles =12 N lizards=67 N possible pairs =66 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **271/285 44/18/7 - 0.04785982**
2011-2012	N alleles =15 N lizards=61 N possible pairs =105 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig

Locus ____747____ Randomization= Fixed-Fixed

Years	Fixed-Fixed
1996-1997	N alleles =16 N lizards=68 N possible pairs =120 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **172/196 25/11/7 + 1.00000000** **211/208 2/1/1 - 0.00000000**
1998-2002	N alleles =15 N lizards=67 N possible pairs =105 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **196/178 9/3/2 - 0.00280694** **184/211 8/5/2 - 0.02148565**
2003-2005	N alleles =14 N lizards=63 N possible pairs =91 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **175/193 41/18/13 - 0.02681384** **178/211 3/1/1 - 0.00000000**
2006	N alleles =13 N lizards=53 N possible pairs =78 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig
2007-2008	N alleles =13 N lizards=59 N possible pairs =78 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **175/199 31/5/4 + 1.00000000**
2009	N alleles =12 N lizards=77 N possible pairs =66 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **175/172 37/15/0 - 0.02952509**
2010	N alleles =12 N lizards=69 N possible pairs =66 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **196/166 13/7/3 - 0.00801665**

2011-2012	N alleles =13 N lizards=61 N possible pairs =78 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **160/199 5/4/2 - 0.00000000** **178/193 4/3/2 - 0.00000000**

Locus ____747____ Randomization= Row Fixed

Years	Fixed-Fixed
1996-1997	N alleles =16 N lizards=68 N possible pairs =120 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **175/172 28/25/5 - 0.02101758**
1998-2002	N alleles =15 N lizards=67 N possible pairs =105 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig
2003-2005	N alleles =14 N lizards=63 N possible pairs =91 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **175/196 41/9/1 - 0.00007733**
2006	N alleles =13 N lizards=53 N possible pairs =78 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **175/172 25/12/1 - 0.00387045**
2007-2008	N alleles =13 N lizards=59 N possible pairs =78 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **175/196 31/8/1 + 1.00000000**
2009	N alleles =12 N lizards=77 N possible pairs =66 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **175/172 37/15/0 - 0.00000170**
2010	N alleles =12 N lizards=69 N possible pairs =66 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **175/172 28/11/1 + 1.00000000**

2011-2012	N alleles =13 N lizards=61 N possible pairs =78 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **175/196 27/12/0 - 0.00009911** **175/166 27/7/0 - 0.01824118** **160/199 5/4/2 - 0.00131691** **178/193 4/3/2 - 0.00031526**

Locus ____839____ Randomization= Fixed-Fixed

Years	Fixed-Fixed
1996-1997	N alleles =11 N lizards=47 N possible pairs =55 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **262/265 24/15/0 - 0.00003224** **262/256 25/5/4 + 1.00000000**
1998-2002	N alleles =10 N lizards=49 N possible pairs =45 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **265/271 16/12/4 - 0.04020088**
2003-2005	N alleles =9 N lizards=57 N possible pairs =36 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **274/290 9/7/3 - 0.01848335** **290/262 7/5/2 - 0.00186371**
2006	N alleles =11 N lizards=58 N possible pairs =55 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig
2007-2008	N alleles =10 N lizards=58 N possible pairs =45 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **271/268 19/18/10 + 1.00000000** **271/265 19/13/7 + 1.00000000** **265/274 13/5/4 - 0.01624288**
2009	N alleles =11 N lizards=77 N possible pairs =55 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **262/253 28/10/4 - 0.00542558** **265/271 18/9/3 - 0.02544106** **290/250 3/2/1 - 0.00000000**

2010	N alleles =9 N lizards=73 N possible pairs =36 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **262/290 27/3/2 - 0.00625503** **271/253 13/7/2 - 0.01393728**
2011-2012	N alleles =9 N lizards=61 N possible pairs =36 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig

Locus ____839____ Randomization= Row Fixed

Years	Row Fixed
1996-1997	N alleles =11 N lizards=47 N possible pairs =55 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **262/265 24/15/0 - 0.00000000** **262/259 24/8/1 - 0.01327716**
1998-2002	N alleles =10 N lizards=49 N possible pairs =45 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig
2003-2005	N alleles =9 N lizards=57 N possible pairs =36 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig
2006	N alleles =11 N lizards=58 N possible pairs =55 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig
2007-2008	N alleles =10 N lizards=58 N possible pairs =45 Sig of matrix =no Sig pairs Na/Nb/Nboth, +/-, Sig **262/268 28/18/4 - 0.00730854** **265/274 13/5/4 - 0.03469687**
2009	N alleles =11 N lizards=77 N possible pairs =55 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **262/265 28/18/0 - 0.00000410** **262/271 28/9/0 - 0.02856360**
2010	N alleles =9 N lizards=73 N possible pairs =36 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **262/265 27/17/0 - 0.00001612** **262/271 27/13/0 - 0.00133250**

2011-2012	N alleles =9 N lizards=61 N possible pairs =36 Sig of matrix =yes Sig pairs Na/Nb/Nboth, +/-, Sig **265/268 22/14/1 - 0.00939345**

Appendix C

The following is an example of a matrix used in this study. The columns represent the infections while the rows represent the alleles. This particular matrix is for the Pmx306 locus for the year grouping 1996-1997. It has 66 infections and 17 alleles.

306alleleanalysisfinaltable19961997.txt

NA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	1	1	1	1	0	1	0	1	1	0	1	0
3	1	0	0	0	0	1	0	0	1	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	0	1	1	1	1	1	0	1	0	0	1	0	0	0	0	1	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	1	0	1	1	1	0	1	1	0	1	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0
7	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	1
9	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	0	1	0	0	0	0	0	1	0	1	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
16	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The following is an example of the printouts that were used from the Pairs analysis (Ulrich, 2008). The example given is for the 1996-1997 time period for the locus Pmx306 using the fixed rows and fixed columns randomization.

CoocPairs.txt file:

306CoocPairs19961997 s random - Notepad

File Edit Format View Help

```
>  File:  306alleleanaly Species:   17 Sites:   66 MatFill: 0.13 Occ:    143  5.00% Confidence limit  Model: fixed - fixed
Index        File                In      SimIn  StdDevIn      Z-In   StIndex    SkewIn LowerCLIn UpperCLIn
Checkerboard 306alleleanaly      83      81.19      3.23      0.56    0.0223     -0.04     75.00     87.00
Combinations 306alleleanaly       0       0.01      0.10     -0.10   -1.0000     10.15      0.00      0.00
C-Score      306alleleanaly   0.019      0.019     0.000     0.596    0.0077     0.048     0.018     0.019
Soerensen    306alleleanaly  0.0706     0.0712    0.0048   -0.1324   -0.0089    0.5449    0.0632    0.0836
Absences     306alleleanaly 1964.662  1964.220     2.234     0.198    0.0002    -0.001  1960.471  1968.603
BR           306alleleanaly      77      78.28      2.43     -0.53    0.5385      0.16     74.00     83.00
Correlation  306alleleanaly  0.1429     0.1397    0.0054    0.5996    0.0233    0.0618    0.1304    0.1501
Variancetest 306alleleanaly  6.4197     6.4197   -0.0001    0.0000    0.0000    0.0000    6.4197    6.4197
```

SigPairs.txt file:

306SigPairs19961997 s random - Notepad

File Edit Format View Help

```
<  File:  306alleleanaly Species:   17 Sites:   66 MatFill: 0.13 Occ:    143 5.000% Confidence limit  Index: c Model: fixed - fixed

#    No           Sp1              Sp2     S1  S2 Com  Obs.Score  Exp.Score  Exp.StDev   Skewness    LowerCL    UpperCL SigZ-Score      Alpha >MeanScore  >CLScore  >BJScore      Alpha
#     1 3              2                31  22   2      0.850      0.406      0.090      0.171      0.238      0.587       4.92 0.00000083       4.92      0.00       4.57 0.00000458
#    32 5              9                19  16   8      0.289      0.585      0.138      0.163      0.355      0.888      -2.15 0.03135121       0.00      0.00       0.00 1.00000000

> Name            SP   CL+   CL-   BM+   BM-   BC+   BC-   BY+   BY-       NR   CL+   CL+
> 306alleleanalys  136     1     1     1     0     0     0     1     0    13600   408   408

& File          Speci   Individ     BRInd       BRZ Cscoreind   CscoreZ Soeren.In  Soeren.Z    AbsInd     AbsZ CheckerIn  CheckerZ   CombInd    CombZ   CorrInd    CorrZ   SchInd      SchZ
& 306alleleanaly   17   143.00    0.5385   -0.5260    0.0077    0.5962   -0.0089   -0.1324    0.0002    0.1975    0.0223    0.5609   -1.0000   -0.1005    0.0233    0.5996    0.0000    0.0000

%         Observed numbers                                                  Expected numbers
% Class    Pairs   CL+   CL-   BM+   BM-   BC+   BC-   BY+   BY-       NR   CL+   CL-
% 0.000        5     0     0     0     0     0     0     0     0      587    15    22
% 0.025        0     0     0     0     0     0     0     0     0        0     0     0
% 0.075        0     0     0     0     0     0     0     0     0        0     0     0
% 0.125        0     0     0     0     0     0     0     0     0        2     1     0
% 0.175        1     0     0     0     0     0     0     0     0       21     0     0
% 0.225        0     0     0     0     0     0     0     0     0       94     3     3
% 0.275        2     0     1     0     0     0     0     0     0      109     4     3
% 0.325        3     0     0     0     0     0     0     0     0      199    11     4
% 0.375        2     0     0     0     0     0     0     0     0      244     7     6
% 0.425        3     0     0     0     0     0     0     0     0      394     5    11
% 0.475       17     0     0     0     0     0     0     0     0     1172    37    37
% 0.525        3     0     0     0     0     0     0     0     0      369    10    11
% 0.575        1     0     0     0     0     0     0     0     0      293     8     7
% 0.625        4     0     0     0     0     0     0     0     0      611    19    23
% 0.675        2     0     0     0     0     0     0     0     0      268     8     7
% 0.725        7     0     0     0     0     0     0     0     0      576    20    17
% 0.775        1     0     0     0     0     0     0     0     0      253     3    12
% 0.825        0     0     0     0     0     0     0     0     0      163     8     5
% 0.875        2     1     0     1     0     0     0     1     0       66     0     2
% 0.925        0     0     0     0     0     0     0     0     0        0     0     0
% 0.975        0     0     0     0     0     0     0     0     0        0     0     0
% 1.000       83     0     0     0     0     0     0     0     0     8179   249   238
```