Iowa Research Online
The University of Iowa's Institutional Repository

University of Iowa
Iowa Research Online

Theses and Dissertations

2009

The Role of chromosomal rearrangements in adaptation in Drosophila americana

Paulina Alejandra Mena
University of Iowa

Copyright 2009 Paulina Alejandra Mena

This dissertation is available at Iowa Research Online: http://ir.uiowa.edu/etd/310

Recommended Citation
Mena, Paulina Alejandra. "The Role of chromosomal rearrangements in adaptation in Drosophila americana." PhD (Doctor of Philosophy) thesis, University of Iowa, 2009.
http://ir.uiowa.edu/etd/310.

Follow this and additional works at: http://ir.uiowa.edu/etd


Part of the Biology Commons

THE ROLE OF CHROMOSOMAL REARRANGEMENTS IN ADAPTATION IN DROSOPHILA AMERICANA

by

Paulina Alejandra Mena

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Biology
in the Graduate College of
The University of Iowa

July 2009

Thesis Supervisor: Associate Professor Bryant F McAllister

ABSTRACT

Natural environments expose organisms to multifarious selective pressures involving numerous aspects of the overall phenotype, therefore eliciting a response from multiple correlated loci. It has been hypothesized that chromosomal rearrangements play a role facilitating local adaptation by establishing new linkage relationships and modifying the recombination patterns between the different chromosomal forms, allowing coordinated adaptation of several loci. The central aim of my work is to test this hypothesis using *Drosophila americana* as a model system. This species segregates several inversions and an X-4 centromeric fusion which makes it an excellent model to study the role of chromosomal rearrangements on local adaptation.

I tested the hypothesis using several approaches. I determined the geographic distribution of the chromosomal rearrangements through sampling of wild populations from a broad geographic range. I found that several of the chromosomal rearrangements exhibit clinal variation. Furthermore, many of these are in linkage disequilibrium (LD) with each other. The X-4 fusion is highly associated with inversions on the X and 4^{th} chromosome. Also, two inversions on chromosome 5 are in strong negative LD.

I studied sequence variation associated with rearrangements of the X using inbred lines. The results show the inversion and the fusion strongly influence variation on this chromosome. Regions of significant population differentiation and linkage with the rearrangements are found interspersed with loci showing neutral variation indicating that in spite of recombination, allelic associations are maintained on this chromosome.

I also extended the analysis to flies directly collected from the wild assessing variation with RFLP throughout the X and 4^{th} chromosome. I found long distance LD among loci on both chromosomes, interspersed with regions not showing this pattern.

I also determined there is significant difference in tolerance to extreme cold temperatures among different arrangements of the X chromosome, although I was unable to establish the role of the X in this phenotype.

In conclusion, the clinal distribution of chromosomal rearrangements and associated genetic variation in conjunction with the detection of islands of LD among the rearrangements and loci on both chromosomes indicate that chromosomal rearrangements are facilitating local adaptation in *D. americana*.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

THE ROLE OF CHROMOSOMAL REARRANGEMENTS IN ADAPTATION IN DROSOPHILA AMERICANA

by

Paulina Alejandra Mena

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Biology
in the Graduate College of
The University of Iowa

July 2009

Thesis Supervisor: Associate Professor Bryant F McAllister

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Paulina Alejandra Mena

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Biology at the July 2009 graduation.

Thesis Committee: _____
Bryant F McAllister, Thesis Supervisor

_____
Trudy Burns

_____
Stephen Hendrix

_____
Ana Llopart

_____
Josep Comeron

To my parents Evelyn Carrasco Gerding and Jaime Mena Lorca

ACKNOWLEDGMENTS

I must first acknowledge my advisor Bryant F. McAllister. He provided both financial support and advice for the duration of my Ph.D. He was tremendously supportive and an excellent adviser, always on our side and wanting us to succeed. Not only is he a role model for excellence in research, but also in terms of teaching undergraduate and graduate students in the classroom as well as in the laboratory.

I would also like to thank the past and present members of the McAllister lab, especially Amy Evans, Sergio Flores and Sara Sheeley. Amy's organization and hard work always kept things going smoothly in the lab. I'd like to thank Sergio, my fellow Chilean, for teaching me how to do *in situ* hybridizations but especially for the advice and support he gave me during the time he spent in the lab. I only wish Sergio could have stayed in the lab longer. My most sincere gratitude goes to Sara for being a great friend, keeping the lab fun, being my partner in crime, and staying late with me so many nights helping me with whatever she could do. Sara proofread many pages and heard my talks on numerous occasions; I will owe her forever. But she is not only supportive in the academic sense, she is also a true friend. I hope she does take me home to her parents some day.

There are also people outside of my lab who deserve my gratitude. In particular I would like to thank the members of my Ph.D. committee. They have always wanted me to succeed and invested their time and efforts to help me do so. Thank you for taking the time to read this dissertation and give me ideas and constructive criticism to improve it.

Also, special thanks to my fellow graduate students that took the time to proofread chapters of this dissertation. Thanks to Rebecca Hart-Schmidt, CJ Jenkins, Josh Peloquin and Shannin Zevian.

I had the good fortune of being surrounded by my family during the time I spent in Iowa City. Thanks to my brothers Fernando and Marcelo, and my sister in law Loreto

who were always there for me, each one in their own special way. They made Iowa City seem like home away from home.

Eternal love and appreciation goes to my parents. Without them I would have never even thought about pursuing a career in academics. The way they raised my brothers and I taught us the importance of education as the foundation on which to build a just society. Now that I have achieved my own education, and will be a professor myself, I will try to make my contribution educating future generations.

Finally, I would like to thank all the teachers and professors that have transmitted passion for science and education to me. I hope someday I am able to have the same impact on my students.

ABSTRACT

Natural environments expose organisms to multifarious selective pressures involving numerous aspects of the overall phenotype, therefore eliciting a response from multiple correlated loci. It has been hypothesized that chromosomal rearrangements play a role facilitating local adaptation by establishing new linkage relationships and modifying the recombination patterns between the different chromosomal forms, allowing coordinated adaptation of several loci. The central aim of my work is to test this hypothesis using *Drosophila americana* as a model system. This species segregates several inversions and an X-4 centromeric fusion which makes it an excellent model to study the role of chromosomal rearrangements on local adaptation.

I tested the hypothesis using several approaches. I determined the geographic distribution of the chromosomal rearrangements through sampling of wild populations from a broad geographic range. I found that several of the chromosomal rearrangements exhibit clinal variation. Furthermore, many of these are in linkage disequilibrium (LD) with each other. The X-4 fusion is highly associated with inversions on the X and 4th chromosome. Also, two inversions on chromosome 5 are in strong negative LD.

I studied sequence variation associated with rearrangements of the X using inbred lines. The results show the inversion and the fusion strongly influence variation on this chromosome. Regions of significant population differentiation and linkage with the rearrangements are found interspersed with loci showing neutral variation indicating that in spite of recombination, allelic associations are maintained on this chromosome.

I also extended the analysis to flies directly collected from the wild assessing variation with RFLP throughout the X and 4th chromosome. I found long distance LD among loci on both chromosomes, interspersed with regions not showing this pattern.

I also determined there is significant difference in tolerance to extreme cold temperatures among different arrangements of the X chromosome, although I was unable to establish the role of the X in this phenotype.

In conclusion, the clinal distribution of chromosomal rearrangements and associated genetic variation in conjunction with the detection of islands of LD among the rearrangements and loci on both chromosomes indicate that chromosomal rearrangements are facilitating local adaptation in *D. americana*.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION

Two main requirements for natural selection to take place are that there must be variation within a species, and that this variation must be stably transmitted from the parents to the progeny (Darwin 1859). Unfortunately for Darwin, the underlying form of the heritable variation remained a mystery during his time. Dobzhansky found large rearrangements of *Drosophila pseudoobscura* chromosomes that varied between strains of the same species. Later he realized that these intraspecific differences in chromosomal rearrangements in *D. pseudoobscura* were the raw material for evolution to occur and populations to become adapted to their environment (Dobzhansky 1937).

Chromosomal rearrangements have been implicated in adaptation and speciation in a wide variety of taxa (Coghlan et al. 2005). In vertebrates, the role of chromosomal rearrangements in evolution has been studied for many decades (White 1978), although most of the evidence is based on observations that closely related species differ in chromosome number or morphology. Direct evidence for the role of rearrangements in adaptation in vertebrates has been sparse. However, it was recently found in the common shrew (Banaszek et al. 2009) that individuals that are heterozygous for a Robertsonian fusion have higher fitness than homozygotes.

Invertebrates, and in particular insects, are by far the group that has been most extensively studied in terms of the role rearrangements have played in adaptation and speciation. For instance, inversions have been implicated in speciation of three sympatric populations of the mosquito *Anopheles gambiae*. These differ in their inversions and the stability of these differences is evidence of assortative mating consistent with these populations being reproductively isolated incipient sympatric species (Coluzzi et al. 2002). Also, it has been proposed that inversions contribute to isolation between *Drosophila pseudoobscura* and its close

relative *D. persimilis* because the genomic regions associated with hybrid sterility are found within inversions (Noor et al. 2001).

Chromosomal rearrangements have also been implicated in adaptation within populations. For instance, a Robertsonian fusion in *Drosophila americana* involving chromosome X and the autosomal 4 shows clinal variation and evidence indicates that the cline is maintained by selection (McAllister 2002; Vieira et al. 2001; Vieira et al. 2006; McAllister et al. 2008). Inversion polymorphism in *Anopheles gambiae* has been associated with differences in *Plasmodium* infection rates (Petrarca and Beier 1992). Also, in the grasshopper *Keyacris scurra* two polymorphic inversions are associated with differences in body size and viability (White 1973). Furthermore, many *Drosophila* species show inversion variation correlated with environmental gradients, exhibiting clinal variation (e.g. Ananina et al. 2002; Rako et al. 2006; Anderson 1987).

There are two main hypotheses of how inversions increase in frequency in a population due to adaptation. Dobzhansky's coadaptation model (1970) proposed that loci within the inversion have epistatic effects on fitness. That is to say, combinations of loci are coadapted and in combination generate higher fitness than expected from the sum of their independent effects. An inversion suppresses recombination with other chromosomal arrangements and therefore contributes to the maintenance of linkage disequilibrium among alleles. Kirkpatrick and Barton (2006) proposed an alternative model that does not require loci within the inversion to have epistatic interactions. Under this model an inversion will be favored in a population if it contains two or more locally adapted alleles with additive effects on fitness. Moreover, these two models are not necessarily mutually exclusive adaptive mechanisms. Although the models differ theoretically, the underlying mechanisms are currently indistinguishable in terms of sequence variation.

Both models used to explain the increase in frequency of an inversion in a population due to selection involve variation in at least two genes, which implies that adaptation to the environment involves many genes. This is consistent with the multifarious nature of selection acting on numerous aspects of the phenotype simultaneously (Johnson and Kliman 2002). Correlated selective pressures from the environment will select for optimal character combinations. Selection will generate linkage disequilibrium among alleles that are either coadapted or simply locally adapted to the same conditions. However, segregation and recombination work against these associations by reducing linkage disequilibrium each generation (Falconer and Mackay 1996; Futuyma 1998). In cases where alleles at two or more loci have correlated effects on fitness, the alleles experience similar selective pressures and under these circumstances it could be advantageous to minimize recombination between these loci (Sinervo and Svensson 2002). If these genes are not closely linked, a suppressor of recombination between these alleles would be beneficial because it would decrease the occurrence of recombinants between them (Kimura 1956; Feldman et al. 1997; Pepper 2003). Maintenance of positive correlations among alleles is the essence of the coadapted gene complex.

Tight linkage between coordinately selected loci minimizes segregational and recombinational load (Hurst 1999). However, if selected loci are not tightly linked, chromosomal rearrangements could facilitate selection through alteration of linkage relationships and recombination patterns, thus protecting linkage disequilibria generated by selection. This way, populations are able to adapt to the local conditions even if there is gene flow with other populations. This is especially relevant when a species has a broad distribution. Populations are exposed to distinct environments and become adapted to the local conditions. Clines for chromosomal rearrangements are a manifestation of this local adaptation (Hoffmann and Rieseberg 2008). Clines for chromosomal inversions are often interpreted in terms of climatic

selection (Lee et al. 2002), but the selective factors involved are poorly understood. Latitudinal clines for inversion frequencies are repeated in different hemispheres and continents (Kimbas and Powell 1992). For example, *Drosophila subobscura* originated in Europe, where there are documented clines for inversions (Sole et al. 2002). Inversion clines that mirror the European clines have been established along the west coasts of South and North America. (Prevosti et al. 1988). Also, a change in inversion frequency related to the direct or indirect effects of temperature shifts due to global warming has been reported for *D. subobscura* (Balanya et al. 2006).

The overarching hypothesis of this study is that chromosomal rearrangements facilitate local adaptation. The correlated selective pressures from the environment cause a coordinated response, therefore resulting in linkage disequilibrium among alleles under selection. Chromosomal rearrangements will aid in maintaining these allelic associations. This hypothesis was tested using *Drosophila americana* as a model system.

1.1 Model System

D. americana is an excellent system to study correlated selection and the role that chromosomal rearrangements play in adaptive evolution. The main feature that makes it suitable for such an investigation is that this species segregates multiple chromosomal rearrangements (Hsu 1953; Blight 1955) and natural selection has been implicated in the maintenance of some of these (McAllister 2002; Evans et al. 2007; McAllister et al. 2008). If the hypothesis is correct and chromosomal rearrangements do facilitate adaptation, I would expect to find these chromosomal rearrangements distributed in a clinal fashion as an indication that selection is favoring certain arrangements in different parts of the species' range. Another piece of evidence that chromosomal rearrangements are facilitating adaptation would be to find associations among loci and rearrangements in spite of the opportunity for recombination because this would be an indication that selection is acting to maintain the linkage disequilibrium.

D. americana is a species endemic to North America. It occupies a large range east of the Rocky Mountains (Patterson and Stone 1952). Although the north-south limits of the species distribution are unclear it appears to be bound in the south by the Gulf of Mexico while the northern limit appears to be close to 45°N (estimated from Hsu 1953 and McAllister lab collections). This species is not a commensal to humans, which is an advantage to studying its natural population structure because human activities are unlikely to affect short-term dynamics. *D. americana* occurs in riparian habitats and is associated with willow (*Salix exigua* and *S. nigra*). Blight and Romano (1953) reported finding larvae within slits in the bark on the trunk of sandbar willow (*S. exigua)* during fall collections and they also recovered pupal cases during winter collections, however they reported that only parasitoid wasps emerged from these cases. Other than the association with willow, little is known about the life history of this species.

D. americana belongs to the virilis subgroup (Caletka and McAllister 2004). Both *D. americana* and *D. novamexicana* comprise the North American lineage of the virilis subgroup which diverged from a Palaeartic ancestor approximately 3 million years ago (Hilton and Hey 1996, 1997; Spicer and Bell 2002; Caletka and McAllister 2004). *D. americana* and *D. novamexicana* are currently allopatric species, with their ranges separated by the Rocky Mountains. It has been estimated that these two species diverged less than 0.4 mya (Caletka and McAllister 2004).

In *D. americana* there are two alternative karyotypes. The ancestral karyotype consists of two pairs of acrocentric chromosomes, one pair of metacentric chromosomes, a pair of sex chromosomes and the dot chromosomes. Conversely, the derived karyotype has a centromeric fusion of the X chromosome and the acrocentric chromosome 4, which corresponds to Muller element B (Throckmorton 1982). The alternative karyotypic forms are often treated as separate species (Orr and Coyne 1989; Pitnick et al. 1997; Powell 1997; Spicer and Bell 2002; Garbuz et al. 2003) or subspecies (Throckmorton 1982; Hilton and Hey 1996, 1997). However, more recent

examination of wild collected flies from throughout the species range has shown that the frequency of the two arrangements of the X shift along a latitudinal gradient within the United States, where the derived fused X-4 arrangement is found at higher frequency in northern populations, and the ancestral unfused arrangement is more prevalent in the southern part of the species range (McAllister 2002; McAllister et al. 2008).

Both arrangements are present in most populations, indicating that the chromosomal polymorphisms are distributed over a large geographic range. Frequencies of the X-4 fusion vary from 100% in the northernmost collection site (44.1°N) to 0% in the southernmost collection site (30.76°N) and are strongly correlated with latitude. The width of the cline is approximately 623 km and has a sigmoid shape, characteristic of a balance between selection and migration (McAllister et al. 2008).

One major piece of evidence that indicates the cline for these alternative arrangements are maintained by natural selection is the finding that the distribution of the X-4 fusion is unique relative to variation in other regions of the genome. Neutral loci surveyed throughout the genome do not show population differentiation among geographically distinct populations or variation correlated with latitude. This is an indication that gene flow is sufficient to homogenize neutral sequence variation throughout most of the genome (McAllister and McVean 2000; McAllister 2002, 2003; Schäfer et al. 2006). The lack of population structure in this species suggests that the cline for the X-4 fusion is maintained by natural selection favoring different karyotypes and the loci associated with them in different parts of the geographic gradient. Furthermore, this species also segregates for several paracentric inversions although their distribution and the role adaptation has had shaping their distribution has not been studied in the same detail as the X-4 fusion.

We hypothesize that the chromosomal rearrangements present in this species, including inversions and the X-4 centromeric fusion are facilitating coordinated evolution for adaptation of

populations to local environments through the maintenance of coadapted sets of locally adapted alleles in high linkage disequilibrium.

1.2 Objectives

The overall objective of this study is to investigate the role chromosomal rearrangements play in adaptation to the environment. I hypothesize that chromosomal rearrangements are selected for because they facilitate the maintenance of allelic associations beneficial in a local environment.

I did this by determining each of the following:

Geographic Distribution of Chromosomal Variation: To date there is little information about polymorphic inversions in *D. americana*. The work of Hsu (1952) and Blight (1955) describe the presence of polymorphic inversions in all chromosomes and the data suggest that some of these might also show clinal distributions, although an extensive survey is needed. Therefore, I examined the geographic distribution and possible associations among the main polymorphic chromosomal rearrangements found in *D. americana*. I found that in addition to the already established cline of the X-4 fusion, there are clines for multiple inversions. Also, I found strong linkage disequilibrium between chromosomal rearrangements. The lack of population structure in this species in conjunction with the clinal distribution and associations among rearrangements is indicative of natural selection acting to maintain this pattern.

Sequence Variation Associated with Alternative Chromosome X Arrangement: Another objective of this study is to use patterns of sequence variation to understand the history of the chromosomal rearrangement involving the X. Polymorphic inversions can have dramatic effects on patterns of sequence variation. Nevertheless, most of the previous studies that examined variation on chromosome X of *D. americana* did not take into consideration the possible effects that the polymorphic inversion Xc could have on the observed patterns. As expected, I found that the polymorphic inversion on the X has major effects on sequence variation. From the variation

patterns I conclude this inversion has been segregating in the population for an extended period of time and that the X-4 fusion appears to be a secondary rearrangement involving the preexisting Xc polymorphism.

Geographic Distribution of DNA Variation Associated with Chromosomal Rearrangements of the X and 4th Chromosome: The clinal distribution of the chromosomal rearrangements suggests that natural selection affects the alternative arrangements through associated alleles; therefore it follows that variation at selected loci on these chromosomes will be in linkage disequilibrium with the rearrangements. Consequently, variation at these genes will also be correlated with the environmental gradient, similar to what is observed for the chromosomal rearrangements. I assayed variation throughout chromosome X and 4 using RFLP in a large and widely distributed sample of *D. americana*. The results indicate that there are islands of linkage disequilibrium interrupted by regions where variation is not associated with the chromosomal rearrangements throughout both chromosomes. These findings lend support to the hypothesis that *D. americana* is under correlated selection and chromosomal rearrangements contribute to the maintenance of beneficial allelic associations favored within specific environments.

Phenotypic Differentiation between Chromosomal Arrangements of Chromosome X: Distribution of the X-4 fusion and *In(X)c* along a latitudinal gradient suggest that traits that are responsible for adaptation to extreme cold and hot temperatures are likely candidates for association with these rearrangements. Consequently, it is of interest to establish if there is any phenotypic differentiation, such as tolerance to cold or hot temperatures between the alternative X chromosome karyotypes. To achieve this, I performed assays for tolerance to extreme heat and cold temperatures on lines with alternative chromosomal arrangements. There were no observed differences among lines for heat tolerance. Differences were observed in tolerance to cold

temperature; however the assay used did not allow me to establish a role for the X chromosome and its rearrangements in these differences.

The compilation of the findings from each of the aspects investigated indicates strong support for the hypothesis that *D. americana* is under correlated selective pressures and chromosomal rearrangements have facilitated adaptation to these conditions by preserving beneficial allelic combinations.

CHAPTER 2

GEOGRAPHIC DISTRIBUTION OF CHROMOSOMAL VARIATION IN *DROSOPHILA AMERICANA*

2.1 Introduction

Understanding the genetic basis of adaptation to local environmental conditions is a major focus of evolutionary biology. One approach to study adaptation is to focus on species that occupy a large geographic range because they are often subjected to diverse environmental conditions. Distinct populations respond to unique selective pressures imposed by the environmental conditions they encounter, often resulting in clinal variation for certain adaptive traits. For example latitude may be associated with many traits that respond to the corresponding gradient in climatic conditions. In *Drosophila* clinal variation has been reported for traits such as body size and development time (James and Partridge 1995; Jameset al. 1995; Bitner-Mathe and Klaczko 1999; Huey et al. 2000; Loeschcke et al. 2000; Hallas et al. 2002; Delong and Bochdanovits 2003; Gilchrist et al. 2004; Norry et al 2006; Sambucetti et al. 2006; Arthur et al. 2008), thermal resistance (Hoffmann et al. 2002; Ayrinhac et al. 2004), timing of reproduction (Mitrovski and Hoffmann 2001), diapause (Schmidt et al. 2005; Schmidt and Conde 2006), starvation resistance (Robinson et al. 2000) and egg size (Azevedo et al. 1996). These and other quantitative traits that respond to the environment in a coordinated fashion imply that a several loci are involved in local adaptation. Therefore, this strong environmental gradient generates a genome-wide response. Coordinated selection pressures create associations among alleles that can be evidenced through patterns of linkage disequilibrium. However, recombination and independent assortment disrupt associations favored by selection, impeding the formation of a coadapted genome.

Chromosomal rearrangements may facilitate local adaptation by establishing new linkage relationships and modifying rates of recombination between different chromosomal forms. A

theoretical analysis by Kirkpatrick and Barton (2006) demonstrated that a chromosomal inversion is favored by selection and will increase in frequency if it suppresses recombination between two or more loci coordinately, but independently, adapted to the same environmental gradient. In contrast, Dobzhansky's (1947, 1954, 1970) coadaptation model proposes that inversions represent coadapted alleles where fitness benefits come from interactions between loci within the inversion i.e. epistasis.

Inversion polymorphisms, and to a lesser degree other chromosomal rearrangements such as chromosomal fusions have been studied in numerous species and are generally found to exhibit variation associated with latitude, altitude or annual seasons (Hoffmann et al. 2004). This may be an indication that chromosomal rearrangements are selected upon because they coordinate additive genetic variation in response to these gradients of selection.

Several features of the widely-distributed North American species *Drosophila americana* make it an excellent model to study effects of coordinated natural selection on the distribution and association among different chromosomal rearrangements. This species segregates several chromosomal rearrangements and it has been suggested that some of these are maintained by local selection (McAllister 2002; Evans et Al. 2007;McAllister et al. 2008). Also, there is evidence that indicates this species has occupied its current range for an extended period of time and has a stable demographic history (Caletka and McAllister 2004). A stable demographic history is relevant for the study correlated selection on a genome because it should minimize demographic effects created by recent colonization and population expansions that also can create patterns of linkage disequilibrium that are not products of natural selection. Previous analyses have focused primarily on a Robertsonian fusion of the X and fourth chromosome (Muller element A and B, respectively) that segregates in natural populations and is distributed along a latitudinal cline in North America (McAllister 2002; McAllister et al. 2008). While the frequency of the X-4 fusion in a population is strongly correlated with latitude, most genes show

no differentiation among geographically distinct populations, indicating that gene flow is sufficient to homogenize neutral sequence variation throughout most of the genome (McAllister and McVean 2000; McAllister 2002, 2003; Schäfer et al. 2006). The contrasting patterns of variation indicate that the cline is maintained by natural selection favoring different karyotypes and the loci associated with them in different geographic regions.

In addition to the X-4 fusion, several common paracentric inversions also segregate in populations of *D. americana*, although only limited information is available on the distribution and associations between rearrangements. The comprehensive study of chromosomal variation by Hsu (1952) surveyed inversions in small samples representing a wide geographic range, but provided a limited assessment of their frequencies across this range. Both the X and 4th chromosomes contain inverted arrangements in *D. americana*, and the inversions are distributed similarly to the X-4 centromeric fusion, yet the associations among arrangement have not been quantified over the species' range. On the other hand, Blight (1955) examined a series of large samples from a population near St. Louis, Missouri and simultaneously determined inversion status and centromeric arrangement. Common inversions in this population, *In(X)c* and *In(4)ab*, were strongly associated with the X-4 fusion in wild flies. However, the persistence of these associations outside of this population is unknown.

Association among rearrangements maintained throughout the species range is evidence for the presence of coordinated selective pressures. Alternatively, these associations could have been established as a result of the series of events that gave rise to each of the rearrangements i.e. if an inversion occurred on a chromosome that already had another rearrangement, these will be found in high association merely as a consequence of historical events. However, DNA sequence variation studies of loci in and near these inversions have revealed that each of these inversions most likely originated independently and on chromosomes with no other rearrangements (unfused X and 4th chromosomes) (Evans et al. 2007; Chapter 3). Both Xc and 4a inversions

show a signature of natural selection and sequence variation that indicates that they occurred previous to the X-4 fusion (Evans et al. 2007 and Chapter 3), while 4b, the small inversion within 4a occurred after (Evans et al. 2007). Xc and 4a were present before the occurrence of the X-4 fusion, and they may have been distributed in clines similar to current geographic patterns.

There is also an interesting pattern involving the inversions on Chromosome 5. This chromosome has two inversions; *In(5)a* and *In(5)b*. *In(5)a* is a large submedian inversion and according to Hsu (1952) it is found in the southern part of the species range. *In(5)b* is a small subterminal inversion found in the northwestern localities surveyed by Hsu (1952). However, in this study there is no measure of the linkage disequilibrium between these inversions or the possible association with other rearrangements. Blight's work (1955) shows that there is strong linkage disequilibrium between inversion 5a and 5b in the Missouri population that he studied. The 5 standard chromosomes were very rare in the population studied, while the hypothetical arrangement carrying both inversions was completely absent in the sample. The pattern of linkage disequilibrium between these two inversions seems to mirror the pattern observed for inversions 4ab and Xc on the X-4 fused chromosomes. While coordinated selection has shaped variation along all these chromosomes, the rearrangements on chromosome 5 differ from those on the X and 4 because there is a negative association between the derived inversions.

The observed distribution and association between inversions indicate coordinated evolution. It has been hypothesized that these rearrangements preserve locally adapted variants and that the association between them is shaped by coordinated selection. Given that there is suppressed recombination within and among rearrangements, there is also the possibility of coadaptation between loci through fitness epistasis. Moreover, differentially adapted populations with distinct karyotypes such as those in *D. americana* can serve as precursors to speciation (McAllister et al. 2008).

Therefore, the objective of this study is to determine if the patterns of chromosomal variation are consistent with chromosomal rearrangements facilitating coordinately adaptation. I carried this out by characterizing the geographic distribution of chromosomal variation and test for associations between rearrangements throughout most of the species' range. I did this through the examination of wild flies collected from several localities within a broad geographic range. The analysis of chromosomes in natural samples provides the opportunity to directly examine the products of natural selection because unfit genotypes will be eliminated from the population, as opposed to lines maintained in the laboratory that could accumulate variation that may not survive under natural conditions.

Specifically, I determined the distribution of the chromosomal rearrangements in order to establish if any of these show clinal variation similar to the already documented cline of the X-4 fusion. Also, I measure any associations among chromosomal rearrangements throughout the species range. Linkage disequilibrium among rearrangements could be of significance because linkage disequilibrium can be caused by coordinated selection favoring certain allelic combinations under specific environmental conditions. Populations from the extremes of the species range face very environmental conditions such as different levels of winter severity and other climatic variables and it is expected that they have evolved distinct strategies to cope with these differences. Establishing that certain rearrangements vary in relation to latitude, longitude, or other rearrangements is relevant because it reflects a role for chromosomal rearrangements in shaping variation throughout the genome.

2.2 Materials and Methods

2.2.1 Fly Samples

I sampled populations of *D. americana* throughout the central United States from a geographic region bounded by 30° to 40° N latitude and 98° to 83° W longitude (Table 2.1). I

obtained the geographic coordinates of each sample site using a handheld Magellan Map 410 GPS receiver. The method of collection is described in detail in McAllister et al. (2008). Frequency of the X-4 fusion ranges from 0% to 100% over this geographic region (McAllister et al. 2008), allowing a robust analysis of its possible associations with inversions. I used two inbred lines, each expressing a recessive mutant eye color and having known chromosome arrangements, as laboratory standards. ML97.5-pur has a dark purple eye phenotype. In addition to the inversions that are fixed between *D. americana* and *D. virilis*, this line contains one inversion (5a) relative to the standard arrangement of the virilis species group (Hsu 1952) and has the ancestral unfused X and 4. NN97.4-red has a bright red eye phenotype. This line also shares a common set of inversions fixed between *D. americana* and *D. virilis*, and in addition, this line contains inversions Xc, 4ab and 5b and has the derived X-4 centromeric fusion. I mated each wild-caught male to one of these standard laboratory strains with known karyotypes and I used the F1 progeny to determine the centromere and inversion status for each wild-caught fly.

2.2.2 X-4 Centromere Status Determination

I identified the arrangement of the X in each wild-caught male as being fused or independent of the 4th chromosome. This assessment was based on the transmission of microsatellite alleles for loci on the 4th chromosome as described in Evans et al. (2007). Briefly, I crossed each wild-caught male to a standard laboratory line of *D. americana*. I pooled the F1 progeny in groups of 6 males and 6 females and the pattern of inheritance of heterozygous microsatellite markers distributed throughout the 4th chromosome was assessed as sex linked or autosomal, indicating the presence of the X-4 fusion or an unfused X, respectively.

2.2.3 Cytological Preparations

I determined the presence of inversions for all chromosomes of each wild caught male through the analysis of polytene chromosomes of F1 progeny produced from the cross of each

wild caught male to a standard inbred line of *D. americana*. I performed polytene chromosome squashes on the salivary glands of at least 5 F1 third-instar larvae of each wild caught male using standard methods for fixing in 45% glacial acetic acid and staining in acetic orcein (Kennison, 2000). I identified each chromosome based on its unique morphology and banding pattern (Gubenko and Evgen'ev 1984). I sexed larvae before the dissection by determining the size of the gonad. Testes are easily visible under the microscope. I only used female F1 progeny to determine the presence of inversions on the X chromosome inherited from each wild-caught male. The arrangement of the 4th chromosome was also specifically associated with the sex of each larva. In cases where the wild-caught male has the X-4 centromeric fusion, differences between female and male F1 progeny can arise due to sex-specific transmission of the 4th chromosome. For each of the other chromosomes, I used five F1 larvae to infer the diploid genotype of the wild caught male with an error rate of 0.031. If I was unable to determine the presence of inversions because of the low quality of the squashes, the wild caught male was eliminated from the analysis. If I used less than 5 squashes, I did not use this male to determine the diploid genotype.

2.2.4 Statistical Analysis

I typed each male collected for inversions and centromere status; therefore this provides information regarding the frequency of the inversions for all chromosomes. Furthermore, centromeric arrangement is not evident in preparations of polytene chromosomes of *D. americana*, thus the combined analysis provides inversion data within the context of the alternative arrangements of the centromere of the X. I performed a logistic regression to measure clines (north-south as well as east-west) for alternative karyotypes in addition to the already documented north-south cline of the X-4 fusion. For each chromosomal rearrangement, I fitted logistic regressions on the number of successful events (number of inverted or fused chromosomes) per trials (number of chromosomes typed) for each site. Latitude and longitude

were evaluated as explanatory variables in logistic regression models. I performed all the statistical analysis using Proc Genmod in SAS v9.1 (Cary, NC). I performed a likelihood ratio test to determine the model that best fit the data.

I determined whether there is linkage disequilibrium among arrangements through a Fisher's exact test. Also, in the case of chromosome 5 I calculated the expected frequencies of the genotypes based on the frequencies of the individual inversions and performed a χ^2 test to determine if there are significant differences between the expected and observed genotype frequencies. For chromosome 5 I also determined if there is a deviation from Hardy-Weinberg equilibrium. For this I performed a χ^2 test to determine if there were significant differences among the observed and expected genotype frequencies. I was able to do this because there are only 3 genotypes in the population (with the exception of the very rare standard arrangement): 5a-/5a-, 5a-/5-b and 5-b/5-b.

2.3 Results

2.3.1 Distribution of the X-4 fusion

Consistent with previous reports (McAlister 2002, McAllister et al. 2008), the X-4 fusion exhibits clinal variation with high frequencies of this arrangement in the northern localities and low frequency at the more southern sites (figure 2.1). The alternative arrangements of the X are distributed over a wide geographic range with only the most southern locality containing a single arrangement. This locality corresponds to CI and is near the southern edge of the species range (McAllister et al. 2008).

2.3.2 Distribution of Inversion Xc

In(X)c is a large paracentric inversion. It was found in perfect association with the X-4 fusion. Of 321 X chromosomes typed, 264 contained In(X)c and were fused with chromosome 4, while 57 maintained the standard gene order and were not fused to the 4th chromosome.

Interestingly, the other possible karyotypes were completely absent from the sample. This implies that *In(X)c* has exactly the same distribution as the X-4 fusion; therefore it shows the same clinal variation (figure 2.1). Latitude used as a predictor variable significantly improves the log likelihood of a logistic regression model predicting the frequency of *In(X)c* in the sample (χ^2=150, DF: 1, p<0.0001). However, the inclusion of longitude does not improve the fit, therefore we conclude that latitude and not longitude is a good predictor of the frequency of these arrangements. Furthermore, the predicted frequencies of the X-4 fusion and *In(X)c* using only latitude closely match the observed frequencies obtained from the sampling (figure 2.1). Inversion Xc shows a sigmoidal shape cline with frequencies changing more drastically with latitude in the middle of the range and stabilizing towards the extremes.

2.3.3 Distribution of Inversion 4ab

In(4)ab is an inversion complex consisting of a small paracentric inversion within a larger submedian inversion. This inversion complex never occurs within unfused 4th chromosomes and only occurs within a subset of X-4 fused chromosomes. Of the 57 unfused X and 4th chromosomes examined here, none had the 4ab inversion. On the other hand, of the 264 fused X-4 chromosomes, 97 had the 4ab inversion. A Fisher's exact test shows that there is a strong association between the 4ab inversions and the X-4 fusion (p<0.0001)

The 4ab inversion complex is found exclusively on the X-4 fused chromosomes; therefore it occurs exclusively in northern populations. The southern localities surveyed (CI, RB, LR and BB) completely lack *In(4)ab* and the southernmost locality where this inversion complex is present, albeit in very low frequency (0.18), is CF (latitude 38.1°N). However, this inversion is always found at low frequency. The highest frequency observed for all the 4th chromosomes typed in this collection was 0.3, which corresponds to the SV07 site (Figure 2.2). When I only consider 4th chromosomes attached to the X, the highest frequency observed is 0.76 in the NN population. Latitude and longitude are good predictors of the frequency of this inversion

complex (estimates of β: 0.25, 95% CI:0.03-0.48 and -0.03, 95% CI: -.08- -0.02 respectively). A likelihood ratio test revealed that the model that best fit the data set included latitude and longitude as predictor variables (χ^2=6, df:2, p<0.05)

2.3.4 Distribution of Inversion 5a and 5b

In(5)a is a geographically widespread large submedian inversion (Table 2.4) with highest frequencies in the southern populations (Figure 2.3). It is fixed in all the southernmost populations (CI, RB, LR and BB). *In(5)b* is a very small subterminal inversion present in more northern and western populations (Figure 2.3). The likelihood ratio tests revealed that the model that fits bests the inversion frequency data is one that includes latitude and longitude (estimates of β: 0.09, 95% CI: 0.03-0.14 and -0.13, 95% CI: -0.33- -0.06 respectively) as the predictor variables ((χ^2=10, df:2, p<0.05).

The two inversions on chromosome 5 show a very strong pattern of linkage disequilibrium. These inversions are never found on the same chromosome and chromosomes without either inversion are extremely rare (4 of 644 chromosomes typed). A Fisher's exact test shows that they are in high linkage disequilibrium (p<0.0001) (Table 2.2). As indicated previously, there is a large deficit of chromosomes with both inversions and chromosomes with no inversions at all, while those having single inversion are overrepresented. A χ^2 test reveals there is a highly significant difference between the observed and expected frequencies of each haplotype. (χ^2 =593.1, df=2, p<0.001).

The level of negative linkage disequilibrium is especially surprising because several populations are polymorphic for both inversions and heterozygous individuals were observed. Of the 273 individuals where I inferred the arrangement of both 5th chromosomes, 39 individuals were double heterozygotes (5a-std/5std-b). The observed diploid genotypes deviate significantly from the Hardy-Weinberg expectation due to a deficit of heterozygotes (χ2=75, df=2, p=0.001) (Table 2.3). It is noteworthy that there is most likely a bias against typing an individual as

heterozygote because of the sampling procedure. When 5 F1 larvae are typed for the inversion there is approximately a 3% chance of mistakenly identifying a heterozygote as a homozygote. However, the probability increases to 12% error rate when only 3 F1 larvae are used to determine the diploid genotype. Cases where I examined fewer than 5 F1larvae were eliminated from this analysis.

2.4 Discussion

The strong association among several of the rearrangements found throughout the range of *D. americana* in combination with the clinal distribution that they exhibit provides evidence that this species' genome has been under correlated selective pressures imposed by the environmental gradients present over the geographic range it inhabits. These correlated selective pressures have resulted in a genome wide response affecting variation at chromosome X, 4 and 5, implying many loci are involved in adaptation to the local conditions. Chromosomal rearrangements facilitate maintenance of linkage disequilibrium between locally adapted loci and therefore are favored in populations connected by gene flow. Demographic effects on the population are not responsible for the distribution or association between arrangements because it has been shown that control loci, not linked to rearrangements show neutral variation and no population differentiation (McAllister and McVean 2000; McAllister 2002, 2003; Schäfer et al. 2006), indicating that gene flow within the species is sufficient to homogenize variation at neutral loci.

A possible non-adaptive scenario that can explain the observed pattern of distribution and association between arrangements involves the historical events that gave rise to the rearrangements currently found in *D. americana*. In other words, if a rearrangement occurs on a chromosome that already contains a rearrangement these will be in linkage disequilibrium. However, evidence from analysis of sequence data have shown that this is not the case for

associations found in *D. americana* (Evans et al. 2007; Chapter 3). These will be explored case by case next.

2.4.1 Rearrangements on Chromosomes X and 4

The data presented here confirm the previously established north-south cline for the X-4 fusion (McAllister et al. 2008). However, I also show that the cline is not independent of the large submedian inversion Xc located on the X chromosome. The extremely high association between the centromeric fusion and Xc had been previously reported by Blight (1955), but only a single population was examined. My data confirm this association and demonstrate its presence throughout the species range. Blight typed 1863 X chromosomes and only 5 were observed as being unfused with the Xc inversion. The presence of these rare chromosomes is evidence that recombination occurs in the region between the proximal Xc inversion breakpoint and the centromere. It is therefore remarkable that this association is maintained because even low levels of recombination would break down this association suggest the presence of association is evidence that natural selection is responsible. Hsu (1953) reported that Xc was found at high frequencies in northern populations and at low frequencies in southern populations, but in this study the status of the centromere as fused and unfused was not determined. Instead lines were classified by subspecies and therefore it was assumed that flies collected from the north would have the X-4 fusion, while southern populations would lack it. This assumption was based on the understanding that *D. americana* consisted of two karyotypically defined taxa (Patterson and Stone 1952; Throckmorton 1982) with distinct distribution and a narrow hybrid zone.

The nested inversions 4ab are associated with the X-4 fusion, and therefore also with *In(X)c*. Blight (1955) found this same pattern of association: of 1863 chromosomes typed, unfused X and 4 with *In(4)ab* chromosomes were not found. However, one case of this arrangement was found in the progeny of a laboratory cross, which most likely resulted from a crossover event between a fused X-4 with *In(4)ab* and unfused 4 with the standard arrangement

in the region between the proximal inversion breakpoint and the centromere. As rare as this is, it is evidence that recombination can breakdown associations among rearrangements and the fact that they are maintained indicates natural selection preserves the observed associations.

Furthermore, inversions on both the X and the 4th chromosome are linked through the centromeric fusion. Although there is evidence of recombination between the centromere and the proximal breakpoints of both inversions, in wild populations they are almost always associated.

Associations between rearrangements like these could be due to the sequence of events that gave origin to them. In this case, the observed associations would result from both inversions occurring on a fused X-4 chromosome. *In(4)ab* is found only on a subset of the X-4 fused chromosomes and therefore is always associated with the X chromosome, consistent with the inversion originating on an X-4 fused chromosome. However, this inversion is also present in *D. novamexicana*, an allopatric sister species of *D. americana* that does not possess the X-4 fusion (Hsu 1952), indicating that the inversion was present in the common ancestor of *D. americana* and *D. novamexicana*. Also, analysis of sequence variation of loci located around the distal breakpoint suggests this is a relatively old inversion (Evans et al. 2007). The same is true for *In(X)c*. Even though it has a perfect association with the X-4 fusion, cytological and sequence data indicate this inversion was present before the fusion occurred. The Xc is also present in *D. novamexicana* (Hsu 1952), indicating it existed in the common ancestor of both species. Also, sequence variation within the inversion indicates it has been segregating in the population for an extended period of time because variation levels are similar in both chromosomal classes, an indication that the variation lost in inverted chromosomes due to the selective sweep that increased its frequency in the population has been recovered through recombination between both chromosomal classes (Chapter 3).

Both the inversions on the X and the 4th chromosome show signatures of selection. Analysis of sequence variation associated with *In(4)ab* has revealed evidence of two independent

selective sweeps among chromosomes with the inverted arrangement (Evan et al. 2007). One of these sweeps appears to result from the coupling of the inverted 4th chromosomes to the X-4 fusion. Evans et al. (2007) found evidence for a second selective sweep involving 4b affecting only X-4 fused chromosomes with 4a. Becoming sex linked may result in coadaptation through intra locus sexual conflict (Rice 1987). An inversion that restricts recombination in a sex linked region will allow the accumulation of sex specific alleles. Evans et al. (2007) hypothesized that feminizing selection has driven coadaption between loci because the X-4 fused chromosome complex is present more often in females and will be cotransmitted with other female factors. In addition, sequence variation associated with *In(X)c* has revealed that coadaptation extends to chromosome X, *In(X)c* and the associated loci, specifically in loci located near the proximal inversion breakpoint (Chapter 3).

2.4.2 Rearrangements on Chromosome 5

The most remarkable feature observed for the rearrangements on chromosome 5 is the extreme linkage disequilibrium between 5a and 5b. Chromosomes with both inversions were never observed. Also, only a few chromosomes were observed without either inversion. This is consistent with a previous study by Blight (1955). Of 2431 chromosomes 5 typed from the wild only 4 had the standard arrangement (found in the heterozygous form) and no chromosomes with both inversions were found. In addition, a standard chromosome 5 was recovered from a laboratory cross, most likely as a product of a crossover event in a heterozygous female. However, the other expected crossover product (a chromosome with both inversions) was never recovered.

Interestingly, these inversions are located very distantly from one another on the chromosome (figure 2.4), suggesting that there is ample opportunity for recombination between these inversions. Also, many populations in this survey are polymorphic for these inversions and heterozygotes are not uncommon. All this suggests that recombination between these inversions

should be possible. However, recombinants are never found in natural populations, implying that natural selection is acting to maintain these in this linkage disequilibrium. These data indicate that 5 standard and especially 5ab are unfit karyotypes.

Historical and demographic events can be ruled out as the cause for the negative association between these rearrangements because there is sufficient gene flow among populations and the possibility for recombination between the inversions to break down the linkage disequilibrium between them. Therefore, the pattern observed in chromosome 5 is consistent with coordinated evolution. That is, the inversions carry locally adapted alleles and the association between these inversions is shaped by coordinated selection, resulting in linkage disequilibrium that maintains alleles adapted coordinately together. However, there is also the possibility that the observed pattern is the result of non environmental factors. For example, mating preferences could be involved, where positive assortative mating results in flies preferring to mate with individuals with the same arrangements (or some unknown phenotype associated with the chromosomal arrangements) (DeVoogd et al. 1995; Guelbeogo et al. 2005) .

2.4.3 Western Populations and Other Inversions

Northwestern populations show a unique pattern in terms of presence of rare inversions (Hsu , 1952). In this study, inversions 4ab and 5b increase in frequency in the northwestern samples. In addition to this, Hsu (1952) describes several inversions that are unique to this region that I did not find in the present study. These inversions seem to be found at low frequencies but most are fixed in *D. novamexicana*. Inversion Xd was reported for some strains from Nebraska, while 2b, 2c and 3a were also reported for western strains of *D. americana* and each of these is fixed in *D. novamexicana*. The reason why the western populations of *D. americana* have a high number of unique inversions is unknown. There could be environmental conditions in this area that make these advantageous. On the other hand, these populations could have undergone a

period of isolation that resulted in the accumulation of these inversions. Or, it could be a result of a combination of these two scenarios.

2.4.4 Conclusions

The association of inversions described in this study with geographic gradients and with each other in *D. americana* are evidence that chromosomal arrangements facilitate local adaptation through coordinated selective pressures that affect traits involving many loci. The genome as a whole responds to these kinds of selective pressures, although, recombination in general will tend to disrupt adaptation by breaking apart favorable associations. Chromosomal rearrangements alter the linkage relationships and the recombination patterns and therefore allow a coordinated response involving many loci.

Table 2.1 Frequency of chromosomal rearrangements found in each sample site.

Site ID	Latitude	Longitude	N	Xc/fusion	4ab	5a	5b
IR04	41.7	-91.7	45	1	0.44	0.79	0.20
CI05	30.7	-91.4	10	0	0	1	0
RB05	32.5	-89.9	18	0.11	0	1	0
LR05	34.7	-89.9	24	0.41	0	1	0
BB05	36.4	-89.3	23	0.65	0	1	0
CF05	38.1	-88.9	11	0.90	0.18	0.86	0.04
DI05	40.4	-89.9	51	0.96	0.43	0.48	0.51
DA06	32.5	-87.8	15	0	0	1	0
FG06	30.7	-84.8	10	0	0	1	0
BU06	31.9	-85.0	5	1	0	1	0
MK07	38.9	-85.8	27	0.92	0.11	0.94	0.05
SV07	39.6	-93.1	13	1	0.38	0.84	0.15
OC07	38.3	-87.3	15	0.93	0.2	0.93	0.06
WS07	40.7	-82.0	17	1	0.37	0.97	0.02
OR07	41.6	-83.2	6	1	0.5	0.83	0.16
II07	41.1	-87.5	7	0.85	0.14	0.93	0.06
SC07	40.1	-95.2	7	0.85	0.28	0.71	0.28
NN07	42.7	-98.0	24	0.95	0.76	0.36	0.58
DN07	41.3	-97.4	23	1	0.56	0.72	0.28

Note: The number in each locality name corresponds to the year the collection was made. N is the total number of individuals that were assayed. Site locations can be observed in figure 2.2.

Table 2.2 Observed and expected frequencies of chromosome 5 haplotypes based on the frequency of each inversion.

	5a-	5ab	5-b	5--
Observed	515	0	125	4
Expected	415	100	25	104
χ^2 =593.1, df=2, p<0.001				

Table 2.3 Observed and expected genotypes under HW equilibrium for chromosome 5.

Genotype	Observed	Expected
5a-/5a-	97	86
5a-/5-b	21	44
5-b/5-b	86	6
χ2=75, df=2, p=0.001		

Figure 2.1 Observed and predicted frequencies of the Xc inversion plotted against latitude.

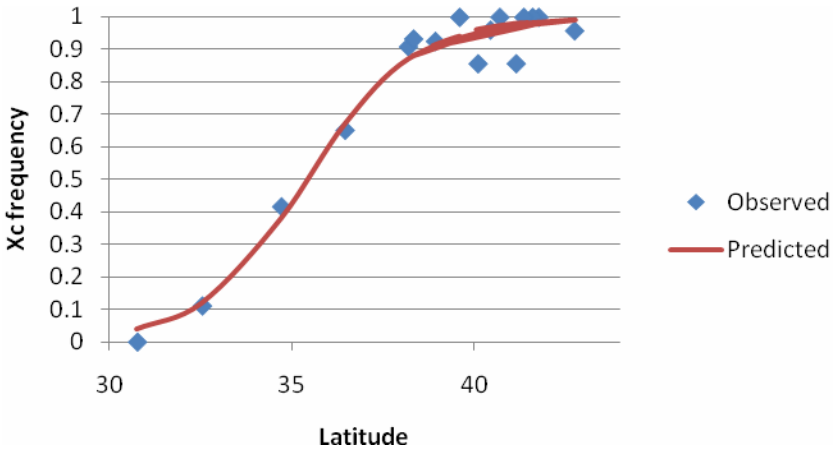


Figure 2.2 Frequency of inversion 4ab. Filled portions of each circle represent the proportion of chromosomes with the 4ab inversion at each location.

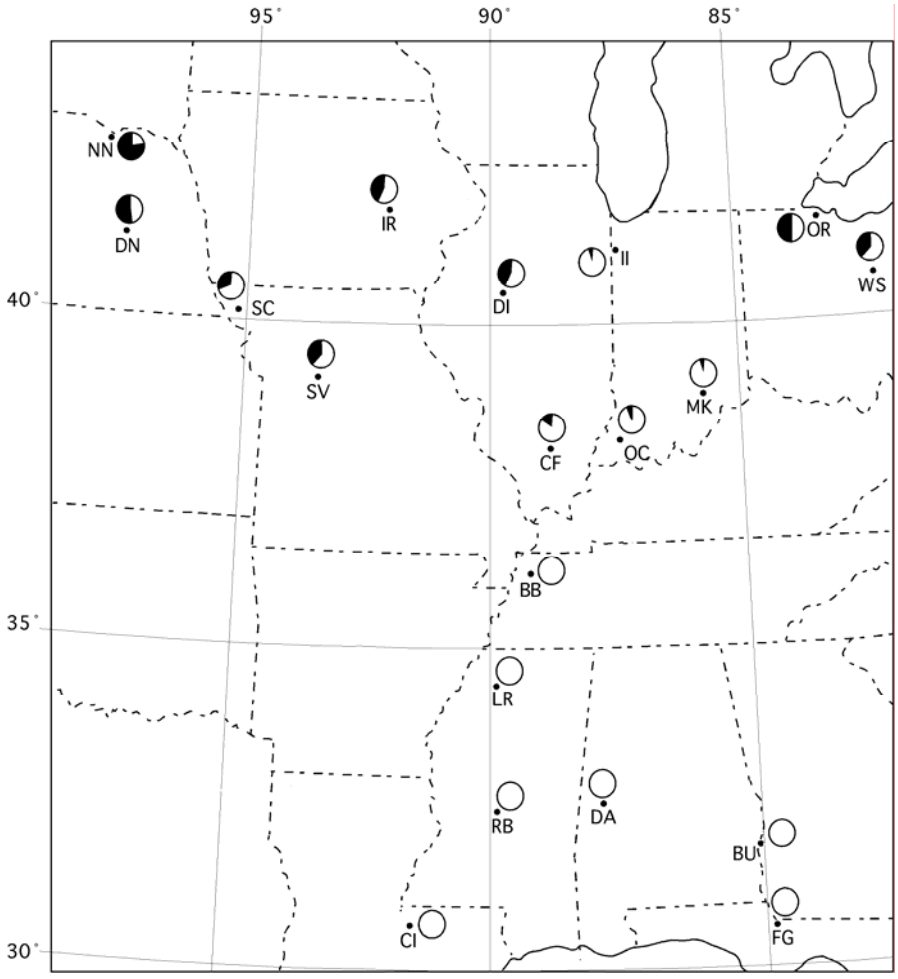


Figure 2.3 Frequency of inversion 5a and 5b The filled part of each circle represents the proportion of chromosomes with the 5b inversion, while the open part represents the proportion that has 5a.The 4 chromosomes with standard arragement from populations NN and CF were ommited from this map.

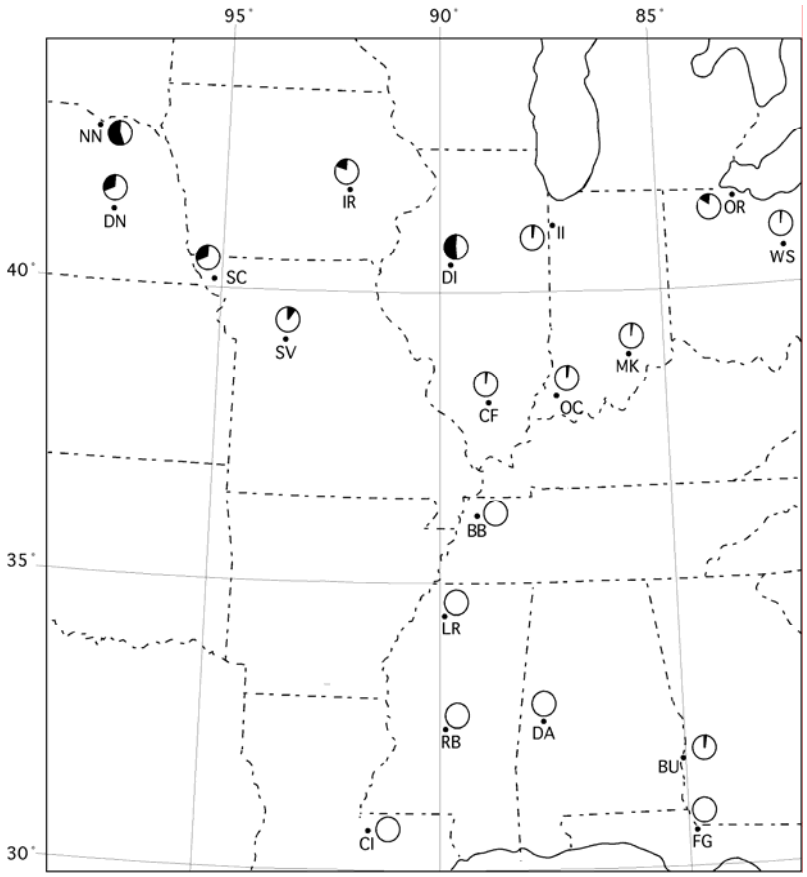
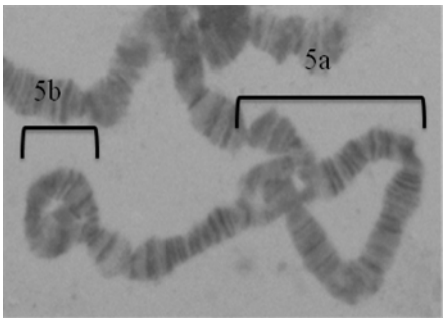


Figure 2.4 Polytene chromosome of chromosome 5 from a heterozygote 5a-/5-b. 5a is a large submedian inversion that corresponds to the large inversion loop, while 5b is a small subterminal inversion and corresponds to the small inversion loop at the tip of the chromosome.



CHAPTER 3
SEQUENCE VARIATION ASSOCIATED WITH ALTERNATIVE X-CHROMOSOME ARRANGEMENTS

3.1 Introduction

In the quest to understand adaptation considerable progress has been made identifying individual genes of large effect that are the subject of intense natural selection, such as genes contributing to insecticide resistance in mosquitoes and other insects (Scott 2000; Hemingway 2000; Daborn 2002). Resistant and sensitive alleles exhibit geographic and seasonal variation correlated with differences in insecticide use, and therefore variable selection pressures (Lenormand et al. 1999). The selective force in this case is intense and ultimately derived from human intervention; however genes responsible for direct responses to natural selective forces are also known (Nachman et al. 2003; Hoekstra et al. 2004). Cases of a single locus response to intense selection may be deceiving because most adaptive evolution is likely to result from multiple alleles of small effect (Orr and Coyne 1992). Quantitative trait loci (QTL) studies generally reveal many loci affecting putatively adaptive traits (Bradshaw et al. 1998; Hawthorne and Via 2001; Zhong et al. 2005; Morgan and Mackay 2006). The recombinational landscape influences adaptation when variation at two or more loci has correlated effects on fitness (Sinervo and Svensson 2002). If the loci are not closely linked, suppression of recombination between them is advantageous because it facilitates the maintenance of associations among alleles (Kimura 1956; Feldman et al. 1997; Pepper 2003). The resulting positive correlation among alleles is the essence of the elusive “coadapted gene complex”, defined as a group of alleles that have higher fitness when they occur together than when alone.

Chromosomal inversions may facilitate adaptation involving multiple loci because they reorganize linkage relationships and strongly reduce recombination between different gene arrangements (Roberts 1976). According to Dobzhansky’s model of coadaptation (Dobzhansky

1970) an inversion protects sets of coadapted alleles with high fitness and thus derives a selective advantage in the population. More recent theoretical work (Kirkpatrick and Barton 2006) has shown that an inversion can spread in a population even without epistatic interactions between alleles, as long as the inversion captures at least two coordinately but independently locally adapted alleles at separate loci. They also show that even if they do not have higher fitness when they occur together, the inversion can spread, making the conditions favoring an inversion in a population much less stringent than previously thought.

Molecular population genetic studies of inversions as coadapted gene complexes have focused on examining patterns of nucleotide variation within and between chromosomal arrangements (Andolfatto et al. 2002). Because inversions dramatically reduce recombination in heterozygotes, they can have drastic effects on patterns of DNA sequence variation. The two common features of sequence data obtained from loci within or near inversions are high differentiation between the chromosomal types (Andolfatto et al. 2001; Navarro et al. 1997; Hoffmann et al. 2004) and elevated linkage disequilibrium in relation to the alternative chromosomal arrangements (Andolfatto et al. 2001; Schaeffer et al. 2003). Some exchange can still occur between inverted and standard chromosomes in the form of rare multiple crossover events (Ashburner 1989) and gene conversion events (Chovnick 1973). Nonetheless, suppression of recombination by the inversion protects linkage disequilibrium between distantly separated alleles, especially near the inversion breakpoints (Navarro et al. 1997; Schaeffer and Anderson 2005). The establishment of an inversion polymorphism in a population also entails a partial sweep and as a result will remove some of the variation initially present in the population (Navarro et al. 2000). The time that has elapsed since the occurrence of an inversion will have a large impact on the patterns of polymorphism throughout the inverted region (Andolfatto et al. 2001). Nonetheless, if there is selection, islands of linkage disequilibrium will be maintained throughout the inverted region (Schaeffer et al. 2003).

Drosophila americana, a member of the virilis species group, provides an excellent system to investigate the relationship between adaptive genetic variation and genome arrangement because it segregates several chromosomal rearrangements (Chapter 2). Previous analyses have primarily focused on a Robertsonian fusion of the X and fourth chromosome (Muller element A and B, respectively). The X-4 centromeric fusion segregates in natural populations and is distributed along a latitudinal cline in North America (McAllister 2002; McAllister et al. 2008). While the frequency of the X-4 fusion in a population is strongly correlated with latitude, most genes examined to date show no differentiation among geographically distinct populations, indicating that gene flow is sufficient to homogenize neutral sequence variation throughout most of the genome (McAllister and McVean 2000; McAllister 2002; 2003; Schäfer et al. 2006; McAllister et al. 2008). These data are consistent with the hypothesis that the cline is maintained by natural selection acting on alleles associated with the different chromosomal arrangements. The X chromosome also segregates for a large submedian inversion (*In(X)c*). In chapter 2 I found that this inversion is in strong linkage disequilibrium with the X-4 fusion and shows the same clinal distribution.

Sequence variation on the X chromosome exhibits strong associations with the alternative centromeric arrangements. The strongest association identified thus far is at the *fused1* (*fu1*) locus (Vieira et al. 2001). Sequences are differentiated between chromosomal classes and exhibit strong linkage disequilibrium with the arrangement of the centromere. Also, nucleotide variation is lower for X-4 fused chromosomes at this locus, which is evidence that this arrangement has undergone a selective sweep. In addition, latitudinal clines for amino acid replacements in *fu1* were revealed among the X-4 fusion chromosomes. These patterns may arise from selection directly on *fu1* (Vieira et al. 2001). Interestingly, the sequence differentiation at *fu1* is not observed at several gene regions located closer to the centromere (Vieira et al. 2003, 2006). This discordance was recognized by Vieira et al. (2006) as possibly being due to inversion Xc. The

original analysis of inversions in *D. americana* by Hsu (1952) described lines containing the X-4 fusion and absent in lines with the unfused X (*D. americana americana* and *D. a. texana* respectively as described by Hsu (1952)). Blight (1955) determined chromosomal arrangement in large samples collected at several localities near St. Louis, Missouri (USA) and found a strong association between the X-4 fusion and *In(X)c* in this population where approximately 85% of the X chromosomes are fused to the 4th. In chapter 2 I found that this association is maintained throughout the species' range and therefore Xc shows the same clinal variation as the X-4 fusion.

Nucleotide variation at *fu1* resembles patterns of variation due to an inversion and even though polymorphic inversions can have dramatic effects on patterns of sequence variation, none of the previous studies have directly examined the effects that inversion Xc may have had in shaping patterns of variation on the X chromosome. Therefore, the objective of this study is to directly contrast patterns of sequence variation at the base of chromosome X and within the inversion using a set of cytologically characterized lines. I use patterns of variation along the chromosome to examine the hypothesis that these rearrangements have distinct evolutionary histories. More clearly understanding the history of events that have affected chromosome X will help focus studies identifying targets of natural selection. Although Vieira et al. (2006) have argued that the base of X and 4th chromosome are the regions that harbor the advantageous variants that maintain the cline for the X-4 fusion, I show that *In(X)c* is likely to have been the original complex upon which the X-4 fusion arose. Therefore, at least one target of selection is expected to be localized within or near the inversion breakpoints. Furthermore, I argue that the strong association observed between *In(X)c* and the centromeric fusion with 4th chromosome is a product of coadaptation, because rare exchange is evident within the complex. It is also noteworthy that there is a nested inversion also highly associated with the X-4 fusion on the 4th chromosome (Blight 1955; Evans et al. 2007; Chapter 2). The origin of the association among this inversion complex and the X-4 fusion appears to have been a target of positive selection

(Evans et al. 2007). The high association observed between the inversions on both chromosomes and the centromeric fusion indicates coadaptation among loci throughout the X and 4th chromosome and reveals that natural selection acts to maintain the arrangements in linkage disequilibrium.

3.2 Materials and Methods

3.2.1 Flies

I used a set of 24 highly inbred lines of *D. americana* (11 to 16 generations of sib mating) to perform a coordinated analysis of cytological and sequence variation. These lines were originally collected at six different geographic localities. The sites of collection and abbreviation for these populations are as follows: Monroe, Louisiana (ML97); Floodgate Park, Arkansas (FP99); Howell Island, Missouri (HI99); Gary, Indiana (G96); Niobrara, Nebraska (NN97) and Ottawa National Wildlife Refuge, Ohio (OR01).

3.2.2 X-4 and Inversion Status

I inferred the arrangement of the centromere as fused or unfused X and 4 for each of the inbred lines by determining the pattern of inheritance (sex-linked or autosomal transmission) of microsatellite markers (Schlotterer 2000; and McAllister 2003) or a visible eye mutation on chromosome 4 (McAllister 2002).

I typed all the inbred lines for all inversions by crossing a female of each inbred line to a standard laboratory line with a visible eye color mutation (ML97.5 Purple). This line is homozygous for the standard *D. americana* arrangement of the X. Standard *D. americana* corresponds to the Xab arrangement (henceforth *D. americana* standard) because these fixed inversions distinguish the X relative to the standard arrangement of *D. virilis*. I analyzed the polytene chromosomes for at least 5 F1 female larvae from each of these crosses. I used standard methods for obtaining orcein-stained salivary gland preparations (Kennison, 2000). The X

chromosome was identified by its banding pattern. If *In(X)c* is present, an inversion loop is observed on the X chromosome.

3.2.3 Recombination Estimates

Recombination estimates between the centromere and *In(X)c* were obtained from females heterozygous for both the Xc and the X-4 fusion arrangement of the X chromosome as described in McAllister and Evans (2006). Reciprocal crosses of the inbred lines ML97.5 and G96.23 were used. ML97.5 has an unfused X and 4 and the standard *americana* arrangement of the X, while G96.23 has the X-4 centromeric fusion and *In(X)c*. Three loci on the X chromosome (2 microsatellites and 1 RFLP) were used to determine the presence of recombinant genotypes. V68-06.2 (Orsini et al. 2004) is located in cytological band 5D of the *virilis* genome. This corresponds to the distal region of the chromosome outside the distal Xc breakpoint. Msat34 (Orsini et al 2004) is located inside the inversion within band 11B and *fu1* (Vieira et al. 2001) is located between the centromere and the proximal inversion breakpoint.

3.2.4 In situ Hybridizations

I performed *in situ* hybridizations with some of the loci used in the sequence analysis, as the location of these with respect to the chromosomal rearrangements is critical. I biotin labeled each probe by performing a PCR reaction using the PCR product as template and replacing the dTTP with 0.015mM unmodified dTTP and 0.01mM of biotin-16-dUTP. I precipitated and resuspended the product in hybridization buffer. I performed *in situ* hybridization on salivary gland polytene chromosome according to Gavrila et al. (2001).

3.2.5 DNA Extractions, PCR Amplification and DNA Sequencing

I used a single male from each inbred line for DNA extraction using a DNeasy Tissue Kit (Qiagen) following the manufacturer's instructions. I obtained sequence from the following gene regions on the X chromosome from the 24 inbred lines: *V1-33, fused 1, nonA, zeste*, and *period*.

These genes are located throughout the chromosome and provide a broad assessment of sequence variation of regions in and near the rearrangements, as well as loci removed from them. I also obtained sequences for a subset of the inbred lines (2 or 4 of each chromosomal arrangement) for several loci in the regions where the inversion breakpoints are thought to be located. I estimated the location of the breakpoints by using loci of known position to orient the appropriate scaffold of the *D. virilis* sequence. I sequenced loci at the breakpoints in a subset of the inbred lines to confirm the patterns observed at the loci sequenced in the full set of inbred lines. These gene regions are: *cac*, GJ19439, GJ18688, *norpA* and GJ17064. I directly sequenced PCR products with both reverse and forward primers using an ABI 3100 or 3730 (Applied Biosystems). I manually edited sequences using Sequencher (Genecodes).

3.2.6 Sequence Analyses

I did the analysis for all sequences of each gene in addition to classifying sequences into groups defined by chromosomal arrangement (fused vs. unfused X and 4 and std vs. *In(X)c*). I calculated standard measures of sequence variation including silent nucleotide diversity, π (Tajima 1983) and heterozygosity, θ (Watterson 1975) using DnaSP v4.10 (Rozas et al. 2003). I also obtained measures of departure from neutrality. The Hudson-Kreitman-Aguade (HKA) test (Hudson et al. 1997) compares divergence to polymorphism between a control locus and a test locus. I used *zeste* and *period* concatenated as the control locus while I used the *D. virilis* sequence to estimate divergence. I performed this test for all chromosomes pooled together as well as for each chromosomal arrangement. The D_T statistic (Tajima 1989) compares numbers of segregating sites to the frequency spectrum of polymorphic sites. Also, I estimated population differentiation, F_{st} (Hudson et al. 1992) and net sequence divergence, Da (Nei 1987) among groups for each gene using DNAsp (Rozas et al 2003) and ProSeq2 (Filatov, 2002).

3.3 Results

3.3.1 Association of the Centromere Status and *In(X)c*

Of the 24 inbred lines that I typed for their centromere and inversion status, 11 had an unfused X with the standard arrangements, 12 had a fused X-4 with *In(X)c* and only one line has the X-4 fusion and the standard arrangement of the X. This nearly perfect association between the X-4 fusion and Xc inversion is highly unlikely to be due to chance (Fisher's exact test: $p < 0.0001$). I also observed inversions of other chromosomes. Chromosome 4 has an inversion complex consisting of a small paracentric inversion within a large submedian inversion: *In(4)ab*. This nested inversion also has a very strong association with the X-4 fusion (Fisher's exact test: $p=0.02$). It was never found on free 4th chromosomes, however not all fused X-4 chromosomes have this inversion, consistent with results from chapter 2.

I also identified two different inversions in chromosome 5. One consists of a large submedian inversion: *In(5)a* and the other a very small subterminal inversion: *In(5)b*. There is the extreme negative linkage disequilibrium between them (Fisher's exact test, $p>0.0001$). The same chromosome never had both inversions, and chromosomes without either inversion were also not observed. One line was typed as being polymorphic: one chromosome carrying the 5a inversion, while the other chromosome had the 5b inversion. It is noteworthy that this line has maintained this polymorphism in spite of 16 generations of sibmating and being maintained in the laboratory for more than a decade.

3.3.2 In situ Hybridization

Location of markers varied with respect to the inversion examined (Figure 3.1 and 3.2). *v1-33* is located at the base of the euchromatic arm of the X, very close to the centromere at section 19C of the *virilis* cytological map. *Fu1* is located in section 18, outside the inversion between the centromere and the proximal breakpoint. Both *cac* and *nonA* are located within the

inverted region. *In situ* hybridization localizes *cac* at 15A of the standard *virilis* map, whereas *nonA* localizes to 15D. Two independent inversions (a and b) have rearranged those genes captured at opposite ends of the region inverted by Xc. Owing to their opposing distribution within Xc, *nonA* is located closer to the centromere of the standard arrangement, and towards the distal end of inverted chromosomes, and *cac* is located closer to the centromere of the inverted arrangement. *Period* and *zeste* are located in the distal region of the chromosome in sections 6 and 7 respectively. Estimating the position of each gene is not a trivial matter since there are two fixed inversions between *D. virilis* and *D. americana* with unknown breakpoints that complicate the orientation of each scaffold relative to the *D. americana* chromosome.

3.3.3 Recombination

Recombination estimates between the centromere and 3 loci distributed throughout the X chromosome were obtained for females heterozygous for the alternative arrangements of the X, which indicates the effects of these rearrangements on recombination rates in heterotypic females (Figure 3.2). No recombinants were recovered for the interval between the centromere and *fu1* (n= 179) and between *fu1* and msat34, located inside the inversion (n=333). Six recombinant chromosomes were recovered for the interval between msat34 and v68-06.2, located in the distal region of the chromosome outside the inversion (n=335). These results reveal a large region of the X, from the centromere and throughout the inversion, showing a very low level of recombination between the different chromosome forms.

3.3.4 Sequence Data Analysis

The analysis of several loci throughout the X chromosome allowed me to characterize the variation associated with the alternative chromosomal arrangements, providing a comprehensive overview of the influence of the chromosomal rearrangements on patterns of variation.

The general pattern revealed by the sequence analysis of 10 loci distributed along the X chromosome is that the chromosomal rearrangements have a major impact on the patterns of sequence variation (Table 3.1 and 3.2). Loci such as *zeste* and *period,* (Table 3.1) being located at the distal part of the chromosome away from the inversion, show levels of variation comparable to that observed for loci in other regions of the genome (Hilton and Hey 1996, 1997; McAllister and Charlesworth 1999; McAllister and McVean 2000) . Also they show no differentiation between chromosomal forms. On the other hand, loci inside the inversion, near the breakpoints and in close proximity to the centromere, show lower variation on the X-4 fused chromosome as well as population differentiation and divergence between chromosomal types. Each chromosomal region will be explored in more detail next:

Control region: *zeste* and *period* are located at the distal end of the chromosome, removed from the centromeric fusion and the Xc inversion. As expected, these loci do not show any evidence that sequence variation is influenced by these rearrangements (Table 3.1).

Centromeric region: V1-33 is located in the proximal region of the X near the centromere. This locus shows extremely low overall variation and complete lack of variation on the derived X-4 fused chromosomes. It is noteworthy that the divergence between *D. virilis* and *D. americana* at this locus is similar to what is observed at other loci in this study, therefore the low level of variation cannot be attributed to low mutation rate in the region. I performed the Hudson-Kreitman-Aguade (HKA) test for all chromosomes pooled together and it shows a significant value for v1-33, which indicates a significant deficit of polymorphism at v1-33 (Table 3.4). Also, v1-33 shows a significantly negative Tajima's D, indicating an excess of low frequency variants, normally attributed to a recent selective sweep or population expansion (Table 3.5). Fay and Wu's H statistic (Fay and Wu 2000) also takes into account the frequency spectrum of the segregating mutations, specifically the skew in the frequency of derived variants. A negative value indicates an excess of derived alleles. Only the v1-33 locus shows a significant

negative value (-1.48, p=0.041) when all chromosome types are pooled together. No other significant value is found in any of the other loci analyzed when chromosome are pooled together or separated by rearrangement. Fu and Li's D and F statistic (Fu and Li 1993) compare the observed number of singletons with a neutral model. A significantly negative value indicates an excess of rare polymorphisms. Again v1-33 is the only locus that shows a significantly negative value (-3.28, $p<0.02$ and -3.45, $p<0.02$ respectively) when all chromosomes are pooled together, indicating a significant excess of singletons. Other loci do not show departure from neutrality. As expected from the low variation, there is no significant differentiation or divergence between chromosomal forms (Table 3.3).

Region between the centromere and proximal inversion breakpoint (*Fu1)*: This gene is located in the region between the centromere and the proximal inversion breakpoint. This locus shows a different pattern compared to closer to the centromere. Although the level of overall variation is still slightly reduced in comparison with the control loci, variation at *fu1* is much higher than observed at v1-33. The most remarkable observation with respect to the levels of variation is the marked difference in the levels of variation between chromosomal forms. It appears that most of the variation found can be attributed to the ancestral unfused X, while the derived X-4 fused chromosomes show a much reduced level. This is also revealed by the HKA test (Table 3.4). It is not significant when it is performed using all chromosomes, however, arrangements are analyzed separately, there is a significant deficit of polymorphism for the X-4 fused chromosomes. Consistent with previous studies (Vieira et al. 2001), there is high differentiation and divergence and 6 fixed differences between chromosomal forms (Table3.3).

Inside inversion Xc: *nonA* is located within the limits of the inversion. This locus shows levels of variation similar to control loci. Also, there is no difference in the levels of variation between the chromosomal forms (Table 3.1). As expected, the HKA test and Tajima's D do not reveal any significant departure from neutrality (Table 3.4 and 3.5 respectively). There, is

however, significant differentiation and high divergence between chromosomal forms. (Table 3.3) There are no fixed differences between the arrangements and there is evidence of a gene conversion event between inverted and standard chromosome at this locus (sequence FP99.52).

Inversion breakpoints: I sequenced the loci used for these regions only sequenced in a subset of inbred lines (2 to 4 for each arrangement) (Table 3.2). GJ18688, GJ19439 and *cac* are located close to the proximal inversion breakpoint, while *norpA*, and GJ17064 are located near the distal breakpoint. Within the region near the proximal breakpoint all loci show differentiation and high divergence between chromosomal forms (Table 3.3). Each has fixed differences between chromosomal forms and very few shared polymorphisms. Although I sequenced few lines, variation is similar to levels observed at the *fu1* locus and loci tend to show lower variation on the derived arrangement.

In the region near the distal breakpoint the pattern is different. *NorpA* shows no variation among all chromosomes analyzed (Table 3.2). Consequently, there is no differentiation or divergence between chromosomal forms at this locus (Table 3.3). On the other hand, GJ17064 shows low levels of variation, similar to levels observed for the loci at the proximal inversion breakpoint. Also at this locus there is significant differentiation and high divergence between chromosomal forms (Table 3.3).

3.4 Discussion

The analysis of a set of 24 inbred lines characterized for their chromosomal rearrangements reveals a strong association between the X-4 centromeric fusion and inversion Xc. This is supported by the analysis of field collections (Chapter 2) which shows that recombinants between the two observed types do not exist in natural populations of *D. americana* (321 typed chromosomes: 57 unfused/stdX, 256 fused/Xc). There is an extremely high positive correlation between the X-4 fusion and *In(X)c* and therefore, it is crucial to consider the effects of both rearrangements on patterns of DNA variation on chromosome X.

Analysis of 10 loci distributed throughout the X chromosome reveal that rearrangements of this chromosome have a large impact on the nucleotide variation at nearby loci.

The gene region located nearest the centromere (v1-33) could be informative with respect to the history of events that have affected the centromere. The divergence between *D. virilis* and *D. americana* at v1-33 is comparable to all the loci analyzed in this study, which suggests a similar mutation rate. Therefore, the overall low variation cannot be attributed to a reduced mutation rate. However, it could be due to low levels of recombination in this region, because levels of variation and recombination are positively correlated. Regions of low recombination exhibit low variation due to selective sweeps and/or background selection (Charlesworth and Charlesworth, 1998). The results of the tests for departure from neutrality indicate this region has been affected by a selective sweep that has eliminated variation at this locus. Two possible hypotheses may account for the lack of divergence between chromosomal forms. It could be due to the fusion between the X and the 4 being a recent event, so the populations of chromosomes with different arrangements have not had sufficient time to diverge. On the other hand, this lack of divergence could be a result of recombination between the chromosomes with alternative arrangements, which has resulted in homogenization of variation. Even though this region appears to have low levels of recombination based on the low observed variation and the estimation of recombination in the lab, other loci at the base of the X chromosomes, such as *para* (Vieira et al. 2003), do show evidence of recombination between chromosomal forms. If this is the case, and there has been sufficient recombination to homogenize variation between the chromosomal forms, this marker is not reliable to estimate the relative age of the X-4 fusion. Only a locus that does not have recombination between the populations of chromosomes with alternative karyotypes can be used to reliably infer the age of the centromeric fusion.

The *nonA* gene region is positioned within inversion Xc , close to the inversion breakpoint. Consequently, analysis of the sequence variation in this gene can aid in

understanding the history of the inversion. The lack of population structure in *D. americana* (McAllister and McVean 2000; McAllister 2002, 2003; Schäfer et al. 2006) indicates that the significant differentiation between inverted and standard arrangements observed at this locus is likely due to decreased recombination in this region in inversion heterozygotes (Andolfatto et al. 2001). Also the net silent divergence between chromosomal forms is 0.0027. When this is contrasted with the lack of divergence observed at the centromere, this region's higher divergence indicates that the inversion occurred prior to the fusion. Furthermore, that the levels of overall variation and the levels for each chromosomal arrangement are comparable to the control loci, in conjunction with evidence of a gene conversion event between a chromosome with the inverted arrangement and a standard chromosome, indicate that the inversion has been segregating in the population for an extended period of time.

The *fu1* locus is near the proximal inversion breakpoint of Xc. However, this gene is located outside the inversion, between the centromere and the proximal breakpoint. The low variation observed at this locus, especially in the fused chromosomes, in conjunction with a significant HKA test revealing a deficit of polymorphisms in the fused chromosomes are indicative of a selective sweep that has only affected the fused/inverted chromosomes and has resulted in reduced variation in this population of chromosomes. Also, the significant differentiation between chromosomal forms is likely due to reduced recombination between chromosomal forms and/or to selection acting to maintain an association between *fu1* haplotypes and the chromosomal rearrangements. Moreover, there is high silent net divergence between chromosomal forms (D_a=0.046), 6 fixed differences and only 1 shared polymorphism between chromosomal types. Therefore, the selective sweep also resulted in high differentiation and divergence between the alternative chromosomal forms.

The three loci in the region where the proximal inversion breakpoint is predicted show a distinct pattern of differentiation and fixed differences between the chromosomal types even

though I only surveyed them in a subset of the inbred lines. The most extreme pattern of reduced overall variation is observed at *cac*. It has only 6 segregating sites. 4 of these are fixed differences between the chromosomal types. GJ19439 is located 600kb from *cac* (Figure 3.2) and as expected shows a similar but less pronounced pattern. GJ18688 is located 800kb from *cac* (Figure 3.2) and at this locus variation is higher and there is no difference in variation levels between chromosomal types. The expected pattern is that loci closest to the breakpoint will show the most extreme differentiation and most fixed differences because recombination is highly restricted in this area, and therefore homogenization is not possible. Also, the inverted chromosomes will show less variation as a product of the selective sweep that brought the derived arrangement to its current frequency. As the distance from the breakpoint increases, recombination also increases. This results in homogenization of the variation and therefore differentiation and divergence between chromosomal forms will decrease with distance. Furthermore, variation levels are expected to be similar in the two chromosomal types.

NorpA and GJ17064 are the loci closest to the region where the distal inversion breakpoint is predicted to be located. GJ17064 shows a pattern of differentiation and fixed differences between the chromosomal types similar to those observed at the proximal inversion breakpoint and within the inversion. However, levels of variation are similar to those observed for control loci and there is no difference in variation for the different chromosomal types, similar to the level observed at *nonA*. Surprisingly, *norpA* is located 550kb from GJ17064 and shows a very different pattern. *NorpA* shows no variation at all, therefore there are no observable differences between the chromosomal types. The pattern observed at *norpA* is inconsistent with patterns of variation observed at other loci at the breakpoints and within the inversion. Low variation can be attributed to low levels of recombination; however, there is no reason to believe that recombination would be restricted in this area. The complete lack of variation at this locus

suggests that any effects the inversion could have had on its variation has been erased by more recent events, such as a selective sweep affecting all chromosomal types.

The analysis of these 10 gene regions in this set of cytologically characterized inbred lines reveal a high association between the centromere and haplotypes at *fu1* and *In(X)c*. This is remarkable considering that recombination could easily break down these associations, and therefore provides evidence that natural selection is maintaining the associations between DNA variation on the X and its chromosomal rearrangements. Further evidence that natural selection is preserving linkage disequilibrium is that the association pattern is not maintained throughout the chromosome. *Para* is located at the base of the X between *fu1*and the centromere. Variation at this locus resembles the pattern observed at *nonA*. Significant differentiation between chromosomal forms but similar levels of variation between them (Vieira et al.2003), indicate that recombination has homogenized variation. *Elav* is located within the inversion and does not show any population differentiation and levels of variation comparable to control loci. Also variation is similar between the different chromosomal types (Vieira et al.2003, Maside and Charlesworth 2007). These combined data indicate that recombination is sufficient to homogenize variation and disrupt linkage disequilibrium, but selection acts to maintain the favorable associations.

Using sequence variation studies in combination with cytological data it is possible to infer a model for the sequence of events that gave rise to the patterns observed. Inversion Xc is fixed in the closely related species *D. novamexicana* (Hsu, 1952). Consequently, one can infer that this inversion was present before these two species diverged approximately 380,000 years ago (Caletka and McAllister 2004). The data from *nonA*, the locus within the inversion, also argues that this inversion has been segregating in the population for a prolonged period of time. In addition, the low divergence between chromosomal forms at the region proximal to the centromere suggests that the centromeric fusion of the X and the 4 occurred more recently than

the inversion event. Under this scenario the inversion occurred on an unfused X chromosome, and later the fusion occurred on one of these inverted X chromosomes. For this to be plausible the inverted chromosomes must have been present in the population at intermediate frequency at that time. The rise in frequency of this inversion could have occurred by drift or natural selection. It is conceivable that the inversion rose in frequency because it contained coadapted or locally adapted genes. Under this scenario *In(X)c* could have been distributed in a clinal fashion similar to its current distribution (same as the distribution of the X-4 fusion (Chapter 2)). Hypothetically it was advantageous to maintain an association between the coadapted or locally adapted alleles found in the inversion with alleles around the centromere of the X and/or 4th chromosome. Chromosome 4 also segregates for inversions consisting of a small paracentric inversion within a large submedian inversion. This inversion complex is found exclusively on 4th chromosomes that are fused to the X. The fusion allowed maintenance of these two chromosomes, and in some instances their inversions, in linkage disequilibrium; therefore it subsequently rose in frequency culminating in its current geographic distribution. Hence, under this model, this selective sweep is responsible for the high association between the centromeric fusion, *fu1* haplotypes and the Xc inversion observed in this study. Through DNA analysis of loci on the 4th chromosome, Evans et al. (2007) found evidence of natural selection acting on the nested inversion *In(4)ab,* and therefore this inversion can also play a role in coadaptation with the X chromosome..

Given the strong association between the centromere and the inversion I predict that the genes harboring the advantageous variants will not only be at the base of the X and the 4th chromosome, but also inside the inversion, particularly in the region proximal to the centromere in the inverted arrangement. This is the only way to explain the high association between the centromere and the inversion, given that even extremely low levels of recombination would be sufficient to break up this association. The very marked patterns of variation observed at the loci

closest to the region that contains the proximal inversion breakpoint corroborate the hypothesis that at least one of the targets of selection will be located in this region. Currently *In(X)c* is a barrier to free recombination between unfused and fused chromosomes. However, there must have been a period of time when unfused Xc and fused Xc chromosomes were segregating in the population simultaneously. This creates the possibility of free exchange between fused and unfused chromosomes in the gene regions contained in this inversion. Therefore, neutral loci located in the inversion had the chance to start recovering variation and homogenize with the standard chromosomes. This explains the observed pattern at *nonA*, located in the inversion. This gene is moved away from the centromere in chromosomes with the inverted arrangement and both chromosomal types have similar levels of variability. The same is true for *elav* (Vieira et al. 2003) located within the inversion removed from the breakpoints. On the other hand the *fu1* locus is an example of this non-neutral pattern. This gene shows evidence of a selective sweep on the fused/inverted chromosomes that has resulted in high differentiation between the chromosomal forms and reduced variation on the fused/inverted chromosomes.

However, a model in which the effects of the selective sweep that took the X-4 fusion to its current frequency have been erased through recombination implies that the sequence variation at the base of the X chromosome may also be influenced by additional forces such as other selective sweeps. Under these circumstances loci near the base of the X cannot be reliably used to estimate the history of the centromere. Consequently, it cannot be concluded from any of the available sequence data that the X-4 fusion has a recent origin.

Whether or not the centromeric fusion is recent, there is still an extremely high association between the centromere, haplotypes at *fu1* and the inversion which implies the influence of natural selection. We hypothesize that the inversion contains coadapted or locally adapted alleles and it is highly beneficial to maintain these alleles in association with alleles near the centromere either on the X or 4^{th} chromosome. Therefore we would expect to find gene

regions that have been or are the target of selection within the inversion, most likely near the breakpoints, in particular within the region that is placed closer to the centromere in the inverted karyotype. These genes would probably show a pattern similar to that observed at *fu1,* with highly significant differentiation between chromosomal forms and lower diversity in the inverted chromosomes. The fact that the *Xc* inversion, the centromeric fusion and the 4ab inversion complex are in high linkage disequilibrium despite the fact that very low levels of recombination can break this association down, is evidence that these inversions are kept together because natural selection favors this combination. These inversions may carry independently locally adapted loci, or certain loci in these inversions may have epistatic interactions and therefore will be maintained in association. Evans et al. (2007) hypothesize the selective sweeps that have affected *In(4)ab* could be related to the fact that the 4th chromosome experiences sex linked transmission as a result of the centromeric fusion. Given that the X-4 fusion is positively correlated with latitude, the 4th chromosome is sex linked in higher frequency in the northern part of the species range (where the fusion is nearly fixed) and therefore these populations would be more susceptible to sex-specific selective pressures. Consequently, the nested chromosomal inversion complex *In(4)ab* suppresses recombination between the 4th chromosome fused to the X (neo-X chromosome) and the 4th chromosome that remains free (neo-Y) allowing the opportunity for accumulation of sex specific alleles on the respective chromosomes (Evans et al. 2007).

These two different inversions on the X and the 4 can be considered as locally adapted variants, with the associations between them shaped by coordinated selection. Given suppressed recombination within and among rearrangements, they are also subject to coadaptation through fitness epistasis. Both of the clines for these inversions are positively associated with latitude and were formed prior to the X-4 fusion. The X-4 fusion acts as a suppressor of recombination between the inversions, consistent with coordinated evolution. However, the specific X linkage of the *4ab* inversion complex indicates coadaptation through intralocus sexual conflict.

Table 3.1 Summary statistics for the loci sequenced in all inbred lines divided by chromosomal arrangement.

		n	sites	(s)	H	(Hd)	π/site	θ/site	
v1-33									Centromeric region
	Fused	13	1688	0	1	0	0	0	
	Unfused	11	1688	9	6	0.71	0.0009	0.0017	
	Total	24	1688	9	6	0.37	0.0004	0.0034	
fu1									Between cen. and inversion
	Fused	13	729.35	7	6	0.76	0.0008	0.0010	
	Unfused	11	729.35	53	11	1	0.0056	0.0073	
	Total	24	729.35	65	17	0.92	0.0055	0.0070	
nonA									Within inversion
	Fused	13	549	35	13	0.99	0.0161	0.0166	
	Unfused	11	549	37	11	1	0.0154	0.0192	
	Total	24	549	52	24	0.99	0.0171	0.0205	
zeste									Distal end of the chromosome
	Fused	13	417.32	67	13	1	0.0095	0.0139	
	Unfused	11	417.32	76	11	1	0.0132	0.0165	
	Total	24	417.32	94	24	1	0.0109	0.0160	
period									
	Fused	13	595.5	92	13	1	0.0199	0.0213	
	Unfused	11	595.5	85	11	1	0.02	0.0223	
	Total	24	595.5	117	24	1	0.0198	0.0227	

.

Note: n; number of sequences, sites: number of silent sites (noncoding and synonymous) in each locus, s: number of segregating sites, H: number of haplotypes, Hd: haplotype diversity, π/site: nucleotide diversity per site, θ/site: heterozygosity per site.

Table 3.2 Summary statistics for the loci sequenced in a subset of inbred lines at the inversion breakpoints divided by chromosomal arrangement.

		n	sites	(s)	H	(Hd)	π/site	θ/site	
cac									
	Fused	2	657	0	1	0	0.003	0.003	Proximal inversion breakpoint region
	Unfused	2	657	2	2	1	0	0	
	Total	4	657	6	3	0.83	0.005	0.0091	
GJ19439									
	Fused	2	948	4	2	1	0.0042	0.0042	
	Unfused	2	948	5	2	1	0.0052	0.0052	
	Total	4	948	16	4	1	0.0096	0.0168	
GJ18688									
	Fused	4	619	6	3	0.83	0.0051	0.0064	
	Unfused	4	619	8	4	1	0.0064	0.0064	
	Total	8	619	17	7	0.94	0.0117	0.0129	
norpA									Distal inversion breakpoint region
	Fused	3	355	0	1	0	0	0	
	Unfused	3	355	0	1	0	0	0	
	Total	6	355	0	1	0	0	0	
GJ17064									
	Fused	5	334	6	5	1	0.007	0.0179	
	Unfused	4	334	4	3	0.83	0.005	0.0011	
	Total	9	334	9	8	0.97	0.006	0.0269	

Note: n; number of sequences, sites: number of silent sites (noncoding and synonymous) in each locus, s: number of segregating sites, H: number of haplotypes, Hd: haplotype diversity, π/site: nucleotide diversity per site, θ/site: heterozygosity per site

Table 3.3 Population differentiation (Fst) silent net divergence (Da).

Locus	F_{st} Fused/unfused	N° polymorphism		N° of fixed differences	Da between chromosomal forms
		Shared	Exclusive		
V1-33	0.016	0	9	0	0
Fu1	0.466***	1	58	6	0.0046
Cac	0.8	0	2	4	0.0067
GJ19439	0.608	0	9	7	0.0068
GJ18688	0.641*	1	12	4	0.0118
nonA	0.125**	22	31	0	0.0027
norpA	-	0	0	0	0
GJ17064	0.522**	1	10	3	0.0048
Zeste	-0.022	51	46	0	-0.0012
Period	-0.017	62	62	0	-0.0013

Note: *, 0.01<p<0.05; **, 0.001<p<0.01; ***, p<0.001

Table 3.4 HKA test.

		χ^2	p-value
v1-33	All	8.54	0.0035**
	Fused	10.21	0.0014**
	Unfused	5.01	0.0251*
fu1	All	0.01	0.9064
	Fused	0.64	0.0109*
	Unfused	0.002	0.9671
nonA	All	0.79	0.3716
	Fused	0.73	0.3924
	Unfused	0.54	0.4615

Note: *, 0.01<p<0.05; **, 0.001<p<0.01. *Zeste* and *period* were concatenated and used as control. *D. virilis* were used to calculate divergence.

Table 3.5 Tajima's D for each locus.

	All	Unfused	Fused
v1-33	-2.20**	-1.88*	-
fu1	-0.83	-1.06	0.07
nonA	-0.73	-0.96	-0.19
zeste	-1.34	-1.04	-1.34
period	-0.70	-0.37	-0.47

Note: *, $0.01<p<0.05$; **, $0.001<p<0.01$

Figure 3.1 *In situ* hybridization on polytene chromosome X of heterozygote std/*In(X)c*.

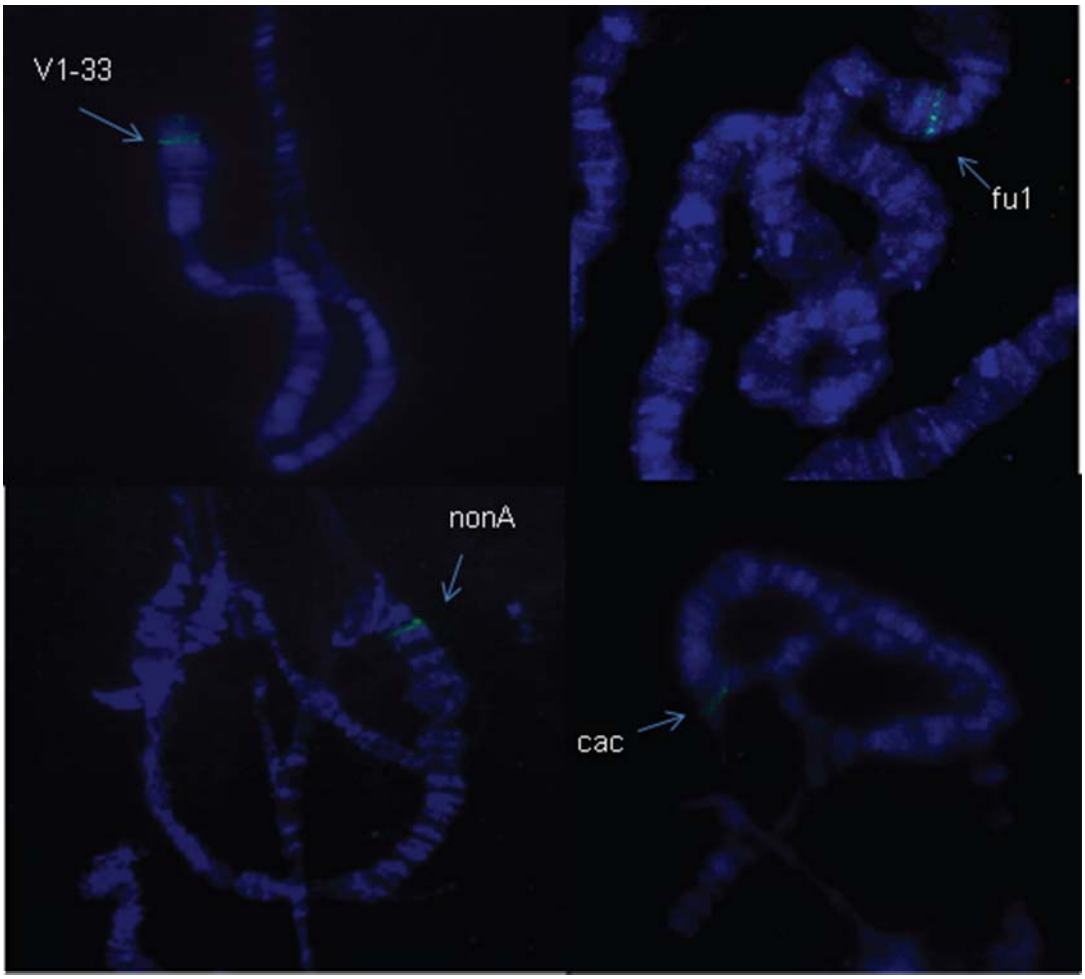
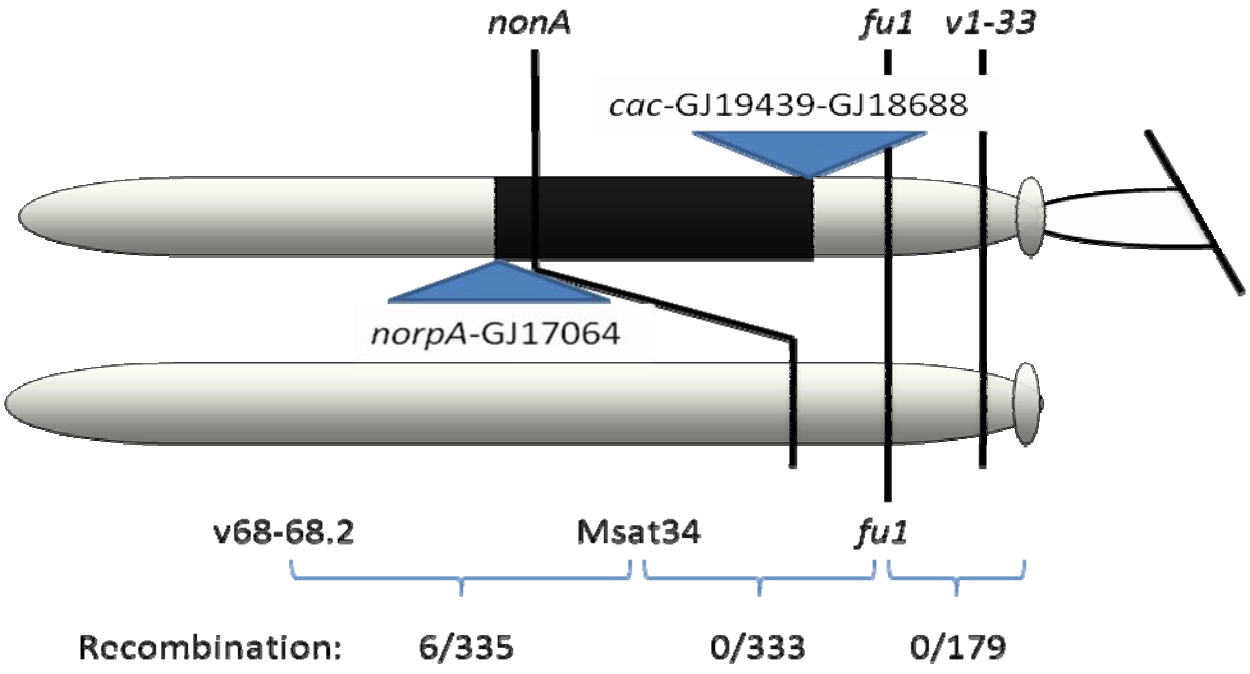


Figure 3.2 Representation of the location of all genes used for the sequence analysis. Also shown are the loci used to estimate recombination. The number of recombinants obtained for each interval are shown on the bottom.



CHAPTER 4

GEOGRAPHIC DISTRIBUTION OF DNA VARIATION ASSOCIATED WITH CHROMOSOMAL REARRANGEMENTS ON CHROMOSOMES X AND 4

4.1 Introduction

Chromosomal rearrangements such as inversions restrict recombination between chromosomal arrangements, resulting in a positive association between alleles even if they are not tightly linked. For this reason it has been hypothesized that inversions could play a role in facilitating adaptation involving two or more loci (Dobzhansky 1970; Kirkpatrick and Barton 2006). Although recombination is restricted between chromosomal arrangements, exchange still occurs in the form of multiple crossovers and gene conversion (Hasson and Eanes 1996; Andolfatto et al. 2001). Despite the potential for genetic exchange between the chromosomal arrangements, genetic markers located in or around rearrangements often show significant associations with inversions and with each other (Prakash and Lewontin 1968, 1971; Weeks et al. 2002; Kennington et al. 2006). Such associations indicate that natural selection favors beneficial allelic combinations (Schaeffer et al. 2003). Further evidence for the role of selection is apparent when these associations are maintained at loci, not within the inversion, but in other regions of the chromosome. In addition, correlated selective pressures could generate associations among separate chromosomal rearrangements and potentially extend the associations through a larger portion of the genome.

In *D. americana* it has been established that the X-4 fusion (McAllister 2002; McAllister et al. 2008) and inversions Xc and 4ab (Chapter 2) show clinal variation. This, in combination with the lack of population structure in this species, is a strong indication that natural selection is generating this distribution gradient.

Furthermore, DNA sequence variation studies on chromosomes X and 4 have found that the chromosomal rearrangements strongly influence patterns of nucleotide variation. In chapter 3

I reported strong differentiation between chromosomal types and linkage disequilibrium with the rearrangements at loci within the inversion and proximal to the inversion breakpoints. However, *fu1* also showed this association pattern and it is located between the inversion proximal breakpoint and the centromere. Moreover, there is extreme linkage disequilibrium between the Xc inversion and the centromere. A similar pattern has been described for chromosome 4 (McAllister 2003; Evans et al. 2007). Inversion 4ab is strongly associated with the X-4 fusion and the *bib* locus, located between the proximal inversion breakpoint and the centromere. Variation at this gene is in strong linkage disequilibrium with the inversion. This pattern is particularly striking considering that other loci in the region between the proximal inversion breakpoint and the centromere do not show the same association (McAllister 2003; Evans et al.2007).

The objective of this study is to establish whether DNA sequence variation on the X and 4^{th} chromosome shows association with the rearrangements throughout the species range. The DNA variation studies currently available were done using isofemale or inbred lines from a limited number of geographic sites (Vieira et al. 2001; McAllister 2002; McAllister 2003; Vieira et al. 2003; Vieira et al. 2006; Evans et al. 2007). This study expands upon those results with flies collected directly from the wild over a broad geographic area, covering a large part of the species' known range. Also, a unique feature of this study is that I assessed variation on both chromosomes X and 4 for each individual. It can be inferred that sequence variation will show association between chromosomes X and 4 from the high associations observed in previous studies of individual chromosomes (McAllister 2003; Chapter 2). However, no study to date has addressed this point directly. If these associations are maintained along the geographic gradient, it is an indication that the DNA variation on these chromosomes is also clinally distributed, providing further evidence that coordinated selection is acting on *D. americana* at multiple loci throughout the genome.

4.2 Materials and Methods

4.2.1 Collection Localities

A detailed list of the collection sites including latitude and longitude can be found in Table 4.2 As indicated previously, these collection sites cover a large portion of the known range of this species and include samples throughout the cline for the X-4 fusion. The collection methodology is found in the materials and methods section of chapter 2.

4.2.2 Flies

The wild caught males I used for the study of the geographic distribution of chromosomal rearrangements (chapter 2) were also used for this study in order to determine the geographic distribution of the genetic variation associated with the chromosomal rearrangements of the X and 4. Therefore I had previously determined the status of the centromere, as either fused or unfused, and the presence or absence of inversion Xc and 4ab. In order to genotype chromosomes directly from the wild, I crossed each male to the *D. virilis* line V46 and I froze the F1 hybrids for RFLP analysis. I extracted DNA in 96 well plates using the Qiagen DNA extraction kit following the manufacturer's instructions.

4.2.3 RFLP Analysis

I used the following genes for this analysis: Chromosome X: *para*, *fu1*, *cac* and *nonA*, Chromosome 4: *chico*, *nmd*, *geep*, *bib* and *tim*. (Figure 4.1). I determined the location of these through *in situ* hybridization or from the *D. virilis* genome sequence. Table 4.1 lists the primers and annealing temperatures for the amplification and the restriction enzyme used to digest each product. Each gene region was digested with restriction enzymes that recognize polymorphic sites identified from sequence data. I electrophoresed digested fragments on a 2% agarose gel. I screened a total of 15 polymorphic restriction sites, however only 11 were polymorphic in this sample.

4.2.4 Statistical Analysis

For each polymorphic restriction site I determined whether there was an association with latitude by calculating Spearman's correlation coefficient between the frequencies of the variants and latitude for each site. A significant correlation with latitude indicates the variants show clinal variation. Also, I estimated linkage disequilibrium between each restriction site and the inversion measured as D' and Fisher's exact test. I also estimated linkage disequilibrium among all loci and chromosomal rearrangements to determine whether there is long distance linkage disequilibrium among loci throughout both chromosomes. I performed pairwise comparisons using Fisher's exact test with a Bonferroni correction for multiple tests.

4.3 Results

I extracted DNA from 288 F1 hybrid females. The karyotype (fused or unfused X and 4 and presence of absence of *In(X)c* and *In(4)ab*) for each chromosome was identified via the analysis presented in chapter 2. Therefore, for each chromosome I have karyotypic and RFLP data (see appendix) which allows me to determine the patterns of linkage disequilibrium among these loci and the chromosomal rearrangements.

4.3.1 Genetic Variation on the X Chromosome

Digestion of the *para* gene product with the TfiI restriction enzyme resulted in a single polymorphic cut site identified as P159. No correlation between the frequency of the different variants and latitude or longitude at this locus (Table 4.2). Also, at this locus there is no significant association between its variants and the *In(X)c* and the X-4 fusion (Fisher's exact test $p=0.09$) (Table 4.3).

Digestion of the *cac* gene product with the XhoI restriction enzyme resulted in a single polymorphic cut site (CA188). There is a significant association between the variants at this site and the *In(X)c* and the X-4 fusion (Fisher's exact test $p<0.0001$) (Table 4.3). Furthermore, there

is a significant correlation between the two alterative variants at this restriction site and latitude ($p<0.0001$), therefore showing clinal variation (Table 4.2).

Digestion of the *fu1* gene product with the ClaI restriction enzyme results in a single polymorphic cute site (F1633) (Vieira et al 2001). There is a significant association between variants at this site and the chromosomal rearrangements of the X (Fisher's exact test $p<0.0001$) (Table 4.3). Also, variants at this site show a significant correlation with latitude ($p=0.0017$), therefore showing clinal variation (Table 4.2).

Lastly, the *nonA* gene product digested with the BstNI restriction enzyme results in a single polymorphic cut site (N300). The variants at this site also show significant association with the X-4 fusion and the Xc inversion (Fisher's exact test $p<0.0001$) (Table 4.3). Moreover, the variants at this site are correlated with latitude ($p=0.0004$), indicating that these also show clinal variation (Table 4.2).

There is widespread linkage disequilibrium among loci on chromosome X. (Figure 4.2). The only locus that does not show association with the rearrangements or the other loci is *para*.

4.3.2 Genetic Variation on the 4th Chromosome

Of the loci assessed on chromosome 4, I discarded *nmd* and *geep* because they had little or no variation at the sites assayed. The *bib* gene product digested with the BbrI restriction enzyme resulted in two polymorphic restriction sites in this sample: site B501 and B206 (McAllister 2003). Variation at site B501 was not in association with inversion 4ab, as indicated by a Fisher's exact test ($p=0.68$) (Table 4.4). On the other hand this site does show correlation with latitude ($p=0.0058$) (Table 4.2), indicating clinal variation. Variation at site B206 is highly associated with the inversion 4ab, as indicated by a Fisher's exact test ($P<0.0001$) (Table 4.4). Furthermore, variation at this site also shows correlation with latitude ($p<0.0001$) (Table 4.2).

I digested the *chico* gene product with 6 different restriction enzymes. MseI and RsaI did not reveal polymorphic sites in this sample. The other 4 enzymes cut at one polymorphic site

each. The variation at the sites assessed by Fnu4HI (site C2168) and DdeI (site C2250) are associated with In(4)ab as revealed by Fisher's exact test (p=0.0001 and p=0.001 respectively) (Table 4.4). However, they do not show a significant correlation with latitude or longitude (Table 4.2). I also found that variation at the sites assessed by the restriction enzymes Nlalll (site C2105) and Tsp45I (site C2486) were associated with *In(4)ab* (Table 4.4). Furthermore, they both also show a significant correlation with latitude ($p<0.001$ and $p=0.01$ respectively) but no correlation with longitude (Table 4.2).

There is extensive linkage disequilibrium between chromosomal rearrangements and RFLP variants in chromosome 4 (Figure 4.2). However, variation at two loci does not show association with the rearrangements or other loci (*tim* and *bib*501).

The *tim* gene product was digested with two restriction enzymes, however PstI failed to reveal any polymorphism in this sample. PvuII on the other hand cut at one polymorphic site (T281) (McAllister 2002). Variation at this site does not show significant association with the inversion 4ab, nor does it show correlation with latitude.

4.3.3 Patterns of Linkage Disequilibrium Along Chromosome X and 4

In addition to the already established linkage disequilibrium among chromosomal rearrangements (Chapter 2), the analysis reveals that there is long distance linkage disequilibrium among loci and rearrangements (Figure 4.2). Loci on chromosome X such as *nonA* and *fu1* show significant association with *In(4)ab* and loci on the chromosome 4 (variants at *bib* and *chico*) after a Bonferroni correction.

On chromosome 4, in addition to *bib*206 and *chico* sites 2105 and 2486 being in linkage disequilibrium with the inversion 4ab, the two *chico* sites are also in linkage disequilibrium with each other. No other significant linkage disequilibrium was revealed in this sample.

4.4 Discussion

Chromosomal rearrangements may be responsible for the formation and maintenance of coadapted gene complexes (Dobzhansky 1949). Selection maintains allelic association within such complexes because they will act harmoniously within local populations, but not when mixed with allelic combinations from other populations (Futuyama 1986). Therefore, the reduced recombination caused by the rearranged chromosome in heterozygotes is critical to the maintenance of coadapted complexes.

Sequence variation studies in *D. americana* have revealed the features of coadapted gene complexes on both chromosome 4 (McAllister 2003; Evans et al. 2007) and chromosome X (chapter 3). Allelic variation shows high association with the rearrangements that affect these chromosomes and also with each other, even among loci not necessarily in close proximity. However, these studies were carried out using isofemale or inbred lines collected from a few localities from a limited part of the species range. Consequently, this study examined wild caught flies, which assesses variation present in natural populations. Also, I include a large sample with flies collected from a broad geographic area that encompasses a large part of the species' range. This allowed me to establish whether or not the associations among chromosomal rearrangements and allelic variation are maintained throughout the cline, which would indicate that allelic variation shows clines similar to those already documented for chromosomal rearrangements.

Several variants on the X show association with chromosomal rearrangements and other loci. This is consistent with studies of sequence variation on the X (Vieira et al. 2001; Vieira et al. 2003; Vieira et al. 2006; Chapter 3) where the same loci assayed here showed significant population differentiation between chromosomal arrangements. The *para* locus however did not show association with any of the rearrangements or other loci indicating that the linkage disequilibrium observed on chromosome X does not extend throughout but is interrupted by

regions that are not in association with other loci. This is strong evidence that selection is acting to maintain these allelic associations because the lack of association with a locus in an area between loci that are in linkage disequilibrium implies that recombination is sufficient to homogenize variation. However, homogenization will not occur if selection favors associations among certain alleles, which would create islands of differentiation between chromosomal classes (Schaeffer et al. 2003).

A similar pattern is observed on chromosome 4, where variants at several loci are in high association with the chromosomal rearrangements and other loci. The linkage disequilibrium between site 205 at *bib* and *In(4)ab* was previously established by McAllister (2003). Here we find that this association extends to variants at *chico,* which are also in linkage disequilibrium with *In(4)ab*. However, variants at *bib* (site 501) and *tim* do not show association with any other loci or rearrangement. This is an indication that the allelic associations observed are preserved in spite of ample opportunity for homogenization of variation through recombination. This is evidence that the linkage patterns cannot be attributed solely to remnants of complete linkage disequilibrium within the inversion at the time of its formation (Aquadro et al. 1991).

Furthermore, associations are not merely within each chromosome. There is significant linkage disequilibrium among variation on chromosome X and 4. This is consistent with the strong association among chromosomal rearrangements throughout both chromosomes found in chapter 2, however, no study had previously assayed variation on both chromosomes simultaneously to determine whether there was also association among alleles. This long distance linkage disequilibrium interrupted by regions that do not show this pattern is further evidence that natural selection is maintaining these associations.

In summary, this study shows that the associations between allelic variation and chromosomal rearrangements previously found using isofemale and inbred lines from limited geographic sampling is conserved in this large, widely distributed sample of wild caught flies. I

found that variation at certain alleles exhibit similar clinal variation as the one described for the chromosomal rearrangements, but most importantly this study provides strong evidence that *D. americana* is under coordinated selection that has resulted in islands of linkage throughout both chromosomes which are preserved in the face of recombination.

Table 4.1 Primers and restriction enzymes used for each gene regions used for the analysis of allelic variation.

Chromo	Locus	Primers	AT °C	Restriction Enzymes
X	*para*	paraF1/paraR2	60	TfiI
	fu1	FU2F/FU41R	58	ClaI
	cac	cac4F/cac4R	56	XhoI
	nonA	nonAF/nonAR	60	BstNI
4	*chico*	chicoF3/chicoR	60	MseI, RsaI, Fnu4HI, DdeI, Nlalll, Tsp45I
	nomad	nomid-amF/nomid-R	62	RsaI
	geep	geepF/geepamR	60	BstUI
	bib	bibamF2/bibgenR2	58	BbrPI
	tim	tim genF/tim amR	60	PstI, PvuII

Table 4.2 Allele frequencies of the digested allele for restriction sites in *para, cac, fu1, nonA, chico, bib and tim.*

Site	Lat	Lon	N	P159	CA188	F1633	N300	B501	B206	C2168	C2250	C2105	C2486	T281
FG	30.7	30.7	9	0.1	0.2	0.18	0.36	0.22	0	0.65	0.5	0	-	0.67
CI	30.7	30.7	7	0.29	0	0	0	0.29	0	1	0.33	0	-	0.5
BU	31.9	31.9	16	0	0	0	0.3	0.19	0	1	0.83	0	-	0.46
DA	32.5	32.5	7	0	0	0	0.3	0	0	0.86	0.88	0	-	0.89
WR	34.5	34.5	14	0.1	0	0.13	0.45	0.14	0	0.91	1	0	-	0.55
LR	34.7	34.7	10	0	0.5	0.14	0.15	0	0	0.64	0.1	0.1	-	1
BB	36.4	36.46	21	0	0.6	1	0.64	0.05	0	0.67	0.53	0	-	0.65
OC	38.3	38.3	9	0	0.78	1	0.67	0	0.11	0.58	0.57	0.18	0.04	0.38
MK	38.9	38.9	12	0.33	0.6	0.77	0.62	0.17	0.17	0.58	0.61	0.21	0.07	0.79
SV	39.6	39.6	7	-	-	-	-	-	-	0.14	0.14	0.57	-	-
DI	40.4	40.4	16	0.5	1	0.94	0.75	0	0.38	-	-	-	-	0.56
WS	40.7	40.7	13	0	1	1	0.89	0	0.38	0.62	0.62	0.3	0.24	0.77
II	41.1	41.1	18	-	-	1	-	-	-	0.61	0.61	0.33	0.31	-
DN	41.3	41.3	10	0.5	1	1	1	0	0.6	0.63	0.67	0.44	0.35	0.7
OR	41.6	41.6	6	0.25	0.8	0.8	0.5	0	0.17	0.9	0.89	0.75	0.4	0.33
NN	42.7	42.7	17	0.75	1	1	1	0	1	0.71	0.71	0.59	0.62	0.67
Spearman Correlation Coefficient				0.46	0.89 *	0.75 *	0.81 *	0.65 *	0.88 *	-0.04	0.1	0.92 *	0.85 *	0.08

Note: The prefix for each RFLP locus indicates the region and the number indicates the position within the gene. N, smallest number of chromosomes for any locus of the sample; several additional chromosomes were analyzed at most loci.

Table 4.3 Association between RFLP alleles and chromosomal types of the X.

	para		*fu1*		*cac*		*nonA*	
	uncut	Cut	uncut	Cut	uncut	cut	uncut	cut
In(X)c	10	29	3	57	0	41	30	11
Std	3	26	27	1	27	0	5	20
D’	0.46		0.95		1		0.62	
	p=0.09		p<0.0001		p<0.0001		p<0.001	

Note: p-values from Fisher’s exact test.

Table 4.4 Association between RFLP alleles and chromosomal types of the 4.

	*Chico*2168		*Chico*2250		*Chico*2105		*Chico*2486		*Bib*205		*Bib*501	
	uncut	Cut	uncut	cut	uncut	cut	uncut	cut	uncut	cut	uncut	Cut
In(4)ab	5	35	6	34	11	27	18	23	0	9	9	0
Std	43	69	50	55	101	6	67	0	59	1	56	3
D’	0.60		0.61		0.75		1		1		1	
	p=0.002		p=0.0001		p<0.0001		p<0.0001		p<0.0001		p=0.64	

Note: p-values from Fisher’s exact test

Figure 4.1 Representation of the gene location on the X and 4th chromosome.

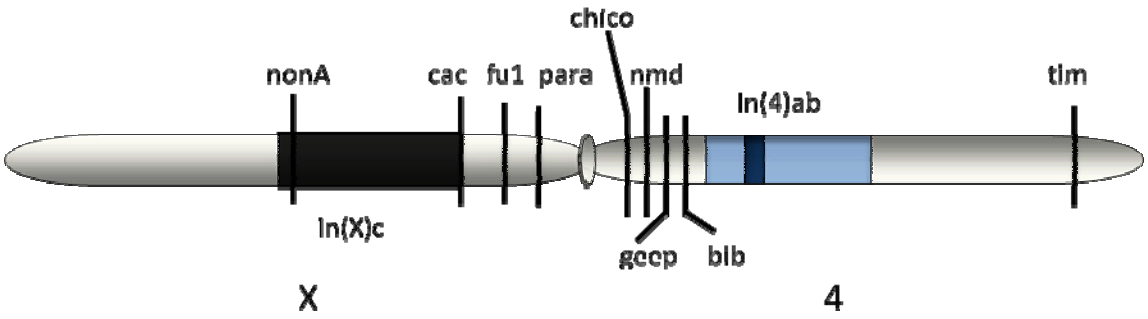
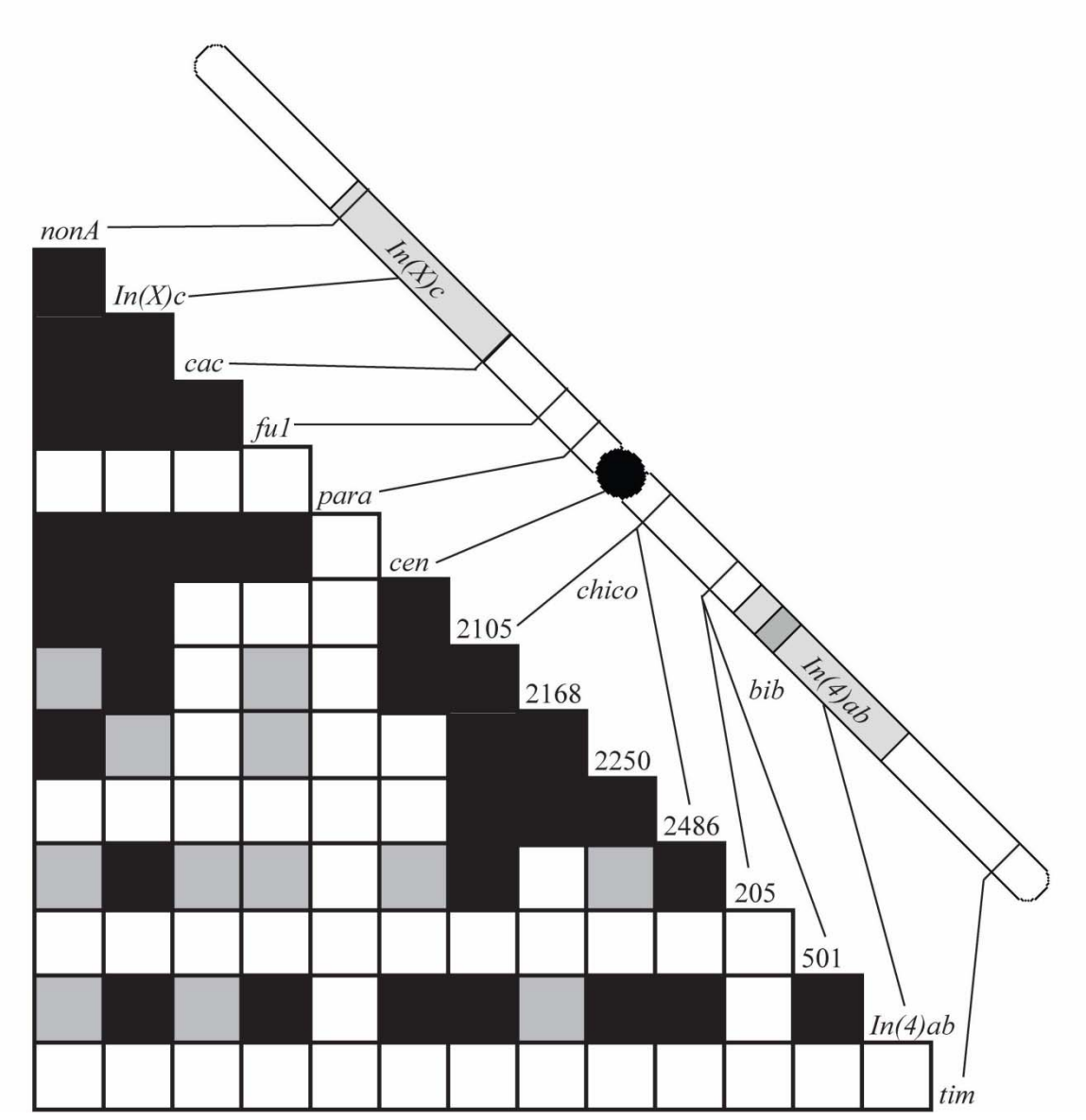


Figure 4.2 Linkage disequilibrium plot for markers on the X and 4th chromosome. Black boxes represent a significant association between two sites using the Fisher's exact test with Bonferroni correction ($p<0.0005$). Grey boxes represent a significant association before the Bonferroni correction ($0.0005<p<0.05$). White boxes represent a nonsignificant Fisher's exact test.



CHAPTER 5
PHENOTYPIC DIFFERENTIATION BETWEEN CHROMOSOMAL ARRANGEMENTS OF CHROMOSOME X

5.1 Introduction

It has been frequently proposed that chromosomal rearrangements are involved in adaptation to environmental conditions (e.g. Dobzhansky 1970). A major piece of evidence for their connection to adaptation is that there are many reported cases of chromosomal rearrangements having distributions correlated with environmental gradients (e.g. Ananina et al. 2004; Rako et al. 2006; Anderson 1987), while neutral variation does not show the same pattern. Another strong piece of evidence is that there are many reported cases of parallel clines forming in different continents and hemispheres (e.g. Ayala et al. 1989; Prevosti et al. 1990; Knibb 1982).

Chromosomal rearrangements may play a role in facilitating local adaptation because they alter recombination patterns and are capable of maintaining linkage relationships among groups of coevolved genes (Hoffmann et al. 2004). It has been theoretically demonstrated that inversions are advantageous in a population when local selection coordinately establishes associations among multiple alleles (Kirkpatrick and Barton 2006). An inversion contributes to the maintenance of these associations by restricting recombination in heterozygotes and therefore protecting the set of coordinately adapted variants from disruption. This is particularly relevant for species that inhabit a broad geographic range and therefore are exposed to very different environmental conditions. Distinct populations adapt to local conditions; however, gene flow from surrounding, differentially adapted populations will break down favorable allelic combinations, thus impeding long-term adaptation. Chromosomal rearrangements allow maintenance of these allelic associations even in the face of recombination between differentially adapted chromosomes.

There are many documented cases of clinal variation for phenotypes attributed to local adaptation to different parts of a species' range. For example, in *Drosophila,* traits such as body melanisation, ovariole number and fecundity (Azevedo et al. 1996; Rajpurohit et al. 2008), body size (James et al. 1995), egg size (Azevedo et al. 1996), development time (James and Partridge 1995), competitive ability (James and Partridge 1998), timing of egg production (Hoffmann and Mitrovski 2001) and heat and cold resistance (Hoffmann et al. 2002) all show clinal variation. In some cases, variation in traits is associated with inversions. Phenotypes such as resistance to cold and hot temperatures (Weeks et al. 2002; Anderson et al. 2003), longevity (Rodriguez et al. 1999), developmental time (Fernandez Iriarte and Hasson 2000) and body size (Rako et al. 2006), among others, have been associated with gene regions within inversions segregating in *Drosophila* species.

In *D. americana,* several chromosomal rearrangements exhibit distributions associated with geographic gradients. In particular, the distribution of X-4 fusion (McAllister 2002; McAllister et al. 2008) and the Xc inversion are strongly associated with latitude. In chapters 3 and 4 I established that high levels of linkage disequilibrium exist among certain loci on the X and 4th chromosome. The associations present at separate loci are maintained with alternative chromosomal arrangements, which indicate that selection acts to maintain these favorable allelic associations. This clinal distribution of the chromosomal rearrangements and allelic variation suggests that climatic factors are the selective force maintaining this pattern.

Any environmental factor associated with the environmental gradient could potentially be linked to the clinal distribution of the alternative chromosomal rearrangements. McAllister et al. (2008) collected climatic data from weather stations located in proximity to collection sites and found that of the available climatic variables, the best predictor of the frequency of the X-4 fusion was extreme minimum temperature in January. This indicates that temperature, in

particular winter severity, is likely to represent an environmental factor that shapes the observed frequencies of the rearrangements throughout the species' range.

Hori and Kimura (1998) found interspecific differences in cold hardiness in *Drosophila.* They used cold stupor (knock-down temperature) and cold tolerance (lethal temperature) as a measure of cold hardiness for several species. Significant differences among species in cold hardiness were found and the ability to survive cold temperatures was correlated with the local environment of each species. However, significant variation within species was not observed. Gilbert et al. (2001) used chill coma temperature as a measure of cold tolerance. This assay consists of exposing an adult fly to near-zero temperatures, which induces a comatose state. Flies are transferred to room temperature and the time required to recover is recorded for each individual. Similar to other tests for cold tolerance, the time required to recover from the comatose state varies among *Drosophila* species depending on their geographic range (tropical or temperate climates) (Gilbert et al. 2001)

A comparable pattern is observed for heat tolerance (Kimura 2006). Interspecific differences in heat tolerance are observed, but this contrast with low levels of intraspecific variation. Interestingly, there is no correlation between heat tolerance and latitudinal distribution of the species (Kimura 2006). However, heat tolerance was higher in species inhabiting openlands or the forest canopy than in those inhabiting the forest understory (Kimura 2006).

Hoffmann et al. (2002) demonstrated latitudinal variation in resistance to high temperatures (measured as knockdown time) when comparing several strains of *D. melanogaster* collected along a latitudinal cline in Australia. Sorensen et al. (2001) was also able to find variation in thermal tolerance among natural populations of *D. buzzatii* originating from high- and low-temperature environments; however no intraspecific variation in thermal tolerance has been reported in the virilis species group. Garbuz (2003) compared several strains of *D. virilis*

and *D. lummei* and found no significant intraspecific variation in their response to heat shock temperatures.

We hypothesize that the alternative chromosomal rearrangements of *D. americana* and the alleles associated with them confer to flies the ability to tolerate different extreme temperatures because they are each adapted to different parts of the geographic range. Strong population differentiation at loci associated with the chromosomal rearrangements has been established previously (McAllister 2002; Vieira et al. 2006; Chapters 3 and 4). These differentiated loci or genes nearby could be coding for the adaptive traits that are under selection. Therefore, there is a possibility that intraspecific differences in temperature tolerance could be found in *D. americana* if populations from different parts of the species range are assayed.

Preliminary unpublished data from the McAllister lab demonstrated that there is a clear effect of the latitude from which the lines were derived on the time required for a fly to recover from a cold coma. That is, lines derived from flies collected from northern localities recovered more quickly from the cold coma than lines originating from southern localities. Hence, the objective of this study is to determine if there is a relationship between tolerance to extreme temperatures and genetic variation of alternative chromosomal forms of the X found in *D. americana*. I hypothesize that flies with the arrangements more prevalent in the southern part of the species' range (unfused X and 4) will have higher tolerance to extreme hot temperatures, while flies with the arrangement more prevalent in the north (fused X-4 and Xc inversion) will better tolerate extreme cold temperatures.

5.2 Materials and Methods

5.2.1 Cold Tolerance

I used a cold coma assay assess any differences in cold tolerance between distinct *D. americana* genotypes. This assay consists of induction of coma by placing flies in individual

vials in a water bath at a low temperature for an extended period of time. Later the fly is removed from the water bath and placed at room temperature and the time it takes to upright itself (recovery time) is recorded. This was initially carried out at several coma temperatures in order to optimize the assay. As with the heat tolerance experiments, the eye mutant lines NN97.4 Red and ML97.5 Pur were used.

The following temperature and exposure time combinations were investigated: 0.5^{o}C for 16 hours, 0^{o}C for 16 hours, -0.5^{o}C for 16 hours, -1^{o}C for 16 hours, -1.5^{o}C for 16 hours, -2^{o}C for 16 hours, -5^{o} C for 6 hours and -10^{o}C for 6 hours. An ANOVA was performed for each assay to determine if there were significant differences in the log transformed recovery time between lines. I used the log recovery time (instead of the raw data) to normalize the skewed measures of the recovery time. The log recovery time was used as the dependent variable and line and sex were used as independent variables. The temperature/exposure time combination of 0.5^{o}C for 16 hours had the highest difference among lines and the least variance; therefore it was chosen to continue the experiments at a larger scale using other inbred lines.

Cold Coma Recovery Time in Inbred Lines: For a first approximation of differences in cold hardiness between flies with different chromosomal arrangements, I performed the cold coma recovery assay on several inbred lines with different chromosomal arrangements. This assay was performed on a total of 15 inbred lines. Of these, 5 lines have unfused X and 4 with a standard arrangement of X, 9 lines have fused X-4 and *In(X)c*, and 1 line has a fused X-4 and a standard X. There is only 1 line available with this rare karyotype. Presumably this karyotype arose in the lab in an isofemale line that was initially a heterozygote, as a product of a recombination event between a standard unfused and a fused Xc chromosome. ML97.5 Pur and NN97.4 Red lines were included in each assay as a control. An ANOVA was performed to determine if there were significant differences in the recovery time among the alternative karyotypes. The log transformed recovery time was used as the dependent variable, while the

chromosomal arrangement of the X and sex were used as the independent variables. I used a Tukey test to determine which pairs were significantly different in the ANOVA.

Effects of Chromosome X on Cold Coma Recovery Time: Of the inbred lines assayed in the previous section, I selected the lines with the most extreme phenotypes for further experiments. Several inbred lines with fast recovery times were each crossed with an inbred line with a slow recovery time. I assayed the parental lines, the F1s from both reciprocal crosses and the F2s from the backcrosses for their recovery time. An ANOVA was used to determine if there were significant differences in the log recovery time among parentals, F1s and F2s. The log recovery time was used as the dependent variable while line (or cross in the case of F1s and F2s) and sex were used as the independent variables.

5.2.2 Heat Tolerance

In order to determine if there are any intraspecific differences in the ability to tolerate extreme heat within *D. americana*, I assayed two inbred eye color mutant lines of *D. americana* for heat tolerance. These two lines originate from opposite ends of the geographic distribution of the species and differ in their chromosomal arrangement. ML97.5 Purple was collected in Monroe, Louisiana and has an unfused X and 4 and the standard *D. americana* arrangements for chromosomes X, 2, 3, and 4, while chromosome 5 has a small subterminal inversion (5b). NN97.4 Red was collected in Niobrara, Nebraska and carries the fused X-4 arrangement, along with inversions on chromosome X (*In(X)c*), 4 (*In(4)ab*) and 5 (*In(5)a*). These flies were anesthetized with CO2 shortly after they eclosed to separate them by sex. However, they were not anesthetized again during the rest of the assays in order to minimize effects the CO_2 could potentially have on the thermotolerance (Nilson et al. 2006).

Basal thermal tolerance assay: Basal thermotolerance is the tolerance of intense exposure to high temperatures in naïve (never exposed to extreme temperatures) adult flies. Following an

aging period of 7-10 days, 35 to 50 flies were transferred to empty vials (no more than 10 individuals per vial) and immersed in a water bath for 30 minutes at temperatures ranging from 39 to 44°C. Subsequently, they were placed in a vial with food in an incubator at 22°C (where the flies were originally reared). For the analysis all flies exposed to each heat shock temperature were pooled together. Tolerance was determined as the proportion of flies that can walk 48 hours after the heat shock.

Inducible thermotolerance: Inducible tolerance is the change in the thermotolerance when flies are exposed to mild hyperthermia before they are exposed to heat shock temperatures. For this assay, the 35-50 (7 -10 day old) flies of the mutant lines were exposed to a pre-treatment that consisted of being placed in a water bath at 25°C for 1 hour or 35°, 36° or 37°C for 30 minutes. These flies were exposed to heat shock temperatures ranging from 39° to 42°C for 30 minutes. As with the previous experiments, tolerance was determined as the proportion of flies that can walk 48 hours after the heat shock.

5.3 Results

5.3.1 Cold Tolerance

The cold coma assays done at several temperatures and exposure time revealed that as expected, as incubation temperature decreased, the time required for flies of both lines to recover from the cold coma increased. At incubation temperatures of -0.5°C and higher, flies from the NN97.4 Red line recover significantly faster from the cold coma than flies from the ML97.5 Pur line (0.5°C: ANOVA F_2=15.3, p=0.0001; 0°C:ANOVA F_2=15.6, p=0.0001; -0.5°C: ANOVA F_2=12.0, p=0.0006). At -1°C there are no significant differences in the recovery time among the lines (ANOVA F_2=1.77, p=0.18). The trend is reversed at incubation temperatures lower than -1°C, where ML97.5 Pur has faster recovery time that NN97.4 Red. (Figure 5.1). However, at -

5^{o}C and -10^{o}C the assays were not able to be completed because the flies took too long to recover or died from the exposure to these low temperatures.

The treatment that gave the highest difference in recovery time among lines in the predicted direction, was the easiest to score (in terms of the time it took for the flies to recover) and had the lowest variance in the recovery time was the 0.5^{o}C for 16 hours temperature/time combination. Therefore this was the treatment that I used for the subsequent experiments.

Cold Coma Recovery Time in Inbred Lines: The mean recovery time of lines with the unfused arrangement of the X is higher than it is for lines with the fused X-4 arrangement, while the recovery time for X-4 fused Xc chromosomes is slightly lower than for X-4 fused standard chromosomes (Table 5.1 and 5.2). There are significant differences in the log recovery time among the chromosomal arrangements of the X (ANOVA F_2=27.9, $p<0.0001$). A Tukey test revealed a significant pairwise difference only between the fused X-4 with the Xc inversion and unfused X and 4 karyotypes.

Effects of Chromosome X on Cold Coma Recovery Time: Five pairs of inbred lines with significantly different recovery times were selected from the previous experiment for these assays. If the X chromosome contains genes that are responsible for the phenotypic differences between the parental lines I expect to find a significant difference between the F1 males of the reciprocal crosses because they will differ in their X chromosome, while the rest of the genome will be similar. Also, I expect to find significant differences between the F2 females from the backcross of the reciprocal crosses because these will differ in their X chromosome. The following is a list of the inbred line pairs that were used and a summary of the results obtains for each assay.

FP99.16 (unfused X and 4 and standard X) and G96.23 (fused X-4 with *In(X)c)*: For the F1 reciprocal cross, an ANOVA shows that there is a significant effect of line and the interaction between line and sex (ANOVA line or cross F_5=4.5, $p<0.004$; sex F_1=0.7, $p=0.37$; line*sex

F_5=4.6, p=0.0004). Furthermore, a Tukey test shows there is a significant difference in recovery time between the F1 males from the reciprocal crosses, and no difference between the F1 females, results consistent with the X having an effect on the recovery time. However, parental line males show the inverse pattern (FP99.16 males recover faster than G96.23 males) (Table 5.3).

In the case of the assay for the F2 reciprocal crosses there is also an effect of line and the interaction between line and sex(ANOVA line or cross F_5=23.8, p<0.001; sex F_1=0.7, p=0.39; line*sex F_5=4.6, p=0.0004), however a Tukey test shows there is no significant difference between the recovery times of the F2 females (Table 5.3).

FP99.50 (fused X-4 with *In(X)c*) and ML97.5 (unfused X and 4 and standard X): For the F1 reciprocal cross, an ANOVA shows that there is a significant effect of line, sex and the interaction between line and sex (ANOVA line or cross F_5=14.2, p<0.0001; sex F_1=25.2, p<0.0001; line*sex F_5=3.9, p=0.009). However, a Tukey test shows there is no significant difference between the F1 males or females of the reciprocal crosses. Backcross assays were not performed (Table 5.4).

OR01.46 (unfused X and 4 and standard X) and FP99.16 (unfused X and 4 and standard X): For the F1 reciprocal cross, results from an ANOVA show that there is a significant effect of line and interaction between line and sex (ANOVA line or cross F_5=8.2, p<0.0001; sex F_1=2.6, p=0.1; line*sex F_5=13.1, p<0.0001). Also, a Tukey test shows there are significant differences between F1 males from the reciprocal crosses, and no significant differences between F1 females of the reciprocal crosses. The back crosses were not performed.

G96.45 (fused X-4 with *In(X)c*) and ML97.5 (unfused X and 4 and standard X): For the F1 reciprocal cross, results from an ANOVA show that there is a significant effect of line, sex and interaction between line and sex (ANOVA line or cross F_5=38.1, p<0.0001; sex F_1=38,

$p<0.0001$; line*sex $F_5=2$, $p=0.05$). However, there is no significant difference between the F1 males or females from the reciprocal crosses (Table 5.6).

In the F2 reciprocal crosses assays there is also a significant effect of line and of interaction between line and sex (ANOVA line or cross $F_5=61.8$, $p<0.0001$; sex $F_1=1.98$, $p=0.1$; line*sex $F_5=5.2$, $p=0.0001$). Furthermore, there is a significant difference between the F2 females from the reciprocal crosses.

FP99.50 (fused X-4 with *In(X)c*) and G96.13 (fused X-4 with *In(X)c*): For the F1 reciprocal cross, results from an ANOVA show that there is a significant effect of line and sex, but not of the interaction between line and sex (ANOVA line or cross $F_5=21$, $p<0.0001$; sex $F_1=27.8$, $p<0.0001$; line*sex $F_5=2.6$, $p=0.02$). Also, there is no significant difference between F1 males of the reciprocal crosses (table 5.7).

5.3.2 Heat Tolerance

Basal thermotolerance assay: All flies were alive and walking at temperatures lower than 39^{o}C and all were dead at temperatures higher than 41^{o}C. There was no difference in the 50% survival among the different lines. LD_{50} was around 40.8^{o}C for both lines and both sexes. Also, there was no consistent trend of one line having higher survival than the other (Figure 5.2).

Inducible thermotolerance assay: The original expectation was that exposing the flies to a mild hyperthermia before the heat shock would increase the thermotolerance in the flies. However, this did not occur and the LD_{50} was still close to 40.8^{o}C for both lines at all pretreatment temperatures (Figures 5.3 and 5.4).

5.4 Discussion

Because the alternative chromosomal arrangements of the X are distributed clinally due to the action of selection (McAllister 2002; McAllister et al. 2008; Chapter 2) and DNA sequence variation on the X is highly associated with the chromosomal rearrangements, showing

a pattern consistent with the presence of coadapted gene complexes harbored in the rearrangements (Chapters 3 and 4), the objective of this study was to determine whether there is also phenotypic differentiation. I was able to establish that there are intraspecific differences in the time required to recover from a cold induced coma.

The experiments performed to optimize the cold coma assay revealed a reversal from the expected pattern. At the higher temperatures used for the incubation the observed differences in the recovery time were in the predicted direction. The NN97.4 Red line recovered significantly faster from the cold coma compared to the ML97.5 Pur line. This is the expected result since the Red line originated from the northern part of the species range, and therefore is presumably adapted to tolerate colder temperatures, while the Pur line originated from the southern part of the species range and is expected to be less cold tolerant. However, at lower incubation temperatures the trend is reversed and the ML97.5 Pur line is able to recover more promptly from the coma than the NN97.4 Red line. The reason for this unexpected switch is unknown but could be evidence that the mechanisms used by the organism to cope with extreme cold temperatures are not the same at all temperatures.

The cold coma assays of the set of inbred lines with diverse X chromosome karyotypes revealed a significant difference between fused and unfused X and 4^{th} chromosomes. However, there was no significant difference between fused X-4 and the standard or inverted arrangements of the X, but it is noteworthy that there was only one line available with the rare karyotype of fused X-4 with the standard X, therefore there were far fewer flies assayed for this category.

In the case of the assays to determine the effects of the X chromosome on the recovery results were mixed, therefore it is not possible to conclude unequivocally whether the X chromosome influences recovery time. The rearing conditions of the flies used for these experiments were standardized but there were several limitations. The goal was to breed flies under the same conditions in order for the flies to be the same size and age at the time of the

assay. The same number of males and females were placed in a vial to mate and lay eggs for the same number of days. This was done to keep the density in each vial standardized among lines thus the flies used for the assay would have similar size (under the assumption that flies would lay the same number of eggs per vial). However, in spite of these precautions there is visible variation in the size and weight of the flies and this could have an impact on the ability of the fly to respond to extreme temperatures. Ideally, each fly used for the assays should have been weighed and this used as a variable that could possibly affect the recovery time. However, at the time that these assays were preformed a balance with the precision necessary to detect differences among the fly mass was unavailable. In conclusion, there are intraspecific differences in the ability to tolerate cold temperatures, in terms of recovery from a cold coma. Unfortunately, with the available data it is not possible to conclusively say if the X chromosome plays a role in these differences.

In terms of tolerance to heat the assays failed to show intraspecific differences. These negative results may be attributed to the fact that average temperatures in a certain geographic regions may not reflect the actual conditions that organisms are exposed to. Higher latitudes have lower average temperatures, but in reality maximum temperatures can be as high as they are in the lower latitudes. Hence, flies must to be able to tolerate similar heat shock temperatures throughout the species' distribution range. The lack of a gradient for extreme high temperatures is corroborated by findings of McAllister et al. (2008). Here, weather data from the last 30 years was collected from weather stations located near the *D. americana* collection sites. The minimum and maximum temperatures for each season are relevant because they can be associated with biological thresholds for survival and/or reproduction. Consistent with these results McAllister et al. (2008) found that winter minimum temperature sufficiently explained the distribution of the X-4 fusion, while the maximum summer temperatures did not improve the

fit. This indicates that although the flies are subject to a gradient in terms of the coldest temperatures that they must tolerate, the gradient is not mirrored by high temperatures.

Although I was unable to establish an association between the chromosomal rearrangements of the X and phenotypes relevant to adaptation to the environmental gradient, this does not mean that the associations do not exist. Evidence that the chromosomal rearrangements are protecting sets of coadapted alleles (Evans et al. 2007; Chapters 2,3 and 4) imply the rearrangements are harboring alleles that affect phenotypes involved in adaptation to local conditions, therefore there must be an association between the chromosomal rearrangements and phenotype.

Table 5.1 Recovery time means for each inbred line.

Line	Karyotype	Mean RT (s)	95% CI
FP99.2	Unfused Std X	195.5	154.8-235.1
FP99.16	Unfused Std X	357.5	239.8-475.2
FP99.34	Fused std X	164.7	121.7-207.7
FP99.50	Fused Xc	113.2	56.4-170.0
G96.13	Fused Xc	377.5	260.6-494.4
G96.23	Fused Xc	92.8	70.6-115
G96.45	Fused X	68.7	32-105.4
G96.70	Fused Xc	164.8	78.6-251
HI99.12	Unfused Std X	204.1	162.2-246
HI99.4	Fused Xc	191.0	148.6-233.4
ML97.5	Unfused Std X	223.8	178.7-268.9
NN97.8	Fused Xc	160.7	103.8-217.6
OR01.46	Unfused Std X	114.8	64.1-165.5
OR01.50	Fused Xc	141.9	106.7-177.1
OR01.92	Fused Xc	188.7	118.2-259.2

Table 5.2 Recovery time means for each karyotype.

Karyotype	N	Mean RT	95% CI
Fused X-4, Std X	46	131.4	102.3-160.5
Fused X-4, Xc	533	154.3	137.4-171.2
Unfused X and 4, Std X	366	248.2	220-276.2

Table 5.3 Recovery time for the F1 and F2 reciprocal crosses between FP99.16 and G96.23.

Cross	Sex	n	Mean RT	95% CI
FP99.16	females	24	69.4	21.4-117.4
	Males	24	27.3	15.3-39.3
G96.23	females	23	36.6	19.6-53.6
	Males	24	70	36-104
FP99.16xG96.23	females	23	26.3	11.3-41.3
	Males	24	32.7	15.7-49.7
G96.23xFP99.16	females	24	74.3	26.3-122.3
	Males	24	16	13.6-18.4
(FP16xG23)xFP16	females	24	80	51-109
	Males	24	69	41-97
(G23xFP16)xG23	females	24	65	38-92
	Males	24	134	73-195

Table 5.4 Recovery time for the F1 reciprocal crosses between FP99.50 and ML97.5.

Cross	Sex	n	Mean RT	95% CI
FP99.50	Female	22	214.5	126.5-302.5
	Male	24	69.9	37.9-101.9
ML97.5	Female	20	361	304-418
	Male	24	214	167-261
FP99.50xML97.5	Female	22	350.7	268.7-432.7
	Male	23	177.7	89.7-265.7
ML97.5xFP99.50	Female	20	361	243-479
	Male	24	241.8	187.8-295.8

Table 5.5 Recovery time for the F1 reciprocal crosses between FP99.16 and OR01.46.

Cross	Sex	n	Mean RT	95% CI
FP99.16	Female	22	42.9	22.9-62.9
	Male	24	62.3	45.3-79.3
OR01.46	Female	24	176.5	121.5-231.5
	Male	24	44.5	27.5-61.5
FP99.16xOR01.46	Female	22	66.1	40.1-92.1
	Male	23	87.4	63.4-111.4
OR01.46xFP99.16	Female	20	43.6	26.6-60.6
	Male	24	36	18-54

Table 5.6 Recovery time for the F1 and F2 reciprocal crosses between ML97.5 and G96.45.

Cross	Sex	n	Mean RT	95% CI
ML97.5	Females	24	127.3	90.3-164.3
	Males	24	51.4	301.4-71.4
G96.45	Females	23	25.2	19.2-31.2
	Males	24	20.9	10.9-30.9
ML97.5xG96.45	Females	23	120.9	86.9-154.9
	Males	24	73.3	31.3-115.3
G96.45x ML97.5	Females	24	43.9	17.9-69.9
	Males	24	26.9	19.9-33.9
(ML5xG45)xML5	Females	24	113	108-118
	Males	24	65.7	61.7-69.7
(G45xML5)xG45	Females	24	33	11-55
	Males	24	62.7	61.7-63.7

Table 5.7 Recovery time for F1 reciprocal crosses between FP99.50 and G96.13.

Cross	Sex	n	Mean RT	95% CI
FP99.50	Female	22	24.7	21.5-27.9
	Male	23	32.6	11.6-53.6
G96.13	Female	24	199.4	135.4-263.4
	Male	24	54.9	23.9-85.9
FP99.50xG96.13	Female	22	131.2	53.2-209.2
	Male	24	33	10-56
G96.13xFP99.50	Female	22	72.6	38.6-106.6
	Male	24	82.3	13.3-13.3

Figure 5.1 Cold coma recovery time means for NN97.4 Red and ML97.5 Pur lines at different incubation temperatures.

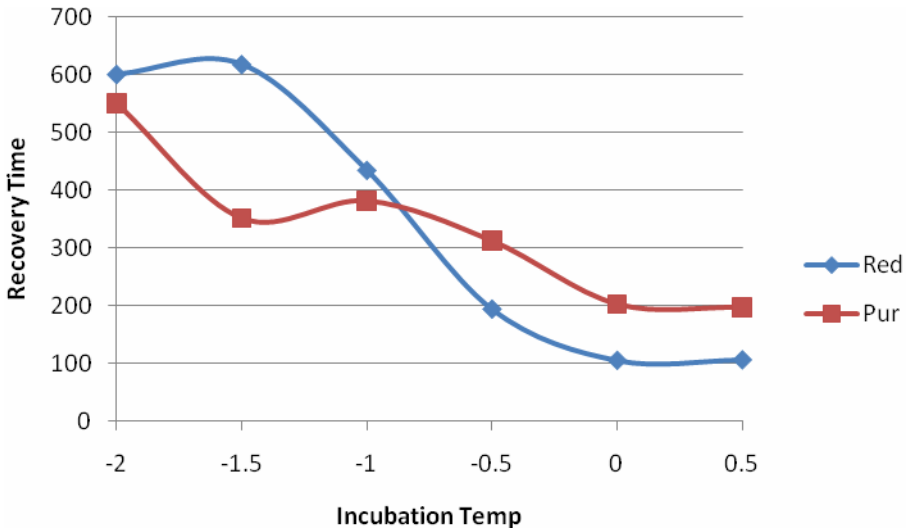


Figure 5.2 Survival rates of NN97.4 Red and ML97.5 Pur lines at different heat shock temperatures.

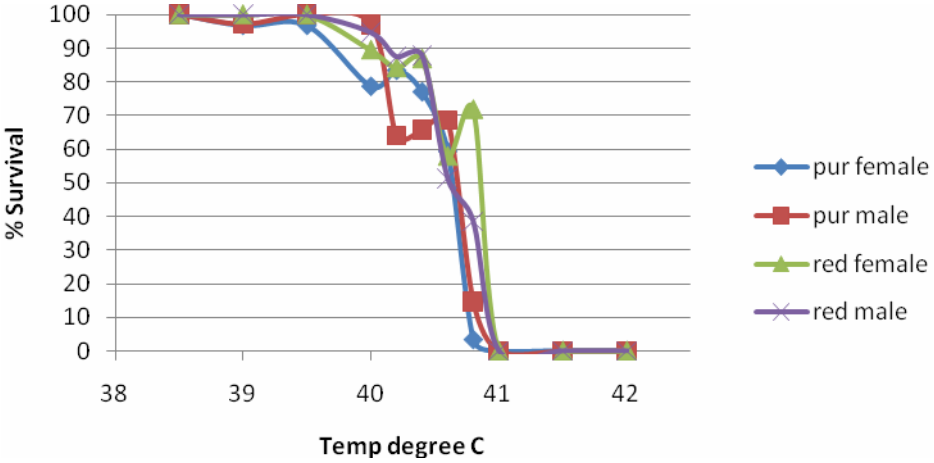


Figure5.3 Survival rates of NN99.4 Red line at different temperatures after flies were exposed to 4 different pretreatment temperatures.

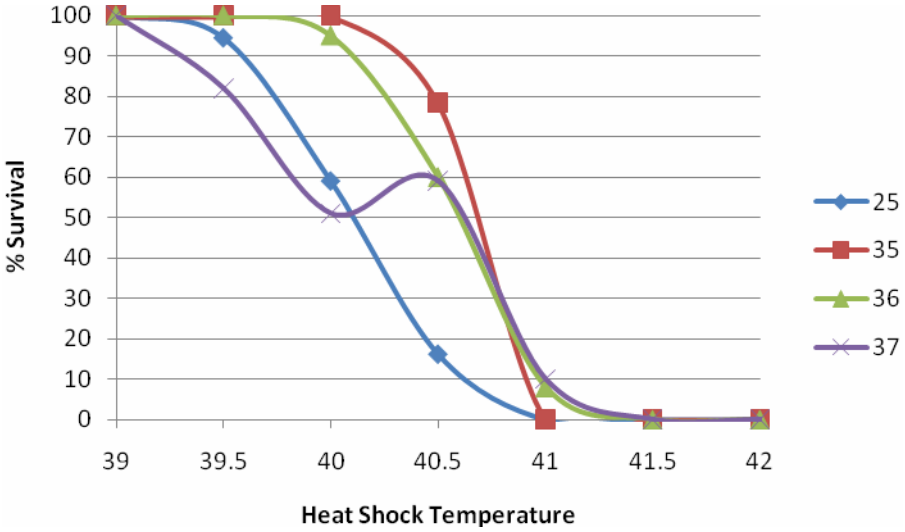
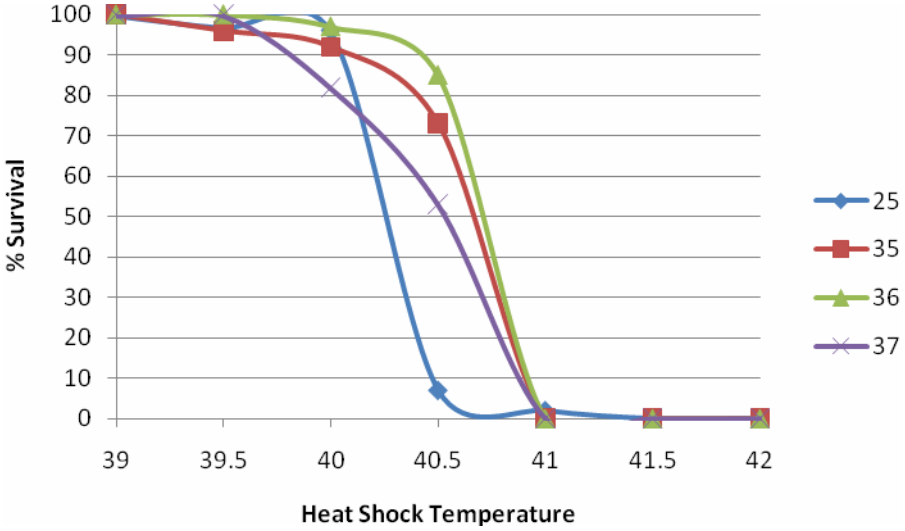


Figure 5.4 Survival rates of ML97.5 Pur line at different temperatures after flies were exposed to 4 different pretreatment temperatures.



CHAPTER 6

CONCLUSIONS

Darwin hypothesized that in order for natural selection to occur populations need to have natural variation, and this variation needs to be stably transmitted from generation to generation (Darwin 1859). However, the nature of this variation remained elusive. Because of the early availability of cytological techniques, chromosomal rearrangements gave the first insights into the source of the phenotypic variation (Dobzhansky 1937).

Adaptation to a local environment will most likely involve several traits and these can be coded for by many loci that can respond to the selective pressures in a coordinated fashion. Chromosomal rearrangements can provide means for the organism to become adapted to the local conditions because they can involve large stretches of DNA and therefore include multiple loci, changing their linkage relationships and recombination patterns. As a result, populations will be able to maintain locally adapted or coadapted alleles at high association, even in the face of gene flow with other populations (Dobzhansky 1970; Kirkpatrick and Barton 2006).

The central aim of this work was to determine what role chromosomal rearrangements have played in adaptation to the environment using *D. americana* as a model system. The available techniques are diverse and allow exploration of this question with approaches that include not only the historically used cytological study of natural variation of the chromosomal rearrangements, but also analysis of the effects that these rearrangements have on the DNA sequence variation in hopes of ultimately finding the loci and phenotypes involved in the adaptation.

D. americana has been extensively studied in terms of the geographic distribution of the X-4 fusion. Even though it was originally thought these chromosomal variants represented distinct species or subspecies (Orr and Coyne 1989; Pitnick et al. 1997; Powell 1997; Spicer and Bell 2002; Garbuz et al. 2003; Throckmorton 1982; Hilton and Hey 1996, 1997), recent studies

involving extensive sampling have established that *D. americana* is a single species. The X-4 fusion is a polymorphic chromosomal rearrangement that shows clinal distribution as a result of natural selection favoring different frequencies of these karyotypes along with associated DNA sequence variants in different parts of the geographic gradient (McAllister 2002 and McAllister et al 2008). However, the same effort has not been put into the study of other polymorphic chromosomal inversions found in this species.

In the present study I investigated in depth the geographic distribution of several polymorphic inversions present in 3 chromosomes of *D. americana* by sampling populations from a broad geographic range. Two major findings emerge from this sampling. First, I found not only is there a cline for the already established X-4 fusion, but there are also polymorphic inversions that show distributions correlated with the geographic gradient. And since there is no population structure for neutral loci in *D. americana* (McAllister and McVean 2000; McAllister 2002, 2003; Schäfer et al. 2006) this is evidence that natural selection, and not demographic history, is responsible for the distribution pattern of these chromosomal rearrangements. Second, this sampling revealed high linkage disequilibrium among several of these rearrangements and these strong associations are attributed to natural selection favoring certain combinations of chromosomal arrangements. These two findings combined are strong evidence that the *D. americana* genome is under correlated selective pressure and chromosomal rearrangements are aiding adaptation by creating and maintaining linkage of the alleles that are involved in local adaptation.

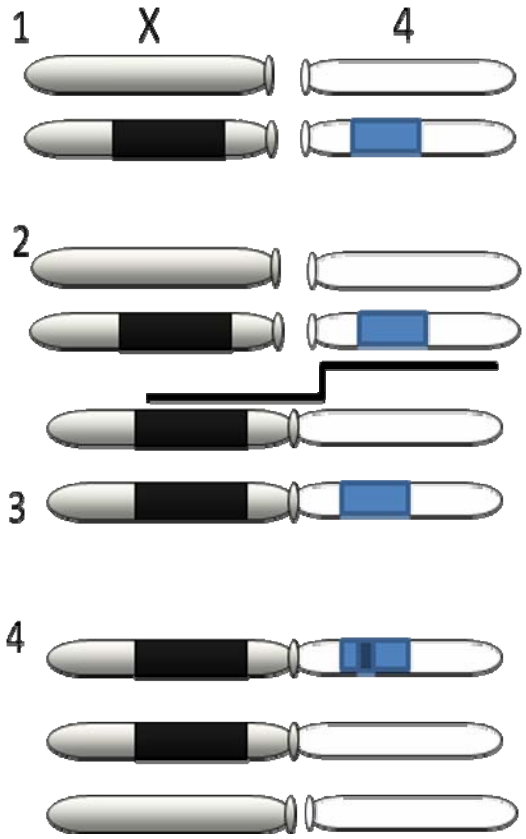
The DNA sequence variation studies done on inbred lines, as well as on flies collected directly from the wild provide further support for the hypothesis that chromosomal rearrangements facilitate local adaptation. It was found that sequence variation at certain loci is highly associated with the chromosomal rearrangements but this association is interrupted by regions that are not in linkage disequilibrium with the rearrangements. Models of coadaptation

show it is selectively advantageous to limit recombination among coordinately selected alleles (Kimura 1956; Kirkpatrick and Barton 2006). Strong linkage disequilibrium among isolated regions of DNA such as the ones demonstrated in this study provide robust evidence for coadaptation (Schaeffer et al 2003).

The combination of the findings from this study with previous reports allows me to propose a model of the historical events that have given rise to the patterns observed for the X and the 4th chromosome (Figure 6.1). Sequence data on the X, combined with the fact the *D. novamexicana*, a sister species to *D. americana* is fixed for the Xc inversion, indicate that this inversion has been segregating in the population for an extended period of time, possibly exhibiting a similar clinal distribution to the one currently observed. The X-4 fusion subsequently arose on an X chromosome with the inverted arrangement and increased in frequency due to selection. Eventually the X-4 fused *In(X)c* chromosome completely replaced the Xc unfused karyotypes resulting in the extreme linkage disequilibrium observed currently between the fusion, the inversion and the associated DNA variation. On the other hand, sequence variation on chromosome 4 indicates that inversion 4a arose on a chromosome 4 free from the X. A subsequent recombination event between an *In(4)a* unfused chromosome and a standard X-4 fused chromosome between the centromere and the proximal inversion breakpoint gave rise to the X-4 fused *In(4)a*, *In(X)c* karyotype. The high association observed among haplotypes at loci located on the 4 and the X and the inversions on both chromosomes indicate that natural selection is favoring these particular allelic associations. A small inversion (4b) nested within In(4)a has been implicated in a subsequent selective event related to the fact that chromosome 4 becomes sex linked through the centromeric fusion, therefore exposing it to feminizing selection that can favor an inversion that would restrict recombination with the autosomal 4.

In conclusion, the study of chromosomal and sequence variation on chromosome X and 4 of *D. americana* indicate the effects of correlated selection, where chromosomal rearrangements have played a major role in maintaining linkage disequilibrium among alleles throughout both chromosomes, protecting a coordinately adapted gene complex containing locally adapted alleles from exchange with other chromosomal forms. This has resulted in islands of linkage disequilibrium between certain gene regions and the chromosomal rearrangements. Regions of differentiation between chromosomal classes product of this differential adaptation could be precursor to the development of reproductive isolation between locally adapted populations, which allows for the possibility that local adaptation sets the foundation upon which separate species could eventually emerge (Dieckmann and Doebeli 1999). Although there is no available DNA sequence data for chromosome 5, the extreme linkage disequilibrium and distinct distribution of *In(5)a* and *In(5)b* indicate there could be a similar pattern of differentiation between chromosomal types due to local adaptation that could also be setting the stage for the evolution of reproductive isolation and ultimately speciation.

Figure 6.1 Evolutionary model of the history of rearrangements on the X and 4th chromosome. 1: inversions on the X and 4th chromosome arise on unfused chromosomes. 2: The X-4 centromeric fusion occurs between an X with *In(X)c* and a chromosome 4 with the standard arrangement. 3:A recombination event between in proximal inversion breakpoint of *In(4)a* and the centromere give rise to the fused Xc-4a karyotype. 4: Karyotypes currently present in the population. *In(4)b* arises within *In(4)a.*



REFERENCES

Ashburner M. 1989. Drosophila: A Laboratory Handbook. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Ayrinhac A., V. Debat, P. Gibert, A. Kister, H. Legout, B. Moreteau, R. Vergilino, and J. David. 2004. Cold adaptation in geographical populations of Drosophila melanogaster: phenotypic plasticity is more important than genetic variability. Func. Ecol. 18:700.

Balanya J., J. M. Oller, R. B. Huey, G. W. Gilchrist, and L. Serra. 2006. Global genetic change tracks global climate warming in Drosophila subobscura. Science 313:1773-1775.

Blight W. C. 1955. A cytological study of linear populations of Drosophila americana near St. Louis, Missouri. Washington University, St. Louis, MO.

Blight W. C., and A. Romano. 1953. Notes on a Breeding Site of Drosophila americana Near St. Louis, Missouri. Am. Nat. 87:111.

Bradshaw H. D., K. G. Otto, B. E. Frewen, J. K. McKay, and D. W. Schemske. 1998. Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (Mimulus). Genetics 149:367-382.

Caletka B. C., and B. F. McAllister. 2004. A genealogical view of chromosomal evolution and species delimitation in the Drosophila virilis species subgroup. Mol. Phylogenet. Evol. 33:664-670.

Charlesworth D., and B. Charlesworth. 1998. Sequence variation: looking for effects of genetic linkage. Curr. Biol. 8:R658-61.

Chovnick A. 1973. Gene conversion and transfer of genetic information within the inverted region of inversion heterozygotes. Genetics 75:123-131.

Coghlan A., E. E. Eichler, S. G. Oliver, A. H. Paterson, and L. Stein. 2005. Chromosome evolution in eukaryotes: a multi-kingdom perspective. Trends Genet. 21:673-682.

Daborn P. J., J. L. Yen, M. R. Bogwitz, G. Le Goff, E. Feil, S. Jeffers, N. Tijet, T. Perry, D. Heckel, P. Batterham, R. Feyereisen, T. G. Wilson, and R. H. ffrench-Constant. 2002. A single P450 allele associated with insecticide resistance in Drosophila. Science 297:2253-2256.

Darwin C. 1859. On the Origin of Species by Means of Natural Selection. John Murray, London.

Dieckmann U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400:354-357.

Dobzhansky T. 1970. Genetics of the Evolutionary Process. Columbia University Press, New York.

Evans A. L., P. A. Mena, and B. F. McAllister. 2007. Positive selection near an inversion breakpoint on the neo-X chromosome of Drosophila americana. Genetics 177:1303-1319.

Fay J. C., and C. I. Wu. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.

Feldman M. W., S. P. Otto, and F. B. Christiansen. 1996. Population genetic perspectives on the evolution of recombination. Annu. Rev. Genet. 30:261-295.

Filatov D. A. 2002. ProSeq: A software for preparation and evolutionary analysis of DNA sequence data sets. Molecular Ecology Notes 2:621.

Fu Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693-709.

Futuyma D. 1998. Evolutionary Biology. Sinauer Associates, Sunderland, Massachusetts.

Garbuz D., M. B. Evgenev, M. E. Feder, and O. G. Zatsepina. 2003. Evolution of thermotolerance and the heat-shock response: evidence from inter/intraspecific comparison and interspecific hybridization in the virilis species group of Drosophila. I. Thermal phenotype. J. Exp. Biol. 206:2399-2408.

Gavrila L., A. A. Ecovoiu, and L. M. Georgescu. 2001. Localizing genes in Drosophila melanogaster polytene chromosomes by fluorescence in situ hybridization. J. Cell. Mol. Med. 5:74-78.

Gubenko I., and M. B. Evgenev. 1984. Cytological and linkage maps of Drosophila virilis chromosomes. Genetica 65:127.

Hawthorne D. J., and S. Via. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412:904-907.

Hemingway J., and H. Ranson. 2000. Insecticide resistance in insect vectors of human disease. Annu. Rev. Entomol. 45:371-391.

Hilton H., and J. Hey. 1997. A multilocus view of speciation in the Drosophila virilis species group reveals complex histories and taxonomic conflicts. Genet. Res. 70:185.

--- 1996. DNA sequence variation at the period locus reveals the history of species and speciation events in the Drosophila virilis group. Genetics 144:1015-1025.

Hoekstra H. E., K. E. Drumm, and M. W. Nachman. 2004. Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. Evolution 58:1329-1341.

Hoffmann A. A., and L. H. Rieseberg. 2008. Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? Annual Review of Ecology Evolution and Systematics 39:21-42.

Hoffmann A. A., C. M. Sgro, and A. R. Weeks. 2004. Chromosomal inversion polymorphisms and adaptation. Trends Ecol. Evol. 19:482-488.

Hoffmann A. A., A. Anderson, and R. Hallas. 2002. Opposing clines for high and low temperature resistance in drosophila melanogaster. Ecol. Lett. 5:614.

Hori Y., and M. T. Kimura. 1998. Relationship between cold stupor and cold tolerance in Drosophila (Diptera: Drosophilidae). Environ. Entomol. 27:1297.

Hudson R. R., M. Slatkin, and W. P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132:583-589.

Hudson R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics 116:153-159.

Hurst L. D. 1999. The evolution of genomic anatomy. Trends Ecol. Evol. 14:108-112.

Kimura M. 1956. Rules for Testing Stability of a Selective Polymorphism. Proc. Natl. Acad. Sci. U. S. A. 42:336-340.

Lee B. S., N. W. Kim, and N. R. Rim. 2002. Chromosome inversion - Environment relationships in Korean populations of Drosophila melanogaster. Korean Journal of Genetics 24:9-20.

Maside X., and B. Charlesworth. 2007. Patterns of molecular variation and evolution in Drosophila americana and its relatives. Genetics 176:2293-2305.

Maside X., A. W. Lee, and B. Charlesworth. 2004. Selection on codon usage in Drosophila americana. Curr. Biol. 14:150-154.

McAllister B. F. 2003. Sequence differentiation associated with an inversion on the neo-X chromosome of Drosophila americana. Genetics 165:1317-1328.

--- 2002. Chromosomal and allelic variation in Drosophila americana: selective maintenance of a chromosomal cline. Genome 45:13-21.

McAllister B. F., and G. A. McVean. 2000. Neutral evolution of the sex-determining gene transformer in Drosophila. Genetics 154:1711-1720.

McAllister B. F., and B. Charlesworth. 1999. Reduced sequence variability on the Neo-Y chromosome of Drosophila americana americana. Genetics 153:221-233.

McAllister B. F., S. L. Sheeley, P. A. Mena, A. L. Evans, and C. Schlotterer. 2008. Clinal distribution of a chromosomal rearrangement: a precursor to chromosomal speciation? Evolution 62:1852-1865.

Morgan T. J., and T. F. C. Mackay. 2006. Quantitative trait loci for thermotolerance phenotypes in Drosophila melanogaster. Heredity 96:232-242.

Nachman M. W., H. E. Hoekstra, and S. L. D'Agostino. 2003. The genetic basis of adaptive melanism in pocket mice. Proc. Natl. Acad. Sci. U. S. A. 100:5268-5273.

Navarro A., A. Barbadilla, and A. Ruiz. 2000. Effect of inversion polymorphism on the neutral nucleotide variability of linked chromosomal regions in Drosophila. Genetics 155:685-698.

Nilson T. L., B. J. Sinclair, and S. P. Roberts. 2006. The effects of carbon dioxide anesthesia and anoxia on rapid cold-hardening and chill coma recovery in Drosophila melanogaster. J. Insect Physiol. 52:1027-1033.

Orr H. A., and J. A. Coyne. 1992. The Genetics of Adaptation - a Reassessment. Am. Nat. 140:725-742.

Orsini L., S. Huttunen, and C. Schlotterer. 2004. A multilocus microsatellite phylogeny of the Drosophila virilis group. Heredity 93:161-165.

Pepper J. W. 2003. The evolution of evolvability in genetic linkage patterns. BioSystems 69:115-126.

Powell J. R. 1997. Progress and Prospects in Evolutionary Biology: The Drosophila Model. Oxford University Press, New York.

Prevosti A., G. Ribo, L. Serra, M. Aguade, J. Balana, M. Monclus, and F. Mestres. 1988. Colonization of America by Drosophila subobscura: Experiment in natural populations that supports the adaptive role of chromosomal-inversion polymorphism. Proc. Natl. Acad. Sci. U. S. A. 85:5597-5600.

Rozas J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Schaeffer S. W., and W. W. Anderson. 2005. Mechanisms of genetic exchange within the chromosomal inversions of Drosophila pseudoobscura. Genetics 171:1729-1739.

Schaeffer S. W., M. P. Goetting-Minesky, M. Kovacevic, J. R. Peoples, J. L. Graybill, J. M. Miller, K. Kim, J. G. Nelson, and W. W. Anderson. 2003. Evolutionary genomics of inversions in Drosophila pseudoobscura: evidence for epistasis. Proc. Natl. Acad. Sci. U. S. A. 100:8319-8324.

Schafer M. A., L. Orsini, B. F. McAllister, and C. Schlotterer. 2006. Patterns of microsatellite variation through a transition zone of a chromosomal cline in Drosophila americana. Heredity 97:291-295.

Scott T. W., A. C. Morrison, L. H. Lorenz, G. G. Clark, D. Strickman, P. Kittayapong, H. Zhou, and J. D. Edman. 2000. Longitudinal studies of Aedes aegypti (Diptera : Culicidae) in Thailand and Puerto Rico: Population dynamics. J. Med. Entomol. 37:77-88.

Sinervo B., and E. Svensson. 2002. Correlational selection and the evolution of genomic architecture. Heredity 89:329-338.

Sole E., J. Balanya, D. Sperlich, and L. Serra. 2002. Long-term changes in the chromosomal inversion polymorphism of Drosophila subobscura. I. Mediterranean populations from southwestern Europe. Evolution 56:830-835.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

--- 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437-460.

Throckmorton L. H. 1982. The virilis species group. in M. Ashburner, H. L. Carson and J. N. J. Thompson, eds. The Genetics and Biology of Drosophila. Academic Press, New York.

Vieira C. P., P. A. Coelho, and J. Vieira. 2003. Inferences on the evolutionary history of the Drosophila americana polymorphic X/4 fusion from patterns of polymorphism at the X-linked paralytic and elav genes. Genetics 164:1459-1469.

Vieira C. P., A. Almeida, J. D. Dias, and J. Vieira. 2006. On the location of the gene(s) harbouring the advantageous variant that maintains the X/4 fusion of Drosophila americana. Genet. Res. 87:163-174.

Vieira J., B. F. McAllister, and B. Charlesworth. 2001. Evidence for selection at the fused1 locus of Drosophila americana. Genetics 158:279-290.

Watterson G. A. 1975. On the number of segregation sites. Theor. Pop. Biol. 7:256.

APPENDIX

Table A.1 RFLP and chromosome arrangement genotypes for each individual.

	nonA	Xc	cac	fu1	para	X-4	C2105	C2168	C2250	C486	Bib501	Bib206	4ab	tim
CI0519	C	N	T	T	T	A	T	C	C	N	C	T	N	C
CI0507	C	G	T	T	T	A	T	C	C	N	T	T	G	T
CI0527	C	G	N	T	C	A	T	C	C	N	T	T	G	T
CI0515	C	G	T	T	C	A	T	C	N	N	N	N	G	C
CI0525	N	N	T	T	C	N	T	C	N	N	T	T	N	C
CI0521	N	G	N	N	N	A	T	C	C	N	C	T	G	T
CI0517	N	G	N	N	C	A	T	C	C	N	T	T	G	T
CI0501	N	G	N	N	C	A	N	C	C	N	T	T	G	C
LR0503	N	G	T	T	C	A	T	C	C	N	T	T	G	C
LR0509	T	A	C	C	C	G	T	T	T	N	T	T	G	C
LR0513	N	A	C	T	C	G	T	T	T	N	T	T	G	C
LR0545	C	G	T	T	C	A	T	C	N	N	T	T	G	N
LR0549	C	G	T	T	C	A	C	C	T	N	T	T	G	C
LR0547	N	A	C	T	C	G	T	T	T	N	N	N	G	C
LR0517	C	A	C	C	C	G	T	T	T	N	T	T	G	C
LR0527	N	G	N	T	C	A	T	C	T	N	T	T	G	C
LR0541	C	G	N	N	C	A	T	C	T	N	T	T	G	C
LR0525	C	G	T	T	C	A	T	C	T	N	T	T	G	C
LR0539	C	G	N	T	C	A	N	C	T	N	T	T	G	C
BB0555	T	N	N	C	C	N	T	C	C	N	T	T	N	C
BB05105	T	N	N	C	C	N	T	C	C	N	T	T	N	C
BB05103	N	A	C	C	C	G	T	N	T	N	T	T	G	T
BB0547	C	G	T	T	C	N	T	C	C	N	T	T	G	C
BB05101	N	G	T	T	C	A	T	C	C	N	C	T	G	T
BB0569	N	N	N	C	C	N	T	T	N	N	T	T	N	C

Table A.1 Continued

BB0559	T	N	N	C	C	N	T	C	C	N	T	T	N	T
BB0573	N	N	N	N	C	N	T	C	C	N	T	T	N	N
BB0501	C	N	N	C	C	N	T	C	C	N	T	T	N	C
BB0503	T	N	N	C	C	N	T	C	N	N	T	T	N	C
BB0505	N	G	T	N	C	A	T	C	N	N	T	T	G	T
BB0507	C	N	N	N	C	N	T	C	N	N	T	T	N	C
BB0509	T	N	C	C	C	N	T	C	T	N	T	T	N	C
BB0511	N	A	C	N	C	G	T	C	T	N	T	T	G	C
BB0529	C	G	N	N	C	A	T	C	C	N	N	N	G	C
BB0513	T	N	N	N	C	N	T	C	T	N	T	T	N	C
BB0527	T	A	C	N	C	G	T	C	T	N	T	T	G	C
BB0531	N	N	N	N	C	N	T	C	N	N	T	T	N	T
BB0533	T	A	C	C	C	G	T	C	T	N	T	T	G	N
BB0537	C	N	N	C	C	N	T	C	T	N	T	T	N	T
BB0523	T	A	C	N	C	G	T	T	T	N	T	T	G	T
RB0505	N	G	T	T	C	A	T	C	C	N	T	T	G	C
RB0507	N	G	T	N	C	A	T	C	C	N	T	T	G	T
RB0503	N	G	T	T	C	A	T	C	C	N	T	T	G	T
WR0631	C	N	N	N	C	N	T	C	N	N	C	T	N	C
WR0641	T	N	T	T	C	N	T	C	N	N	T	T	N	N
WR0643	T	N	N	N	C	N	T	C	N	N	T	T	N	T
WR0645	T	N	N	N	C	N	T	T	N	N	C	T	N	T
WR0647	C	N	N	N	C	N	T	N	N	N	T	T	N	C
WR0651	T	N	N	N	T	N	T	N	C	N	T	T	N	C
WR0605	C	N	N	T	C	N	T	N	N	N	T	T	N	T
WR0627	N	N	N	T	C	N	T	C	N	N	T	T	N	C
WR0609	C	N	N	T	C	N	T	C	N	N	T	T	N	C
WR0639	C	G	N	T	C	N	N	C	N	N	T	T	G	N
WR0601	N	N	N	N	N	N	T	C	N	N	T	T	N	N

Table A.1 Continued

WR0611	N	N	N	C	N	N	T	C	N	N	T	T	N	T
WR0615	C	N	N	T	N	N	T	C	C	N	T	T	N	C
WR0623	T	N	N	T	N	N	T	C	C	N	T	T	N	T
MK0601	T	N	T	T	N	A	T	T	C	N	T	C	N	C
MK0603	C	N	T	T	N	A	T	C	C	N	T	T	N	N
MK0605	T	A	C	C	N	G	T	T	T	N	C	T	N	C
MK0713	N	N	N	N	N	N	N	N	N	T	T	T	N	C
MK0711	N	N	N	N	C	N	N	C	C	T	T	T	N	C
MK0709	T	N	N	C	C	N	N	T	T	T	C	T	N	C
MK0739	C	N	N	C	T	G	N	C	C	T	T	T	N	T
MK0737	T	N	N	C	N	N	N	T	T	T	T	T	N	C
MK0745	C	N	N	T	T	N	N	C	N	T	N	N	N	C
BA0603	C	N	N	T	C	N	T	N	N	N	T	T	N	T
BA0637	C	N	T	T	C	N	T	C	N	N	T	T	N	T
BA0633	N	N	N	N	C	N	T	N	C	N	C	T	N	C
BA0601	C	G	T	T	C	A	T	N	C	N	T	T	G	C
BA0605	C	G	T	T	C	A	T	C	N	N	T	T	G	T
BA0609	N	N	T	T	C	N	T	C	N	N	T	T	N	N
BA0617	C	N	T	T	C	N	T	C	T	N	T	T	N	T
BA0621	C	N	N	N	C	N	N	C	N	N	C	T	N	T
BA0625	T	N	N	T	C	N	N	N	N	N	T	T	N	N
BA0627	N	N	N	T	C	N	N	C	N	N	T	T	N	T
BA0635	C	N	N	T	N	N	T	C	N	N	T	T	N	C
BA0639	T	G	N	T	N	N	T	C	C	N	T	T	G	C
BA0641	N	N	N	T	N	N	T	C	C	N	T	T	N	C
BA0645	N	G	N	N	N	N	T	C	C	N	T	T	G	N
BA0647	T	G	N	T	N	A	T	C	N	N	C	T	G	C
BA0649	N	N	N	N	N	N	T	C	N	N	T	T	N	T
DA0601	N	G	N	T	C	A	T	C	C	N	T	T	G	T

Table A.1 Continued

DA0609	N	G	T	T	C	A	T	C	C	N	T	T	G	C
DA0607	C	G	T	T	N	A	T	T	T	N	N	N	G	C
DA0615	C	G	T	N	N	A	T	C	C	N	T	T	G	C
DA0619	T	G	N	T	N	A	T	C	C	N	T	T	G	C
DA0621	T	G	N	T	N	A	T	C	C	N	T	T	G	C
DA0633	C	G	N	T	N	N	T	N	C	N	T	T	G	C
DA0635	C	G	N	T	N	A	T	C	C	N	T	T	G	C
FG0641	N	G	T	T	C	A	T	C	N	N	C	T	G	C
FG0645	C	N	N	N	C	N	T	C	N	N	T	T	N	C
FG0647	N	N	N	N	C	N	T	C	N	N	C	T	N	N
FG0649	N	N	N	N	C	N	T	C	N	N	T	T	N	T
FG0653	N	N	N	N	C	N	N	C	N	N	T	T	N	C
FG0655	C	N	N	N	C	N	N	N	C	N	T	T	N	T
FG0605	N	G	T	C	N	N	T	T	T	N	N	N	G	N
FG0609	C	G	T	T	N	A	T	T	T	N	T	T	G	C
FG0611	C	N	N	N	N	N	T	C	C	N	T	T	N	C
FG0613	C	N	N	C	N	N	T	C	N	N	T	T	N	C
FG0637	N	A	C		T	G	N	C	C	T	T	C	G	C
FG0629	C	G	T	T	C	A	N	N	C	T	N	N	G	T
FG0627	T	N	N	T	C	N	N	N	N	N	T	T	N	N
FG0625	T	N	N	T	N	N	N	N	N	T	T	T	N	T
FG0621	T	N	N	T	N	N	N	N	N	N	T	T	N	C
FG0619	T	G	N	T	C	A	N	C	C	T	T	T	G	C
FG0617	N	G	N	N	N	A	N	N	N	N	T	T	G	T
FG0615	C	N	N	T	N	N	N	C	C	T	Ta	N	N	T
OR0703	N	A	C	C	C	G	N	N	N	T	T	T	G	T
OR0705	N	N	N	C	C	G	N	C	C	C	T	T	N	T
OR0709	N	A	C	C	C	G	C	C	C	T	T	C	A	C
OR0713	C	N	N	T	N	G	N	C	N	T	T	T	N	C

Table A.1 Continued

OR0715	N	A	C	N	N	G	T	C	C	C	N	N	A	N
OR0717	T	N	C	C	T	G	T	C	C	T	T	T	N	T
OC0715	N	N	N	N	N	N	T	T	T	T	T	T	N	N
OC0719	T	A	C	C	C	N	N	C	C	T	T	T	A	T
OC0725	N	N	C	C	C	N	N	N	T	T	T	T	N	C
OC0739	T	N	C	C	N	N	N	C	C	T	T	T	N	T
OC0741	C	N	C	C	C	N	N	C	C	T	T	T	N	T
OC0747	N	N	N	C	N	N	N	C	C	T	N	N	N	N
OC0749	N	N	N	C	N	N	N	N	T	T	T	C	N	T
OC0751	N	N	N	C	N	N	N	N	N	T	N	N	N	N
OC0753	N	N	N	C	N	N	N	C	T	T	T	T	N	N
OC0757	N	A	C	C	N	G	N	T	T	T	C	T	G	C
WS0701	N	A	C	C	C	G	N	N	C	T	T	C	A	C
WS0707	N	A	N	C	N	G	N	N	C	T	T	C	A	T
WS0709	N	N	C	C	C	N	N	N	C	T	T	T	N	C
WS0713	T	A	C	C	C	G	N	C	C	T	T	T	G	T
WS0703	T	A	N	C	N	G	N	N	T	T	N	N	A	C
WS0705	T	A	N	C	C	G	N	N	T	T	N	N	A	N
WS0717	T	A	C	C	C	G	N	T	C	T	T	T	G	T
WS0723	T	N	C	C	N	N	N	N	C	T	T	T	N	C
DI0581	N	A	N	C	N	G	N	C	C	C	T	C	A	C
DI0589	N	A	C	C	T	G	N	T	T	T	T	T	G	C
DI05103	N	A	C	C	C	G	N	T	T	T	T	T	G	C
DI0543	N	A	N	C	N	G	N	C	N	T	T	T	G	T
DI0507	T	A	C	C	T	G	N	C	C	T	T	T	G	T
DI0513	T	A	C	C	T	G	N	N	N	T	T	C	N	T
DI0521	C	A	C	C	C	G	N	C	C	T	T	T	G	C
DI0533	N	A	N	C	N	G	N	N	N	T	T	C	A	C
DI0511	T	A	C	C	N	G	N	C	C	T	T	T	G	T

Table A.1 Continued

DI0505	T	N	C	C	N	N	N	T	N	T	T	C	N	T
DI0599	T	A	N	C	N	G	N	T	N	T	T	T	G	T
DI05105	T	A	C	C	N	G	N	C	N	T	T	C	A	C
DI05113	N	G	N	T	N	A	N	N	N	T	T	T	G	C
DI0515	N	N	N	N	N	N	N	N	N	T	N	N	N	N
DI0527	C	A	C	C	N	G	N	T	N	T	T	T	G	C
DI0531	N	A	N	C	N	G	N	N	N	N	T	T	G	T
DI0517	N	A	C	C	C	G	N	C	N	C	T	C	A	C
II0705	N	N	N	N	N	N	T	T	T	T	N	N	N	N
II0707	N	A	N	N	N	G	C	C	C	T	N	N	A	N
II0709	N	N	N	N	N	G	T	C	C	T	N	N	N	N
II0711	N	N	N	N	N	N	T	C	C	T	N	N	N	N
II0713	N	A	N	N	N	G	C	C	C	T	N	N	G	N
II0715	N	N	N	N	N	N	C	C	C	T	N	N	N	N
II0719	N	N	N	N	N	G	C	C	C	C	N	N	N	N
II0735	N	N	N	N	N	N	T	T	T	T	N	N	N	N
II0733	N	N	N	N	N	G	T	C	C	T	N	N	N	N
II0723	N	A	N	N	N	G	T	T	T	T	N	N	G	N
II0721	N	N	N	N	N	N	T	T	T	T	N	N	N	N
II0743	N	A	N	N	N	G	T	C	C	T	N	N	G	N
II0727	N	N	N	N	N	N	C	T	T	T	N	N	N	N
II0731	N	A	N	N	N	N	T	T	T	T	N	N	G	N
II0741	N	G	N	N	N	N	T	C	C	T	N	N	G	N
SV0711	N	A	N	N	N	G	T	T	T	T	N	N	G	N
SV0723	N	A	N	N	N	G	T	C	C	T	N	N	G	N
SV0705	N	A	N	N	N	G	T	C	C	T	N	N	G	N
SV0717	N	A	N	N	N	G	C	C	C	C	N	N	A	N
SV0709	N	A	N	N	N	N	T	C	C	T	N	N	G	N
SV0721	N	A	N	N	N	G	T	T	T	T	N	N	G	N

Table A.1 Continued

SV0701	N	A	N	N	N	G	T	T	T	T	N	N	A	N
NN0715	N	A	N	N	N	G	C	C	C	C	N	N	A	N
NN0719	N	N	N	N	N	A	T	C	C	C	N	N	N	N
NN0721	N	A	N	N	N	G	C	C	C	C	N	N	A	N
NN0725	N	A	N	N	N	G	T	T	T	T	N	N	G	N
NN0727	N	A	N	N	N	G	C	C	C	C	N	N	A	N
NN0729	N	A	N	N	N	G	C	C	C	C	N	N	A	N
NN0731	N	A	N	N	N	G	T	T	C	T	N	N	G	N
NN0733	N	A	N	N	N	G	C	C	C	C	N	N	A	N
NN0737	N	A	N	N	N	G	C	C	C	C	N	N	A	N
NN0701	N	A	N	N	N	G	T	T	T	T	N	N	G	N
NN0703	N	A	N	N	N	G	T	T	T	T	N	N	A	N
NN0707	N	A	N	N	N	G	C	C	C	C	T	C	A	C
NN0709	N	A	N	N	N	G	T	C	C	T	N	N	G	T
NN0711	N	A	N	N	N	G	C	C	C	C	T	C	A	C
NN0713	N	A	N	N	N	G	C	C	C	C	N	N	A	N
NN0717	N	A	N	N	N	G	C	C	C	C	N	N	A	N
SC0705	N	A	N	N	N	N	T	T	T	T	N	N	A	C
SC0709	N	A	N	N	N	N	C	C	C	C	N	N	A	N
SC0711	N	A	N	N	N	N	T	T	T	T	N	N	G	N
SC0707	N	A	N	N	N	N	T	C	C	T	N	N	G	N
SC0713	N	A	N	N	N	N	T	C	T	T	N	N	G	N
DN0743	N	A	N	N	N	G	C	C	C	C	N	N	A	N
DN0713	N	A	N	N	N	G	T	C	C	T	N	N	G	N
DN0705	N	A	N	N	N	G	C	C	C	C	N	N	A	N
DN0725	N	N	N	N	N	G	N	T	N	T	N	N	N	N
DN0711	N	A	N	N	N	G	C	C	C	C	N	N	A	N
DN0733	T	A	C	C	N	G	T	T	T	T	T	T	G	T
DN0739	N	A	N	N	N	N	C	C	C	C	N	N	A	N

Table A.1 Continued

DN0727	N	A	N	N	N	G	T	T	T	T	N	N	G	N
DN0729	N	A	N	C	N	G	T	T	T	T	T	T	G	C
DN0731	N	A	N	N	N	G	T	T	T	T	N	N	G	C
DN0749	N	N	C	C	N	G	C	C	C	C	T	C	N	C
DN0753	N	N	C	C	N	G	C	C	C	C	T	C	N	C
DN0755	T	A	C	N	C	G	C	C	C	C	T	C	A	C
DN0757	N	A	N	N	N	G	T	C	C	T	N	N	G	N
DN0761	T	N	N	C	T	G	T	T	T	T	T	T	N	C
DN0765	T	A	N	C	N	G	T	T	T	T	T	T	G	T
DN0767	N	N	N	N	N	G	C	C	C	C	T	C	N	C
DN0769	N	A	N	N	N	N	T	C	C	C	N	N	A	N
DN0717	N	A	N	N	N	G	T	C	C	C	T	C	A	T
WS0711	N	A	N	C	N	G	T	C	C	T	T	T	G	C
WS0719	T	N	C	C	T	N	T	T	T	T	T	T	N	C
WS0727	C	A	N	C	N	G	T	T	T	T	T	T	G	C
WS0731	T	N	N	C	N	N	T	C	C	T	T	T	N	C
WS0737	N	A	N	C	N	G	T	T	T	T	T	T	G	C
WS0739	T	N	N	C	N	G	C	C	C	C	T	C	N	C
WS0743	N	A	N	N	N	G	C	C	C	C	T	C	A	N
WS0745	N	A	N	N	N	N	T	T	T	T	T	N	G	N
WS0749	N	A	N	N	N	G	C	C	C	T	N	N	A	N
WS0751	N	A	N	N	N	G	T	C	C	T	N	N	G	N
WS0753	N	N	N	N	N	N	C	C	C	C	N	N	N	N
WS0755	N	A	N	N	N	G	T	T	T	T	N	N	G	N
WS0757	N	A	N	N	N	N	T	C	C	T	N	N	G	N
WS0759	N	A	N	N	N	G	T	T	T	T	N	N	G	N
WS0721	N	N	N	N	N	N	C	C	C	T	N	N	N	N
WS0729	N	A	N	N	N	G	T	C	C	T	N	N	A	N
OC0705	N	A	N	N	N	G	T	C	C	T	N	N	G	N

Table A.1 Continued

OC0717	T	N	C	C	C	G	T	C	C	T	T	T	N	C
OC0703	N	A	N	N	N	G	T	T	T	T	N	N	G	N
OC0729	N	A	C	N	N	G	T	T	T	T	N	N	G	N
OC0713	N	A	N	N	N	G	T	T	T	T	N	N	A	N
OC0711	N	N	N	N	N	G	T	T	T	T	N	N	N	N
OC0723	T	A	C	C	C	G	T	C	C	T	T	T	G	T
OC0709	N	G	N	N	N	A	T	C	C	T	N	N	G	N
OC0737	C	N	N	C	C	G	T	T	T	T	T	T	N	T
OC0731	N	A	N	N	N	G	T	C	C	T	N	N	G	N
OC0727	N	A	N	N	N	N	C	C	C	T	N	N	G	N
OC0735	N	A	C	N	N	G	C	C	C	T	N	N	G	N
OC0701	N	A	N	N	N	G	T	C	C	T	N	N	A	N