THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOLOGY

MOLECULAR EVOLUTIOINARY GENETICS OF THE *DROSOPHILA PSEUDOOBSCURA* THIRD CHROMOSOME REARRANGEMENTS.

IAN STUART LEOPOLD
SPRING 2014

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biology
with honors in Biology

Reviewed and approved* by the following:

Stephen W. Schaeffer
Professor of Biology
Thesis Supervisor

Bernhard Lüscher
Professor of Biology and Professor of Biochemistry and Molecular Biology
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

The genetic forces that establish chromosomal rearrangements in natural populations are not clearly understood. An analysis of nucleotide and amino acid variation of genes in a Drosophila chromosome that is naturally polymorphic for different gene arrangements generated by a series of paracentric inversions is used to investigate the molecular genomic mechanisms that establish chromosomal rearrangements in natural populations. Cytogenetic analysis is used to identify the different gene orders using the banding and puffing patterns of the polytene chromosomes isolated from the salivary glands of *D. pseudoobscura* larvae. The different gene arrangements or gene orders are named for the location where the chromosomes were originally collected. Chromosomes of 50 *D. pseudoobscura* strains collected from the southwestern United States and Mexico were classified via cytogenetic analysis. A subset of these strains was collected from Pablo Etla in Oaxaca, Mexico. The initial analysis identified two gene arrangements in this sample, Tree Line and Cuernavaca. The Cuernavaca arrangement is derived from the Santa Cruz arrangement by a single inversion step and should generate a single large inversion loop in a Santa Cruz/Cuernavaca heterozygote. Cytogenetic analysis of Santa Cruz/Cuernavaca heterozygotes revealed that the putative Cuernavaca strain produced a looping pattern consistent with two rather than one inversion differences between the strains. The strains carrying the putative Cuernavaca arrangement were determined to carry the Paxtepec arrangement.

After the cytogenetic identifications were complete, the genomic sequences of the 50 flies from five different gene arrangements of *D. pseudoobscura* were generated using NextGen sequencing. Using this extensive data set, the pattern and organization of nucleotide and amino acid variation can be examined to determine the evolutionary forces responsible for the origin and establishment of the chromosomal rearrangements. An analysis of variable amino acid sites in the coding sequences of genes from the third chromosome inferred the ancestral and mutation-derived changes to the amino acid sequences. The analysis wished to detect amino acid changed chemical classes that were at high

frequency within particular gene arrangements, which might be associated with an adaptive event acting on the chromosomal variants. The majority of amino acid substitutions changed stayed within chemical class (non-polar, polar, acidic, or basis). Of the amino acid changes that went between chemical classes, polar to nonpolar mutations were the most common type. A gene ontology (GO) analysis on proteins with a high frequency of non-conservative chemical changes to identify any functions of the genes or their homologues in *Drosophila melanogaster* that might help to explain the stable clinal pattern formed by the third chromosome arrangements found in the southwestern United States.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to first thank Dr. Stephen Schaeffer for his efforts, not only for his training me in the technique of salivary extraction and squashes, but also as a tireless mentor and friend for the duration of the time I have had the privilege of working under his tutelage. The hours he has put in to ensure that I completed this research and thesis are a testament to his commitment to not only me, but all of his students and colleagues. Also Dr. Schaeffer should be credited for the useful reference figure found in appendix B.

I would like to thank Zachary Fuller for the figure found in appendix C and for his knowledge and expertise in computer programming that allowed me to quickly and effectively analyze the extensive sequence data I was able to help generate. Thanks to Dominique Cowart for her teaching of basic sequencing and ExoSap techniques. Dr. Andre was integral in the vetting process my laboratory skills when I first began to work for Dr. Schaeffer and also for his guidance in learning proper DNA extraction technique.

Thank you Dr. Bernard Lüscher for your quick responses to my emails and being flexible with the timeframe needed to properly review my thesis.

Chapter 1

Introduction

1.1 Polytene Chromosomes and Chromosomal Inversions

The study of polytene chromosomes from the salivary glands of Drosophila larvae has a storied and well documented history. The method of observing these chromosomes was developed as an improvement over studying translocations and cytological maps through observation of metaphase chromosomes. With the advent of this new method, it was discovered that the chromosomes of older larvae can undergo somatic synapsis where homologous chromosomes unite into one apparent element, band for band. However, if the chromosomes were heterozygous and containing inverted or deleted regions, the chromosomes reflect the changes in their structure and are easily categorized and annotated cytologically. The ability to directly identify large scale deletions and inversions on chromosomes of *Drosophila melanogaster* allowed researchers at the time to directly study the problems associated with these chromosomal "aberrations" for the first time. In this novel work done by Painter (1934), the procedure for staining and squashing the salivary glands was outlined and is generally the same method used in laboratories today to examine the evolution of gene order and map genomic contigs to the Drosophila genome (Schaeffer et al. 2008). It was concluded from this work that there was sufficient consistency in the patterns of banding and puffs in the salivary gland chromosomes that chromosomal aberrations could be repeatedly identified (Painter, 1934).

From Painter's discovery of polytene chromosomes in *D. melanogaster*, intensive study of related fruit fly species began in *D. pseudoobscura, D. persimilis*, and *D. simulans* to infer their cytogenetic relationships to each other. It was through these initial studies of chromosomal polymorphism that Stutevant and Dobzahansky (1938) were able to construct phylogenetic trees describing the evolutionary

origin of these gene arrangements from banding patterns in the polytene chromosomes. Later studies of chromosome structure found that although duplications and deficiencies are observable though polytene chromosomes, they never attained as high a frequency in natural populations as inversions and translocations. Translocations are found in lower frequencies than inversions in wild populations of *Drosophila*. This is easily explained as heterozygosity for a translocation results in the production of unbalanced chromosomes in the resultant gametes. Inversions on the other hand have been found in over 66% of *Drosophila* species. Heterozygous inversions act as strong suppressors of genetic recombination because any single cross-over within an inverted region results in chromosomes with deletions or duplications. Crossing over that occurs within paracentric inversions (not including the centromere) generate dicentric or acentric chromosomes that are selectively eliminated in *Drosophila* females during polar body formation in oogenesis. In female Drosophila, this was found not to reduce fertility because only a single product of meiosis ends up in the developing oocyte (Sturtevant and Beadle, 1936). There is no loss of fertility in males due to the formation of unbalanced gametes because very few crossing-over events occur in males. In the case of, pericentric inversions (including the centromere) that undergo crossing over events are monocentric and have no known mechanism of selective elimination from the egg nucleus. Due to the large probability of dominant lethality resulting from pericentric inversions, these mutation events are rare in wild *Drosophila* populations with paracentric inversions being significantly more common. Even when pericentric inversions are found, paracentric inversions typically exist to inhibit exchange within the inverted regions. Different paracentric inversions may be present on the same chromosome arm in distinct, overlapping, or inclusive manners, allowing for the discovery of the phylogenetic relationships between, or present within, species. This concept required the researcher to designate a “standard” arrangement within a species, which could be compared to any other chromosomal variants, although such a choice is arbitrary and without other biological data, the direction of the evolutionary steps cannot be determined (Sturtevant and Dobzhanzky, 1936). This method assumes that inversion mutations are very rare and arise only occur once, which was difficult to prove until the advent

of molecular genetic markers (Aquadro et al. 1991). However, this method can be utilized to construct phylogenetic trees if the effect of recurrent mutations is not strong enough to confuse the analysis of the branching process (Sperlich and Pfriem, 1986).

Different species have varying degrees of inversion mutations across *Drosphila* populations, from almost ubiquitous presence of inversion mutations across most chromosome arms to species that have inversion mutations present on only one or two chromosomes such as *D. pseudoobscura*. This raises the question of why do some chromosomes have inversions and others do not. Although it is restricted in such fashion, *D. pseudoobscura* is still highly polymorphic even though there are a limited number of locations for inversions to take place. It has been proposed that having inversions on one chromosome may inhibit the formation of inversions, although the effect is probably not very strong. *D. pseudoobscura* has only 6 inversions of its total characterized 28 that occur outside of the third chromosome. Geographic patterns of inversion polymorphism depend on a number of different factors and natural selection through environmental factors certainly plays a significant role. Other possible important factors include population structure, size, migration activity, breeding behavior, number of generations per year, and historical effects such as the founder effect (Schaeffer et al. 2008). Some *Drosophila* species show completely homogeneous karyotypes while others such as *D. pseudoobscura* have highly diversified polymorphisms with strong clinal relationships along different climates (Sperlich and Pfriem, 1986).

Dobzhansky and Sturtevant (1938) discovered a rich gene arrangement polymorphism for *D. pseudoobscura* in the southwestern United States. Over 30 different chromosome types have been observed in natural populations (Powell, 1992). Dobzhansky (1962) proposed two different designations for inversion polymorphisms: flexible and rigid. With seasonal, altitudinal or long-term changes would have an effect on arrangement frequencies of the flexible group, while rigid mutations would not exhibit any frequency changes. *D. pseudoobscura* has been shown to be a fly species with flexible type inversions, where arrangements' frequencies change with environmental factors or new advantageous mutations. For instance Standard arrangements tend to be more frequent in summer and low altitudes

while Chiricahua strains proved to be cold adapted, with higher frequencies at higher altitudes. This relationship can differ among localities and over different time periods. Over four decades, the frequencies of Pike Peak arrangements initially increased, then decreased along with a concomitant movement in Tree Line frequencies (Anderson et al. 1991). These long term changes have been interpreted as changes in the local environment or that new adaptive karyotypes have adapted and spread through the populations. This flexibility was replicated under laboratory settings as frequencies of arrangements among the collected flies changed in just a few generations. Unfortunately there is no clear explanation for why a species would have flexible as opposed to rigid inversion polymorphisms. Because the different arrangements are thought to be favored in different environments, a species with a flexible polymorphism pool will be able t o adapt to changing conditions, increasing the populations' fitness as a whole. However if homozygous or heterozygous combinations of genes increase the buffering ability of the population to environmental factors or ecological tolerance, a rigid system of inversion polymorphisms should arise. In *D. pseudoobscura*, it has been shown in laboratory experiments that inversion polymorphism frequencies are maintained through selective advantage of heterozygous individuals to cope with environmental changes (Wright and Dobzhansky, 1946). These results mirror those already observed in nature through the flexible character of its inversion polymorphism frequencies (Sperlich and Pfriem, 1986).

This selective superiority of heterokaryotypes over homokaryotypes can be explained by either dominance or over-dominance theories. The dominance theory purports that heterokaryotypes have higher fitness because deleterious alleles are masked in the heterozygous inversion combinations; therefore the observed heterosis is only due to the dominance of favorable alleles over disadvantageous ones (Crow, 1952). Over-dominance theory supposes that heterozygosity increases the fitness of the heterozygotes beyond that of the corresponding homozygotes. The theory does leave open whether the heterosis is due to the additive accumulation of single gene heteroses or from the higher fitness due to epistatic interactions between the genes within the inverted region. Later experimentation would reveal that *D.*

pseudoobscura populations would have the highest fitness when the mutual facilitation of karyotypes within the polymorphic populations rather than better performance of the existing heterokaryotypes (Dobzhansky, 1948). Polymorphic populations were also calculated to have a higher innate value of increase (r_m) than monomorphic populations through work conducted by Dobzhansky and others (1964). Frequency dependent selection may also play a role in the maintenance of these inversion polymorphisms. Ehrman (1970) showed that the rarest genotype can gain a mating advantage in mixed populations for visible gene markers, polygenic traits and carriers of different gene arrangements. Other species of *Drosophila* have also been shown to have similar frequency-dependent fitness values for specific karyotypes.

Inversion mutations have been proposed to have greater evolutionary effects than just the suppression of recombination in heterokaryotes inside the inversion loop region. Dobzhansky proposed a coadaptation hypothesis stating that stable and repeatable frequency equilibrium were only reached when founder flies were from the same collection area. If flies from different areas were introduced into the same environment, no equilibrium was reached or chromosome frequencies changes in an unpredictable manner (Dobzhansky and Pavlovsky, 1953). From this, it was interpreted that local chromosomes had coadapted for the highest fitness in their local environments, producing highly locally fit heterokaryotypes because flies from different regions might not have coadapted gene arrangements which would therefore not be heterotic. This phenomena was even shown for a species of *Drosophila* with rigid and highly stable inversion polymorphisms. Also it was observed that “triads” of genes do not occur at the same location in nature (Wallace, 1953). That is to say if two karyotypes were each descended from an ancestral line with a different singe inversion step, those two karyotypes would not experience crossing-over effects and could coadapt. However if the ancestral arrangement were added to the locality, crossing-over between each distinct descendant and the ancestral strain could interfere with the coadaptation of those descended arrangements. Enzyme gene loci seem to be exempt from clinal variation in allele frequencies, unlike inversion regions. The only proved instances of clinal enzyme variations have been

due to linkage of a particular allele of an enzyme locus to a certain inversion polymorphism gene arrangement. Since enzyme loci included in or tightly linked to inversions can be part of a selectively active supergene, the presence of certain allozyme alleles associated with inversion polymorphisms provide direct evidence for the coadaptation theory (Sperlich and Pfriem, 1986). The coadaptation model predicts that chromosomes of the same arrangement from different populations will not be identical by descent, however, recent molecular genetic data rejects this model because the *D. pseudoobscura* chromosomes are not genetically differentiated among populations (Schaeffer et al. 2003).

1.2 Mechanisms of Establishing Inversion Mutations

The method through which naturally occurring chromosome aberrations occur has been elucidated only recently through molecular analysis of breakpoint sequences (Richards et al. 2005; Ranz et al. 2007). At the time this review was written however, there were only prevailing hypotheses concerning the origin of inversion mutations. These features can be recreated in the laboratory with the use of mutagens, but the spontaneous natural events that cause such mutations remain a mystery. It has been stated that aberrations such as inversions have adaptive properties due to the capturing of genes and suppression of recombination, but how they originate remains unclear. Two major hypotheses exist to explain the phenomena. The position effect hypothesis assumes that a new inversion mutation creates variation that natural selection acts upon. For instance, the breakpoint mutations disrupt the structure of a gene or alter the expression of genes at the lesion. There would be no immediate heterosis in this model since coadapted blocks of genes would not have had time to occur and would only be positional factors responsible for the establishment of the new polymorphism. The preadaptation hypothesis does not rely on position effects, rather asserting that coadaptive relationships may already have been established prior to any inversions. If the genes are closely linked to each other and selection favors certain combinations of these alleles, then equilibrium could be reached. Recombination at this point would only result in the

production of less fit recombinant alleles while any inversion capturing the genes would be favored since it inhibits recombination in heterokaryotypes (Sperlich and Pfriem, 1986). Evidence supporting both hypotheses had been published at this time, but yields to recent data regarding the nature of the origins of inversion polymorphisms.

One of the most important evolutionary effects of inversion mutations is that they suppress recombination in heterozygous individuals. Although recombination events can occur in inversion heterozygotes, the meiotic products of recombination are selected against because of the production of unbalanced gametes. Large inversions still have some viable recombination events, but they occur orders of magnitude less than in homozygotes. On long timescales however, this recombination supplements mutation as a source of genetic variation within the inversion regions. Similar to other mutations, inversions evolve under selection and random drift. Thus, selection can occur in one of three ways, the first being the generation of structural problems in meiosis as mentioned in chapter 1.1. The second is a breakpoint disrupting an open reading frame or alter gene expression having positive or negative effects. The third is if the inverted region captures sets of selected alleles (Kirkpatrick and Barton, 2006). Inversion regions can be either under or over dominant depending on their size and where they are located. The basis for overdominance is unknown, but it could result from the positional effects of the breakpoints, a locus within the inversion if there is a fixed ancestral allele and an inverted fixed allele, or associative overdominance. This could occur when an inversion captures one or more deleterious recessive alleles. If this inversion is selected for when it is rare in the population, it can spread to a point where recessive homozygotes offset the initial advantage of the inversion resulting in a balanced polymorphism with the properties of conventional overdominance. Meiotic drive (when inheritance of the allele is greater than 50%) can also be a factor and usually occurs when a pair of interacting genes needs to be inherited for the allele to enter a population. Inversions mutations would suppress recombination that could disrupt the system. These mutations are so intriguing because they seem to arise relatively quickly in fly populations and are highly polymorphic in some species.

Geographical variation in inversion frequency serves as a clear sign that such regions are involved in adaptation (Kirkpatrick, 2010). Inversion clines correlated with environmental gradients, such as those found in *D. pseudoobscura* in the southwestern United States, are solid examples of this adaptive capability. Local adaptation describes the situation where different alleles are favored in different environments due to repressed recombination that keeps specific groups of genes together that increase fitness when transmitted together as described in chapter 1.1.

1.3 Chromosomes of *Drosophila pseudoobscura*

Sturtevant and Dobzhansky (1938) were the first researchers to study the inversions found on the *D. pseudoobscura* using the salivary gland polytene chromosome staining method proposed by Painter just a few years before their work. They identified the third chromosome of the five found in the species as having the majority of inversion mutations found in the species. They are also credited with the discovery of inversion polymorphism within a specific region due to their findings of many inversion heterozygotes through their collections (Dobzhansky and Sturtevant, 1938). To generate these heterozygote looping patterns in the laboratory, a standard strain was chosen arbitrarily and collected flies were crossed to it and the progeny's salivary chromosomes to diagnose the strains. Consequently, homozygote chromosomes pair linearly, while other arrangements show the aforementioned looping patters. Through their analyses, the researchers were able to identify the relationships between the strains and characterize cytological regions in which such inversions took place as described in appendix B. A weakness of the phylogeny made only from the gene arrangement karyotypes is that one cannot determine the ancestral chromosome. Molecular genetic data is necessary to infer the ancestral arrangement. Standard, Santa Cruz, or the "hypothetical" gene arrangement was suggested to be the ancestral arrangement from just these chromosome karyotypes and their observed overlapping inversion regions (Dobzhansky and Sturtevant, 1938). The sequences of 18 markers across the third chromosome support

the hypothesis that the Hypothetical chromosome was the ancestral arrangement (Wallace et al. 2011). The initial geographic data for the different arrangements was also collected in this study and which has, with additional study, culminated in the figure in appendix C describing the clinal patterns the different southwestern strains of *D. pseudoobscura* third chromosome arrangements form. Research completed over half a century later concerning the clinal frequencies of *D. pseudoobscura* third chromosome arrangements was able to confirm the stability of the clinal pattern except for the increase of the Tree Line arrangement frequency along the pacific coast (Anderson et al, 1991). As expected from the literature concerning paracentric inversion mutations, the increase in Tree Line arrangement frequency is supposed to be dependent on either a climate change in the niche or an accumulation of more advantageous mutations inside of the inversion region.

Evolutionary genomic studies were also carried out with respect to the inversion regions of these third chromosomes that further reinforced the hypotheses of inversion mutations as strong suppressors of recombination in order to keep positively associated loci among gene arrangements as *D. pseudoobscura* evolved in varied environmental conditions (Schaeffer, et al., 2003). The findings of this study go further in confirming the first premise of the coadaptation model by confirming the predictions that loci will be differentiated between gene arrangements. Unfortunately, the analyses used in the study were unable to detect differences in arrangements among geographical populations using biological or statistical methods. It is possible that the differences went undetected due to the additive properties of the small positive fitness effects of many loci, that the examined regions are not near targets of epistatic selection, or the statistical tests of differentiation and linkage disequilibrium were not of sufficient power to resolve a desired result. The study did find evidence through accumulation of negative Tajima's (1989) D values on the *D. pseudoobscura* third chromosome. Tajima's D is a statistical measure that summarizes the site frequency spectrum of DNA sequence mutations. Negative values imply an excess of low frequency polymorphisms relative to expectation. Therefore this evidence serves as even greater support of the recent rapid expansion and population growth of the species. Fitness values for corresponding to each of

the six ecological niches were also calculated for Standard, Chiricahua, Arrowhead, and Pike Peak along with the estimated amount of loci that contribute to the arrangements' local adaptations to heterozygous environments. It was also concluded that the selective factors that act on the egg, larval, or pupal stages of the fly life cycle drive the diversity of gene arrangements along the six ecological niches (Schaeffer, 2008).

The phylogenetic relationship between the different arrangement inversions as proposed by Sturtevant and Dobzhanzky (1938) was also recently uncovered through DNA sequence analysis carried out by Schaeffer and Wallace (2011) using 18 genetic markers on the *D. pseudoobscura* third chromosome. Their findings are represented in the inversion phylogeny table in appendix C, dating the initial Standard and Santa Cruz arrangements to about 1.38 Ma with the Standard arrangement giving rise to the Pikes Peak and Arrowhead inversions 0.99 and 0.58 Ma, respectively and the Santa Cruz arrangement giving rise to the Tree Line and Chiricahua inversions 1.22 and 0.51 Ma, respectively. It is interesting that both lineages have similarly older and newer inversion arrangement, although it is unclear why the newer arrangements were favored at the times of their advent.

With respect to the genes located around breakpoint regions of the *D. pseudoobscura* third chromosome, work by Wallace, Detweiler, and Schaeffer (2013) has shown that genes located on the distal portion of the chromosome show greater amounts of sequence diversity than those located closer to the centromere (more proximally). It was also observed that high levels of linkage disequilibrium were present in all of the 11 breakpoint regions and between the ends of most of the proximal and distal breakpoints although the central region had the highest overall LD levels since it is the location of the greatest amount of recombination suppression.

For this work, the author classified 50 different strains of *D. pseudoobscura* collected from wild environments via cytogenetic analysis. Then, whole genome sequences of the third chromosome were generated for each of the strains and the amino acid sequences for each of the 2831 protein coding genes were inferred. These sequences were then aligned using MEGA along with out-group sequences and

analyzed for amino acid changes that were fixed for specific arrangements, but were conserved in close relative species and in other *D. pseudoobscura* arrangements. The data were then analyzed to identify aberrant amino acid changes especially due to mutations resulting in transition of the class of amino acid at the mutated position.

Chapter 2

Methods and Materials

2.1 Polytene Chromosome Squashes and Identification of Gene Arrangements

Polytene chromosomes were isolated from the salivary glands of third instar larvae. Approximately twenty *Drosophila pseudoobscura* males and females from designated strains are taken from their respective vials and deposited into a fresh vial with a medium composed of corn meal, molasses, and agar. The flies are allowed to mate and oviposit eggs. After two weeks, third instar larvae begin to emerge from the medium where the flies laid their eggs. The harvested larvae are deposited into Drosophila Ringer's solution (7.5g/L NaCl, 0.35g/L KCl, 0.21g/L $CaCl_2$) to wash off any residual medium off of the specimens. A slide was prepared for the salivary glands by cleaning with silicon paper and placing five drops of Lacto-Aceto-Orcein stain (50mL acetic acid, 2g orcein, 50 mL lactic acid) on the slide in pattern similar on the five side seen on dice. The larvae were dissected under a binocular microscope in a concave slide with 5 mL of 45% acetic acid by volume. The larval salivary glands were dissected with tweezers that held the posterior end of the larva and a dissecting needle just behind the mouth parts on the dorsal side of the larvae and pull the head away from the main body. The salivary glands of the larvae should be apparent on either side of the head behind the mandibles in a "Y-shaped" structure. Using two dissecting needles, the salivary glands were cleaned by removing the fat body and other viscera. The glands were transferred from the acetic acid solution and placed in one of the five drops of stain. A cover slip was placed over the salivary glands. A paper towel was used to draw the excess stain from the cover slip and to begin the squashing procedure. The slide is placed under a cover slip in a paper towel and placed under a 1 kg weight for 30 minutes to allow the salivary gland cells to burst and for the polytene chromosomes to spread out. The capillary action from the paper pulling the dye

out from the slide helps to spread the chromosomes. If the weight is on the slide for too long, the chromosomes will spread out too much and lose their characteristic banding and puffing patterns and look like “spaghetti”. If the weight is not on long enough, the chromosomes will not spread out enough to be diagnosed. Chromosomes were diagnosed and imaged using a phase contrast microscope. Typically, one inspects the chromosomes under 100X magnification and then moves to 400X. Ultimately, chromosomes were viewed under at 1000x using an oil immersion lens. A digital camera attachment on the eyepiece was used to take images of the third chromosome, which has a characteristic double bulb at the end of the chromosome that looks like a guitar’s body. Multiple images files for each chromosome were imported into Adobe Photoshop and a mosaic image was constructed. The resultant image can then be used to diagnose the gene arrangement using the chromosomal banding patterns and puffing patterns as well as loops in the homologous pairs indicating possible inversion regions.

We used genetic crosses to further validate the identify chromosomes. We obtained a collection of strains from San Pablo Etla, Oaxaca, Mexico, which kindly provided by Dr. Therese Markow from the University of California. These strains were made homozygous for the third chromosome using balancer crosses (Dobzhansky and Queal, 1938). Our preliminary diagnoses suggesting that we had two different chromosomes in these strains, Tree Line and Cuernavaca. The names of the chromosomal inversions are based on the locations where these chromosomes were first collected. We crossed the putative Tree Line chromosome strains to laboratory strains that were known to carry the Tree Line arrangement. Thus, offspring of these crosses should not show any loops in a hybrid individual. We did not have any strains that were homozygous for the Cuernavaca arrangement in the laboratory, so we crossed the putative Cuernavaca strains to a Santa Cruz strain, whose gene arrangement is one inversion difference from Cuernavaca and should show a single loop in a Santa Cruz/Cuernavaca heterozygote.

2.2 DNA Extraction, Purification, and Sequencing

Vials of fly stock containing only single strains were taken and the adult flies were subsequently anesthetized using chloroform on a sponge in a plastic vial. After the flies have stopped moving, we waited a minute, then the dead and anesthetized flies were poured out under a dissecting microscope and the males and females were sorted and separated using a fine-tip paintbrush. The males and females were distinguished by their distinctive coloration; the males have orange coloration on the underbelly while the females have white coloration on their ventral thorax. The females were disposed of since they could have been pregnant, which would contaminate the DNA extraction process since the larval DNA would also be present in the sample. The males were then placed into individual vials. The flies were homogenized and their DNA was extracted using the procedure from the Qiagen DNeasy Extraction kit.

The samples were then purified using the ExoSap procedure, prior to sequencing, in order to remove any leftover PCR products including the primers. 5μL of the PCR reaction mixtures were applied to corresponding labeled wells in strip tubes along with 1.5μL of the ExoSap solution. The tubes were then run in the thermocycler using the ExoSap program for about 30 minutes.

After the PCR products have been purified, the samples were submitted to the PSU sequencing core facility in 96 well tube plates with 2μL of the purified DNA product and 2μL of either forward or reverse primer for the corresponding PCR reaction in each well.

2.3 Next Generation Sequencing and Analysis of Amino Acid Polymorphism

A total of 50 strains isochromosomal for the third chromosome were sequenced using an Illumina HiSeq2000 machine. The sequence reads from each strain were mapped to the *D. pseudoobscura* reference genome using BWA (Li and Durbin, 2009). Polymorphic sites were extracted from the aligned reads using GATK (McKenna et al. 2010). After alignment, the data was run through an original program written to identify genes that contained mutations from close relatives of *D. pseudoobscura* whose

homologous genes matched most closely with those of the Arrowhead reference strain as identified through automated blast searches on the Flybase website. The program created an output list of the genes that varied from the ancestral and reference sequences along with providing the ancestral amino acid type, the derived amino acid type, and the position on the gene of the mutation in nucleotides. The genes were then compared by counts of type of amino acid transition and subsequently the genes with statistically significant amounts of mutation, as compared to other mutated genes found in their strain, were identified and subjected to a gene ontology (GO) analysis using the Flybase website.

Chapter 3

Data

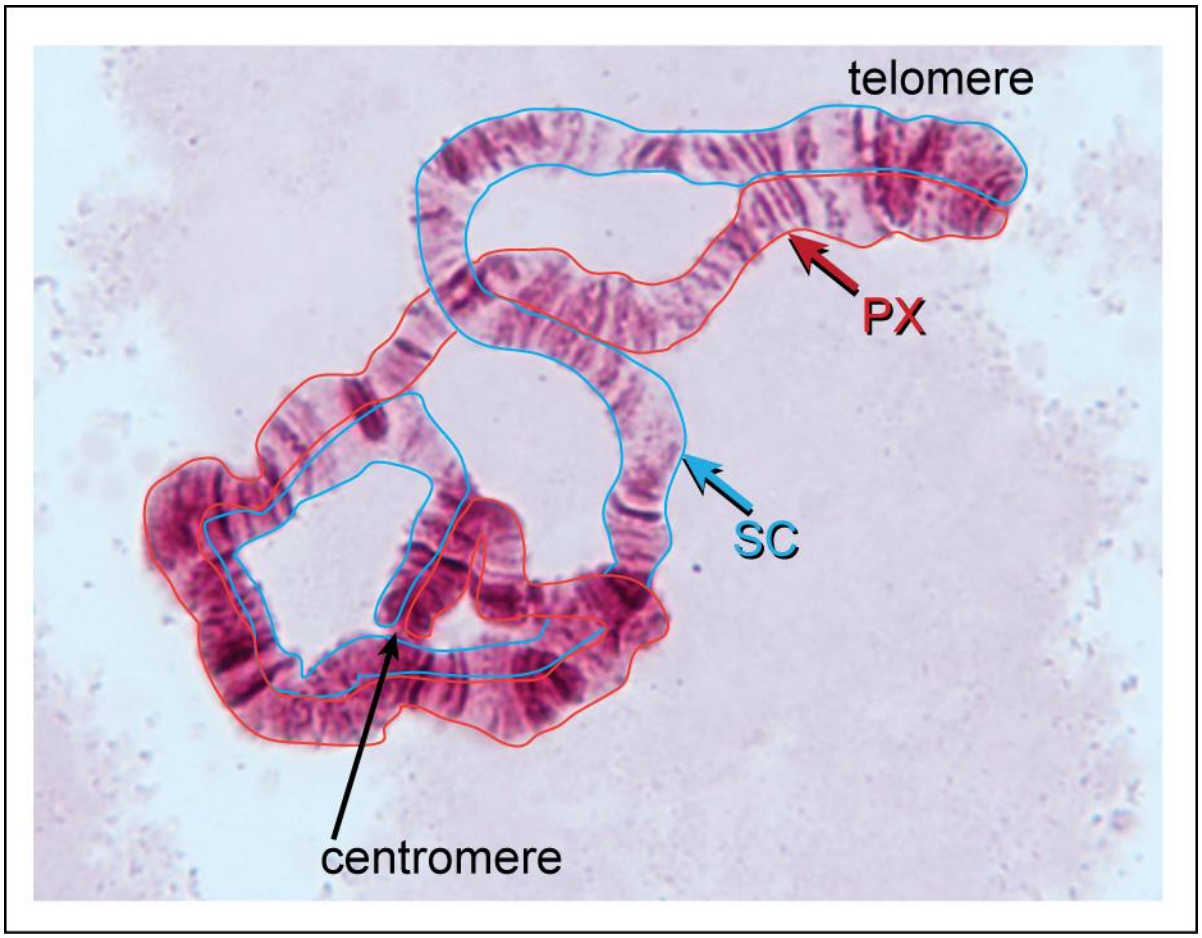


Figure 1: Polytene preparation for a Santa Cruz/"Cuernavaca heterozygote.

This image represents the culmination of cytogenetic analysis for the supposed Cuernavaca strains that were later found to be of the Paxtepec line due to the double loop structure seen above. We expected to find only one loop since there should only be one inversion between the supposed Cuernavaca and Tree Line strains. This result led to more investigation of the typing of the putative "Cuernavaca" line. After crossing a known line with an unknown specimen line, the resulting salivary chromosomes

from the larvae tend to look like this if the procedure is correctly carried out. The labeling was added for ease of viewing, and discerning the two different chromosomes (Olvera et al. 1985).

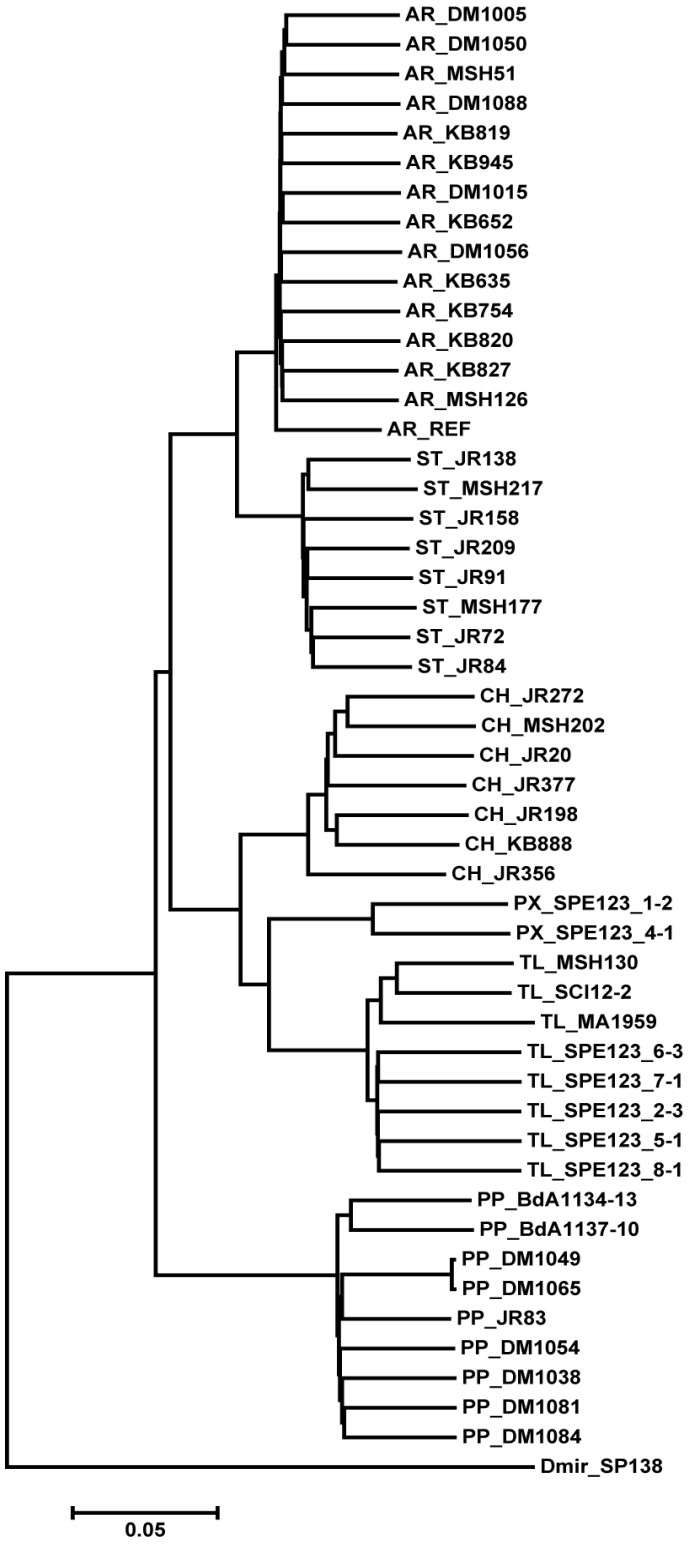


Figure 2: Evolutionary relationships of 50 strains isochromosomal for the third chromosome

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 2.50883450 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al. 2004) and are in the units of the number of base substitutions per site. The analysis involved 50 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1117374 positions in the final dataset. Evolutionary analyses from the sequence data of the 18 marker genes of the *D. pseudoobscura* third chromosome were conducted in MEGA6 (Tamura et al. 2013).

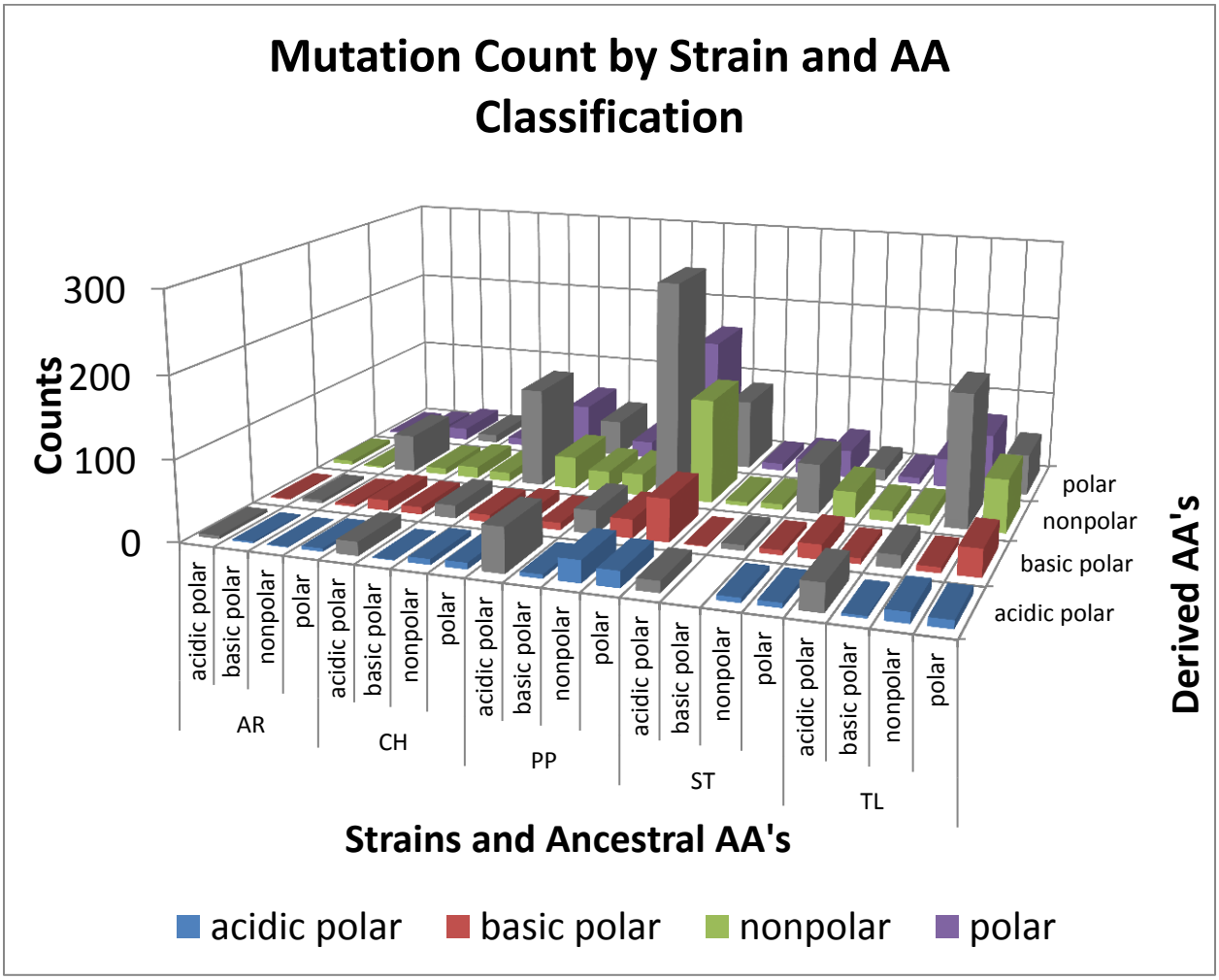


Figure 3: Number of amino acid mutation counts by strain and ancestral to derived amino acids

This graph represents all of the different amino acid transitions found in the data set and arranged by strain, ancestral amino acid class, and the derived amino acid types. Areas labeled with grey signify mutations where the class of amino acid did not change. Pike Peak has the highest levels of polar-nonpolar transitions, although all of the strains had the highest counts of these types of mutation. Further analysis focuses on this specific mutation type since it was the most common and would likely appear in genes with other mutations.

Ancestral Amino Acid by Class		Derived Amino Acid by Class																											
		Acidic Polar		Acidic Polar Total	Basic Polar			Basic Polar Total	Non-polar										Non-polar Total	Polar					Polar Total	Totals
		D	E		H	K	R		A	C	F	G	I	L	M	P	V	W		N	Q	S	T	Y		
Acidic Polar		**72**	**52**	**124**	**5**	**22**		**27**	**16**			**31**	**2**	**1**			**13**		**63**	**23**	**22**			**4**	**49**	**263**
	D		52	**52**	5			**5**	5			17	1				5		**28**	21	1			3	**25**	**110**
	E	72		**72**		22		**22**	11			14	1	1			8		**35**	2	21			1	**24**	**153**
Basic Polar		**4**	**7**	**11**	**10**	**19**	**43**	**72**		**5**		**15**	**4**	**13**	**8**	**9**	**1**	**5**	**60**	**46**	**54**	**9**	**12**	**9**	**130**	**273**
	H	3		**3**		1	7	**8**						7		4			**11**	14	29		1	9	**53**	**75**
	K	1	7	**8**			36	**36**					2		4		1		**7**	31	16	1	9		**57**	**108**
	R				10	18		**28**		5		15	2	6	4	5		5	**42**	1	9	8	2		**20**	**90**
Non-polar		**24**	**30**	**54**	**12**	**7**	**29**	**48**	**102**	**9**	**45**	**52**	**103**	**111**	**45**	**28**	**172**	**4**	**671**	**4**	**26**	**170**	**137**	**10**	**347**	**1120**
	A	8	11	**19**		2		**2**				36		2	1	17	69		**125**			66	82		**148**	**294**
	C						1	**1**			8	4							**12**			14		6	**20**	**33**
	F									2			9	27			1	1	**40**			1	1	4	**6**	**46**
	G	13	11	**24**			14	**14**	32	3							18		**53**	2	2	33			**37**	**128**
	I					3		**3**			7	2		15	11		31		**66**	2		4	15		**21**	**90**
	L		2	**2**	5		7	**12**	1	1	26	1	25		23	11	38	3	**129**		11	7	1		**19**	**162**
	M				1	2	2	**5**					31	18			14		**63**		1		7		**8**	**76**
	P				6		5	**11**	28					16			1		**45**		12	44	30		**86**	**142**
	V	3	6	**9**					41		4	8	38	31	10				**132**			1	1		**2**	**143**
	W									3		1		2					**6**						**6**	
Polar		**20**	**28**	**48**	**51**	**57**	**23**	**131**	**87**	**14**	**15**	**19**	**47**	**35**	**11**	**44**	**3**	**1**	**276**	**64**	**1**	**77**	**63**	**4**	**209**	664
	N	17	1	**18**	7	22		**29**					13						**13**		1	28	10	3	**42**	**102**
	Q		26	**26**	32	21	8	**61**					1	19	1	13			**34**			3			**3**	**124**

		S				1		11	12	25	10	5	18	9	16		18	2	1	104	48			53	1	102	218
		T		1	1	1	13	4	18	62	1		1	23		10	13	1		111	14		44			58	188
		Y	3		3	10	1		11		3	10		1						14	2		2			4	32
	Totals		120	117	237	78	105	95	278	205	28	60	117	156	160	64	81	189	10	1070	137	103	256	212	27	735	2320

Table 1: Counts of ancestral to derived amino acid mutations

The summary data from the table are the basis for figure three. In this representation, specific amino acid changes are displayed along with overall trends of amino acid class mutations.

		Derived Amino Acid Mutation Counts				
Ancestral Amino Acid Mutations by Arrangement		acidic polar	basic polar	non-polar	polar	**Change Totals by Ancestral Class**
AR Change Totals		**7**	**17**	**15**	**24**	**63**
	acidic polar	4	1	5	2	**8**
	basic polar	2	4	2	7	**11**
	non-polar	1	3	47	15	**19**
	Polar	4	13	8	10	**25**
CH Change Totals		**16**	**31**	**65**	**96**	**208**
	acidic polar	17	9	14	9	**32**
	basic polar	1	16	11	25	**37**
	non-polar	7	9	123	62	**78**
	Polar	8	13	40	45	**61**
PP Change Totals		**52**	**85**	**183**	**228**	**548**
	acidic polar	55	9	26	21	**56**
	basic polar	5	29	27	44	**76**
	non-polar	27	23	273	163	**213**
	Polar	20	53	130	88	**203**
ST Change Totals		**12**	**25**	**44**	**62**	**143**
	acidic polar	14	1	5	9	**15**
	basic polar		7	7	18	**25**
	non-polar	6	6	62	35	**47**
	Polar	6	18	32	15	**56**
TL Change Totals		**26**	**48**	**92**	**116**	**282**
	acidic polar	34	7	13	8	**28**

		basic polar	3	16	13	36	**52**
		non-polar	13	7	166	72	**92**
		Polar	10	34	66	51	**110**
	Change Totals By Derived Class		**113**	**319**	**339**	**526**	**1297**

Table 2: Amino acid class changing mutations by arrangement

The table depicts amino acid class data sorted by the different *D. pseudoobscura* arrangements in this study. It is interesting to note that the Pikes Peak strain has at least double the amount of mutations than any other strain. Derived mutations producing non-polar amino acids were the most common, followed by polar mutations.

Gene	**Derived Mutations**				**Grand Total**
	Acidic Polar	**Basic Polar**	**Non-polar**	**Polar**	
GA24967	4	6	21	7	38
GA24455	6	5	14	8	33
GA15872		9	14	10	33
GA21795	5	1	12	11	29
GA12088	2	1	13	8	24
GA18514	4	3	10	7	24
GA14121	2	3	13	5	23
GA20669	2	3	10	7	22
GA11891		3	12	6	21
GA24331		2	11	4	17
GA24495	2		11	4	17
GA16823	4		9	3	16
GA12123	2		8	5	15
GA24314	2	2	4	7	15
GA21341	2		5	7	14
GA24382	1		7	5	13
GA24797		1	9	3	13
GA14934	4	3	5	1	13
GA24782	1	2	6	4	13
GA18089	1	1	6	4	12
GA10623	1		5	6	12
GA15235	1	1	6	4	12
GA25040	1	2	5	4	12
GA15603	1	1	8	2	12
GA24138		2	6	4	12

GA24315	2	1	6	3	12
GA20915	1	1	4	5	11
GA24317		1	6	4	11
GA14224		2	4	5	11
GA11265		1	6	4	11
GA15135	2		6	3	11
GA24975		2	3	5	10
GA12087			6	4	10
GA10653		1	3	6	10
GA25033		2	6	2	10
GA24332	1	3	5	1	10
GA21271	1	1	3	5	10
GA24375		3	4	3	10
GA15581	2	1	2	5	10
GA21027		1	5	4	10
GA18304	3		5	2	10
GA17318	2		2	5	9
GA19331	1		5	3	9
GA24863	1	1	5	2	9
GA15577			6	3	9
GA24371		1	3	5	9
GA11024	1		8		9
GA18464	3		2	4	9
GA24783		1	5	2	8
GA12222	1	1	5	1	8
GA24356	2	1	1	4	8
GA15744			3	5	8
GA24502		1	5	2	8
GA20928	1		4	3	8
GA24788	1	2	5		8
GA24135	1		5	2	8
GA18253	1		4	3	8
GA24824		1	1	5	7
GA24462		1	4	2	7
GA14980		1	6		7
GA15660	2		3	2	7
GA17624			4	3	7
GA10058	1		4	2	7
GA14750		1	5	1	7
GA20796	1		6		7
GA24373	2		3	2	7
GA15707	1	1	4	1	7

GA15265		1	4	2	7
GA21111	1		5	1	7
GA15267	2	1	2	2	7
GA11599	1	1	4	1	7
GA14141		2	3	2	7
GA15737	3		2	2	7
GA17764	2	2	2	1	7

Table 3: Genes with a mutation count in the 95th percentile

The genes listed in this table have a total mutation count in the 95th percentile of all genes with amino acid mutations. The types of derived mutations are also listed along with total count of mutated amino acid sites.

Gene	***D. melanogaster* Homologue**	**Strain**	**AA polar-nonpolar mutations**
GA21795	Tud	PP	23
GA24455	CG33017	PP	22
GA18514	MESR4	PP	17
GA24495	CG30263	PP	15
GA10623	Tou	PP	11
GA15603	CG30046	PP	10
GA15135	mus205	PP	9
GA19331	Mip120	TL	8
GA15744	CG30275	PP	8
GA24975	-	AR	8
GA24502	CG30203	PP	7
GA24863	CG8920	CH	7
GA24852	-	ST	4

Table 4: Unique mutations by strain

This table contains a list of all of the genes, *D. melanogaster* homologues, the strains of the flies and the number of mutations that are unique to their strains. The Pikes Peak strain has the most unique mutations and the highest counts of amino acid changes in each protein for all of the above genes.

***D. pseudoobscura* genes**	**Strains**	***D. melanogaster* homologue**	**AR**	**CH**	**PP**	**ST**	**TL**	**Total**
GA24967	CH, PP, TL	CG30007		13	1	14		28
GA15872	AR,CH,PP,ST,TL	-	7	5	7	3		22

GA12088	PP, TL, CH	ana3		5	11		5	21
GA11891	CH, TL	Ir47a		10			8	18
GA14121	PP, TL	GC16724			11		7	18
GA20669	TL, CH, PP	Gp210		1	4		12	17
GA24331	PP,TL,AR,CH	-	2	4	8		1	15
GA12123	CH,PP,ST,TL	sha		7	4	1	1	13
GA16823	CH,TL	snama		8			4	12
GA24797	CH,PP	CG3004		8	4			12
GA24314	PP,ST	dom			7	4		11
GA24317	PP,ST,AR	CG13724	1		7	2		10
GA11265	PP,TL,CH	stan		2	7		1	10
GA18089	TL, CH	CG4294		3			7	10
GA24138	TL, CH	CG13185		4			6	10
GA24315	ST, AR	lva	3			6		9
GA20915	CH, PP, ST	CG8232		6	2	1		9
GA15577	TL, CH, PP	CG30015		1	1		7	9
GA21271	CH, PP	Lpin		6	2			8
GA24783	CH, TL	CG33476, CG33477, CG12898		6			1	7
GA10058	ST, TL	CG10073, CG10081				4	2	6

Table 5: Matched mutations by strain

This table contains the list of genes with their *D. menalogaster* homologues, and the counts of polar-nonpolar mutations for each strain for the mutations that have matches amongst the other strains.

Chapter 4

Results and Discussion

4.1 Cytogenetic Analysis

The majority of the strains were diagnosed through genetic crosses between newly acquired undiagnosed chromosomes and known chromosomes. These third chromosomes will show looped regions where large scale inversion mutations have taken place and banding patterns identified previously can be used to diagnose the arrangement carried by the unknown strain (Kastritsis and Crumpacker, 1966; Kastritsis and Crumpacker 1967). The putative Cuernavaca strains were analyzed initially though only their gene markers causing the initial confusion concerning their identity. The banding looping pattern obtained through backcrossing Santa Cruz and putative Cuernavaca fly strains was the first step in accurately assessing their identity as shown in figure 1. Cuernavaca and Santa Cruz heterozygotes should only contain one large looped region on their chromosomes. However, it is clear to see from figure 1 that two major inversion events separate these chromosomes because there are two loops observed in the heterozygotes. Therefore, the sequence data was reassessed and the three new strains were determined to carry the Paxtepec arrangement.

4.2 Evolution of Amino Acid Sequences on Chromosomal Inversions

Of the 2831 genes sequenced in this analysis, only 73 were found to have significant numbers of amino acid sequence mutations that were conserved in the outgroups to *D. pseudoobscura*. From the data presented in table 1, it is clear to see that certain types of amino acid transitions have occurred much more often than others. The entire group of top 20 amino acid transition types (pairs of transitions) all had

codon sequences that varied at only one nucleotide from the codon necessary to code for the derived amino acid. These results are to be expected since a single point mutation is much more likely than multiple nucleotide changes. These types of effects are much more likely to have deleterious effects than a single amino acid change. These data are also displayed in a slightly different manner in table 3where…. The genes listed in the table have significantly more transitions to non-polar amino acids than the other three classes, although transitions to polar number as the second most frequent. Table 5 shows summary data for each mutation from ancestral to derived classes of amino acids grouped by third chromosome arrangement. These summary data show that mutations that do not change the overall class of amino acid are more abundant than those that do. This is also to be expected as well because a transition to an amino acid with similar properties will likely have a less profound effect on protein structure and function than a mutation outside of the ancestral class. Overall, Pikes Peak had the highest number of mutations on its third chromosome followed by Tree Line, Chiricahua, Standard, and Arrowhead. Pikes Peak also has the highest mutation counts in each category individually (table 2).

4.3 Radical Amino Acid Changes and Gene Ontology (GO) Analysis

The summary data of amino acid transitions by acidity and polarity reveals a trend in CH, PP, ST, and TL where transitions between nonpolar to polar and polar to nonpolar amino acids occur most frequently. The PP arrangement has the highest incidence of type-homogeneous amino acid substitutions along with nonpolar to polar and polar to nonpolar transitions followed by the TL strain. This relationship may provide interesting to investigate since phylogenetic data generated by Dr. Schaeffer show that PP and TL sequence from the 70B to 74 group position on the third chromosome cluster closely, a phenomena not directly explained by accepted evolutionary trees for the species' strains.

A Gene Ontology (GO) analysis was performed on genes that had high rates of radical amino acid mutations in the different chromosomal arrangements. Of the 13 genes that were selected because

they numbers of mutation over two standard deviations away from the mean, nine had listed gene ontology information and seven of the nine were found only in the Pikes Peak arrangement. Some of the repeated associations of the genes were four with nucleotide or DNA binding activity, two with zinc binding, and two with developmental and lifespan-determining factors. Of the 21 genes selected as above two standard deviations away from the mean sorted by arrangement, 14 had gene ontology information associated with either them or associated *D. melanogaster* homologous genes. Six different genes from this grouping were associated with digestive, DNA repair and ATPase enzyme activity and three others were implicated in cell signaling and cytoskeletal assembly pathways. Also, a group of 4 genes has growth and development associations.

Genes associated with the cell membrane and that have high amounts of transition between polar and non-polar amino acids should see large effects since which residues are able to inhabit the polar cytoplasm versus the non-polar intermembrane space can be integral to protein function. GA24975 has eight polar-non-polar transitions and is a unique mutation to the Arrowhead arrangement and GA11264 has 10 polar-non-polar transitions and is found in Pikes Peak, Tree Line, and Chiricahua arrangements.

Interestingly, the sequence analysis of the single nucleotide polymorphism data from revealed that a section of the third chromosome from 70B-74 departs from the expected phylogenetic relationships expected based on the cytogenetic data, i.e., the sequences of Pikes Peak and Tree Line arrangements which has not been previously observed as is evident by prevailing lineage histories (Dobzhanksy, 1944). Of the genes identified by their polar-nonpolar mutations, GA20669, GA15872, GA12123, GA11265and GA24967 are all found within this region and have mutations for both Pikes Peak and Tree Line arrangements. GA15872 does not have a *D. melanogaster* homologue nor ontology information, but has mutations in all of the *D. pseudoobscura* arrangements. GA24967 is related to neurogenesis (as above), GA12123 is implicated in functions of morphogenesis of antennae and a few other minor developmental characters, GA11265 is implicated in G-protein coupled signal transduction, and GA20669 is implicated via *D. melanogaster* homologue in olfactory learning and some other cellular functions. It is interesting

that this chromosomal region has so many mutated genes involved in development and learning and it could be useful to pursue the similarities between Pikes Peak and Tree Line in this region further for this reason.

Chapter 5

Conclusion

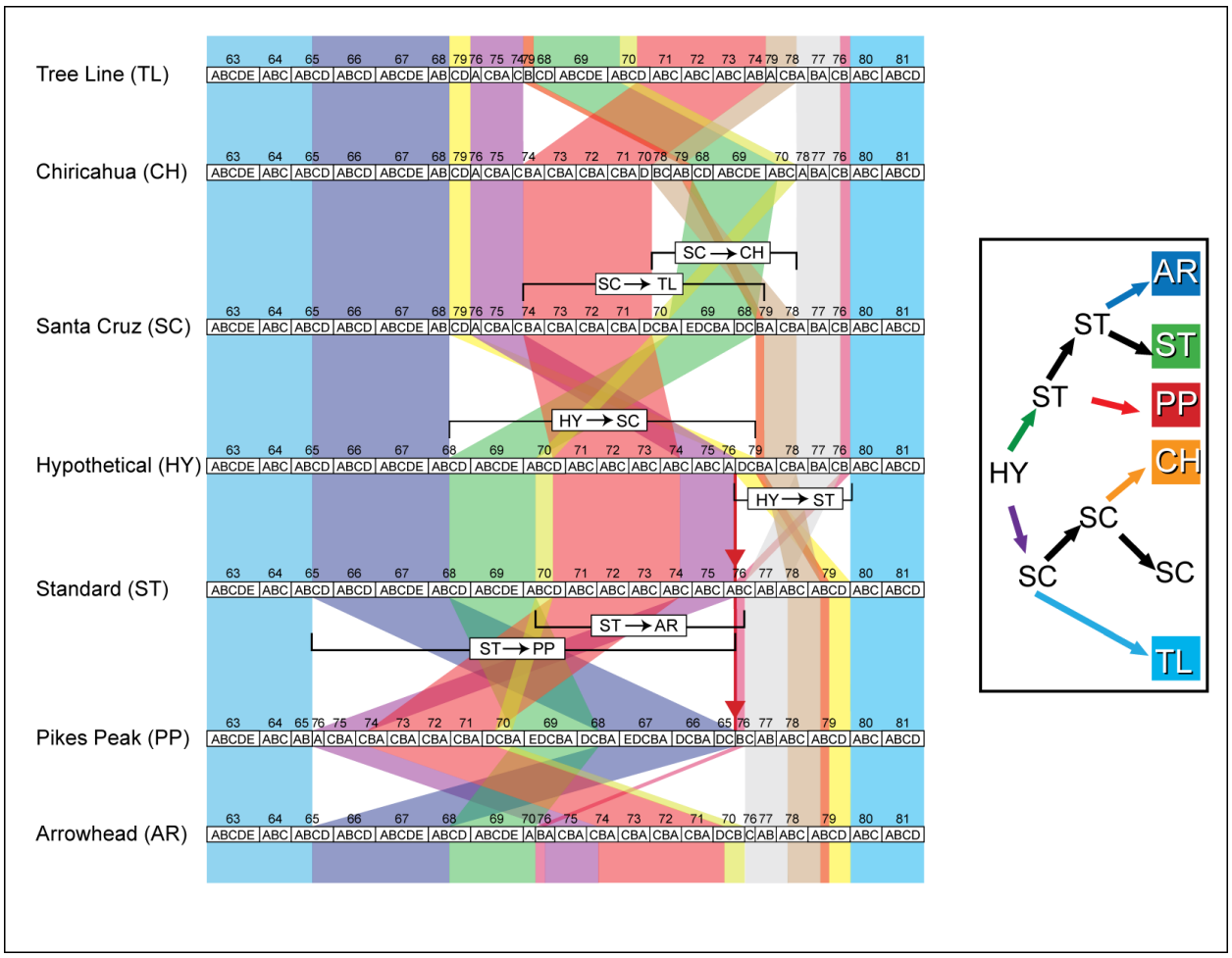
It is clear from the results that there are many genes that have higher frequencies of mutations in either one or a few *Drosophila pseudoobscura* strains. These mutations follow generally expected biochemical trends in that more of the mutations in these genes stay within the same general class of amino acid. Interestingly, of the most divergent genes, the majority of mutations take place via non-polar to polar (and vice versa) transitions. Of the 5 *Drosophila pseudoobscura* strains analyzed, Pikes Peak was found to have the greatest number of mutations followed by Tree Line, Chiricahua, Standard, and finally Arrowhead strains. It is possible that since the Arrowhead and Standard arrangements are relatively newer inversion arrangements, that they have had less time to acquire locally adapted and evolutionarily advantageous loci within their boundaries. The function of their protein products are only hinted at by the gene ontology and further functional studies are necessary to elucidate their potential evolutionary implications with respect to the inversion regions where they are located.

Appendix A: List of Strains and Collection Localities

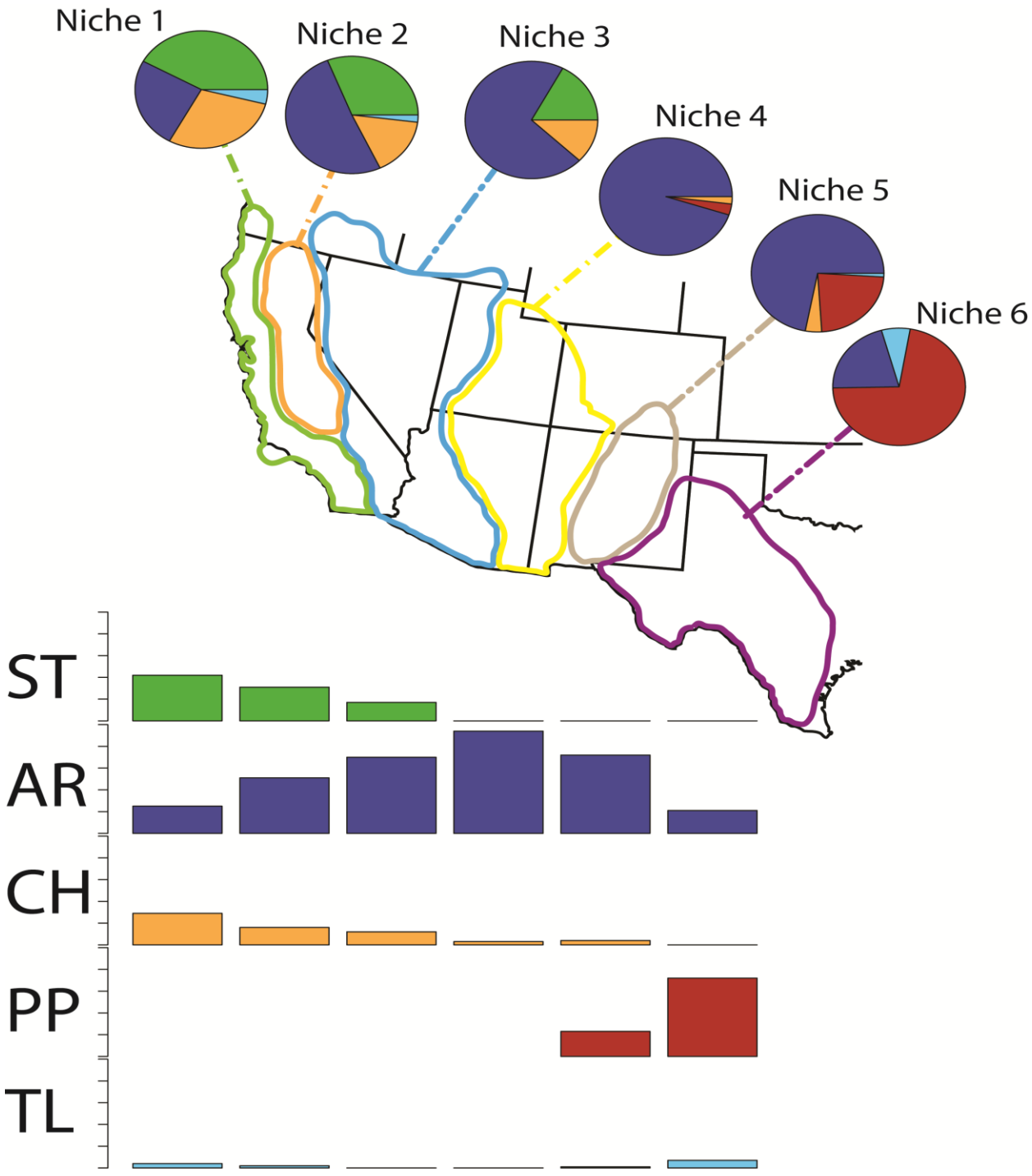
Strain No.	Seq Name	Collection Locality
1.	AR_REF	Mesa Verde, CO
2.	AR_DM1005	Davis Mountains, TX
3.	AR_DM1015	Davis Mountains, TX
4.	AR_DM1050	Davis Mountains, TX
5.	AR_DM1056	Davis Mountains, TX
6.	AR_DM1088	Davis Mountains, TX
7.	AR_KB635	Kaibab National Forest, AZ
8.	AR_KB652	Kaibab National Forest, AZ
9.	AR_KB754	Kaibab National Forest, AZ
10.	AR_KB819	Kaibab National Forest, AZ
11.	AR_KB820	Kaibab National Forest, AZ
12.	AR_KB827	Kaibab National Forest, AZ
13.	AR_KB945	Kaibab National Forest, AZ
14.	AR_MSH126	Mount St. Helena, CA
15.	AR_MSH51	Mount St. Helena, CA
16.	PP_BdA1134-13	Bosque delApache, NM
17.	PP_BdA1137-10	Bosque delApache, NM
18.	PP_DM1038	Davis Mountains, TX
19.	PP_DM1049	Davis Mountains, TX
20.	PP_DM1054	Davis Mountains, TX
21.	PP_DM1065	Davis Mountains, TX
22.	PP_DM1081	Davis Mountains, TX
23.	PP_DM1084	Davis Mountains, TX
24.	PP_JR83	James Reserve, CA
25.	ST_JR138	James Reserve, CA
26.	ST_JR158	James Reserve, CA
27.	ST_JR209	James Reserve, CA
28.	ST_JR72	James Reserve, CA
29.	ST_JR84	James Reserve, CA
30.	ST_JR91	James Reserve, CA
31.	ST_MSH177	Mount St. Helena, CA
32.	ST_MSH217	Mount St. Helena, CA
33.	CH_JR198	James Reserve, CA
34.	CH_JR20	James Reserve, CA
35.	CH_JR272	James Reserve, CA
36.	CH_JR356	James Reserve, CA
37.	CH_JR377	James Reserve, CA
38.	CH_KB888	Kaibab National Forest, AZ
39.	CH_MSH202	Mount St. Helena, CA
40.	TL_MA1959	Mather, CA
41.	TL_MSH130	Mount St. Helena, CA
42.	TL_SCI12-2	Santa Cruz Island, CA
43.	TL_SPE123_2-3	San Pablo Etla, Oaxaca, Mexico

44.	TL_SPE123_5-1	San Pablo Etla, Oaxaca, Mexico
45.	TL_SPE123_6-3	San Pablo Etla, Oaxaca, Mexico
46.	TL_SPE123_7-1	San Pablo Etla, Oaxaca, Mexico
47.	TL_SPE123_8-1	San Pablo Etla, Oaxaca, Mexico
48.	PX_SPE123_1-2	San Pablo Etla, Oaxaca, Mexico
49.	PX_SPE123_4-1	San Pablo Etla, Oaxaca, Mexico
50.	Dmir_SP138	

Appendix B: Inversion Map



Appendix C: Definition and Frequency of Niches



Works Cited

1) SCHAEFFER, S. W., A. BHUTKAR, B. F. MCALLISTER, M. MATSUDA, L. M. MATZKIN *et al.*, 2008 Polytene chromosomal maps of 11 Drosophila species: The order of genomic scaffolds inferred from genetic and physical maps. Genetics **179:** 1601-1655.
2) PAINTER, T. S., 1934 A new method for the study of chromosomal aberrations and the plotting of chromosomal maps in *Drosophila melanogaster*. Genetics **19:** 175-188.
3) STURTEVANT, A. H., and G. W. BEADLE, 1936 The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. Genetics **21:** 544-604.
4) DOBZHANSKY, T., and A. H. STURTEVANT, 1938 Inversions in the chromosomes of *Drosophila pseudoobscura*. Genetics **23:** 28-64.
5) POWELL, J. R., 1992 Inversion polymorphisms in *Drosophila pseudoobscura* and *Drosophila persimilis*, pp. 73-126 in *Drosophila Inversion Polymorphism*, edited by C. B. KRIMBAS and J. R. POWELL. CRC Press, Ann Arbor, MI.
6) AQUADRO, C. F., A. L. WEAVER, S. W. SCHAEFFER and W. W. ANDERSON, 1991 Molecular evolution of inversions in *Drosophila pseudoobscura*: The amylase gene region. Proceedings of the National Academy of Sciences USA **88:** 305-309.
7) ANDERSON, W. W., J. ARNOLD, D. G. BALDWIN, A. T. BECKENBACH, C. J. BROWN *et al.*, 1991 Four decades of inversion polymorphism in *Drosophila pseudoobscura*. Proceedings of the National Academy of Sciences USA **88:** 10367-10371.
8) WRIGHT, S., and T. DOBZHANSKY, 1946 Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain populations of *Drosophila pseudoobscura*. Genetics **31:** 125-156.

9) CROW, J. F., 1952 Dominance and overdominance, pp. 282-297 in *Heterosis : a record of researches directed toward explaining and utilizing the vigor of hybrids*, edited by J. W. GOWEN. Iowa State College Press, Ames.

10) DOBZHANSKY, TH., R.C. LEWONTIN, and O. PAVLOVSKY, 1964 The capacity for increase in chromosomally polymorphic and monomorphic populations of *Drosophila pseudoobscura*. Heredity **19**: 597-614.

11) DOBZHANSKY, T., 1948 Genetics of natural populations. XVIII. Experiments on chromosomes of *Drosophila pseudoobscura* from different geographic regions. Genetics **33:** 588-602.

12) EHRMAN, L., 1970 The mating advantage of rare males in *Drosophila*. Proceedings of the National Academy of Sciences USA **65**: 345-348.

13) WALLACE, B., 1953 On coadaptation in *Drosophila*. American Naturalist **87:** 343-358.

14) SPERLICH, D., and P. PFRIEM, 1986 Chromosomal polymorphism in natural and experimental populations, pp. 257-309 in *The Genetics and Biology of Drosophila 3e*, edited by M. ASHBURNER, H. L. CARSON and J. N. THOMSON. Academic Press, New York, NY.

15) DOBZHANSKY, T., and O. A. PAVLOVSKY, 1953 Indeterminate outcome of certain experiments on *Drosophila* populations. Evolution **7:** 198-210.

16) SCHAEFFER, S. W., P. GOETTING-MINESKY, M. KOVACEVIC, J. PEOPLES, J. L. GRAYBILL *et al.*, 2003 Evolutionary genomics of inversions in *Drosophila pseudoobscura*: Evidence for epistasis. Proceedings of the National Academy of Sciences USA **100:** 8319-8324.

17) KIRKPATRICK, M., and N. BARTON, 2006 Chromosome inversions, local adaptation and speciation. Genetics **173:** 419-434.

18) RICHARDS, S., Y. LIU, B. R. BETTENCOURT, P. HRADECKY, S. LETOVSKY *et al.*, 2005 Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene and *cis*-element evolution. Genome Research **15:** 1-18.

19) WALLACE, A. G., D. DETWEILER and S. W. SCHAEFFER, 2011 Evolutionary history of the third chromosome gene arrangements of *Drosophila pseudoobscura* inferred from inversion breakpoints. Molecular Biology and Evolution **28:** 2219-2229.

20) RANZ, J. M., D. MAURIN, Y. S. CHAN, M. V. GROTTHUSS, L. W. HILLIER *et al.*, 2007 Principles of genome evolution in the *Drosophila melanogaster* species group Public Library of Science Biology **5:** 1366-1381.

21) KIRKPATRICK, M., 2010 How and why chromosome inversions evolve. PLoS Biol **8:** e1000501.

22) KIRKPATRICK, M., and N. BARTON, 2006 Chromosome inversions, local adaptation and speciation. Genetics **173:** 419-434.

23) DOBZHANSKY, T., and A. H. STURTEVANT, 1938 Inversions in the chromosomes of *Drosophila pseudoobscura*. Genetics **23:** 28-64.

24) ANDERSON, W. W., J. ARNOLD, D. G. BALDWIN, A. T. BECKENBACH, C. J. BROWN *et al.*, 1991 Four decades of inversion polymorphism in *Drosophila pseudoobscura*. Proceedings of the National Academy of Sciences USA **88:** 10367-10371.

25) SCHAEFFER, S. W., P. GOETTING-MINESKY, M. KOVACEVIC, J. PEOPLES, J. L. GRAYBILL *et al.*, 2003 Evolutionary genomics of inversions in *Drosophila pseudoobscura*: Evidence for epistasis. Proceedings of the National Academy of Sciences USA **100:** 8319-8324.

26) SCHAEFFER, S. W., 2008 Selection in heterogeneous environments maintains the gene arrangement polymorphism of *Drosophila pseudoobscura*. Evolution **62:** 3082-3099.

27) WALLACE, A. G., D. DETWEILER and S. W. SCHAEFFER, 2011 Evolutionary history of the third chromosome gene arrangements of *Drosophila pseudoobscura* inferred from inversion breakpoints. Molecular Biology and Evolution **28:** 2219-2229.

28) WALLACE, A. G., D. DETWEILER and S. W. SCHAEFFER, 2013 Molecular Population Genetics of Inversion Breakpoint Regions in Drosophila pseudoobscura. G3: Genes|Genomes|Genetics **3:** 1151-1163.

29) LI, H., and R. DURBIN, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:** 1754-1760.

30) MCKENNA, A., M. HANNA, E. BANKS, A. SIVACHENKO, K. CIBULSKIS *et al.*, 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20:** 1297-1303.

31) DOBZHANSKY, T., and M. L. QUEAL, 1938 Genetics of natural populations. II. Genic variation in populations of *Drosophila pseudoobscura* inhabiting isolated mountain ranges. Genetics **23:** 463-484.

32) OLVERA, O., R. F. ROCKWELL, M. E. DE LA ROSA, M. I. GASO, F. GONZALEZ *et al.*, 1985 Chromosomal and behavioral studies of Mexican Drosophila. III. Inversion polymorphism of *D. pseudoobscura*. Heredity **76:** 258-262.

33) SAITOU, N., and M. NEI, 1987 The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution **4**:406-425.

34) TAMURA, K., M. NEI, and S. KUMAR, 2004 Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences USA **101**:11030-11035.

35) TAMURA, K., G. STECHER, D. PETERSON, A. FILIPSKI, and S. KUMAR, 2013 MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution **30**: 2725-2729.

36) KASTRITSIS, C. D., and D. W. CRUMPACKER, 1966 Gene arrangements in the third chromosome of *Drosophila pseudoobscura* I. Configurations with tester chromosomes. Heredity **57:** 150-158.

37) KASTRITSIS, C. D., and D. W. CRUMPACKER, 1967 Gene arrangements in the third chromosome of *Drosophila pseudoobscura* II. All possible configurations. Heredity **58:** 112-129.

38) ANDERSON, W. W., J. ARNOLD, D. G. BALDWIN, A. T. BECKENBACH, C. J. BROWN *et al.*, 1991 Four decades of inversion polymorphism in *Drosophila pseudoobscura*. Proceedings of the National Academy of Sciences USA **88:** 10367-10371.

39) DOBZHANSKY, T., 1944 Chromosomal races in *Drosophila pseudoobscura* and *Drosophila persimilis*. Carnegie Inst. Washington Publ. **554:** 47-144.

40) STURTEVANT, A.H. and TH. DOBZHANSKY 1936 Inversions in the third chromosome of wild races of *Drosophila pseudoobscura*, and their use in the study of the history of the species. Proceedings of the National Academy of Sciences USA **22**:448-450

ACADEMIC VITA

Ian Leopold
1441 Catlin Way, Dresher, PA 19025
Ileopold21@gmail.com

Education

B.S. in Biology, 2014. The Pennsylvania State University, University Park, PA

Honors and Awards

- Schreyer Honors Scholarship, The Pennsylvania State University, 2010-2014
- Phi Kappa Sigma Chapter Merit Scholarship, Phi Kappa Sigma International Fraternity, 2013-2014

Association Memberships/Activities

- Genetics Society of America, Undergraduate member
- Order of Omega
- Alpha Epsilon Delta

Professional Experience

- Research Assistant, Dr. Stephen W. Schaeffer, The Pennsylvania State University, February 2011 – Present
- Laboratory Assistant, Dr. Carmen Sapienza, Temple University, May 2011 – August 2011 & June 2012 – July 2012
- Laboratory Assistant, Dr. Rugang Zhang, Fox Chase Cancer Center, November 2008 – August 2009

Professional Presentations

- Poster presentation of Molecular evolutionary genetics of *Drosophila pseudoobscura* third chromosome, Genetics Society of America meetings, Washington D.C., April 2013