Spring 2010

Investigating patterns of parallel genetic change in repeated adaptation

Sara Lynn Sheeley
University of Iowa

INVESTIGATING PATTERNS OF PARALLEL GENETIC CHANGE IN REPEATED ADAPTATION

by

Sara Lynn Sheeley

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

May 2010

Thesis Supervisor: Associate Professor Bryant F. McAllister

ABSTRACT

The phenomenon of repeated evolution runs counter to expectations about the role of contingency in adaptation. However, many examples of independently acquired similar traits show that evolution sometimes does follow the same path. Factors influencing the probability of such an event include selection, trait complexity and relatedness. Previous investigations of repeated adaptation have primarily focused on low-complexity traits subject to strong selection. Studies of systems with varying levels of trait complexity, selection, and relatedness are needed to evaluate the relative contributions of these factors. The series of studies reported here 1) establishes a system for inquiry into the role of parallel adaptation among hosts and parasites and 2) provides an assessment of the role of parallel genetic change in the evolution of a complex trait.

In Chapter 2, I show that all-female broods in a line of *Drosophila borealis* are caused by infection with a male-killing strain of *Wolbachia* that is very closely-related to another male-killing strain infecting a geographically and evolutionarily distant species of *Drosophila*. This host-parasite system, together with two other known male-killing *Wolbachia* strains infecting *Drosophila* provides a framework for investigating the role of parallel evolution in the independent acquisition of the male-killing trait among *Wolbachia*, as well as in the adaptation of divergent hosts to similar male-killing parasites.

In Chapters 3-5, I investigate the role of parallel genetic change in a complex trait in two species of *Drosophila* by searching for evidence of adaptation in the *Drosophila americana* homologs of genes thought to underlie adaptation to climate in *Drosophila melanogaster.* In Chapter 3, I investigate the *D. americana* homolog of *Alcohol dehydrogenase* (*Adh*). In contrast with *D. melanogaster*, which segregates functionally distinct variants in *Adh* that represent local adaptation to climate, *D. americana* segregates little variation. This is surprising, especially because *Adh* of *D. americana* is

found near a polymorphic chromosomal rearrangement that does segregate geographically-structured alleles across the species' range. In Chapter 4, I report similarities at the *Phosphoglucomutase* (*Pgm*) locus in the two species, including a shared excess of nonsynonymous variants and the presence of clinal alleles. However, while variation at *Pgm* of *D. melanogaster* is proposed to underlie local adaptation, variation at *Pgm* of *D. americana* appears to be predominantly neutral. In Chapter 5, I investigate the role of positive selection in sequence evolution in the *D. americana* homologs of a group of genes thought to underlie local adaptation to climate in *D. melanogaster*. The two species share a large geographic range and exhibit levels of sequence variation that indicate a similar effective population size, but *D. melanogaster* appears to undergo more frequent fixation of advantageous alleles. Approximately half of all amino acid divergence in *D. melanogaster* is attributable to positive selection, but I find no signs of positive selection in the investigated genes in *D. americana*. Overall, the results reveal little or no parallel evolution at the single genes analyzed. This lack of parallel evolution is likely a result of the high complexity of adaptation to climate as well as contingency.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

INVESTIGATING PATTERNS OF PARALLEL GENETIC CHANGE IN REPEATED ADAPTATION

by

Sara Lynn Sheeley

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

May 2010

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

Sara Lynn Sheeley

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Biology at the May 2010 graduation.

Thesis Committee: _____
Bryant F. McAllister, Thesis Supervisor

_____
Josep Comeron

_____
Stephen Hendrix

_____
Ana Llopart

_____
Toshihiro Kitamoto

To Kent Morrie Sheeley and Joy Ann Sheeley

ACKNOWLEDGMENTS

First and foremost, I'd like to thank my advisor, Dr. Bryant McAllister, for his continued patience and support. Bryant's guidance was indispensable as I learned and grew as a scientist. I would also like to thank the members of my committee: Steve Hendrix, Josep Comeron, Ana Llopart and Toshi Kitamoto for their insight and assistance. Thanks to Gary Gussin for serving on my committee, always making sure I had a spot teaching Fundamental Genetics, and for many helpful discussions regarding my future. In addition to Gary, I'd like to thank Bob Malone and Brenda Leicht for being excellent teaching mentors.

Thank you to the members of the McAllister Lab, both past and present. Thanks to Amy Evans, Yaz Ahmed, Rebecca Hart-Schmidt, Josh Peloquin and Paulina Mena. Special thanks to Paulina for being a constant source of support, both professionally and personally. It has been a lot quieter without her here, and I bet Bryant will miss us when we're both gone--whether or not he ever admits it.

Thanks to the rest of my bitches: Chrissy Freisinger, Shannin Zevian and CJ Jenkins for listening to me complain, especially as I wrote this thesis. Thanks to the Biology Department as a whole. The department has been nothing but supportive and enthusiastic during my time here. Thanks in particular to Phil Ecklund for his constant help navigating university and departmental guidelines.

Last but not least, thanks especially to my family. While I think my sanity has suffered some during graduate school, I'm certain it would be considerably worse without them. Thanks to Nathan & Katherine, Trevor & Crystal, Taylor, Clayton, Brielle, Ashlyn and to Mom & Dad. Thanks for letting me vent about school when I wanted to, and letting me avoid the subject when I didn't. Thanks in particular to Mom and Dad for their constant love and for abundant vegetables. Everything I have achieved is attributable in large part to their unvarying support.

ABSTRACT

The phenomenon of repeated evolution runs counter to expectations about the role of contingency in adaptation. However, many examples of independently acquired similar traits show that evolution sometimes does follow the same path. Factors influencing the probability of such an event include selection, trait complexity and relatedness. Previous investigations of repeated adaptation have primarily focused on low-complexity traits subject to strong selection. Studies of systems with varying levels of trait complexity, selection, and relatedness are needed to evaluate the relative contributions of these factors. The series of studies reported here 1) establishes a system for inquiry into the role of parallel adaptation among hosts and parasites and 2) provides an assessment of the role of parallel genetic change in the evolution of a complex trait.

In Chapter 2, I show that all-female broods in a line of *Drosophila borealis* are caused by infection with a male-killing strain of *Wolbachia* that is very closely-related to another male-killing strain infecting a geographically and evolutionarily distant species of *Drosophila*. This host-parasite system, together with two other known male-killing *Wolbachia* strains infecting *Drosophila* provides a framework for investigating the role of parallel evolution in the independent acquisition of the male-killing trait among *Wolbachia*, as well as in the adaptation of divergent hosts to similar male-killing parasites.

In Chapters 3-5, I investigate the role of parallel genetic change in a complex trait in two species of *Drosophila* by searching for evidence of adaptation in the *Drosophila americana* homologs of genes thought to underlie adaptation to climate in *Drosophila melanogaster.* In Chapter 3, I investigate the *D. americana* homolog of *Alcohol dehydrogenase* (*Adh*). In contrast with *D. melanogaster*, which segregates functionally distinct variants in *Adh* that represent local adaptation to climate, *D. americana* segregates little variation. This is surprising, especially because *Adh* of *D. americana* is

found near a polymorphic chromosomal rearrangement that does segregate geographically-structured alleles across the species' range. In Chapter 4, I report similarities at the *Phosphoglucomutase* (*Pgm*) locus in the two species, including a shared excess of nonsynonymous variants and the presence of clinal alleles. However, while variation at *Pgm* of *D. melanogaster* is proposed to underlie local adaptation, variation at *Pgm* of *D. americana* appears to be predominantly neutral. In Chapter 5, I investigate the role of positive selection in sequence evolution in the *D. americana* homologs of a group of genes thought to underlie local adaptation to climate in *D. melanogaster*. The two species share a large geographic range and exhibit levels of sequence variation that indicate a similar effective population size, but *D. melanogaster* appears to undergo more frequent fixation of advantageous alleles. Approximately half of all amino acid divergence in *D. melanogaster* is attributable to positive selection, but I find no signs of positive selection in the investigated genes in *D. americana*. Overall, the results reveal little or no parallel evolution at the single genes analyzed. This lack of parallel evolution is likely a result of the high complexity of adaptation to climate as well as contingency.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION

> Wind back the tape of life to the early days of the Burgess Shale; let it play again from an identical starting point, and the chance becomes vanishingly small that anything like human intelligence would grace the replay.
>
> Stephen J. Gould, Wonderful Life: The Burgess Shale and the Nature of History

As unlikely as it may be on the scale referred to by Gould, the phenomenon of repeated adaptation continues to attract the attention of evolutionary biologists. Independently derived traits with phenotypic similarity are well known, including such classic examples as the evolution of wings in birds, insects and bats and the repeated loss of pigmentation in cave-dwelling organisms. Often, attempts are made to distinguish such instances of repeated adaptation as either 'convergent' or 'parallel,' based upon the ancestral state of each organism; *Parallel evolution* occurs when the same trait arises independently from the same ancestral state in two lineages, while *convergence* describes independent occurrence of the same trait from distinct ancestral states (Simpson 1961).

Interestingly, as knowledge of the genetic basis of traits increases, these definitions become more difficult to apply. Even in cases of known parallel evolution, the underlying genetic architecture of the acquired trait may differ in some lineages (genetic convergence). For instance, populations of the ancestrally light-colored rock pocket mouse *Chaetodipus intermedius* have independently acquired adaptive melanism in populations living on lava flows. So, while the ancestral state and the acquired phenotype of all melanic populations is likely the same, suggesting parallel change, convergence is apparent at the genetic level (Hoekstra & Nachman 2003).

When the genetic basis of a trait is unknown, the evolutionary relatedness of the two species frequently guides inference of parallelism versus convergent evolution. Parallel evolution is generally inferred for independently derived traits in closely related species, whereas convergence is generally inferred for distantly related taxa. In cases like

the very closely related populations of rock pocket mice, though, this classification will fail. Because of this difficulty, some scientists have suggested that the dichotomy of parallel and convergent evolution is unnecessary, and that all repeated evolution should be referred to as convergence (Arendt & Reznick 2007). However, identification of the level of homology at which change is parallel (e.g. phenotype, genetic pathway, gene, amino acid or nucleotide change) remains an important pursuit. It is imperative, though, to qualify observations of parallel evolution in terms of the level of homology.

When Gould theorized that the repeated acquisition of human-level intelligence was highly improbable, he was likely correct, but what is it that makes that chance 'vanishingly small?' Gould implicates contingency - a combined result of chance and history that causes the path of life to vary among iterations. In his extreme example of the evolution of human intelligence, contingency makes parallel adaptation very unlikely. However, parallel adaptation is observed frequently in the form of homologous change among lineages.

Gould wrote of the potential for repeated phenotypic evolution, but perhaps even more extraordinarily, parallel adaptation at the genetic level is becoming well-known. For example, multiple highland lineages of waterfowl exhibit adaptive changes in their hemoglobin that are not shared with more closely-related lowland lineages (McCracken et al. 2009) and the enzyme responsible for C_4 photosynthesis in grasses evolved at least eight times from the same non-C_4 enzyme (Christin et al. 2007). Parallel evolution is not restricted to change in very closely related lineages; Sodium ion channels in garter snakes and pufferfish have independently evolved increased resistance to toxins via similar amino acid changes (Feldman et al. 2009; Jost et al. 2008) and changes in the same gene, *Mcr1*, have been implicated in adaptive pigmentation change in birds, mammals, reptiles and fishes (Mundy 2005; Nachman 2005; Rosenblum et al. 2004; Gross et al. 2009).

The accumulation of examples of parallel adaptation shows that under some circumstances, the tape of life does repeat itself. It is therefore of interest to understand

the factors that influence parallel adaptive change. Perhaps the most obvious contributor is selection. In each of the cases mentioned above, the trait of interest is one that clearly provides a similar selective advantage in the two lineages. Independent change to the same character state in two lineages suggests positive selection, whether or not the ancestral state was the same (Zhang & Kumar 1997). Simulations confirm that parallel evolution at the nucleotide level is twice as likely under selection as under neutrality (Orr 2005) and empirical analysis of protein sequence evolution indicates that only certain mutational paths are probable under positive selection (Weinrich et al. 2006). Purifying selection also influences parallel adaptation by constraining amino acids to certain character states, thus increasing the probability of parallel change (Rokas & Carroll 2008). At the sequence level, parallel evolution results in homoplasy, or the presence of the same character state that is not due to identity by descent. A recent survey of eight genomes revealed twice as much homoplasy at the amino acid level as would be expected under neutral evolution. The observed homoplasy consisted mostly of substitution of amino acids with similar properties, implying that parallel and convergent change is constrained by purifying selection. However, the quantity of homoplasious sites implicates a role of positive selection (Rokas & Carroll 2008).

A second factor that influences the probability of parallel adaptation is homology. The more similar the substrates for evolutionary change, the more changes have the potential to be parallel. While similar substrates likely increase the probability of parallel evolution, it is important to note that no level of evolutionary relatedness guarantees parallel change. For instance, while the *Mcr1* gene is responsible for adaptive pigmentation differences among distantly-related taxa, similar pigmentation changes in some closely-related mice have a different genetic basis (Hoekstra et al. 2006).

Other factors influencing parallel evolution are trait and gene specific. If many genes underlie a trait, it is more likely that the response to selection in two lineages will involve changes in different genes (Gompel & Prud'homme 2009). Even for complex

traits, certain genes are expected to undergo adaptive change more frequently than others; Genes that have large mutational targets are more likely to be targets of selection, and within genes, mutations with the least deleterious pleiotropic effects are most likely (Weinrich et al. 2006; Gompel & Prud'homme 2009).

Outstanding questions in repeated evolution include the frequency, levels of homology and factors influencing parallel change. For the most part, analysis of parallel evolutionary change has focused on traits of low complexity like the examples listed above. In order to understand the forces influencing the likelihood of parallel change, a multitude of systems that vary in selection, complexity, and relatedness must be investigated. Herein I contribute to our knowledge in two ways: 1) by establishing a system in which to investigate parallel evolution in a host-parasite system and 2) by direct assessment of the role of parallel sequence change in adaptation to climate in *Drosophila*.

In Chapter 2, I present a characterization of an embryonic male-killing *Wolbachia* infection in *Drosophila borealis*. *Wolbachia* are common intracellular parasites of insects, and are capable of eliciting a number of reproductive phenotypes in their hosts, including embryonic male-killing (Hilgenboeker et al. 2008; Stouthamer et al. 1999). The male-killing phenotype has evolved independently in at least five lineages of *Wolbachia* infecting insects (Dyer & Jaenike 2004), but the genetic basis of the trait is unknown in all cases. Previous to this characterization, two male-killing lineages of *Wolbachia* with independent origins were known to infect *Drosophila* hosts. The newly-characterized male-killing *Wolbachia* described here is very closely related to a lineage of male-killing *Wolbachia* infecting *Drosophila innubila*. Together, these two closely-related *Wolbachia* represent an independent origin of male killing relative to the third male-killer infecting *Drosophila bifasciata.* Characterization of this third male-killing lineage of *Wolbachia* infecting a *Drosophila* host provides a framework to investigate parallel evolution in both the hosts and parasite.

Within this framework, adaptation in both parasites and hosts can be investigated to determine the level of homology at which the species have adapted in parallel. In the parasite, the phenotype is parallel; phylogenetic analysis indicates that each male-killing lineage likely evolved from an ancestral state that caused cytoplasmic incompatibility (Dyer & Jaenike 2004). It remains to be seen, however, whether the genetic basis of the male-killing trait is the same among *Wolbachia* lineages. For the hosts, there are undoubtedly selection pressures that are common to hosts of male-killing parasites. In particular, the closely-related *Wolbachia* in *D. borealis* and *D. innubila* are very efficient male-killers, completely eliminating male offspring of infected females in the lab. In order for these two host species to avoid extinction, some flies must be able to produce male offspring, either by avoiding infection or by suppressing the male-killing phenotype. Investigation in the hosts would begin by identifying the phenotypes that allow male survival, and could ultimately lead to identification of the underlying genetic changes caused by selection imposed by the male killer.

Chapters 3-5 provide a direct assessment of potential parallel adaptation to climate between *Drosophila melanogaster* and *Drosophila americana*. Numerous surveys of variation in *D. melanogaster* have identified genes proposed to contribute to local adaptation to the newly-colonized temperate portions of the species' range. Many of the identified genes are involved in central or toxin metabolism, and it is hypothesized that they segregate variation in response to selection that varies with temperature. It is unknown, however, if local adaptation in other species exposed to similar temperature gradients are likely to involve these genes or genes in the same pathways.

Drosophila americana provides an ideal system in which to investigate possible parallel adaptation to a temperature gradient because it is endemic over the same latitudinal range in North America that is now colonized by *D. melanogaster*. In addition, *D. americana* is known to maintain variation in a chromosomal polymorphism consisting of the fusion of the X and 4th chromosomes (Muller elements A and B) in response to

selection across its range. The fusion is nearly fixed in populations throughout the Upper Mississippi River Valley and nearly absent in populations near the Gulf Coast (McAllister et al. 2008). Unstructured variation in putatively neutral markers on the 4th chromosome demonstrates the maintenance of the latitudinal cline for the chromosomal polymorphism despite high levels of gene flow (McAllister 2002). In addition to the X-4 fusion cline, *D. americana* exhibits long-distance linkage disequilibrium among the fusion, multiple inversions on the X and 4th chromosomes and loci in and around those chromosomal features. In contrast, unlinked variation segregates among these linked features. This is a further indication that gene flow is sufficient to randomize these arrangements, but selection maintains them (McAllister 2003; Evans et al. 2007; Mena 2009).

Chapter 3 is an investigation of the selective pressures acting on the *Alcohol dehydrogenase* (*Adh*) locus of *D. americana*. The *Adh* locus is perhaps the most well-understood example of adaptive variation within a single gene in *D. melanogaster*. Geographically-structured coding and noncoding variation at *Adh* of *D. melanogaster* on multiple newly-colonized continents is associated with kinetic differences in ADH function as well as phenotypic differences in ethanol tolerance (Oakeshott et al. 1982; Laurie et al. 1991). Furthermore, artificial selection for heat and cold tolerance affects *Adh* allele frequency, and natural clines have shifted in response to climate change (Kamping & van Delden 1999; Umina et al. 2005). Variation at *Adh* of *D. melanogaster* is almost certainly selectively maintained. It remains to be seen, however, whether maintenance of variation at *Adh* is a common mode of local adaptation.

I investigate patterns of variation at *Adh* in *D. americana* in order to determine whether there is evidence that allelic differences contribute to local adaptation. The *Adh* locus is of particular interest in *D. americana* because it is near the base of the fourth chromosome, and recombination between *Adh* and the centromere is very low in female flies heterozygous for the polymorphic fusion of the X and 4th chromosomes. This

reduced recombination means that parallel adaptation via allelic differences at *Adh* could contribute to maintenance of the chromosomal fusion polymorphism.

Unlike previous work in *D. americana* that has focused on portions of the genome where recombination is restricted by chromosomal polymorphism, the final two chapters of my thesis include the evaluation of potential parallel evolution in presumably freely-recombining regions of the genome. Free recombination should allow selection to work on individual adaptive variants in these regions.

Chapter 4 is a comprehensive evaluation of the adaptive significance of variation at the *Phosphoglucomutase* (*Pgm*) locus of *D. americana*. *Pgm* of many species is highly variable at the amino acid level, and that variation is frequently proposed to be locally adaptive (Carter & Watt 1988; Hoffman 1985; Leigh Brown 1977; Pogson 1991; Verrelli & Eanes 2000; 2001a; 2001b). Because *Pgm* lies at a branchpoint in the central metabolic pathway, allelic differences potentially alter the flux of glucose through the pathway and may result in fitness tradeoffs in different environments. In *D. melanogaster,* PGM variation is proposed to be locally adaptive based on a series of studies that report excess amino acid variation (Verrelli & Eanes 2000), clinal segregation of PGM alleles (Verrelli & Eanes 2001a) and kinetic differences among PGM alleles (Verreli & Eanes 2001b).

Given the frequent reports of high levels of variation at *Pgm* and the evidence for selective maintenance of variation in *D. melanogaster*, it is possible that maintenance of variation at this locus is a common mode of local adaptation among species. I investigate patterns of sequence and associated phenotypic variation in *D. americana* in order to examine the selective forces acting on the *Pgm* locus. By expanding analysis of *Pgm* to *D. americana*, I contribute to an understanding of the forces influencing variation at that locus across species and investigate a potentially frequent site of parallel adaptation. *D. americana* provides an ideal comparator to *D. melanogaster* because it shares a similar life history, potentially resulting in similar selection pressures, but a different demographic history.

Chapter 5 is an evaluation of the role of positive selection on the evolution of genes in *D. americana* whose homologs are proposed to segregate variation that underlies local adaptation in *D. melanogaster*. While parallel change can occur in the absence of positive selection, its presence makes repeated evolution twice as likely (Orr 2005). Additionally, positive selection is required to classify repeated evolution as adaptive. The presence or absence of positive selection on the analyzed genes will determine whether the genes have the potential to be responsible for parallel adaptation to a common climatic gradient. Multiple methods are utilized to ascertain the impact of positive selection on sequence evolution. In this chapter I use both intra- and interspecific data to determine the role of positive selection and therefore the potential for parallel adaptation between *D. americana* and *D. melanogaster*.

Together, these analyses will shed light on the circumstances under which evolution is repeated and those conditions under which it is precluded by contingency. Only with data from a variety of natural systems will we be able to understand the forces influencing repeated evolution.

CHAPTER 2

MOBILE MALE-KILLER: SIMILAR *WOLBACHIA* STRAINS KILL MALES OF DIVERGENT *DROSOPHILA* HOSTS[1]

2.1 Abstract

Wolbachia are capable of eliciting a variety of reproductive phenotypes from their hosts, including the production of all-female progeny through embryonic male-killing. To date, phylogenetic analyses indicate six independent acquisitions of the ability to kill male embryos among *Wolbachia* strains infecting insects. Of these six strains, only one appears to have experienced horizontal transmission between host species while maintaining a male-killing phenotype. The rarity of male-killing *Wolbachia* and their disjunct phylogenetic relationships is surprising given the apparently common occurrence of horizontal transfer involving *Wolbachia* strains causing other phenotypes. A male-killing *Wolbachia* strain examined here in *Drosophila borealis* represents a second case of apparent horizontal transmission, based on its close relationship to a male-killing strain in a distantly related *Drosophila* species. The results reported here demonstrate that this *Wolbachia* has maintained a stable phenotype in *D. borealis* over a period of at least 50 years and that a similar strain elicits the same male-killing phenotype in a second *Drosophila* species, indicating that male-killing may be a stable long-term strategy. Sampling bias and/or a lack of suitable hosts are discussed as possible causes of the low frequency of male-killers identified among *Wolbachia* strains.

2.2 Introduction

Insects are hosts to a diverse assemblage of microorganisms, some of which are able to manipulate host reproduction to their own advantage. Perhaps the best studied of

---

[1] This chapter has been published in its entirety. See: Sheeley SL, McAllister BF. (2009) Mobile male-killer: similar *Wolbachia* strains kill males of divergent *Drosophila* hosts. Heredity. 102(3):286-292.

these reproductive parasites is the α-proteobacteria *Wolbachia*. *Wolbachia* inhabit ~66% of insect species as well as some isopods and nematodes (Hilgenboeker et al. 2008). They cause reproductive phenotypes such as cytoplasmic incompatibility between infected and uninfected hosts, feminization of genetic males, parthenogenesis or embryonic male-killing (Stouthamer et al. 1999). In each of these cases manipulation of host reproduction increases the fitness of the maternally transmitted microbe. The evolutionary history of *Wolbachia* is characterized by rampant horizontal transmission among hosts and repeated phenotypic shifts among cytoplasmic incompatability-inducing, feminizing, parthenogenesis-inducing and male-killing strains. Male-killing *Wolbachia* strains are observed only rarely, and this phenotype appears to be independently derived in most cases. Rare origin of this unique phenotype may reflect instability of the male-killing phenotype with strains frequently changing to a different phenotype, an inability of male-killing lineages to successfully sustain infection in novel hosts resulting in strain extinction, or a host-dependence with few host species eliciting a male-killing response.

Among *Wolbachia* phenotype classes, male-killers impose the greatest fitness cost on their hosts due to the embryonic lethality of all male offspring (Charlat et al. 2003). Selection favors suppression of male-killing caused by the bacteria but evidence for suppressors has been observed in only one male-killing infection (Hornett et al. 2006). Theory suggests that maintenance of an equilibrium frequency of infection by a male-killer is possible only if there is an increase in fitness experienced by female offspring of infected female hosts. Benefits of having an infected mother are attributed to reductions in sibling competition, inbreeding, and egg cannibalism (Hurst 1991). In addition, parasite characteristics such as transmission efficiency affect the level of infection within a host population and are influenced by abiotic factors such as temperature. All of these parameters must reach a delicate balance in order to maintain an infection without parasite fixation resulting in complete loss of males and extinction of both host and endosymbiont.

While horizontal transmission of *Wolbachia* is common, evidence of horizontal transmission of a male-killing strain of *Wolbachia* has been found in only one case of two distantly related butterfly hosts (Dyson et al. 2002). The presence of additional clades of male-killing *Wolbachia* would indicate that male-killing can be maintained over long time periods and among different hosts, and that the lack of observed horizontal transmission may be due to both the scarcity of appropriate hosts and biased sampling of male-killing strains. To date, only two strains of male-killing *Wolbachia* have been identified within the genus most widely used as an insect genetic model, *Drosophila*, and phylogenetic evidence shows that the strains have independent origins (Dyer and Jaenike 2004). Incidences of female-biased sex ratios have been reported in several other *Drosophila* species that have yet to be attributed to a specific causal agent. One factor influencing the observed rarity of male-killing *Wolbachia* among *Drosophila* is the inherent difficulty of discovery. Lines that are host to male-killing strains would likely be lost with normal maintenance, so stock center screens to identify the presence of bacterial symbionts, including Wolbachia, fail to identify male-killing strains (Mateos et al. 2006).

Carson (1956) originally described a line of *D. borealis* that produced a female-biased sex ratio. A recent isolate of this species also produced all female progeny in laboratory culture and preliminary screening of this line revealed the presence of *Wolbachia*. Our characterization of this line demonstrates this to be a third male-killing *Wolbachia* infection within the genus *Drosophila*. Analysis of ovarian bacterial flora shows almost exclusively *Wolbachia* and the presence of that infection is associated with a 50% reduction in egg hatch and the loss of male progeny. Phylogenetic analysis of DNA sequences of *wsp* and multilocus sequence typing system (MLST) genes shows that the male-killing strain in *D. borealis* forms a common clade with the male-killing *Wolbachia* strain infecting *Drosophila innubila*. Presence of closely related strains of male-killing *Wolbachia* in two distantly related hosts suggests maintenance of the male-killing phenotype through at least one incidence of horizontal transmission.

2.3 Materials and Methods

A line of *D. borealis* was established with offspring of a single female collected near OshKosh, Wisconsin (44°6.13'N, 88°55.26'W). This female produced 100% female offspring. The line was maintained in the laboratory by outcrossing with males from another line collected at the same locality. Approximately nine generations of outcrossing were performed prior to this study, thus homogenizing the nuclear genetic background in the two lines.

2.3.1 Survey of ovarian bacterial flora

The causal relationship between the presence of *Wolbachia* and the sex ratio trait was investigated by identifying microorganisms in the ovaries of affected flies, followed by antibiotic curing of the infection. Ovarian DNA was isolated from sexually mature virgin female flies from the all-female and normal sex ratio line and previously published general primers were used to amplify eubacterial ribosomal DNA (Weisburg et al. 1991). Amplified DNA was cloned into a TOPO TA vector (Invitrogen, Carlsbad, CA) and clones were sequenced. The resulting sequences were used to design a restriction digestion assay specific for the *Wolbachia* strain present, and additional clones were typed.

2.3.2 Antibiotic curing of sex ratio trait

Female flies from each line were crossed with males from the normal sex ratio line and allowed to oviposit either on standard cornmeal media or media containing 0.025% tetracycline. Resulting female offspring were crossed twice more to normal males, and these two generations (egg to adult) were also reared on each food type. After each generation, offspring were sexed and counted. DNA was extracted from ovaries of treated and untreated females from each line and *Wolbachia*-specific primers for the *wolbachia surface protein* (*wsp*) gene were used to determine *Wolbachia* infection status.

Third generation offspring from the tetracycline-treated all-female line were used to establish a cured line derived from the all-female line.

2.3.3 Analysis of egg viability

In order to differentiate between feminization and male-killing as the cause of sex ratio distortion, a comparison of egg viability was carried out among all-female, normal and cured flies. Female flies from each line were mated *en masse* to normal males in population cages and allowed to oviposit on grape-juice agar supplemented with live yeast. Eggs were collected twice daily from each line and arrayed on fresh grape-juice agar plates. After 48 hours, the proportion of eggs hatched was assayed for each collection by visually inspecting each egg. Egg hatch data was analyzed by carrying out a logistic regression to test the effects of line on egg hatch proportion while controlling for effects of the blocked design created by collecting eggs over several days (PROC GLIMMIX, SAS v9.1). Egg hatch proportion from the cured line was compared pairwise to that of the all-female and normal sex ratio lines.

2.3.4 Phylogenetic analysis of *Wolbachia* sex ratio distorters

To determine the phylogenetic position of the *Wolbachia* strain in *D. borealis*, portions of the *wolbachia surface protein* (*wsp*) and the *Wolbachia* MLST genes (*ftsZ*, *coxA, gatB, hcpA,* and *fbpA*) were amplified from ovarian DNA of the all-female line using previously published primer sequences (Zhou et al. 1998; Baldo et al. 2006). Sequences of the amplified products were determined using routine methods and reference sequences from other strains were retrieved from GenBank. Accession numbers for sequences from other strains are listed in the appendix (Table A1). Sequences were aligned manually. All parsimony analyses were carried out in PAUP*4.0b10 (Swofford 2002). Model selection for Bayesian analysis was carried out using Akaike Information Criterion as implemented in Modeltest 3.6 (Posada and Crandall 1998). Bayesian

posterior probabilities were generated using Mr. Bayes v3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003).

Wolbachia in insects belong to two deeply divergent clades. Sequences of *wsp* were used to determine the phylogenetic position of the strain infecting *D. borealis* relative to previously described strains. This is the fastest evolving protein known in *Wolbachia*, and has been used extensively for phylogenetic analysis (Zhou et al. 1998). Sequences from both insect-hosted clades were included and sequences of two strains from a distantly-related clade from nematode hosts were used to root the phylogeny. All known male-killing strains were included in the analysis. Parsimony criteria were used to reconstruct a phylogeny with 511bp of sequence from 39 strains of *Wolbachia* using the default heuristic settings in PAUP* 4.0b10 (Swofford 2002). Confidence in nodes was determined by 10,000 bootstrap replicates of the maximum parsimony analysis and by Bayesian posterior probabilities using a GTR+Γ+I model. Posterior probabilities were generated from 300,000 Markov chain Monte Carlo generations with 75,000 discarded as burn in.

A more robust analysis of the phylogenetic position of the strain in *D. borealis* relative to highly similar strains was conducted using *wsp* (460 bp) and MLST genes: *ftsZ* (435 bp), *gatB* (368 bp), *coxA* (403 bp), *hcpA* (444 bp), and *fbpA* (428 bp). An exhaustive parsimony search was performed. Confidence in nodes was determined by 10,000 bootstrap replicates of the maximum parsimony analysis and by Bayesian posterior probabilities using a GTR model. Posterior probabilities were generated from 200,000 Markov chain Monte Carlo generations with 50,000 discarded as burn in.

A multilocus sequence typing (MLST) system has been developed specifically for *Wolbachia* as a complement to phylogenetic analysis (Baldo et al. 2006). The MLST system consists of a database of allelic profiles of five slowly-evolving loci. Each distinct allele at each locus is given a numeric designation, so that each strain has a five-number allelic profile, and each distinct allelic profile is designated as a unique sequence type.

Only sequences that are identical for a given gene will share the same number in their allelic profile for that locus, and only strains whose sequence is identical at all five loci will share the same sequence type. Strains with identical alleles of 3 or more of the five MLST loci share a sequence type complex. In addition, a second database catalogs the amino acid sequences of four hyper-variable regions of *wsp*. This allows detection of allele shuffling by recombination and comparison of the congruence of *wsp* and the MLST loci. Searches on the MLST database were conducted through the website (http://pubmlst.org/wolbachia/) developed by Jolley et al. (2004.)

Inference about the origins of infection among hosts relies on a comparison of their phylogenetic relationships in conjunction with those of the endosymbionts. Given the inferred relationships among the major species group in the genus Drosophila, classification of *D. borealis* and *D. innubila* respectively in the virilis and quinaria species groups indicates a distant relationship between these host species. Specific resolution of their phylogenetic relationship was obtained with available sequence data analyzed within the framework of the reference genome sequences of Drosophila species. Sequences of mitochondrial genes COI and COII of *D. borealis* and *D. innubila* and *D. recens*, and two and three regions of nuclear genes of *D. recens* and *D. innubila*, respectively, were aligned with the corresponding regions from the reference genome sequences of nine Drosophila species. Over 20 kb of additional sequence from nine reference species contained 14 genes that yielded method-independent inference of the consensus topology among 12 species of Drosophila in the analysis of Rasmussen and Kellis (2007). A table listing the data included in this supermatrix is included in the appendix (Table A2). Overall tree topology and branch lengths were obtained using Maximum Likelihood in PAUP under the GTR+Γ+I model, with posterior probabilities for each node estimated in Mr. Bayes.

2.4 Results

2.4.1 Survey of ovarian bacterial flora

A *Wolbachia* infection causing this sex-ratio trait would be expected to be prevalent in the ovaries of affected females. Preliminary sequencing of eight bacterial ribosomal DNA clones from ovarian DNA of the all-female line revealed mostly *Wolbachia* sequences (7 of 8 sequences). The single unique sequence showed significant BLAST to uncultured bacterial strains. The presence of this bacterium was attributed to contamination from gut flora during ovary dissection. A restriction digest assay was designed to differentiate the sequence of the *Wolbachia* strain from ribosomal DNA of a variety of other microorganisms, including the uncultured bacterium and all known types of reproductive parasites. Of 54 additional clones screened, all tested positive as *Wolbachia*. All attempts to amplify bacterial DNA failed with ovarian DNA from the normal sex ratio line.

2.4.2 Antibiotic curing of sex ratio trait

Further correlation of the sex ratio trait with *Wolbachia* infection was provided by their simultaneous loss following treatment with antibiotics. After rearing flies for a single generation on medium containing tetracycline, the normal sex ratio line remained ~50% male, while the all-female line produced no males. The all-female line responded to tetracycline treatment after a second generation of culture in the presence of tetracycline, producing an approximately 1:1 sex ratio. This sex ratio remained stable after a third generation of antibiotic treatment. Non-tetracycline-treated controls continued to produce only female offspring and sex ratio remained at 1:1 in the normal sex ratio line regardless of treatment for the duration of the experiment (Table 2.1). Failure to amplify *Wolbachia* DNA corresponded with loss of the sex ratio trait.

2.4.3 Analysis of egg viability and Y chromosome inheritance

The mode of male loss was ascertained by comparing the egg hatch of the all-female line with that of controls. Male-killing will result in a 50% reduction in hatch, assuming a 1:1 primary sex ratio. Eight collections of 160 eggs per line were made (1280 eggs/line). Egg hatch proportions for normal and cured lines were 0.77±0.02 and 0.78±0.02, while egg hatch proportion for the all-female line was 0.38±0.02 (Figure 2.1). A logistic regression was fitted to the data and showed a significant effect of line on egg hatch ($p<0.0001$). Post hoc pair-wise comparisons showed a significant difference between hatch proportion of all-female and cured lines ($p<0.0001$), but not between normal and cured lines. The approximately 50% reduction in egg hatch in the all-female line is indicative of specific lethality of males. Further indication of male lethality is provided by failure to amplify the Y-chromosome gene *KL-2* from DNA of the all-female line as well as DNA from females of the normal sex ratio line. Amplification was successful with DNA from males of the normal sex-ratio line (data not shown).

2.4.4 Phylogenetic analysis of Wolbachia sex ratio distorters

Horizontal transmission of *Wolbachia* is evidenced by the presence of identical or similar strains in divergent hosts. Phylogenetic incongruence between hosts and endosymbionts indicates independent infection of the two hosts by the same strain, rather than inheritance from a common ancestor. Within 511 bp of aligned *wsp* sequence, only two nucleotide differences are present between the male-killing *Wolbachia* strains infecting *D. borealis* and *D. innubila*. The most parsimonious reconstruction showed the *D. borealis* strain as sister to the male-killing strain of *Wolbachia* infecting *D. innubila*, but owing to the minimal divergence in these short sequences, bootstrap support for the node connecting the *D. borealis* and *D. innubila* strains is low (Figure 2.2). The two

male-killers grouped with five strains exhibiting minimal divergence and mostly present among several divergent clades of *Drosophila.* Consistent with previous results (Dyer and Jaenike 2004), the third male-killing strain of *Wolbachia* in *D. bifasciata* grouped separately from that of *D. borealis* and *D. innubila*. The previously recognized 'A,' 'B' and 'C' clades of *Wolbachia* were resolved with high confidence.

The *D. borealis* male-killer and five other closely related strains were examined in a second phylogenetic analysis. The genes used in this analysis were chosen because they are part of the MLST system for *Wolbachia* and sequences are available for many strains (Baldo et al. 2006). Typically, the MLST genes are not combined with *wsp* sequences in phylogenetic analyses because *wsp* is known to recombine frequently among strains (Baldo et al. 2005; Baldo et al. 2007), leading to incongruence in combined datasets. The strains investigated here, though, are so similar that their relationships have not been influenced by recombination at *wsp.* A partition homogeneity test confirmed that the genes show no incongruence, both among all genes and between *wsp* and the five MLST genes ($p = 1.0$). Although the taxa for this analysis were chosen based on their similarity to the male-killing strain at *wsp,* those strains are also the most closely related from the MLST database, confirming that within this group, *wsp* and the MLST genes share a similar evolutionary history. This analysis consisted of 2539 characters and confirmed the sister relationship between the *D. borealis* and *D. innubila* male-killers with strong support for a common node (Figure 2.3 A). Approximately 3500 bp of sequence was obtained in total from the newly-characterized male-killing strain, and all sequence except that of *wsp* is identical in the two male-killing strains, including *ftsZ* (614 bp), *gatB* (368 bp), *coxA* (403 bp), *hcpA* (444 bp), *fbpA* (428 bp), and 16s rDNA (694 bp).

The male-killing strains infecting *D. borealis* and *D. innubila* have the same MLST allelic profile, and therefore the same sequence type (10) and sequence type complex (STC-13), but differ by one nonsynonymous change in two of the four

hypervariable regions of *wsp*. The MLST database includes another strain (id 83) with an identical MLST allelic profile that only differs from the strain infecting *D. borealis* by a single nonsynonymous change in one hypervariable region (Table 2.2). This strain, identified in *Drosophila munda,*was amplified from a single specimen of unknown sex, and has an unknown affect on the host species (Julie Stahlhut, personal communication). This strain was not included in our phylogenetic analysis, as we cannot draw any conclusions with regard to the evolutionary history of the male-killing phenotype without additional information. Other strains with phenotypic information in the database are more distantly related to the clade of male-killing strains shared between *D. borealis* and *D. innubila*.

The two host species, *D. borealis* and *D. innubila* belong to widely divergent species groups, the virilis group and the quinaria group, respectively. Phylogenetic analysis of these species with ten additional *Drosophila* species, nine of which have complete genome sequences confirmed that *D. borealis* and *D. innubila* are deeply divergent and belong to lineages which diverged prior to the colonization of the Hawaiian archipelago by *Drosophila* (Figure 2.3 B), an event estimated to have occurred 25 to 42mya (Desalle 1992; Powell and Desalle 1995; Tamura et al. 2004).

2.5 Discussion

Females of the all-female line harbor almost exclusively *Wolbachia* in their ovaries and the infection and sex-ratio trait are eliminated upon antibiotic treatment. Since the conclusion of this experiment, the cured line has been maintained over 20 generations, and the sex ratio remains at 50:50. Egg viability is reduced by 50% in the all female line compared to control and cured lines, consistent with complete lethality of male embryos, and there is no evidence of feminization of genetic males as investigated by attempted amplification of a Y-linked gene. Together, these results indicate that the *Wolbachia* strain infecting *D. borealis* is an embryonic male-killer.

Phylogenetic analysis shows that this strain of male-killing *Wolbachia* is very closely related to another male-killing strain infecting *D. innubila*, a species that shared a common ancestor with *D. borealis* 25-42 mya. This close relationship between the parasites of two distantly related hosts is a hallmark of horizontal transmission and may reflect a common clade of *Wolbachia* that has established and maintained a male-killing phenotype in more than one host. Although the available data indicate a similar strain of Wolbachia causes male-killing in these divergent hosts, insufficient sampling and characterization of phenotypic effects in host species infected with this Wolbachia clade may generate a spurious association between these male-killing strains. Switching between phenotypes may occur on such a rapid time-scale that male-killing in *D. borealis* and *D. innubila* results from convergence rather than it being the ancestral mode of transmission of this Wolbachia clade. The clade also infects *D. munda*, but the absence of phenotypic analysis of the infection does not help resolve the persistence of the male-killing phenotype. Further sampling coupled with phenotypic analysis is needed to determine if male-killing is ancestral to this clade and if other phenotypes arise as a consequence of host suppression. This has been observed among populations of a single host species of another male-killer, infecting *Hypolimnas bolina*. In this case, suppression of male-killing in some populations results in expression of a cytoplasmic incompatibility phenotype (Hornett et al. 2008).

Interestingly, the two distantly-related hosts of male-killing strains, *D. borealis* and *D. innubila*, do not share any portion of their natural ranges. *D. borealis* is found in the upper Midwest and Canada, while *D. innubila* is found solely on mountains known as sky islands in the desert Southwest and Mexico. Parasitoid wasps are capable of transferring *Wolbachia* between distantly-related host species (Heath et al. 1999) and parasitoids often harbor *Wolbachia* infections that appear closely related to those of their hosts at the *wsp* locus (Vavre et al. 1999), but our limited knowledge of the ecology of these species precludes identification of the intermediate transmission vector. Gaps in

infection over geographic and phylogenetic space suggest that only certain species may be adequate for sustained infection.

This result, coupled with the presence of a male-killing clade in butterflies (Dyson et al. 2002) indicates that male-killing may be a sustainable strategy in some *Wolbachia* clades and that the male-killing phenotype can be maintained during transmission among different hosts. This has previously been investigated at the intraspecific level, where experimental transfer of the male-killing *Wolbachia* infecting *Drosophila bifasciata* into lines established from historically uninfected populations resulted in successful embryonic killing (Veneti et al. 2004). Recent analysis of another male-killing microorganism, *Spiroplasma*, indicates a reduction in male-killing efficiency with increased divergence between the original and novel hosts (Tinsley and Majerus 2007). Although the biology of *Spiroplasma* and *Wolbachia* male-killers may be very different, the authors observed highly successful male-killing after intragenus transfer, consistent with our finding that this strain of *Wolbachia* is a highly efficient male-killer in both *D. borealis* and *D. innubila*. This male-killing strain appears to be capable of maintaining a male-killing phenotype among hosts and for some time within hosts (evidence from *D. borealis* indicates at least 50 years), so sampling bias and the lack of suitable hosts may explain the rarely discovered examples of male-killing *Wolbachia*, rather than an inherent instability of the phenotype. Characterization of a third male-killing *Wolbachia* infection in *Drosophila* will allow investigation of requirements for sustained infection and allow powerful interspecific comparisons for investigating the mechanism of male-killing. The ability of this strain to selectively kill male embryos in two divergent species indicates that it may be possible for it to successfully infect and kill male embryos of *D. melanogaster*, potentially unleashing its extensive genetic toolkit to identify the mechanism of male killing.

Table 2.1 Sex ratio (males/females) of offspring after tetracycline (Tet) treatment in the normal sex ratio (wt) and all-female (♀) line.

	Generation 1		Generation 2		Generation 3	
Tet	-	+	-	+	-	+
wt	0.48 ± 0.03	0.54 ± 0.03	0.53 ± 0.07	0.52 ± 0.04	0.46 ± 0.05	0.51 ± 0.04
♀	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.53 ± 0.12	0.00 ± 0.00	0.50 ± 0.04

Table 2.2 *Wsp* and MLST allelic profiles of *D. borealis* and the two most closely-related isolates from the *Wolbachia* MLST database.

	Wsp HVR				MLST						
Host	1	2	3	4	*gatB*	*coxA*	*hcpA*	*ftsZ*	*fbpA*	ST	ST-C
D. borealis	1	unique	13	70	1	1	1	3	2	10	13
D. munda	1	10	13	70	1	1	1	3	2	10	13
D. innubila	1	10	13	10	1	1	1	3	2	10	13

Note: HVR, hypervariable region; ST, sequence type; ST-C, sequence type complex.

Figure 2.1Egg hatch proportion for control, cured and all-female lines. Dashed line indicates 50% of the average hatch of the control and cured lines. *p < 0.0001

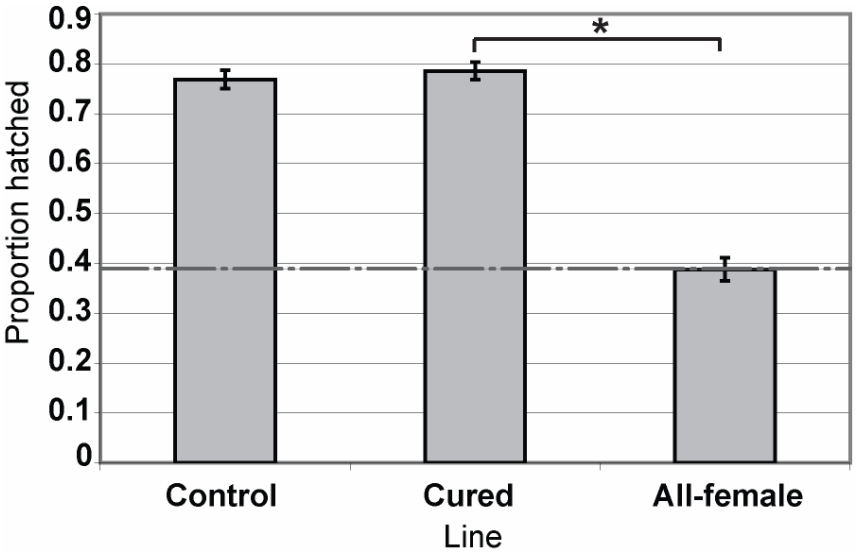


Figure 2.2 Most parsimonious reconstruction of 39 *Wolbachia wsp* sequences. Bootstrap values are shown for 10,000 parsimony replicates (top) as well as Bayesian posterior probabilities (bottom). Male-killing lineages are shown in bold, and the major divisions of the *Wolbachia* clades are shown as 'A,' 'B,' and 'C'.

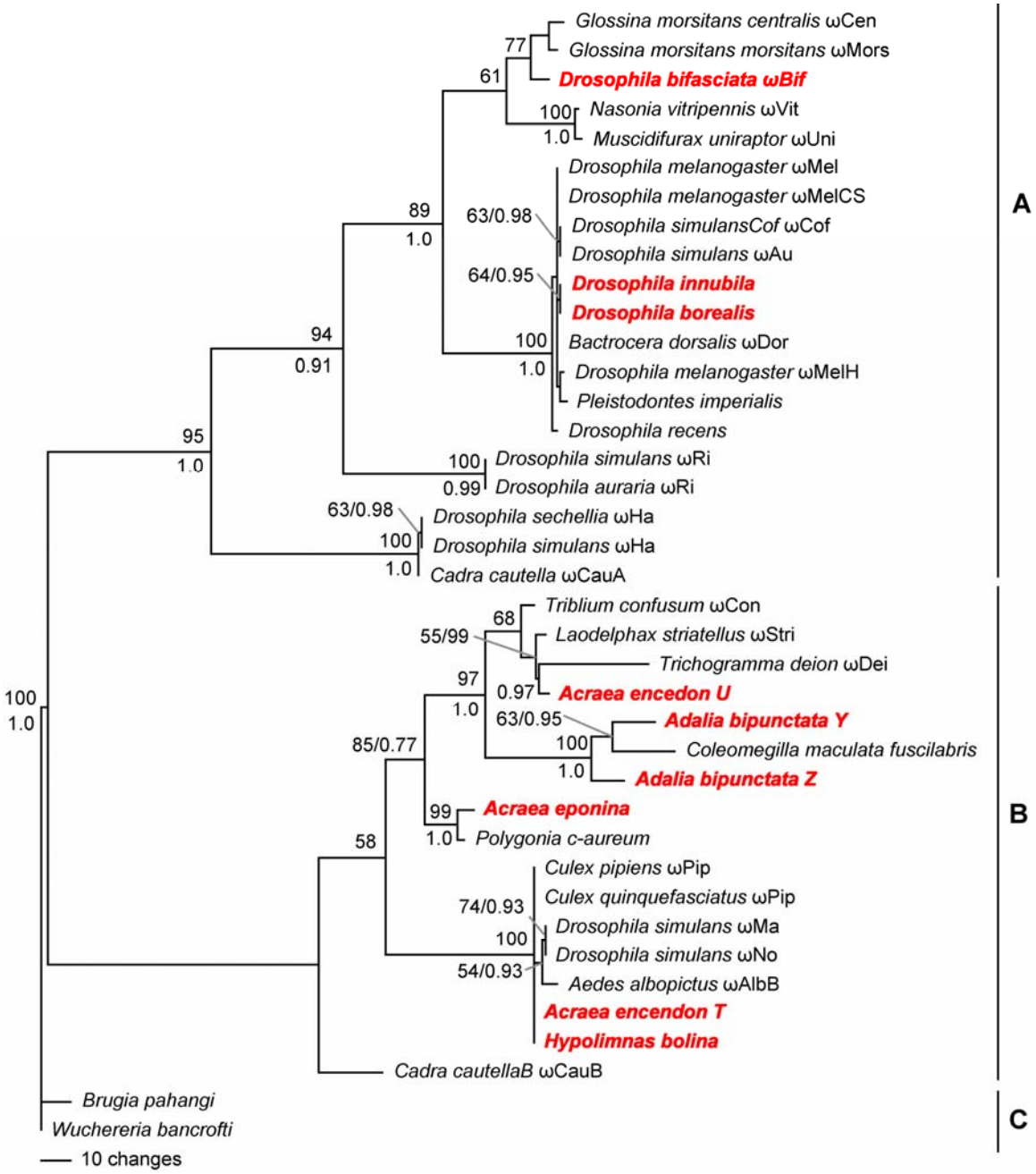
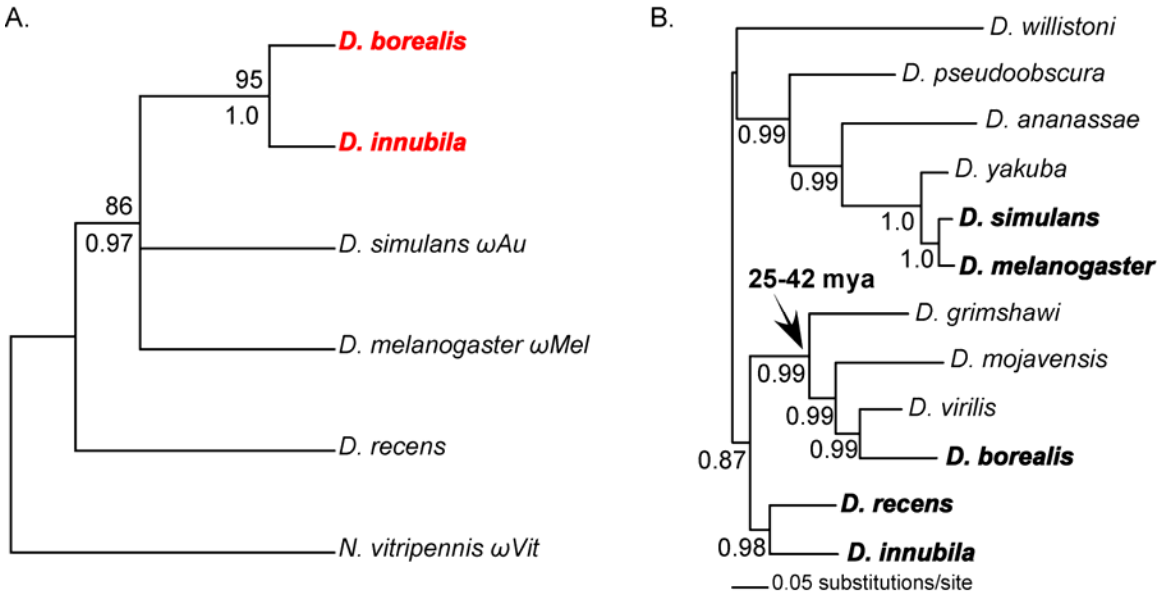


Figure 2.3 Phylogenetic analyses of *Wolbachia* and their hosts. A. Most parsimonious reconstruction of sequences representing the *Wolbachia* strains most closely related to the *D. borealis* male-killing strain. For each strain, partial sequences from five genes were used: *wsp* (460 bp), *ftsZ* (435 bp), *gatB* (368 bp), *coxA* (403 bp), *hcpA* (444 bp), and *fbpA* (429 bp). Bootstrap values for 10,000 parsimony replicates (top) and Bayesian posterior probabilities (bottom) are shown. Male-killing lineages are shown in bold. B. Phylogenetic reconstruction of representative *Drosophila* species. Those species that are host to *Wolbachia* strains included in 'A' are shown in bold. Bayesian posterior probabilities are indicated at each node.



CHAPTER 3

PATTERNS OF NATURAL SELECTION AT THE *ALCOHOL DEHYDROGENASE* LOCUS OF *DROSOPHILA AMERICANA*[2]

3.1 Abstract

Similar outcomes are often observed in species exposed to similar selective regimes, but it is unclear how often the same mechanism of adaptive evolution is followed. Here we present an analysis of selection affecting sequence variation in the *Alcohol dehydrogenase* (*Adh*) gene of *D. americana*, a species endemic to a large climate range that has been colonized by *D. melanogaster*. Unlike *D. melanogaster*, there is no evidence of selection on allozymes of ADH across the sampled range. This indicates that if there has been a similar adaptive response to climate in *D. americana*, it is not within the coding region of *Adh.* Instead, analyses of a combined dataset containing 86 alleles of *Adh* reveal purifying selection on the *Adh* gene, especially within its intron sequences. Frequency spectra of derived unpreferred variants at synonymous sites indicate that these sites are affected by weak purifying selection, but the deviation from neutrality is less drastic than observed for derived variants in noncoding introns. This contrast further supports the notion that noncoding sites in *Drosophila* are often subject to stronger selection pressures than synonymous sites.

3.2 Introduction

Selection can generate convergent evolution of phenotype, but the underlying genetic mechanism is unclear. Traits may respond to similar selection pressures through similar mutational paths, unique variants within the same genes, changes within the same genetic pathways or through entirely different mechanisms. Similar adaptive genetic

---

[2] This chapter has been published in its entirety. See: Sheeley SL and McAllister BF. (2008). Patterns of Natural Selection at the *Alcohol dehyrogenase* gene of *Drosophila americana*. Fly 2(5):242-243.

changes have been documented in a wide range of taxa exposed to the same selective pressures. For a review see Wood (2005). The genetic changes responsible for phenotypic change in these lineages are often within the same gene, and even due to identical changes in amino acid sequence. For instance, convergence of genetic change among grasses that have independently gained the ability to use C_4 photosynthesis is evidenced by positive selection for the same or similar substitutions at 21 amino acids in phosphoenolpyruvate carboxylase (Christin et al. 2007).

The *Alcohol dehydrogenase* (*Adh*) gene/enzyme system of *Drosophila* offers an ideal system for examining patterns of adaptive evolution among lineages. ADH activity and alcohol content of natural larval substrates are positively correlated across species (Mercot et al. 1994), indicating that selection for differences in alcohol metabolism may influence patterns of sequence divergence at the *Adh* locus in many lineages. Analysis of *Adh* of *Drosophila melanogaster* indicates a potential for this gene to respond to differential selection in clinally varying environments by selectively maintaining coding and noncoding polymorphisms affecting ADH quantity and catalytic activity. The well-known Fast/Slow (F/S) allozyme polymorphism is distributed latitudinally on several continents (Oakeshott et al. 1982), and DNA sequences indicate that these independent clines are maintained by balancing selection (Kreitman & Hudson 1991; Berry & Kreitman 1993). The allozyme cline is accompanied on several continents by an intronic sequence polymorphism that influences protein quantity and whose variants are strongly associated with the F/S polymorphism (Laurie et al. 1991). Together, the F/S and intronic polymorphisms cause a 2.5 to 3-fold difference in total ADH activity. Fluctuations in F/S allozyme frequencies are associated with temperature changes (Kamping and van Delden 1999) and the allozyme cline has recently shifted in wild populations in Australia, mirroring climate change (Umina et al. 2005).

In addition to selection on amino acid variants and intron sequence substitutions in *D. melanogaster*, purifying selection on codon usage is evident in many species. A

survey of the *D. melanogaster* genes identified *Adh* as having highly biased codon usage and manipulation of synonymous sites *in vivo* decreased ethanol tolerance (Duret & Mouchiroud 1999; Carlini & Stephan 2003; Carlini 2004). Codon bias at *Adh* is also high in *D. pseudoobscura* and *D. simulans* (Akashi and Schaeffer 1997) but appears relatively low in the virilis species group (Nurminsky et al. 1996).

Here we investigate the possibility of similar adaptive genotypic evolution at the *Adh* locus in two species with a similar latitudinal range in North america, *D. melanogaster* and *D. americana*, an endemic North American species with a distribution that ranges from Louisiana to Wisconsin. Selection apparently maintains a chromosomal polymorphism characterized by a derived fusion of the X and 4th chromosomes (Muller elements A and B) across the species' latitudinal range, with the fusion nearly fixed in northern populations and absent in the south (McAllister et al. 2008). Unstructured variation in putatively neutral markers on the 4th chromosome, including four RFLPs within *Adh,* demonstrates the maintenance of the latitudinal cline for the chromosomal polymorphism despite high levels of gene flow (McAllister 2002). The *Adh* gene is located at the base of Chromosome 4 so it potentially contributes to the cline. Structurally, *Adh* of *D. americana* closely resembles that of other *Drosophila,* with three exons and two small introns. *Adh* of the closely related *D. virilis* shows an expression pattern similar to that of *D. melanogaster,* with distinct larval and adult transcripts originating from proximal and distal promoters (Nurminsky et al. 1996). The range of *D. americana* is similar to the North American portion of the range of *D. melanogaster* and the *Adh* locus provides an opportunity to compare and contrast the selective forces acting on homologous genes in two lineages where they inhabit the same latitudinal range.

In addition to the centromeric fusion polymorphism, segregating inversions are also present on Chromosome 4. The *Adh* gene of *D. americana* is located proximal to a polymorphic nested inversion complex, *In(4)ab,* that is completely associated with the X-4 fusion. *Adh* shows differentiation between inverted and standard chromosomes

(McAllister 2003; Evans et al. 2007), so only *Adh* sequences from unfused chromosomes were included in this analysis to ensure that observed patterns of variation were independent of the chromosomal polymorphisms. Sequences of the *Adh* coding region and introns from 86 unfused 4th chromosomes of *D. americana* from eight wild populations encompassing much of the latitudinal range of the species were examined. These sequences have been accumulated in studies investigating the evolutionary consequences of polymorphic chromosomal rearrangements of the X and 4th chromosomes (McAllister & Charlesworth 1999; McAllister 2003; McAllister & Evans 2006). This large data set provides statistical power to reveal small deviations from neutral expectations within the *Adh* coding region and intervening sequence.

3.3 Materials and Methods

Aligned sequences encompassed an 834-bp region consisting of nearly the entire coding region and including two small introns and 7 bp of the 5' UTR. Sequences of two closely-related species, *D. virilis* and *D. lummei* were used as outgroups. Collection localities and sequence accession numbers are provided in Table 3.1. All SNPs segregating at 10% or greater were tested for latitudinal association using Spearman rank correlations. Frequency spectra of preferred and unpreferred synonymous mutations were compared using a Mann-Whitney U test and the frequency spectrum of derived unpreferred variants was used to estimate selection on synonymous codons (Akashi and Schaeffer 1997). Selection on intronic sites was similarly estimated from the frequency spectrum of derived variants. The region was inspected for significant linkage disequilibrium (LD) among sites using LD plot (Schaeffer et al. 2003). Sequence diversity and heterozygosity were calculated and patterns of variation were examined using Tajima's *D* (Tajima 1989) and Fu and Li's *D* (Fu & Li 1993) as implemented in DNAsp (Rozas et al. 2003) to compare patterns of polymorphism with neutral expectations. Fay and Wu's *H* (Fay & Wu 2000) was calculated to check for an excess of

high frequency derived polymorphism. The McDonald-Kreitman test (MK, McDonald & Kreitman 1991) was carried out using divergence from *Adh* sequence of *D. lummei* to test for non-neutral polymorphism/divergence ratios. Sequences from *D. virilis* were used to perform the HKA (Hudson et al. 1987) test for non-neutral levels of polymorphism at *Adh* compared to a control region, *bib*.

3.4 Results and Discussion

Polymorphism among the *Adh* sequences from *D. americana* consists of 37 segregating synonymous sites, 3 nonsynonymous sites and 15 noncoding sites. Thirteen of these segregating sites were observed at ≥10%, but none are significantly associated with latitude. Unlike *D. melanogaster*, there is no evidence for clinally selected variants in the coding and intronic sequences. Frequency spectra of preferred and unpreferred synonymous mutations differ ($U = 49.00$, $p < 0.05$) and the frequency spectrum of derived unpreferred mutation indicates a scaled selection coefficient ($\gamma = 2N_e s$) of -1.12 against these polymorphisms. Our estimate of selection on synonymous codon usage is consistent with other genes in *D. americana* (Maside et al. 2004). Significant linkage disequilibrium (LD) is apparent only in two groups of rare variants within the second exon. All of the variants exhibiting significant LD are segregating at <20% in the sampled sequences, with the majority <10%, suggesting drift as a cause. Absence of significant LD between pairs of distant sites indicates a lack of segregating co-adapted variants.

Fu and Li's *D* is significantly negative for the entire gene, while Tajima's *D* does not differ from neutral expectations. When coding and noncoding regions were tested separately, both statistics showed negative but insignificant *D*-values in coding regions, and significantly negative values in noncoding, primarily intron, regions ($p < 0.05$, Table 3.2). When the same tests were carried out in a sliding window, they revealed significantly negative values of Tajima's *D* in both introns and the second exon and

significantly negative values of Fu and Li's *D* in the second exon and second intron (Figure 3.1). The frequency distribution of intronic polymorphism reveals about 3-fold stronger purifying selection against derived variants in introns than against derived unpreferred mutations at synonymous sites (γ = -3.32 vs -1.12, Figure 3.2). Fay and Wu's *H*, MK and HKA tests showed no deviation from neutral expectations for the *Adh* region.

These results indicate that polymorphisms within *Adh* of *D. americana* are influenced primarily by weak purifying selection. In particular, the second exon and the two small introns have low neutral mutation rates, as evidenced by significantly negative Tajima's *D* and Fu and Li's *D* values indicating purifying selection. Purifying selection is also evidenced by the lack of divergence in introns. No fixed differences are observed within the 62-bp first intron in comparisons of *D. americana* and *D. lummei*, and only a single derived substitution is apparent in each lineage within the 63-bp second intron, indicating almost complete conservation of an ancestral intron sequence. Purifying selection is most pronounced near the border between the second exon and the second intron and estimates of selection based on the frequency spectrum of polymorphisms reveal stronger purifying selection against derived variants within the two introns than for derived unpreferred variants at synonymous sites.

Although portions of *Adh* have low frequency variation, a non-significant HKA test is consistent with overall polymorphism being proportional to divergence at *Adh*. Furthermore, Fay and Wu's *H* does not show an excess of high frequency derived variants as expected under a hitchhiking scenario. The very low frequency of variants observed in the introns is unexpected for a region where mutations are often assumed to be neutral. In fact, the neutral mutation rate appears extremely low in both intron sequences of the *Adh* gene. This finding, coupled with the finding that synonymous sites are under weak purifying selection indicates that no particular class of sites is evolving neutrally. Similar results have been reported in *D. melanogaster,* where a survey of coding and noncoding regions on the X chromosome showed higher selective constraint

for introns than for synonymous sites (Andolfatto 2005). Pre-mRNA 2° structures could contribute to non-neutral evolution of introns, and both functional analyses of mutants and investigations of linkage disequilibrium in *Adh* introns of other *Drosophila* species indicate compensatory changes that may be involved in 2° structure formation (Kirby et al. 1995; Chen & Stephan 2003; Matzkin 2004). Nevertheless, the lack of LD in the introns of *D. americana* suggests that there are no segregating variants representing compensatory changes in sequence to maintain polymorphism for such a structure. The small size of the introns may be one reason for their low variation, but the depression in noncoding variation has not been reported for other species studied to date, although small introns are conserved across the genus. This survey indicates that for at least some gene regions, although not strictly neutral, synonymous sites more closely fit neutral expectations than noncoding sites. By building upon previous studies of *Adh*, the gene region of *D. americana* could be utilized as a model for understanding the constraints on intron sequences.

The lack of latitudinally structured polymorphism indicates that unlike *D. melanogaster,* the coding region and intervening sequence of *Adh* in *D. americana* does not appear to be a target of local adaptation. Extensive population genetic analysis of *Adh* in *D. pseudoobscura* has also revealed an almost monomorphic protein whose sequence is subject to strong purifying selection (Schaeffer and Miller 1992). While the adaptive response to temperature at the *Adh* locus in *D. melanogaster* appears to be novel, *D. americana* and *D. pseudoobscura* could achieve a similar response to temperature through variation in regulatory regions controlling *Adh* or in other genes contributing to ethanol catabolism or involved in the flux of ethanol into lipid synthesis (Freriksen et al. 1991). For example, variation in ethanol tolerance in *D. melanogaster* is also attributed to sequence changes in *Aldehyde dehydrogenase* (Fry et al. 2008), to regulatory changes in *acetyl-CoA synthetase* (Montooth et al. 2006), and to interactions between temperature and membrane fluidity (Montooth et al. 2006). It is noteworthy that although the climatic

range inhabited by *D. americana* and *D. melanogaster* is similar in North America, their feeding habits are quite different. *D. melanogaster* feeds on rotting fruit substrates high in alcohol, while *D. americana* is a saprophytic species and may encounter lower alcohol levels, so perhaps the two species do not experience similar variation in ethanol exposure over their range.

This investigation of the *Adh* locus in *D. americana* did not reveal any similarities to *Adh* of *D. melanogaster* that could not be attributed to conservation of ancestral function. Although there may be selection on *D. americana* for differential alcohol metabolism or flux to lipid synthesis throughout their range, there is no evidence for environmentally structured variation within the coding or intervening sequence of *Adh*. The targets of selection may be similar to others that have been associated with adaptation for differential alchohol metabolism in *D. melanogaster*, for instance *Aldehyde dehydrogenase* or *acetyl Co-A synthetase*, or *D. americana* may cope with this environmental gradient in an entirely different manner. Genetically similar mechanisms of adaptive changes have been documented among many lineages, and while it is tempting to extend observations of adaptive evolution to related taxa, especially among lineages that are closely related, the assignment of loci responsible for adaptation necessitates species-specific investigation. It remains to be seen how often similar genetic changes underlie adaptive phenotypic convergence. Resolution of this question requires further investigation as well as consistent reporting of positive as well as negative results.

Table 3.1 Collection localities and sequence accession numbers

Locality	N	Latitude	Longitude	Accession Nos.
NN	8	42° 44.9' N	98° 02.6' W	AY340300-AY340307
IR	10	41° 46.7' N	91° 42.9' W	EF136999-EF137008
G	19	41° 33.0' N	87° 22.0' W	AF136680-AF136698
SB	10	41° 29.7' N	91° 09.7' W	EF136989-EF136998
DN	7	41° 22.1' N	97° 29.7' W	AY340293-AY340299
HI	10	38° 39.7' N	90° 40.7' W	EF136979-EF136988
FP	12	34° 11.5' N	91° 04.5' W	EF136969-EF136978, EU99942, EU99943
LP	10	32° 42.7'N	94° 40.0' W	AF136699-AF136708
D. lummei	n/a	n/a	n/a	U26843
D. virilis	n/a	n/a	n/a	U26846

Table 3.2 Measures of sequence variation at *Adh.*

Region	Sites	π^a	θ^b	$D_T{}^c$	$D_{F\&L}{}^d$	$H_{F\&W}{}^e$
Coding	706	0.0078	0.0118	-1.099	-1.861	-0.78
Noncoding	128	0.0056	0.0247	-2.184*	-2.548*	-0.785
Total	834	0.0076	0.0138	-1.467	-2.448*	-1.565

Note: a, pairwise diversity; b, heterozygosity; c, Tajima's D; d, Fu and Li's D; e, Fay and Wu's H; *$p < 0.05$

Figure 3.1 Analyses of sequence variation across the Adh locus. Tajima's *D*, Fu & Li's *D* and Fay & Wu's H statistics were calculated using the total number of mutations in a sliding window of 50 bp and a step size of 10 bp. Significance levels are shown for individual tests without Bonferroni correction. * $p < 0.05$; ** $p < 0.02$

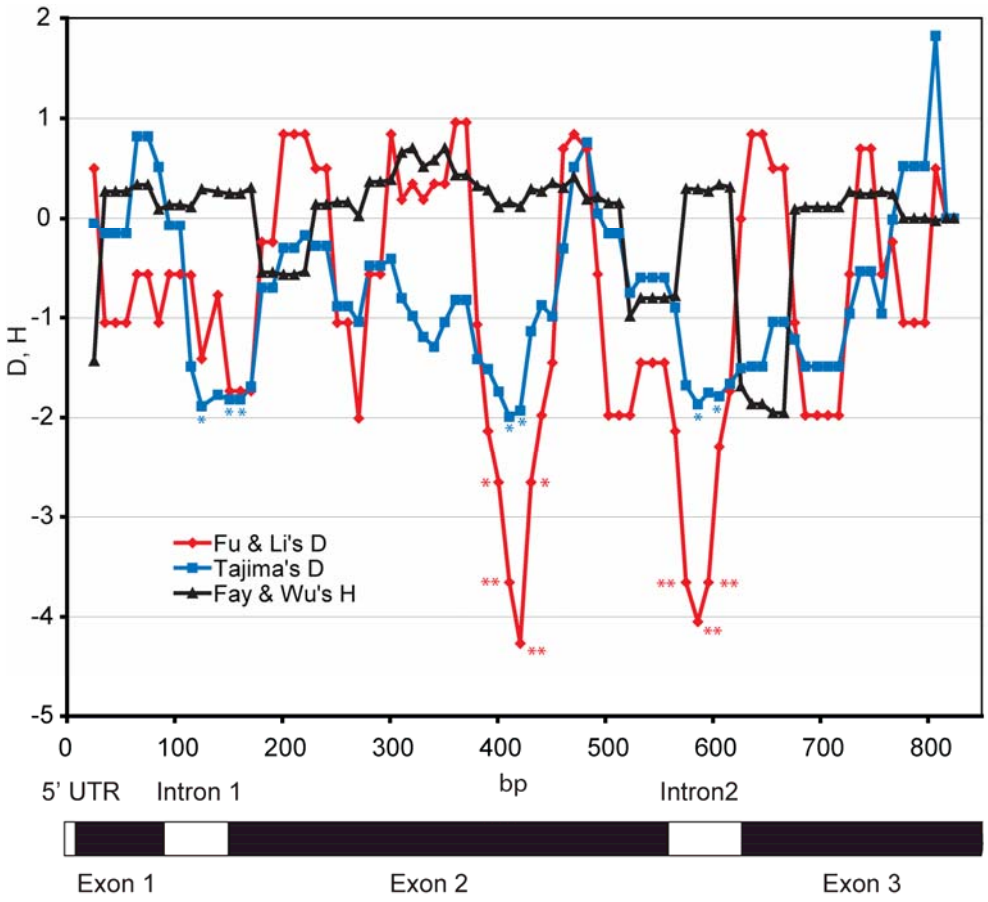
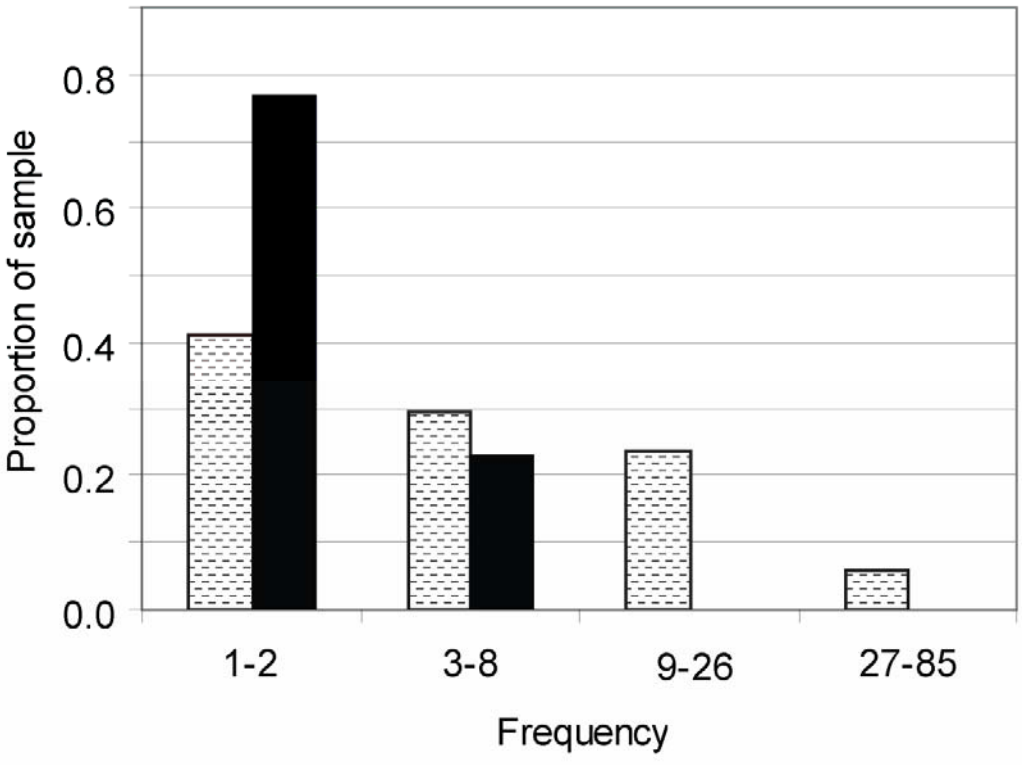


Figure 3.2 Frequency distributions of derived mutations to unpreferred codons at synonymous sites (gray) and derived mutations in introns (black).



CHAPTER 4

AN INVESTIGATION OF VARIATION AT THE *PHOSPHOGLUCOMUTASE* LOCUS OF *DROSOPHILA AMERICANA*

4.1 Abstract

The role of polymorphism in adaptation has been debated ever since the discovery of large amounts of segregating variation within species. It is a continuing challenge to accurately identify those variants that are maintained by selection as opposed to those that segregate due to neutral processes. The *Phosphoglucomutase* (*Pgm*) locus is highly variable in many taxa and allelic differences may affect flux between energy storage and utilization, potentially influencing fitness in varied environments. The extent to which variation at *Pgm* is influenced by selection for locally adapted alleles is unknown in most cases. The most complete analysis of the contribution of variation at *Pgm* to local adaptation to date is in *Drosophila melanogaster*, where kinetically distinct alleles of PGM segregate clinally with latitude. Balancing selection like that proposed for *D. melanogaster* may be characteristic of the *Pgm* locus. Here I present a comprehensive examination of the selective forces acting on variation at *Pgm* of *Drosophila americana*, an endemic species whose range largely overlaps with the recently colonized North American range of *D. melanogaster*. Several aspects of variation at *Pgm* of *D. americana* are consistent with segregation of locally adapted alleles, including excess amino acid polymorphism and correlation of allele frequencies at some noncoding sites with latitude and longitude. In addition, PGM activity and starvation resistance vary in association with alleles at one clinally-segregating site. However, it is likely that the observed patterns are the result of neutral processes. The roles of selective and neutral forces in maintaining variation at *Pgm* of these two species are discussed.

4.2 Introduction

The impact of standing genetic variation on adaptive evolution is an unresolved issue in evolutionary biology. Stimulated by the first assessment of allozyme polymorphism within a species (Lewontin & Hubby 1966), numerous population surveys ultimately revealed extensive protein polymorphism across diverse taxa (e.g. Burns & Johnson 1967; Prakash et al. 1969; Ayala et al. 1972). Realization that such high levels of intraspecific variation persist within gene sequences created the still unresolved controversy over the relative roles of selective and neutral processes in maintaining this variation. Balancing selection arising from differential selective pressures across a species' range may contribute to the maintenance of this variation, but conclusive identification of such responsive adaptive variants remains a challenge.

Several approaches can be used to support the inference that genetic variants are locally adapted to specific environments. Clinality of a variant across geographic or other environmental gradients is one such line of evidence. Many examples of allele frequency clines have been identified. For instance, populations of the killifish, *Fundulus heteroclitus,* along the eastern US coastline maintain two protein isoforms of the metabolic gene *lactate dehydrogenase B* (*Ldh-B*) (Powers & Place 1978), the anemone *Metridium senile* shows an allozyme cline in *glucosephosphate isomerase* (*Gpi*) along the northeastern coast of the US (Hoffmann 1981), and *Drosophila melanogaster* worldwide exhibit latitudinally distributed variants of *Alcohol dehydrogenase* (*Adh*) (Oakeshott et al. 1982).

Clinality of variants is an expected outcome of differential adaptation (Endler 1973), but every clinal variant is unlikely to be a product of local adaptation. Neutral processes can also generate spatially-structured allelic variation. Vasemagi (2006) simulated the effects of isolation by distance on standing variation and found that even modest amounts of isolation create clinal patterns similar to those often attributed to differential adaptation across a species' range. The fact that neutral processes can create

geographic patterns of allelic variation makes the link between the presence of single-locus clines and local adaptation tenuous.

Another line of evidence that segregating variants play a role in local adaptation is through demonstration of a relevant phenotypic difference between alleles. For the allozyme cline at *Adh* of *D. melanogaster*, a 2-3 fold difference in ADH activity between alleles is mediated both by the allozyme polymorphism and a tightly-linked intron sequence polymorphism (Laurie et al. 1991). Likewise, differences in activity are evident for allozymes of LDH-B of *F. heteroclitus* (Dimichele & Powers 1982) and for GPI of *M. senile* (Zamer & Hoffman 1989). For some loci, though, delineation of allele-specific phenotypes can be elusive, as kinetic differences may be cryptic, only manifesting under stressful conditions (Eanes 1999), and observed differences may not correspond to fitness differences, rendering them selectively neutral.

The most important evidence for adaptive significance of sequence variation is a demonstration of realized fitness consequences in natural populations. For *Adh*, genotype is correlated with ethanol tolerance, but the effect differs among populations, indicating the influence of genetic background (Merçot et al. 1994). In *F. heteroclitus, Ldh-b* variation influences a wide variety of traits, including hatching time, mortality under thermal stress and swimming performance (Dimichele & Powers 1982a; 1982b; 1991). Such observations of allelic differences strongly imply that observed variation is locally adaptive rather than neutral, if the conditions altered in the laboratory correspond with selective forces experienced in wild populations.

Under certain conditions, selection to increase the frequency of a locally adapted allele leaves a genomic signature through patterns of variation in nearby neutral polymorphisms. Locally adaptive variants influence the fate of closely linked neutral variants, thus shaping patterns of nucleotide variation throughout the genome. Variants subject to balancing selection may exhibit signs of recent partial selective sweeps, with an excess of intermediate-frequency linked polymorphism. For instance, the nucleotide

site underlying the well-studied allozyme cline in *Adh* of *D. melanogaster* is within a region of elevated silent polymorphism, reflecting linkage of alleles to the site under selection. This narrow region is symmetric around the allozyme polymorphism and breaks down due to recombination on either side (Kreitman & Hudson 1991; Berry & Kreitman 1993). Analysis of nucleotide variation within samples from either end of a cline can also be informative. In addition to the previously described allozyme cline, *F. heteroclitus* exhibits differences in *Ldh-b* expression associated with allelic differences in the 5' regulatory region. Patterns of nucleotide variation in the 5' flanking region are consistent with neutral expectations, but when nucleotide variation in northern populations alone is analyzed, there is an apparent excess of singleton variants (Schulte et al. 1997). This pattern could be explained by recent fixation of a haplotype consisting of an advantageous allele and linked neutral variants (Fu & Li 1993). The strength of the signal associated with balancing selection is dependent on the age of the advantageous allele as well as the recombination rate in the region, and may not be readily apparent for all balanced polymorphism (Charlesworth 2006).

Together, the above-described tools allow a multi-faceted approach for evaluating the adaptive significance of variation. Many of these approaches have been utilized to investigate variation in the *Phosphoglucomutase* (*Pgm*) locus, resulting in the discovery of potentially adaptive variants in many taxa. *Pgm* is a branch-point enzyme which controls the breakdown and storage of glycogen. If fine-tuning of flux-control of glycogen metabolism over different environmental conditions is generally advantageous, I would expect to see variation segregating in parallel with environmental gradients such as temperature. Such variation has been discovered in *D. melanogaster* (Verrelli & Eanes 2001a), but it is unclear whether this is a species-specific adaptive mechanism or whether it is a common feature of PGM in species subjected to similar selection pressures. In *D. melanogaster, Pgm* shows an elevated ratio of polymorphism to divergence at nonsynonymous sites compared with polymorphism and divergence at synonymous sites

(Verreli & Eanes 2000). This pattern indicates either balancing selection or reduced constraint on amino acid sequence. Local adaptation through functionally distinct variants is supported by the clinal distribution of amino acid variants in PGM and by the different activities of these alleles (Verreli & Eanes 2001a; 2001b).

Studies of *Pgm* in distantly-related taxa have generated mixed results. For instance, *Colias* butterflies carrying different alleles at *Pgm* do not have differential flight capacity or survival, indicating a lack of phenotypic difference (Carter & Watt 1988). Likewise, alleles representing a clinal polymorphism at *Pgm* of *M. senile* appear to have similar activities (Hoffman 1985). On the other hand, *Pgm* shows evidence consistent with balancing selection in some species; *Pgm* alleles of the Pacific oyster *Crassotrea gigas* exhibit overdominance in PGM activity, with the highest activity in heterozygotes of the most common allele with any of several less common alleles (Pogson 1991) and alternate *Pgm* alleles in mice are associated with differences in glycogen storage after starvation (Leigh Brown 1977).

Variation in PGM is common, but the frequency with which that variation contributes to local adaptation is unclear. It may be that maintenance of alternate alleles of PGM is a common mode of adaptation to environmental gradients. To test this, I assess genotypic and associated phenotypic variation at *Pgm* in *Drosophila americana,* an endemic that inhabits a latitudinal range with large climatic differences similar to the North American range now occupied by *D. melanogaster*. Selection maintains a chromosomal fusion polymorphism in *D. americana* along its latitudinal range (McAllister et al. 2008). Unstructured variation in putatively neutral markers on the 4th chromosome demonstrates the maintenance of the latitudinal cline for the chromosomal polymorphism despite high levels of gene flow (McAllister 2002). Because *D. americana* is able to selectively maintain polymorphism across its range in the presence of gene flow, it is an ideal species in which to investigate the adaptive significance of variation at *Pgm*.

Here I analyze sequence variation in *Pgm* of *D. americana* and measure PGM activity and starvation tolerance associated with clinally distributed variants. I find that although *D. americana* segregates many nonsynonymous polymorphisms as well as silent clinal polymorphisms at *Pgm* and does exhibit variation in activity and starvation in conjunction with alleles at one clinal site, it is unlikely that this represents adaptive variation. In fact, variation at *Pgm* of *D. americana* appears to largely consist of neutral and mildy deleterious variants. Although *Pgm* is a branch-point enzyme and alleles could theoretically alter the flux of metabolites through the glycolytic pathway, it likely has excess capacity which reduces the impact of selection and allows sub-optimal alleles to persist at low frequency in wild populations.

4.3 Materials and Methods

4.3.1 Population genetic analysis

The *Pgm* region was amplified and sequenced from 38 *D. americana* chromosomes sampled from four populations representing the northernmost and southernmost portions of the species' range: Iowa River, Hawkeye Wildlife Area, IA (IR) and Illinois River at Duck Island, IL (DI) in the north and Pearl River WMA, MS (RB) and Cat Island NWR, LA (CI) in the south (Table 4.1). Samples of wild flies were obtained using methods described in detail in McAllister et al. (2008). Single chromosomes were isolated in hybrids by crossing wild-caught male and female (sperm-depleted) flies to V46, a lab strain of the closely-related *D. virilis* for which a genome sequence is available. Hybrid offspring were isolated and frozen at -80°C. Limited sequencing of inbred *D. americana* lines in the *Pgm* region allowed identification of fixed nucleotide differences relative to the reference sequence from *D. virilis*. These fixed differences were used to design primers that specifically amplify *D. americana* alleles from F1 interspecific hybrids. For most flies, four overlapping *D. americana*-specific primer pairs were sufficient for amplification of an approximately 3000 bp region

including most of the *Pgm* coding sequence and ~800 bp 5' of the coding region. Primers and annealing temperatures are detailed in Table 4.2. Amplified DNA was sequenced on an ABI 3730 (Applied Biosystems) following standard methods and sequences were manually edited with Sequencher (Applied Biosystems). Alignments produced by Sequencher were also manually edited.

Sequence diversity and heterozygosity were calculated and two measures of selection based on polymorphism frequency, Tajima's *D* (Tajima 1989) and Fu and Li's *D* (Fu and Li 1993), were determined for each population, pooled northern and southern populations, and for the dataset as a whole. Coalescent simulations using segregating sites and assuming no recombination were generated to evaluate the fit of Tajima's *D* and Fu & Li's *D* to neutral expectations. Tajima's *D* was also calculated for coding, synonymous, nonsynonymous and silent (synonymous and noncoding) sites separately. A sliding window analysis of Tajima's *D* and Fu & Li's *D* with a window of 150 bp and a step size of 50 bp was used to detect localized deviations from neutral polymorphism frequency. Fay & Wu's *H* (Fay & Wu 2000) was calculated in a sliding window in order to detect regions with an excess of high frequency derived polymorphism indicative of a recent hitchhiking event. Linkage disequilibrium among informative sites was calculated and evaluated using Fisher's exact test. A McDonald-Kreitman test (MK; McDonald and Kreitman 1991) was used to compare ratios of polymorphism to divergence among different types of sites using *D. virilis* sequence as an outgroup. All population genetic analyses were carried out in DnaSP v5.10 (Librado & Rozas 2009).

In order to investigate selection on different types of sites, frequency spectra were produced for derived variants in introns, 5' of the coding region, and within the coding region at nonsynonymous sites and at synonymous sites to unpreferred variants. The derived state was inferred relative to *D. virilis*. Frequency spectra were used to estimate scaled selection coefficients following the method of Akashi and Schaeffer (1997).

4.3.2 Analysis of clinal polymorphism

Frequencies of alleles at polymorphic sites segregating at ≥10% were analyzed with respect to latitude using Pearson and Spearman Rank correlations as implemented in SAS software v.9.2 (Cary, NC). Three polymorphic sites 5' of the coding region that were significant or marginally significant by both Spearman Rank and Pearson correlations were chosen for further analysis. For sites -643 and -644, 19 additional chromosomes were sequenced including nine from a fifth population of intermediate latitude collected at Coldwater River, Mississippi (LR), and ten from the previously-sampled northern populations, IR and DI. This generated an enlarged (N = 57) sequence dataset for a 639 bp region 5' of the coding region. Tajima's D was calculated for the expanded dataset in a sliding window and the scaled selection coefficient was estimated using the frequency spectrum of derived variants.

For site -91, alleles were determined by *D. americana*-specific PCR followed by restriction digest in 12 additional populations encompassing much of the species' range (Table 4.1). Primers and restriction digest conditions are detailed in Table 4.2.

4.3.3 Pgm activity of inbred lines

Twenty-seven inbred lines, generated via 11-14 generations of full-sib mating were genotyped for two of these three polymorphisms that remained clinal in the expanded sample (-643 and -91) to serve as material to investigate functional consequences of *Pgm* nucleotide variation (Table 4.3). Reproductively mature single flies from the genotyped inbred lines were used to measure differences in PGM activity attributable to identified nucleotide differences. Three separate treatments were investigated: six day-old flies reared and aged at 22°C (N=27 lines), six day-old flies reared at 22°C and aged at 11°C (N=17 lines), and 10-11 day-old flies reared and aged 7-8 days at 22°C and starved three days at 22°C (N=17 lines). PGM activity was determined according to the methods of Stam and Laurie-Ahlberg (1982). For each

treatment, PGM activity was assayed for two female and two male flies of each assayed line. Single flies were collected and frozen at -80°C and later homogenized in randomized blocks in 250μL 0.01M KH_2PO_4, 1.0mM EDTA, pH 7.4. Homogenate was centrifuged at 4°C for 1.5 min at 10,000 rpm. The supernatant was removed and used to measure PGM activity and total soluble protein. For PGM activity, absorbance at OD_{340} was measured for three minutes immediately following mixture of 12.5μL of fly homogenate with 212.5μL PGM activity reagent (0.83mM glucose-1-phosphate, 0.05mM glucose-1,6-bisphosphate, 0.5mM NADP, 1.0mM $MgCl_2$, 3.1 U/mL glucose-6-phosphate dehydrogenase, 20 mM Tris-Cl, pH 7.4). Units of PGM activity were calculated as micromoles NADP reduced/minute over the linear portion of the absorbance increase, according to the equation: ([(ΔOD_{340}/min)(dilution)]/[(μM extinction coefficient NADPH)(path length)]). In order to scale PGM activity, total soluble protein was determined using the micro pyrogallol red method (Sigma total protein kit, TP0400). End-point absorbance at OD_{600} was measured after five minutes for 5μL fly homogenate in 250μL total protein reagent and quantified relative to a dilution gradient of human serum albumen. Scaled PGM activity is reported as units PGM/mg soluble protein. Spectrophotometric measurements were made on a SpectraMax M2 Microplate Reader (Molecular Devices).

PGM activity was analyzed in a general linear model using proc GLM in SAS v9.2. A Type III Sum of Squares model was constructed including sex, preparation block, allele at each clinal site, interactions of each allele with sex and interaction between alleles. 95% confidence intervals around the mean activities for each allele were calculated. Correlation coefficients for PGM activity of biological replicates (pairs of males and females from each line/treatment experiment) were used to determine whether results were consistent within genetic backgrounds.

4.3.4 Starvation of inbred lines

Starvation resistance was assayed for the same seventeen lines for which scaled PGM activity was measured in flies starved for three days. Flies of each line were reared at low density at 22°C and aged in single-sex groups of 10-15 flies for 7-9 days on food before transferring to a nutrient-free 2% agar media. Deaths in each vial were recorded each day until all flies died. Mean survival/vial was calculated and compared among alleles for genotyped clinal polymorphisms of interest in an ANOVA using proc GLM in SAS v.9.2.

4.4 Results

4.4.1 Population genetic analysis

Thirty-eight sequences representing almost the entire coding region and introns of PGM, and including the region immediately upstream of this gene were determined for single alleles from two northern and two southern populations of *D. americana*. Each sequence included 873 bp upstream of the start codon through most of the last intron and excluding the last exon for a total of 3052 bp. The sequence of the last exon was not included due to difficulty placing primers in repetitive sequence just 3' of the coding region.

For the sequenced region as a whole, Tajima's *D* and Fu & Li's *D* both exhibit negative values and Fu & Li's *D* was marginally significant (-1.29, $p > 0.10$, and -1.99, $0.10 > p > 0.05$, respectively). This is consistent with neutrality for the majority of variants at *Pgm*, accompanied by a slight inflation of rare mildly deleterious variants, especially singletons, caused by weak purifying selection across the entire gene.

Partitions including only coding, synonymous, nonsynonymous, and silent sites all showed negative Tajima's *D* values, but only nonsynonymous sites show a significantly negative value ($D = -2.26$, $p < 0.01$; Table 4.4). Although all classes of sites segregate variants at low frequency, the greatest excess of low frequency variants and

therefore the strongest purifying selection is at nonsynonymous sites. Analyses in a sliding window reveal reductions in Tajima's *D* and Fu & Li's *D* across the sequenced region, with the most pronounced excess of low frequency polymorphism near the beginning of the third and largest exon (Figure 4.1). Sequences from each population alone and from pooled northern and southern populations also exhibit negative, non-significant Tajima's D and Fu and Li's D values (Table 4.5), so it appears that variants are similarly subject to purifying selection across the species' range.

Selection coefficients estimated from frequency spectra of derived variants at nonsynonymous, unpreferred synonymous, intron and 5' noncoding sites are all negative (Figure 4.2). Among silent sites (unpreferred synonymous, intron and 5' noncoding) derived unpreferred variants at synonymous sites have the largest negative selection coefficient ($\gamma = 2N_e s = -2.33$) while selection coefficients for derived variants in introns and 5' of the coding region are closer to neutrality ($\gamma = 2N_e s = -0.85$ and -0.49, respectively). The estimated selection coefficient for derived nonsynonymous changes shows much stronger purifying selection against amino acid changes ($\gamma = 2N_e s = -9.15$) compared to each class of silent site. Variants at nonsynonymous sites segregate at very low frequency indicative of a greater efficacy of selection against changes in amino acid sequence of PGM. Variation at silent sites appears mostly neutral, but some sites segregating at low frequency are presumably weakly deleterious.

An excess of low frequency polymorphism could also be caused by a recent selective sweep of an advantageous allele that carried linked variants to high frequency, but not completely to fixation. A sliding window calculation of Fay & Wu's H was carried out to check for high frequency derived variants that would remain following such an event. This test did not show an excess of high-frequency derived alleles in any window, so there is no evidence of a recent selective sweep. Furthermore, Fisher's exact tests for linkage disequilibrium are significant after Bonferroni correction for only four

pairs of sites, each pair being separated by only 1-2 bp. Thus there is no widespread linkage that would imply a recent selective sweep event.

Further evidence that variation in *Pgm* is primarily neutral is indicated by the McDonald-Kreitman test. Comparison of polymorphism and divergence at nonsynonymous and synonymous sites does not differ significantly from neutral expectations (Fisher's exact test $p = 0.3$). When nonsynonymous and silent sites are compared, however, a marginally significant deviation from neutrality is observed (Fisher's exact test $p = 0.07$, G test $p = 0.05$). A combined analysis comparing polymorphism and divergence at noncoding, synonymous and nonsynonymous sites also shows a marginally significant inflation of polymorphism at nonsynonymous sites (Table 4.6). An inflation of nonsynonymous polymorphism could be evidence for balancing selection, but combined with the significantly negative Tajima's D for these sites, this implies that they are subject to purifying selection rather than balancing selection. It is interesting that polymorphism at nonsynonymous sites is elevated when noncoding sites are included in the neutral reference class, but not when only synonymous sites are included. This difference can be explained by codon bias that consists of weak purifying selection against changes to unpreferred codons and results in a paucity of fixations relative to polymorphisms at those sites. Because of this phenomenon, noncoding sites may be a better proxy for neutrality than synonymous sites.

4.4.2 Analysis of clinal polymorphism

The sampled sequences segregate 23 nonsynonymous mutations, but only one of these is present at a frequency >10%. As a result, no clinally-distributed amino acid polymorphisms were detected. Among 75 polymorphic silent sites segregating at >10%, Spearman rank correlations of allele frequency with latitude are marginally significant for 8 sites without correction for multiple tests ($p = 0.051$) and significant after a Bonferroni correction for one polymorphic site (site -91, $p < .0001$). Pearson correlations with

latitude are significant for ten sites ($p < 0.05$), six of which overlap with those marginally significant in Spearman rank correlations (Table 4.7).

Three sites 5' of the coding region were chosen for further analysis due to clinal patterns suggestive of local adaptation. Site -91 was chosen because it showed the strongest correlation with latitude and sites -643 and -644 because they exhibit a marginally significant Spearman rank correlation and a significant Pearson correlation with latitude. Sites -644 and -643 sites were further examined in nineteen additional chromosomes typed by sequencing, including nine chromosomes from LR, a population of intermediate latitude and ten additional chromosomes from the northern populations, IR and DI. Additional typing of chromosomes at sites -644 and -643 strengthened the Spearman rank correlation with latitude for site -643 compared to the initial smaller sample ($p = 0.0048$) but not for site -644 (Table 4.8; Figure 4.3). Moreover, sequence surrounding the clinal site at -643 shows an elevation in intermediate frequency polymorphism when Tajima's D is calculated in a sliding window (Figure 4.4). The selection coefficient estimated from the frequency spectra of derived variants in the expanded dataset 5' of the coding region is positive ($\gamma = 2N_e s = 1.38$). Together, the relationship with latitude and the elevation of intermediate frequency polymorphism imply that balancing selection may be acting to maintain variation in this region.

The third clinal site (-91) is about 100 bp 5' of the start codon, following a run of T's of variable length. Sampling of site -91 from 12 additional populations with an RFLP assay did not reveal a latitudinal correlation (Table 4.9, Figure 4.5), but does show a correlation with longitude ($R^2 = 0.58$, $p = 0.0015$; Table 4.9, Figure 4.6). Site -91 was typed by RFLP, so a corresponding set of sequences was not obtained to analyze polymorphism surrounding the site, but the original sequence dataset shows no significant increase in intermediate frequency polymorphism near site -91 (Figure 4.1).

4.4.3 PGM activity of inbred lines

Polymorphism at -643 and -91 was found within twenty-seven genotyped inbred lines. PGM activity scaled by total soluble protein for these lines varied between 0.43 and 1.36 units/mg soluble protein. A Type III sum of squares model of PGM activity including sex, preparation block, alleles at each polymorphic site and their interactions with sex showed that PGM activity of flies reared and aged under standard conditions did not differ significantly between alleles defined by either site (Figure 4.7). Sex contributed significantly to variation in PGM activity ($F = 44.49$, $p < 0.0001$) as did the blocked design of homogenate preparation ($F = 3.37$, $p = 0.003$). Under these conditions, PGM activity is not influenced by variation at either clinal site. A correlation of individuals from each line-sex combination with their biological replicates was not significant, indicating that differences in PGM activity among lines are minimally influenced by the genotype of the line.

Similarly, PGM activity of flies reared under standard conditions and aged at 11°C did not vary significantly between alleles (Figure 4.7). Unlike flies reared and aged at 22°C, there was no significant difference between sexes, but the significant effect of preparation block remained ($F = 3.80$, $p = 0.0149$) and the biological replicates were not significantly correlated. While PGM activity of flies aged at 11°C did not differ according to allele at either clinal site, it is apparent that PGM activity is affected by the treatment relative to flies aged at 22°C. Male flies at 22°C have significantly higher PGM activity than females, but the sexes are not significantly different after aging at 11°C. In addition, total protein, PGM activity and PGM activity scaled by total protein appear lower on average for the flies aged at 11°C, but because the assays were not carried out on the same day, they are not directly comparable.

PGM activity of flies reared at 22°C on food and starved for 3 days did differ between alleles at site -91 ($F = 7.72$, $p = 0.0074$) but not at site -643 (Figure 4.7). PGM activity also differed between sexes ($F = 74.81$, $p < 0.0001$). Biological replicates within

this experiment were correlated ($R^2 = 0.53$, $p < 0.0001$). After 3 days of starvation there is a difference in PGM activity that can be attributed to the clinally varying alleles. Unlike the other two treatments, results are consistent between flies within lines, so genotypic differences among lines are apparent.

4.4.4 Starvation of inbred lines

Starvation time/vial was fit to a type III sum of squares model including sex, each polymorphism, the interaction of each polymorphism with sex and the interaction between the two polymorphisms. Sex had a large effect on starvation time ($F = 175.08$, $p < 0.0001$) with an average male survival 3.6 days less than average female survival (7.83 and 11.42 days, respectively). Both clinal polymorphisms had a significant effect on starvation time within the model (-643: $F = 7.27$, $p = 0.008$; -91: $F = 9.30$, $p = 0.003$), but a Tukey's test shows a significant difference in starvation time only for alleles defined by site -91 (Figure 4.8). The interaction of the two polymorphic sites was also significant ($F = 66.45$, $p < 0.0001$). When each site is considered separately, lines with alternate alleles defined by site -91 have significantly different starvation survival. However, while lines with the derived 'C' allele at site -91 have lower average starvation survival, the relationship between starvation survival and allele is reversed for the subset of lines with an 'A' at site -643.

4.5 Discussion

J. B. S. Haldane wrote in *The Causes of Evolution* (1932): "Related species will vary in similar directions and be subject to similar selective influences. They may therefore be expected to evolve in parallel." Because of repeated observation of high levels of variation across taxa, *Pgm* is an ideal test case for investigating the forces governing variation among species and the resulting potential for repeated local adaptation at this locus. Depending on the degree of homology and the complexity of traits under selection, there may be a finite number of adaptive paths available to

evolution. At the amino acid level, empirical evidence suggests that proteins may be confined to only a few mutational paths, constrained by selection against possible intermediates (Weinrich et al. 2006), and striking examples of parallel adaptation have been recorded in a wide variety of taxa (Wood et al. 2005).

We hypothesized that homologous genes of taxa that inhabit similar environmental gradients may share a common adaptive response by maintaining locally-adapted variants. To investigate this, I carried out a comprehensive evaluation of variation at the *Pgm* locus of *D. americana* with the goal of understanding the roles of selective and neutral processes acting on this gene. Here I compare and contrast those findings with the selective pressures proposed to act on *Pgm* of *D. melanogaster.* The *Pgm* gene of *D. melanogaster* has been examined extensively for its role in local adaptation. Amino acid polymorphisms in PGM of *D. melanogaster* exhibit latitudinal clines (Verrelli & Eanes 2000; 2001a) and these variants have been shown to affect enzymatic activity (Verrelli & Eanes 2001b). Amino acid variation in PGM may be maintained by selection across the species' range for differential flux of glucose between storage as glycogen and utilization via glycolysis in different environments across the species' range.

The results reported here show similarities in patterns of variation in the two species, but the selective pressures shaping those patterns appear to be different. Like *D. melanogaster, D. americana* segregates many amino acid variants of PGM. *D. americana* also segregates potentially locally adaptive clinal variants that are associated with differences in PGM activity. However, unlike *D.* melanogaster, most variation in *Pgm* of *D. americana* appears neutral or weakly deleterious.

At nonsynonymous sites, both species exhibit an elevated ratio of polymorphism/divergence compared to silent sites, reflecting a shared excess of nonsynonymous polymorphism in the two species. Several phenomena could cause this pattern, including balancing selection that favors alternate alleles in different

environments, as is suggested to explain the observations in *D. melanogaster* (Verreli & Eanes 2001a). On the other hand, the excess nonsynonymous polymorphism may reflect relaxed selection on those sites. In *D. americana,* the departure from neutrality for polymorphism at nonsynonymous sites is very small, and there is an excess of low frequency nonsynonymous polymorphism. Together, this indicates that the nonsynonymous polymorphism in *D. americana* is subject to purifying selection rather than balancing selection, although it apparently isn't sufficiently strong to immediately purge new mutations from the population.

Both species segregate clinal variants that could be a result of balancing selection. While the clinal variants in *D. melanogaster* include nonsynonymous sites (Verreli & Eanes 2001a), all the clinal variants are at silent sites in *D. americana*. Of course, both silent and nonsynonymous variants could be adaptive, although they are likely to have effects on expression and enzymatic function, respectively. In many cases where parallel adaptation of homologs has been observed, though, it has occurred either in the coding sequence or in a cis-regulatory region repeatedly for a given gene (Gompel & Prud'homme 2009). Differences in the type of site affected (coding vs. noncoding) doesn't rule out parallel adaptive change, but it is interesting to note that this example would stray from the norm in this respect if clinal sites in both genes were confirmed to be adaptive.

Finally, both species exhibit a difference in PGM activity between flies carrying alternate alleles at clinal sites. In *D. melanogaster,* there is an increase in a derived haplotype associated with higher PGM activity and greater glycogen storage in flies from northern latitudes (Verreli & Eanes 2001a; 2001b) while in *D. americana* there is an increase in a derived allele in eastern populations that is associated with lower PGM activity and faster starvation.

The patterns of sequence polymorphism and associated differences in PGM activity and starvation tolerance in *D. americana* may reflect adaptation to a selective

gradient that varies with geography like the proposed adaptive variation in *D. melanogaster*. On the other hand, the patterns that I see at *Pgm* in *D. americana* may be a result of neutral processes. Two lines of evidence support a neutral origin for the observed variation. First, a sample of isofemale lines from throughout the species range shows no differences in starvation survival that correlate with longitude (unpublished data). Second, there is no detectable increase in intermediate-frequency polymorphism around site -91 that would indicate the action of balancing selection. While there is clearly variation at the *Pgm* locus as well as variation in PGM activity and starvation resistance, it is likely that it is a result of mutation and drift, rather than geographically variable selection.

Previous work indicates that *D. americana* is capable of maintaining adaptive variation across the sampled range (McAllister 2002; McAllister et al. 2008). Given these observations, *D. americana* should be able to maintain adaptive variation at *Pgm,* provided that such adaptive variation exists. It is important to note, though, that all the regions previously identified as segregating variation as a result of selection are in regions of reduced recombination due to their proximity to polymorphic chromosomal rearrangements. *Pgm*, on the other hand, is found in a region of the genome that is monomorphic in chromosomal form. For individual adaptive variants to be maintained in a freely recombining region, the individual selective advantage of an allele must be sufficiently large that it is visible to selection, whereas linked advantageous alleles may interact either additively or epistatically to increase their selection coefficient above the threshold required for maintenance of polymorphism.

Although many alleles of PGM were found in wild *D. americana* populations, they were each found at low frequency and do not appear to be subject to balancing selection. This contrasts with observations of clinal amino acid polymorphism in PGM of *D. melanogaster*. The pattern in *D. americana* is consistent with a mix of neutral and

weakly deleterious variants segregating at PGM. The lack of strong selection indicates that the alleles are nearly equivalent in terms of individual fitness.

Null or low activity alleles for many enzymes are viable in a heterozygous state, indicating that as little as 50% activity is sufficient for survival under some conditions (Burkhart et al. 1984; Gibson et al. 1991). In fact, experimental manipulation of PGM levels in *D. melanogaster* showed that flies with as little as 10% of wild type PGM activity show normal flight performance (Eanes et al. 2006). These observations indicate that PGM has excess capacity, so that it exerts little control over flux of metabolites through the pathway, and ultimately has little effect on variation in phenotype. If this is the case, it may be that the variation in PGM of *D. melanogaster* is not subject to balancing selection, but rather that it may be an ephemeral, non-equilibrium state following colonization of its North American range.

Comparative analyses like the one presented here allow characterization of a baseline state of variation at a particular locus. Differences from that norm may reflect either unique demography or differences in selection pressures across species. For *Pgm*, the baseline state among species appears to include excess amino acid polymorphism, but that polymorphism does not necessarily contribute to differences in fitness. So, while *Pgm* varies in parallel in *D. americana* and *D. melanogaster*, the observed variation does not appear to represent parallel adaptation at the level of the gene. The pervasive variation at *Pgm* might instead be a consequence of neutral processes acting on a gene with conserved excess enzymatic capacity.

Table 4.1Collection sites and sample numbers for each sequence based analysis.

	Site	latitude (°N)	longitude (°W)	Full Sequence	5' Sequence	RFLP
NN	Niobrara River at Hwy. 12, west of Niobrara, NE	42.7	-98.0	--	--	6
DN	Platte River south of Duncan, NE	41.3	-97.4	--	--	9
IR	Hawkeye Wildlife Area, near Iowa River, IA	41.7	-91.7	9	4	--
CI	Cat Island National Wildlife Refuge, LA	30.7	-91.4	9	--	9
DI	Illinois River at Duck Island, IL	40.4	-89.9	10	6	19
LR	Coldwater River at Hwy. 51, MS	34.7	-89.9	--	9	11
RB	Pearl River Wildlife Management Area, MS	32.5	-89.9	10	--	21
BB	Brewer Basin, near Reelfoot National Wildlife Refuge, TN	36.4	-89.3	--	--	14
WR	Wheeler National Wildlife Refuge, AL	34.7	-87.8	--	--	8
DA	North of Demopolis, AL	32.5	-87.8	--	--	15
OC	Patoka National Wildlife Refuge, IN	38.3	-87.3	--	--	16
MK	Muscatatuck National Wildlife Refuge, IN	38.9	-85.8	--	--	16
BU	Bradley Unit of Eufaula National Wildlife Refuge, GA	32.0	-84.8	--	--	18
FG	Sneads and Seminole Lake, FL	30.7	-84.8	--	--	6
OR	Ottawa National Wildlife Refuge, near refuge office, OH	41.6	-83.2	--	--	10
WS	Killbuck Marsh National Wildlife Refuge, OH	40.7	-82.0	--	--	4
Total				38	19	182

Table 4.2 Primers, annealing temperatures and restriction enzymes.

Forward		Reverse		°C	size	enzyme
Da-855f	TGACGTGGAAGGGCTGATGC	Da-147r	ATGCTGACTGCGTGCGAATTTG	58	708	--
-311f	GCCCTTGAATGACATTGGC	Da654r	CGGAATATTTCCAATCAGTTTCCA	58	965	--
161f	AACAGTTRYACGGCACACAACC	Da936r	TAYGAGAGCCTAGAGTTCTCGA	58	775	--
371f	CGTCTCCAGCCTGATACGTC	Da1643r	ATGGCGTCCAGATAGTGAGCG	59	1272	--
Da1530f	GACTTTGGTGGACTRCATCCC	2371r	CCAATGCAATGTCAATCAGC	58	841	--
Da-300f	GAATGCATTGGCACCCTACAC	Da216r	CCACAATCAGGGTGGAGCCC	60	516	HpyAV

Table 4.3 Alleles at sites -643 and -91 for inbred lines used in PGM activity and starvation tolerance assays.

	-643	-91	PGM 22°	PGM 11°	PGMstarve	starve
FP99.16	A	C	x	x	x	x
FP99.52	A	C	x	x	x	x
FP99.34	A	T	x	--	--	--
FP99.50	A	T	x	x	x	x
FP99.10	G	T	x	x	x	x
FP99.2	G	T	x	--	--	--
FP99.46	G	T	x	--	--	--
G96.13	A	C	x	x	x	x
G96.38	A	T	x	--	--	--
G96.70	A	T	x	--	--	--
G96.10	G	C	x	x	x	x
G96.21	G	C	x	x	x	x
G96.23	G	T	x	x	x	x
G96.30	G	T	x	x	x	x
HI99.4	A	T	x	--	--	--
HI99.48	A	T	x	x	x	x
HI99.50	A	T	x	--	--	--
HI99.12	G	T	x	x	x	x
HI99.24	G	T	x	x	x	x
HI99.38	G	T	x	x	x	x
HI99.46	G	T	x	x	x	x
NN97.8	G	T	x	x	x	x
OR01.34	A	C	x	x	x	x
OR01.50	A	T	x	x	x	x
OR01.92	A	T	x	--	--	--
OR01.46	G	T	x	--	--	--
OR01.52	G	T	x	--	--	--
Total			27	17	17	17

Table 4.4 Measures of polymorphism and selection within types of sites.

	π	θ	Tajima's *D*
Synonymous	0.0393	n.a.	-1.03, n.s.
Nonsynonymous	0.0017	n.a.	-2.26, $p < 0.01$
Silent	0.0259	n.a.	-1.19, n.s.
Noncoding	0.0223	0.0344	-1.24, n.s.
Total	0.0172	0.0263	-1.29, n.s.

Note: n.a = not applicable, these values of θ were not calculated due to multiple codon changes. n.s = not significant

Table 4.5 Measures of polymorphism and selection within samples and geographic regions.

	π	Θ	Tajima's *D*	Fu & Li's *D*
IR	0.0164	0.0191	-0.72, n.s.	-1.13, n.s.
DI	0.0191	0.0229	-0.87, n.s.	-1.18, n.s.
RB	0.0180	0.0205	-0.60, n.s.	-0.91, n.s.
CI	0.0187	0.0223	-0.79, n.s.	-1.03, n.s.
North (IR+DI)	0.0175	0.0236	-1.09, n.s.	-1.52, n.s.
South (RB+CI)	0.0181	0.0236	-0.97, n.s.	-1.47, n.s.
Total	0.0172	0.0263	-1.29, n.s.	-1.99, $0.10 > p > 0.05$

Note: n.s. = not significant

Table 4.6 Ratios of fixation to polymorphism at nonsynonymous, synonymous and noncoding sites.

	Fixed	Polymorphic	F:P
Nonsynonymous	3	23	1 : 7.67
Synonymous	25	85	1 : 3.40
Non-coding	88	205	1: 2.33
		Fisher's Exact test p = 0.06	

Table 4.7 Clinally segregating sites at *Pgm*.

	5’						In1	In2			Ex3		In3	
	-829	-774	**-644**	**-643**	-348	**-91**	94	608	707	874	1605	1916	2410	2419
IR	0.11	0.89	**0.56**	**0.11**	0.33	**0.56**	0.11	0.89	0.33	0.11	0.11	0.56	0.33	0.33
DI	0.33	0.89	**0.44**	**0.22**	0.11	**0.33**	0.11	0.89	0.44	0.11	0.11	0.44	0.11	0.33
RB	0.40	0.80	**0.30**	**0.50**	0.00	**0.20**	0.30	0.80	0.60	0.20	0.20	0.30	0.10	0.20
CI	0.40	0.80	**0.30**	**0.50**	0.00	**0.10**	0.30	0.80	0.60	0.20	0.20	0.30	0.10	0.10
$r_{Spearman}$	-0.95	0.89	**0.95**	**-0.95**	0.95	**1.00**	-0.89	0.89	-0.95	-0.89	-0.89	0.95	0.95	0.95
$p_{Spearman}$	0.05	0.11	**0.05**	**0.05**	0.05	**<.0001**	0.11	0.11	0.05	0.11	0.11	0.05	0.05	0.05
$r_{Pearson}$	-0.80	0.98	**0.95**	**-0.98**	0.86	**0.92**	-0.99	0.99	-0.96	-0.99	-0.99	0.95	0.68	0.97
$P_{Pearson}$	0.20	0.01	**0.05**	**0.02**	0.14	**0.08**	0.01	0.01	0.04	0.01	0.01	0.05	0.32	0.03

Note: Sites highlighted in bold were investigated in additional populations.

Table 4.8 Frequency of the derived allele at site -643 in an expanded sample with a Pearson correlation with latitude.

Sample	frequency	latitude	N
IR	0.08	41.7	13
DI	0.27	40.4	15
LR	0.44	34.7	9
RB	0.5	32.5	10
CI	0.5	30.7	10
		$R^2 = 0.89$, $p = 0.015$	

Table 4.9 Frequency of the derived allele at site -91 in an expanded sample of collection sites and their Pearson correlations with latitude and longitude.

Sample	latitude	longitude	frequency	N
NN	42.7	-98.0	0	6
IR	41.7	-91.7	55	9
OR	41.6	-83.2	50	6
DN	41.3	-97.4	17	6
WS	40.7	-82.0	50	4
DI	40.4	-89.9	24	25
MK	38.9	-85.8	29	14
OC	38.3	-87.3	14	14
BB	36.4	-89.3	28	21
LR	34.7	-89.9	0	5
WR	34.7	-87.8	33	13
RB	32.5	-89.9	33	12
DA	32.5	-87.8	25	8
BU	32.0	-84.8	23	13
CI	30.7	-91.4	12	17
FG	30.7	-84.8	16	12
	$R^2 = 0.09$	$R^2 = 0.24$		
	$p = 0.3$	$p = 0.06$		

Figure 4.1 Measures of selection analyzed in a sliding window across *Pgm*. Window length = 150 bp, step = 50 bp.

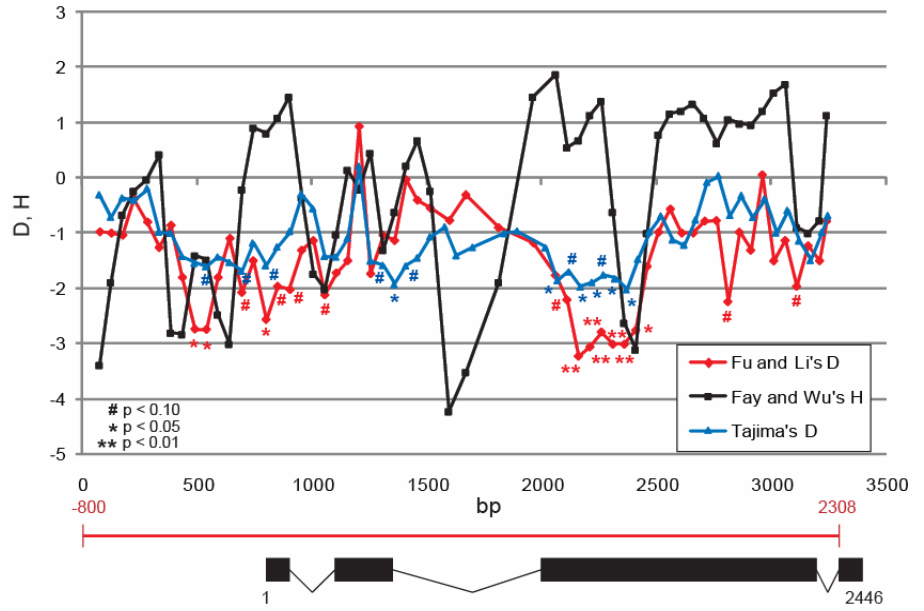


Figure 4.2 Frequency spectra of derived alleles at unpreferred synonomous, intron and sites 5' of the coding region. Estimates of selection based on these spectra are: nonsynonymous $\gamma = 2Nes = -9.15$; unpreferred synonomous $\gamma = 2Nes = -2.33$; intron $\gamma = 2Nes = -0.85$, 5' region $\gamma = 2Nes = -0.49$.

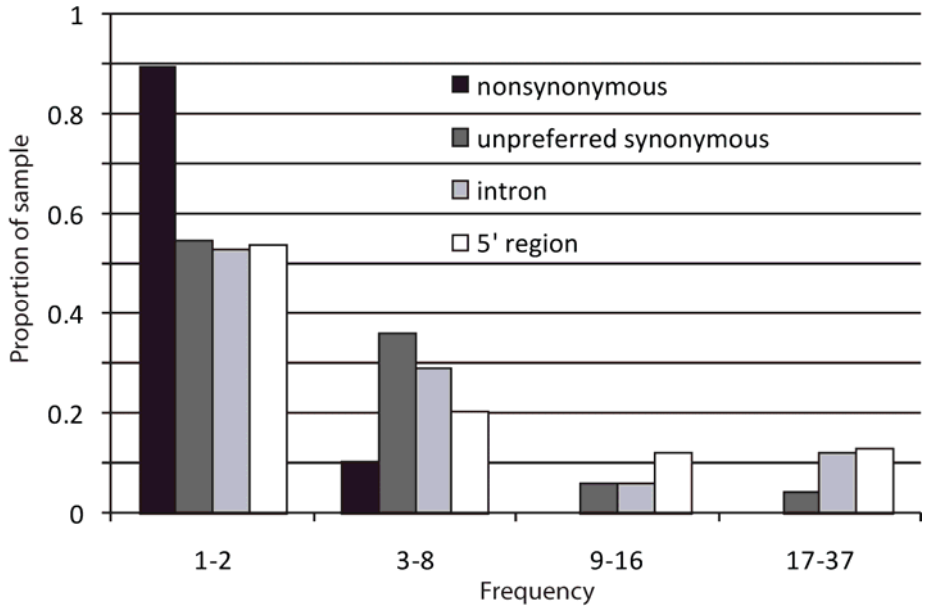


Figure 4.3 Frequency of the derived allele at site -643 plotted against latitude.

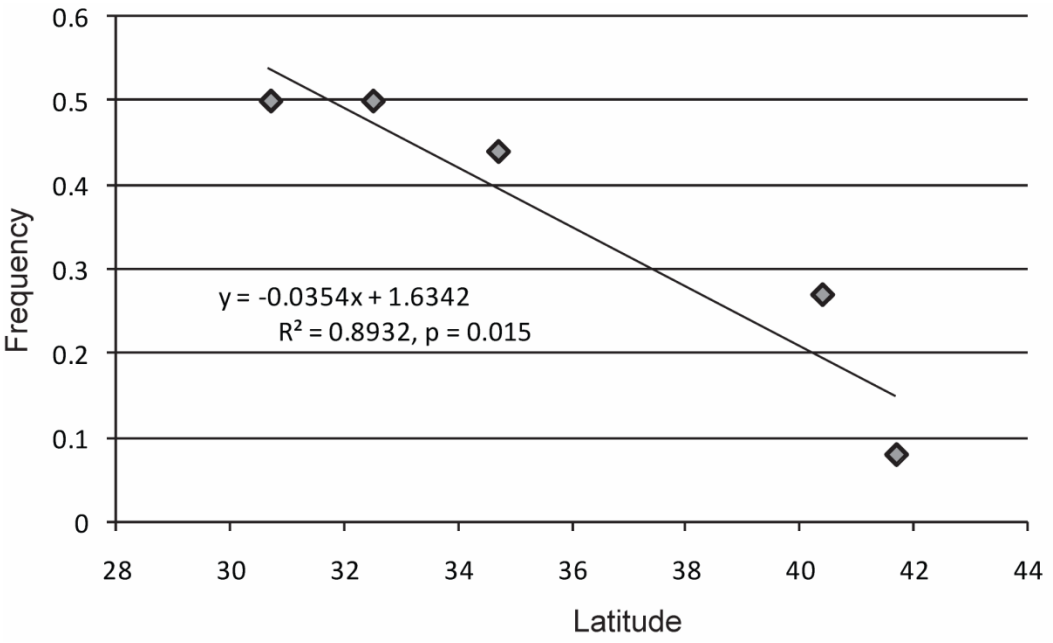


Figure 4.4 Sliding window analysis of Tajima's D for an enlarged sequence dataset in the region 5' of the *Pgm* coding region. Site -643 is marked by an arrow on the x-axis.

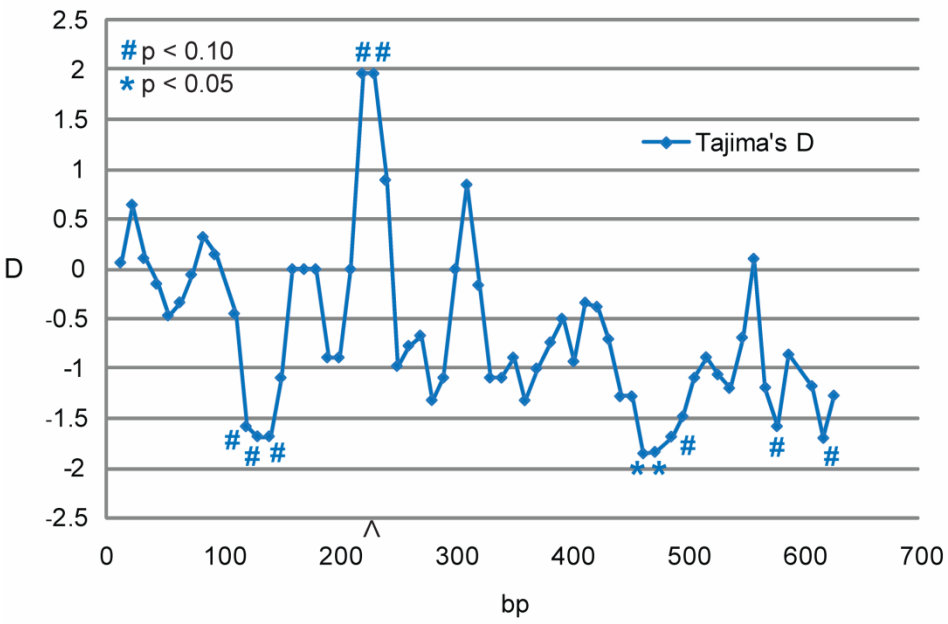


Figure 4.5 Frequency of the derived allele at site -91 plotted against collection latitude.

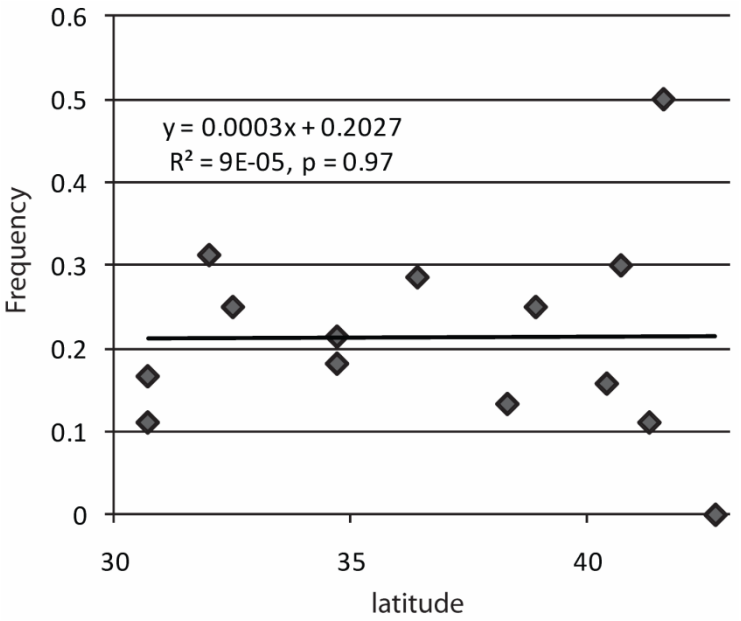


Figure 4.6 Frequency of the derived allele at site -91 plotted against collection longitude.

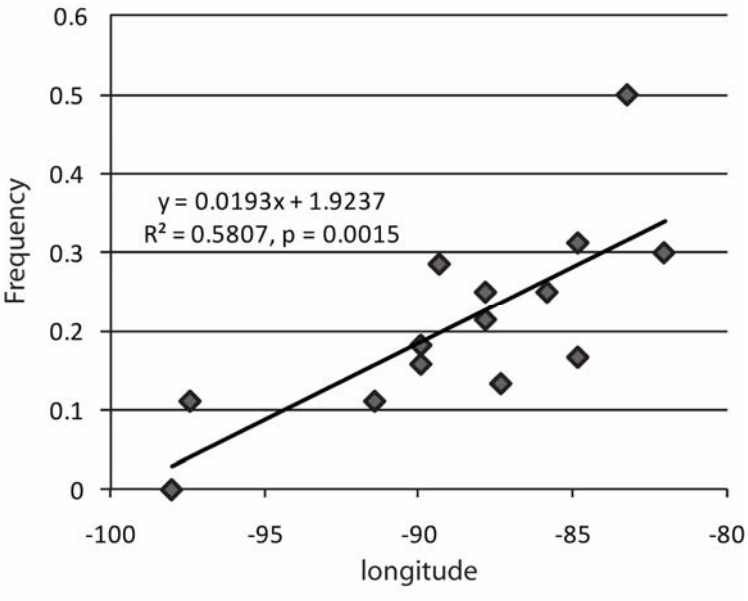


Figure 4.7 Scaled PGM activity of inbred lines. A, D: reared and aged at 22°, B, E: reared and aged at 11°, and C, F: reared, aged and starved at 22°C, categorized by allele at site -643 (A-C) and site -91 (D-F). *p < 0.05.

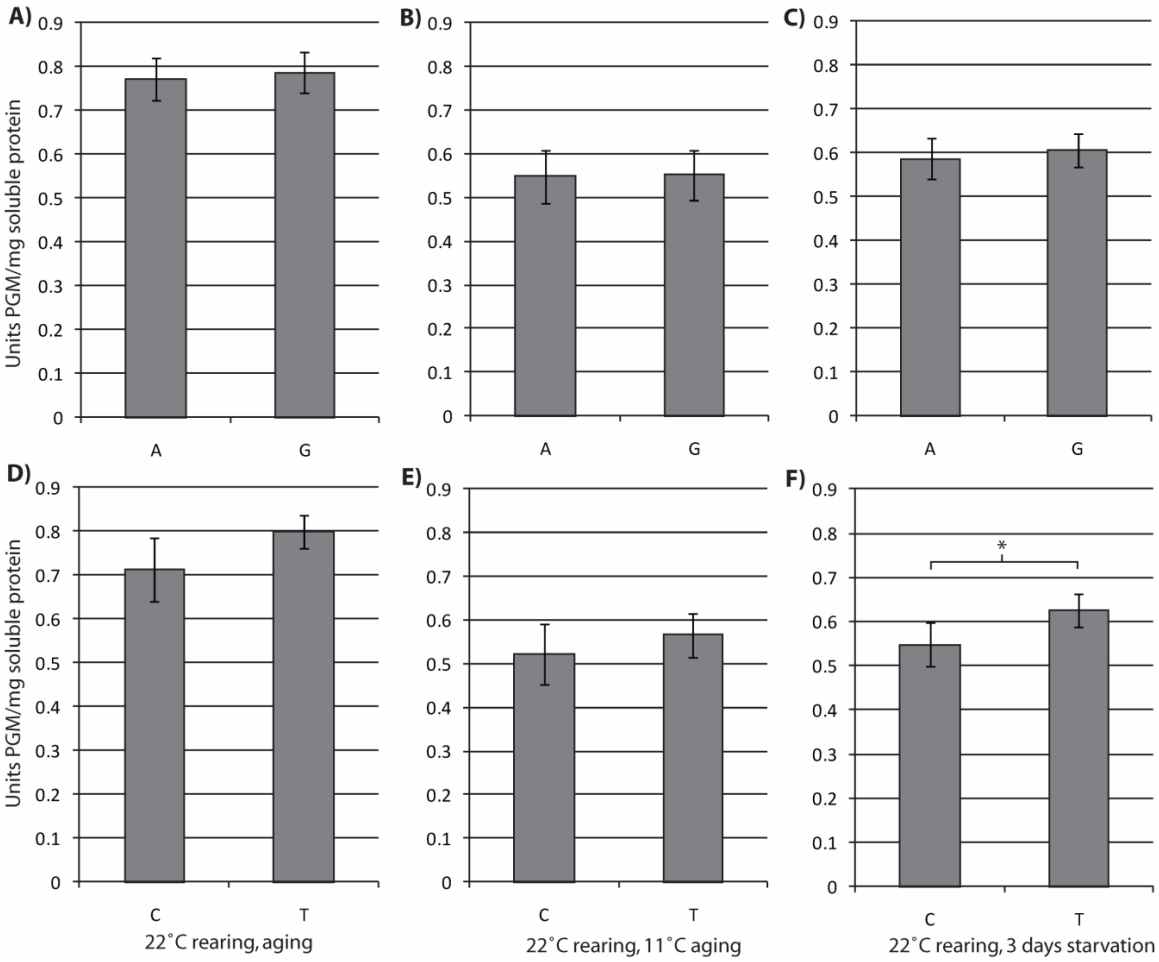
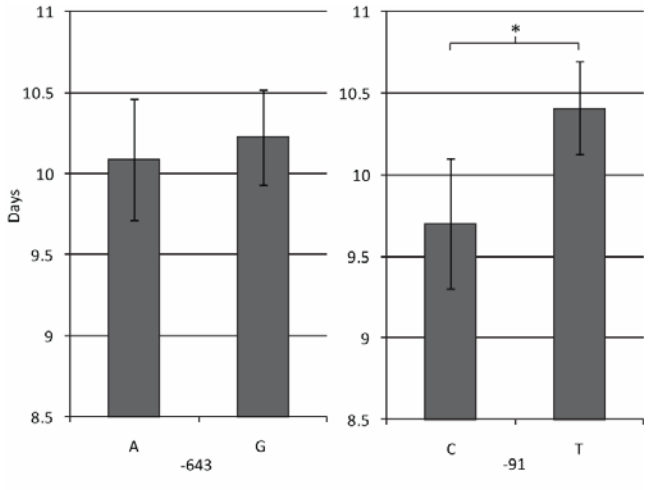


Figure 4.8 Starvation survival of inbred lines categorized by allele at site -643 and -91. *p < 0.05.



CHAPTER 5

THE ROLE OF POSITIVE SELECTION IN THE EVOLUTION OF *DROSOPHILA AMERICANA*

5.1 Abstract

Neutral theory posits that most mutations are invisible to selection, with their frequencies governed by drift, and that the fixation of advantageous variants plays only a small role in evolution. Recently this view of the role of positive selection in evolution has begun to change. Methods utilizing intra- and interspecific data show that a significant portion of sequence change is driven by positive selection. It is unclear, though, if the rate of adaptive evolution is universally high across taxa, because biogeographic history influences estimates. Here I carry out an analysis of the role of positive selection in evolution of an endemic *Drosophila* species, *D. americana*, which has a simple biogeographic history. In contrast with other model species with complicated biogeographic and demographic histories, very little adaptive evolution is evident in *D. americana.* The influence of biogeographic and demographic history on estimates of the rate of adaptive evolution requires that many species be investigated in order to discover any universality among rates.

5.2 Introduction

Evolutionary biologists have long sought to understand the influence of positive selection on variation within and divergence between species. While measurement of polymorphism and divergence is straightforward, evaluation of the proportion of changes that represent adaption is only beginning to be possible. Recent expansion of available sequence data has allowed estimation of the impact of adaptive evolution using methods that utilize intra- and/or interspecific data. Estimates of the rate of adaptive evolution vary among taxa and calculation methods, but in general these analyses identify a

significant proportion of sequence changes in both coding and noncoding regions resulting from positive selection (reviewed in Eyre-Walker 2006).

Intraspecific data alone can be used to detect only recent positive selection, either as a result of partial or complete selective sweeps. Complete selective sweeps are the result of fixation of an advantageous allele. As it becomes fixed, it reduces polymorphism at linked loci by bringing variants to high frequency or fixation (Maynard Smith & Haigh 1974). The extent of the reduction of polymorphism is controlled by local recombination rate and by the strength of selection on the adaptive allele, with recombination rate inversely proportional to the width of the region of reduced polymorphism, and the strength of selection directly proportional (Kaplan et al. 1989; Gillespie 2000; Kim & Stephan 2002). Over time, the region of reduced polymorphism will decay, making old selective events difficult or impossible to identify. In cases of local adaptation, an incomplete or partial selective sweep that brings an allele to fixation in only some populations can result in an excess of intermediate frequency linked variation and regions of differentiation between populations (Tajima 1989).

Detection of complete or partial selective sweeps is carried out by surveying patterns of variation across a genome. Such analyses of polymorphism have been used to pinpoint regions of intermediate and low frequency polymorphism resulting from selective sweeps. For example, surveys of intraspecific variation in humans show an 11% reduction in polymorphism in gene-rich regions of the genome, caused by selection on linked sites (Cai et al. 2009). That selection could be positive (selective sweeps) or negative, with neutral alleles removed as a result of linkage with deleterious polymorphism (background selection). In *Drosophila melanogaster*, a maximum likelihood approach using intraspecific data was used to estimate the fixation of ~160 advantageous alleles on the X chromosome, ~60 of which can be attributed to evolution in European populations, an indication that selective sweeps brought many mutations to fixation in both African and European populations in the last 60,000 years (Li & Stephan

2006; Andolfatto 2007). In both species, it appears as though recent selection has brought advantageous changes to fixation throughout the genome, perhaps as a result of adaptation to newly-colonized habitats.

Rates of adaptive evolution over longer time scales can be inferred from interspecific data alone by comparing divergence at nonsynonymous sites to that at synonymous sites, with the expectation that synonymous divergence is composed of neutral changes and that nonsynonymous divergence will exceed synonymous if some proportion of nonsynonymous substitutions are advantageous. This test lacks power and will only detect positive selection when adaptation is frequent (Eyre-Walker 2006). However, the advantage of this test is that it can be applied in a phylogenetic framework that allows analysis of differences in rates of adaptive evolution among tree branches. Application to divergence between humans and chimpanzees gives an estimate of 0.08% of genes undergoing adaptive evolution (Clark *et al.* 2003) and in *D. melanogaster,* genes involved in reproduction show an excess of nonsynonymous change attributable to positive selection (Civetta & Singh 1995). Usually, though, nonsynonymous divergence is lower than that at synonymous sites, a result of the fact that most nonsynonymous sites are functionally constrained such that nonsynonymous mutations are deleterious (Eyre-Walker 2006). Alternatively, approaches that evaluate positive selection on individual codons (Nielsen & Yang 1998; Suzuki & Gojobori 1999) have resulted in estimates of 75-85% of codons being subject to positive selection within single genes in HIV (Nielsen & Yang 2003).

Expanding an analysis to include both intra- and interspecific data provides the most powerful test of adaptive evolution. Recently, an extension of the McDonald-Kreitman test has been used to estimate the proportion of fixations that are due to positive selection (α) by comparing the divergence for a putatively functional class of sites (e.g. nonsynonymous, intron, etc.) to that expected given polymorphism for that class of sites and the ratio of divergence to polymorphism for reference class of linked neutral sites

(usually synonymous) (Smith & Eyre-Walker 2002). An excess of fixation relative to polymorphism for the functional sites compared to neutral sites indicates that some portion of the fixations at functional sites are adaptive.

This method can be applied to any potentially functional class of sites (e.g. nonsynonymous, regulatory, intron) but it is sensitive to changes in population size and the segregation of weakly negatively selected polymorphism. While humans show evidence of recent selective sweeps in the form of reduced frequency polymorphism, it appears as though only a small proportion of nonsynonymous changes relative to chimpanzees were fixed as a result of positive selection (CSAC 2005; Zhang & Li 2005). In contrast, estimates in *D. melanogaster* and the closely-related *Drosophila simulans* indicate that upwards of half of all nonsynonymous divergence is adaptive (Bierne & Eyre-Walker 2004). Likewise, analysis of bacterial and mouse sequences also show levels of adaptive fixation of about 50% for nonsynonymous changes (Charlesworth & Eyre-Walker 2006; Halligan *et al*. 2010). A similar contrast is evident when this extension of the MK test is applied to noncoding sites. In humans, estimates of the proportion of changes fixed by adaptive evolution in 5' and 3' flanking regions is low, with an estimated α of 10-15% (Keightley *et al*. 2005) while analysis of noncoding sites in *Drosophila* shows frequent adaptive evolution in intronic, 5' and 3' flanking regions (Andolfatto 2005).

The McDonald-Kreitman framework for estimating α rests on assumptions that are potentially violated by differences in demographic history between the two species being compared. For instance, the fraction of new mutations that are neutral is estimated from polymorphism data in one species, and assumed to be the same for the other. This is potentially problematic, given that levels of polymorphism are influenced by population size (Eyre-Walker 2002; Sella et al. 2009) Recent analysis of mice (*M. m. castaneus*) indicates that the apparently slow rates of adaptive change in humans are a result of the small population size of humans, rather than a characteristic of the mammalian lineage

(Halligan *et al*. 2010). However, the relatively small number of taxa for which the rate of adaptive evolution has been estimated makes it difficult to disentangle the influence of demographic history on estimation of the number of fixations driven by positive selection.

Methods that use different types of data to estimate the rate of adaptive evolution allow detection of positive selection at different time-scales and intensities. Difference among estimation methods is one reason why accumulation of data has not resulted in a universal understanding of the rate of adaptive evolution. The examples detailed above where rates of adaptive evolution have been estimated in by multiple methods that utilize intra-and interspecific data as well as a combination show that adaptive evolution can affect polymorphism and divergence in different ways depending on the strength of selection and the age of advantageous variants. In addition, the two species that have been most thoroughly investigated, *Homo sapiens* and *Drosophila melanogaster,* have very complicated biogeographic histories, characterized by bottlenecks during colonization and subsequent expansion. Changes in population size associated with those colonization events likely influences estimates of the rate of adaptive evolution. Estimates from a variety of species are necessary in order to tease apart the influence of methodology and demographics on estimation of rates of adaptive evolution

This study provides an analysis of the role of positive selection in the evolution of coding and noncoding regions using intra- and interspecific data in an additional species of *Drosophila*, *Drosophila americana. D. americana* is endemic to much of the North American range that has been colonized by *D. melanogaster* in the last 10,000 years (Baudry et al. 2004; Haddrill et al. 2005). The effective population size of *D. americana* is estimated to be around 1×10^6, comparable to the value of estimated for African *D. melanogaster*, but unlike *D. melanogaster*, *D. americana* has a relatively uncomplicated biogeographic history. *D. americana* is estimated to have colonized North America ~3 mya (Caletka & McAllister 2004). Following its spread to the current range, the

biogeography of the species has likely been stable, with potential contraction and expansion of range size during and after glaciations events. Throughout this time, the species appears to have maintained a consistently large population size, similar to that of African *D. melanogaster*.

Here I search for signs of positive selection in the *D. americana* homologs of genes that display signs of adaptive sequence variation or are associated with adaptive phenotypic variation in *D. melanogaster.* This includes three metabolic genes: *Pgm, AcCoAS,* and *UGP* for which there is evidence of selection for locally adapted alleles (Verrelli & Eanes 2000; Montooth et al. 2006; Sezgin et al. 2004), and one gene involved in insulin signaling and implicated in diapause variation*, Pi3k92e* (Williams *et al*. 2006).

5.3 Materials and Methods

5.3.1 Sequencing

Female *D. americana* were collected from two locations representing northern (IR, Hawkeye Wildlife Area, near Iowa River, IA 41° 46.76' N, 91° 42.91' W) and southern (CI, Cat Island National Wildlife Refuge, LA, 30° 45.52' N, 91° 28.33' W) populations following the methods described in detail in McAllister *et al*. (2008). For *Pgm* and *Pi3k92e*, polymorphism and divergence data for two additional populations of intermediate latitude were also included in MK tests: DI (Illinois River at Duck Island, IL, 40° 27.26' N, 89° 56.69' W and RB (Pearl River Wildlife Management Area, MS, 32° 32.56' N, 89° 56.38' W).

The female flies were serially transferred until all sperm from insemination in the wild were depleted, and then crossed with males of a lab line of *Drosophila virilis* for which the genome sequence is available (V46). The resulting interspecific hybrid offspring were frozen at -20°C and DNA was later extracted.

Limited sequencing of inbred *D. americana* lines was used to identify fixed differences between *D. americana* and the *D. virilis* line V46. These differences were

used to design primers that specifically amplify from the *D. americana* chromosomes in *D. americana/D. virilis* interspecific hybrids. Most of the coding and intron sequence as well as some flanking noncoding sequence was amplified from *AcCoAS*, *UGP* and *Pi3k92e* in 6-10 chromosomes from interspecific hybrids generated from each sampled population using *D. americana*-specific primers. Sequences were determined on an ABI3730 using standard methods with BigDye chemistry. Resulting sequences were edited in Sequencher (GeneCodes) and aligned in BioEdit (Hall 1999).

5.3.2 Intraspecific analysis

Each gene was analyzed separately as well as in a concatenated dataset that included the three genes listed above and sequences from *Pgm* that were determined for a previous study (Chapter 4). Two measures of polymorphism, sequence diversity (π, Tajima 1983) and heterozygosity (θ, Waterson 1975) were determined for silent, noncoding, synonymous and nonsynonymous sites for each population separately and for the two populations combined.

One expectation of recent partial or complete selective sweeps is an alteration in the frequency of polymorphisms. In order to test whether polymorphism frequencies are consistent with neutral expectations, Tajima's *D* (Tajima 1989) was calculated for each type of site for the concatenated dataset. An excess of intermediate frequency polymorphisms with respect to neutral expectations is a sign of a partial selective sweep and will result in a positive value of Tajima's *D*. A complete selective sweep will generate an excess of low frequency polymorphisms that will result in a negative value of Tajima's *D*. Because excess low frequency polymorphism can also be generated by purifying selection, a test for an excess of high frequency derived polymorphisms brought near fixation due to linkage with a locus under positive selection was also carried out (Fay & Wu's *H,* Fay & Wu 2000). Such an excess results in a significantly negative value of *H* and is consistent with a selective sweep. Coalescent simulations using

segregating sites and assuming no recombination were used to generate confidence intervals to determine significance of Tajima's *D* and Fay & Wu's *H* values.

Balancing selection maintaining locally adapted variants may cause significant differentiation between northern and southern populations. To detect this, differentiation between the northern and southern population was evaluated using Kst. All calculations using intraspecific data were carried out using DNAsp version 5.10 (Librado & Rozas 2009).

5.3.3 Interspecific analysis

Rates of evolution among species were estimated using the ratio of nonsynonymous to synonymous divergence (dN/dS) in the Drosophila subgenus including *D. americana* and those species with available genome sequences including *D. virilis, D. grimshawi* and *D. mojavensis*. This analysis will provide show whether the patterns of adaptation we see in *D. americana* are typical for the subgenus. dN/dS for each branch of a phylogeny of the included species was estimated using a maximum likelihood approach for each gene individually and a concatenated alignment of all four genes. A single *D. americana* sequence was chosen randomly for inclusion in this analysis (CI05.38). Alignments were created manually except in the case of *Pi3k92e*, which has multiple indels among the included species. For *Pi3k92e*, an alignment was generated using MUSCLE software (Multiple Sequence Comparison by Log-Expectation, Edgar 2004).

Two models of evolutionary rate were fit to the data using the codeml application in PAML (Yang 2007). The first model holds dN/dS constant across the tree, while the second allows dN/dS to vary for each branch. The fit of the two models was compared with a log likelihood ratio test in order to determine whether the model allowing evolutionary rate to vary among branches fit the data significantly better than the one-rate model.

5.3.4 Combined analysis

A McDonald-Kreitman test (MK, McDonald and Kreitman 1991) with *D. virilis* as an outgroup was used to test whether the ratio of polymorphism to divergence is consistent between synonymous and nonsynonymous sites. An excess of divergence at nonsynonymous sites is indicative of positive selection. The proportion of changes at nonsynonymous, 5' and 3' flanking and intron sites that is due to positive selection was estimated using an extension of the MK test (Smith & Eyre-Walker 2002). This extension of the MK test relies on neutrality of changes at synonymous sites as well as a lack of purifying selection on the test sites (all other classes). *D. americana* is known to experience codon bias, so some polymorphisms at synonymous sites segregate at low frequency due to weak purifying selection against unpreferred codons. In addition, many polymorphisms at nonsynonymous sites are likely to be deleterious. In both cases, the negatively selected sites contribute an excess of polymorphism compared to divergence with respect to neutrally evolving sites. To remedy this, the data were analyzed using all the polymorphism data as well as after removing singleton polymorphisms.

Estimates of α were calculated in two ways for each class of sites for each gene and in four ways for each class of sites in a dataset including all genes. For each gene, α was calculated using a simple method with all polymorphism data and using the simple method after removal of singleton polymorphism. For the combined dataset, the data were summed across genes and the simple method was applied with and without singleton polymorphisms for each class of sites, and a maximum likelihood approach detailed in Bierne & Eyre-Walker (2004) was carried out with all polymorphism data and after removal of singleton polymorphisms. Maximum likelihood estimates were generated using DoFE (Distribution of Fitness Effects, Eyre-Walker 2005).

5.4 Results

5.4.1 Sequences

Sequences at *AcCoAS* consist of two sections, separated by a region of repetitive DNA within the large first intron. The first sequenced region is 1417 bp long and consists of noncoding sequence 5' of the first exon, the first exon and a portion of the first intron. For this region, 10 chromosomes were sequenced from each population. The second sequenced region includes a portion of the first intron, all of the next four exons and introns and part of the last exon for a total of 2314 bp. This second region was sequenced from 7 IR chromosomes and 6 CI chromosomes. Sequences of *UGP* included the entire coding region and intervening introns as well as sequence 5' and 3' of the coding region for a total of 2741 bp. *UGP s*equences were acquired from 10 IR chromosomes and 8 CI chromosomes. Sequences of *Pi3k92e* include sequence 5' of the coding region and most of the coding sequence and intervening noncoding sequence for a total of 3546 bp from 10 chromosomes each of four populations: IR, DI, RB and CI. Sequences of *Pgm* detailed in Chapter 4 were also analyzed. The concatenated dataset including *Pgm* is 12434 bp long and includes 7 IR and 6 CI chromosomes.

5.4.2 Intraspecific analyses

Silent sequence diversity and heterozygosity is similar for the two populations (IR and CI) at all sequenced loci. If either population had recently fixed an advantageous allele, I would expect to see differences in polymorphism between populations at the affected locus. The small differences in silent polymorphism that do exist between populations do not show a trend of increased polymorphism in either population, and silent site diversity and heterozygosity are nearly identical for the two populations in the concatenated dataset (π = 0.019 & 0.019, θ = 0.021 & 0.020, Table 5.1). A similar level of silent polymorphism in the two populations for the concatenated dataset indicates that effective population size in the two populations is approximately equal, and thus is

unaffected by potential seasonal bottlenecks in northern populations. Alternatively, any difference in silent polymorphism caused by a demographic difference could have been homogenized by gene flow.

Patterns of polymorphism are also consistent among genes, both within this study and compared with published results. In the combined datasets for each gene, silent site diversity varies between 0.011 and 0.025 with a value of 0.019 for the concatenated dataset (Table 5.1). A Mann-Whitney U test indicates that the level of silent diversity in these genes is not significantly different from that in a group of 10 genes on the second and third chromosomes analyzed in another study ($p = 0.19$, Maside & Charlesworth 2007).

The level of polymorphism at synonymous sites is approximately twice that at noncoding sites at each gene and in the concatenated dataset. This is also consistent with previous results in *D. americana* for autosomal loci. As noted above, nonsynonymous polymorphism is very low for all genes (Table 5.1).

Tajima's D for every class of sites for *UGP, Pi3k92e* and both regions of *AcCoAS* is negative (but not significantly so) except for synonymous sites in 3'AcCoAS, where it is near zero. It is reported in Chapter 3 that this is also true for *Pgm,* except that Tajima's D is significantly negative for nonsynonymous sites at that locus. Nonsynonymous sites show a marginally significantly negative value of Tajima's D in the concatenated dataset before correction for multiple tests ($D = -1.513$, $0.10 > p > 0.05$). The negative trend of Tajima's D shows that there are more polymorphisms at low frequency than expected at all types of sites. This is consistent with purifying selection acting on a subset of weakly deleterious mutations in each class. Low frequency polymorphisms could also result from a selective sweep, with neutral mutations rising to high frequency due to linkage with a positively selected allele, but no excess high frequency derived polymorphism was detected using Fay and Wu's *H* at any locus.

There is no differentiation evident between the two populations at any of the sequenced genes (Table 5.2). The lack of differentiation also runs counter to expectations for balancing selection on adaptive variants, and is consistent with the observed lack of intermediate frequency variation that is evidenced by negative values of Tajima's *D.*

5.4.3 Interspecific analyses

Evolutionary rates (dN/dS) were estimated using two different models for each gene and for the concatenated dataset: a model that allows dN/dS to vary among branches and one that fits a single dN/dS value to all branches. A likelihood ratio test shows that the variable rate model fits the data significantly better than a model with a single dN/dS value for three of the four genes, as well as for the concatenated dataset. After correction for multiple tests, the more complex model is still significantly better for *Pi3k92e, AcCoAS* and the concatenated dataset (Table 5.3).

When dN/dS rates are allowed to vary, they show that dN/dS is often very low for the *D. virilis* and *D. americana* branches (Figure 5.1). In the concatenated dataset, dN/dS for *D. grimshawii* and *D. mojavensis* is very similar (0.0563, 0.0558, respectively) and the branch leading to the common ancestor of *D. americana* and *D. virilis* is slightly elevated (0.0663). The elevated dN/dS on the branch leading to the common ancestor of *D. americana* and *D. virilis* coupled with the reduced dN/dS in the *D. americana* and *D. virilis* (0.0225 and 0.0127) suggests that the ancestral amino acid sequence has largely been conserved in *D. virilis* and *D. americana* since their divergence. In addition to a very low dN, the *D. virilis* and *D. americana* lineages both exhibit a low dS (0.06 and 0.059 vs. 0.571 and 0.469 for *D. grimshawii* and *D. mojavensis*)(Figure 5.1). Thus, it appears as though the ancestral coding sequence has changed very little in the *D. americana* and *D. virilis* lineages.

5.4.4 Combined analyses

The proportion of interspecific changes driven by positive selection (α) was estimated using an extension of the MK test that detects an excess of divergence relative to polymorphism at a class of functional sites compared to a class of neutral reference sites. In order to account for the presence of weakly negatively selected polymorphism, analyses were carried out using all polymorphism data and after removing all singleton polymorphism.

For *AcCoAS* and *UGP*, a lack of divergence or polymorphism made it impossible to calculate α for nonsynonymous sites. For the other two loci, *Pgm* and *Pi3k92e*, estimates of α for nonsynonymous sites are negative, regardless of inclusion or exclusion of singleton polymorphism. When the data for all four genes is analyzed together with a simple summation method or a maximum likelihood method, the estimates of α are always negative for nonsynonymous sites, regardless of the inclusion or exclusion of singleton polymorphism. This is consistent with previous estimates of adaptive evolution at nonsynonymous sites when comparing *D. americana* and *D. virilis*. It appears as though very little adaptive divergence has occurred at nonsynonymous sites in *D. americana* since it diverged from *D. virilis*.

Estimates of α for introns are positive for *AcCoAS* and *UGP* with and without singleton polymorphisms and at *Pgm* α is positive for introns with singletons included, but not after they are removed (Table 5.4). For all genes combined, α is positive and varies between 0.01 and 0.13 depending on the method used, but maximum likelihood estimates of 95% confidence intervals show that these values are not significantly different from zero (Table 5.4). Similarly, estimates of α for 5' flanking sites are positive for some genes and methods, but the maximum likelihood estimates for the combined dataset indicates that α is not significantly different from zero. Like nonsynonymous sites, it appears that very little divergence at noncoding sites is adaptive.

5.5 Discussion

The aim of this study was to gain an understanding of the role of positive selection in the evolution of a *Drosophila* species with a large population size and a stable demographic history. Like *D. melanogaster*, *D. americana* has a population size around 1 x 10^6, but it has a very different demographic history. The demographic history of *D. melanogaster* and the closely-related *D. simulans* is characterized by recent colonization of new habitats, resulting in bottlenecks followed by population expansions. *D. americana* has likely maintained a relatively constant biogeography since the species colonized North America ~ 3 mya (Caletka & McAllister 2004), resulting in fewer fluctuations in population size. This study provides an important addition to previous analyses of positive selection by allowing a contrast between similar species with different demographic histories.

Intraspecific analysis of *D. americana* does not show any evidence of recent positive selection resulting in partial or complete selective sweeps. Levels of polymorphism are similar across populations and genes and the frequencies of individual polymorphisms are consistent with most variation being neutral with some interspersed weakly deleterious variants. While the lack of evidence for selective sweeps might be a consequence of the loci chosen for analysis, it may also reflect a lack of recent positive selection in *D. americana.*. *D. melanogaster*, on the other hand, shows evidence of many selective sweeps (Li & Stephan 2006; Andolfatto 2007). This may reflect the difference in the time since colonization for the two species. Adaptation likely occurred during the recent colonization of temperate regions by *D. melanogaster*, while adaptation of *D. americana* populations to the same conditions occurred much longer ago, allowing the patterns of nucleotide variation caused by selective sweeps to diminish with time.

Interspecific analyses indicate that adaptive evolution at nonsynonymous sites slowed even before the colonization of North America by *D. americana.* Compared with other species in the *Drosophila* subgenus*, D. americana* and *D. virilis* appear to have a

very low rate of nonsynonymous change. This finding has been reported elsewhere (Maside & Charlesworth 2007), and it was suggested that it reflects increased purifying selection against nonsynonymous changes in those lineages.

Estimates of α, the proportion of divergence that represents adaptive change, are also consistent with very little adaptive divergence at nonsynonymous sites in the *D. americana* lineage since the split with *D. virilis*. This implies that nearly all nonsynonymous mutations are neutral or deleterious, so that mutations that change amino acid sequence are fixed at the same rate or less often than mutations at neutral sites. If nonsynonymous changes are primarily neutral or deleterious, noncoding changes might instead underlie adaptive differences between the two species. However, estimates of α in intron and 5' flanking regions are also very low. Overall, estimates of α indicate that there is very little excess divergence at any of the potentially functional sites analyzed, indicating that very few observed interspecific differences have been brought to fixation as a result of positive selection.

Intra- and interspecific data indicate that very little adaptive change has occurred since the divergence of *D. americana* and *D. virilis*. There is no evidence of selective sweeps at the analyzed loci, both the *D. americana* and *D. virilis* lineages exhibit a very low rate of amino acid change compared to other species in the *Drosophila* subgenus, and evidence of very slow adaptive divergence extends to noncoding sites. Selective sweeps are not necessarily expected given the small number of loci investigated, but the lack of adaptive divergence shown by an extension of the McDonald-Kreitman test stands in stark contrast to estimates of adaptive evolution in *D. melanogaster*.

One potential influence on α is the presence of segregating deleterious variants at nonsynonymous sites. In populations with a stable demographic history, this increases the ratio of polymorphism to divergence for nonsynonymous sites, and reduces α. The reduction in α in stable populations can be controlled for somewhat by removing singleton polymorphisms from the analysis, as they represent the most strongly

deleterious variants. When a population has undergone recent expansion, though, mutations that were fixed by drift in the smaller ancestral population can be miscategorized as adaptive, inflating values of α (Eyre-Walker 2002). Reduced polymorphism in temperate populations of *D. melanogaster* relative to African populations indicates that the species experienced bottlenecks in population size at the time of colonization, potentially increasing estimates of α (Baudry et al. 2004; Haddrill et al. 2005). While a recent increase in population size may well be responsible for an inflated estimate of α in *D. melanogaster*, changes in population size do not explain the high values of α in *M. m. castaneus,* because polymorphism and divergence were determined using populations from the ancestral range of the species, rather than populations commensal with humans (Halligan et al. 2010).

Given the data presented here, it seems premature to infer that most species undergo adaptive evolution at a rate similar to that of *D. melanogaster*. Analysis has revealed species with relatively high rates of adaptive evolution: some *Drosophila* (Bierne & Eyre-Walker 2004), *M. m castaneus* (Halligan et al.2010) and enteric bacteria (Charlesworth & Eyre-Walker 2006), as well as some species with apparently low rates of adaptive evolution: *D. americana*, hominids (Zhang & Li 2005), yeast (Doniger et al. 2008; Liti et al. 2009) and *Arabidopsis* (Bustamante 2002). Analyses of additional species that differ in demographic history are necessary to establish an understanding of the factors influencing estimates of adaptive evolution in natural populations, and to determine the typical level of adaptive fixation, if it exists.

Table 5.1 Measures of polymorphism for each population and the combined sample.

		IR					CI					total				
Gene		N	L	S	π	θ	N	L	S	Π	θ	L	S	π	θ	T's D
Pgm	sil	9	1878.1	144	0.024	0.028	10	1831.9	173	0.028	n.a.	1793.9	219	0.026	n.a.	-1.08
	nsyn		1170.9	8	0.002	0.003		1169.1	8	0.002	n.a.	1167.1	13	0.001	n.a	-2.05*
	syn		372.1	45	0.039	0.044		372.9	49	0.042	n.a.	371.9	65	0.040	n.a	-0.82
	non		1506	99	0.021	0.024		1459	124	0.024	0.030	1422	154	0.022	0.034	-1.16
UGP	sil	10	1651.6	83	0.015	0.018	8	1591.9	70	0.015	0.017	1581.7	110	0.015	0.020	-1.13
	nsyn		1159.4	2	0.000	0.001		1159.1	0	0.000	0.000	1159.3	2	0.000	0.001	-1.51
	syn		352.6	31	0.026	0.031		352.9	23	0.022	0.025	352.7	41	0.025	0.034	-1.09
	non		1299	52	0.012	0.014		1239	47	0.013	0.015	1229	69	0.012	0.016	-1.12
5'AcCoAS	sil	10	1183.5	38	0.010	0.011	10	1198.4	42	0.011	0.012	1172.4	58	0.011	0.014	-1.01
	nsyn		242.6	0	0.000	0.000		242.6	1	0.001	0.001	242.6	1	0.000	0.001	-1.16
	syn		66.5	2	0.011	0.011		66.4	1	0.003	0.005	66.4	2	0.007	0.008	-1.16
	non		1117	36	0.010	0.011		1132	41	0.011	0.013	1106	56	0.011	0.014	-1.03
3'AcCoAS	sil	7	1349.5	90	0.028	0.027	6	1340.5	82	0.026	0.027	1332.6	111	0.025	0.027	-0.24
	nsyn		978.5	0	0.000	0.000		978.5	0	0.000	0.000	978.5	0	0.000	0.000	n.a.
	syn		308.5	32	0.045	0.042		308.5	29	0.042	0.041	308.6	41	0.044	0.043	0.08
Pi3k92e	sil	10	1356.8	70	0.016	0.018	10	1360.0	65	0.014	0.017	1354.9	98	0.015	0.020	-1.06
	nsyn		2190.2	6	0.001	0.001		2190.0	3	0.000	0.000	2190.1	6	0.001	0.001	-0.35
	syn		641.8	45	0.023	0.025		642.0	42	0.019	0.023	641.9	61	0.021	0.027	-0.82
	non		715	25	0.010	0.012		718	23	0.009	0.011	713	37	0.009	0.015	-1.39
All	sil	7	7130.7	366	0.019	0.021	6	7154.9	334	0.019	0.020	7074.7	511	0.019	0.023	-0.89
	nsyn		5355.4	13	0.001	0.001		5355.2	10	0.001	0.001	5355.4	19	0.001	0.001	-1.51#
	syn		1622.6	132	0.030	0.033		1622.8	110	0.028	0.030	1622.6	174	0.029	0.035	-0.76
	non		5508	145	0.016	0.017		5532	224	0.016	0.018	5452	337	0.016	0.020	-0.95

Note: Sites highlighted in bold were investigated in additional populations.

Table 5.2 Numbers of shared and exclusive polymorphisms and differentiation between northern (IR) and southern (CI) samples.

	shared	exclusive	Kst	p
Pgm	87	162	0.006	0.2
UGP	35	78	0.023	0.06
5'AcCoAS	39	21	-0.009	0.7
3'AcCoAS	58	54	-0.011	0.7
Pi3k92e	64	40	0.005	0.3

Table 5.3Log-likelihood values for models assuming one dN/dS ratio and allowing different dN/dS on each branch.

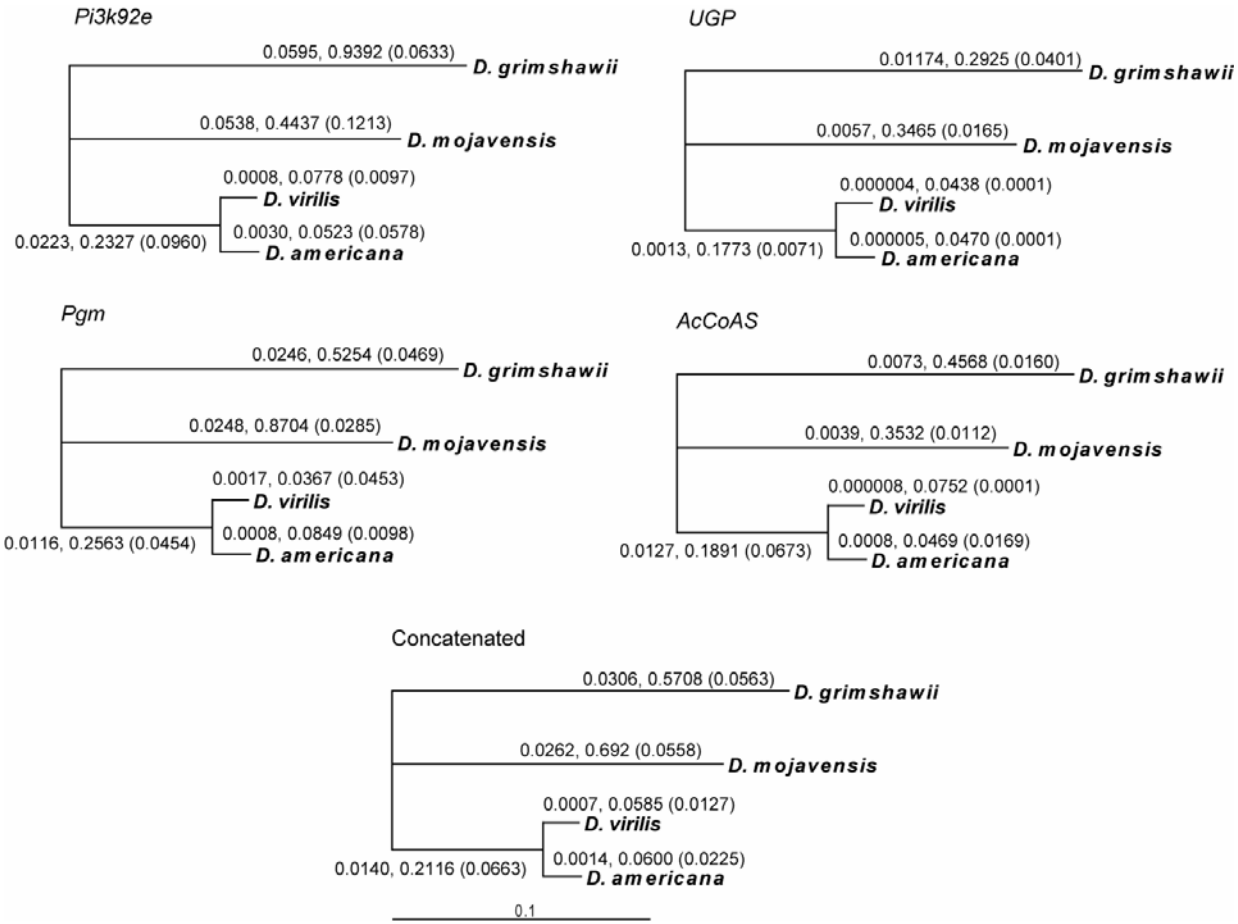
	lnL: one	lnL: free	p
Pgm	-3600.48	-3597.93	0.3
UGP	-3004.96	-2999.49	0.03
AcCoAS	-3248.98	-3240.44	0.002
Pi3k92e	-7204.72	-7192.58	0.0001
all	-17284.7	-17273.8	0.0002

Table 5.4 Estimates of α.

		div	poly(-S)	α	αS	α_{ML}	αS_{ML}
AcCoAS	nsyn	1	0(0)	---	---		
	Syn	17	29(19)	---	---		
	5'	8	10(5)	0.27	0.44		
	Int	84	77(41)	0.46	0.56		
UGP	nsyn	0	2(0)	---	---		
	Syn	22	41(20)	---	---		
	5'	4	13(7)	-0.74	-0.93		
	Int	35	51(29)	0.22	0.09		
	3'	9	5(1)	0.7	0.88		
Pi3k92e	nsyn	4	11(4)	-1.36	-0.24		
	Syn	67	78(54)	---	---		
	5'	18	23(10)	-0.09	0.31		
	Int	17	30(26)	-0.51	-0.9		
Pgm	nsyn	3	23(7)	-1.26	-0.12		
	Syn	25	85(52)	---	---		
	5'	41	97(62)	0.3	0.27		
	Int	36	108(83)	0.12	-0.11		
All	nsyn	8	36(11)	-1.53	-0.24	-1.16 (-4.25, 0.04)	-0.03 (-1.83, 0.62)
	Syn	71	143(84)	---	---	---	---
	5'	172	266(179)	-0.13	-0.07	0.11 (-0.31, 0.40)	0.22 (-0.21, 0.51)
	Int	131	233(145)	0.13	0.06	0.13 (-0.19, 0.38)	0.01 (-0.42, 0.31)

Note: Simple summation estimates for each class of sites for each gene and the combined dataset, with (α) and without (αS) singleton polymorphism, and maximum likelihood estimates with 95% confidence intervals for the combined dataset using the method of Bierne and Eyre-Walker (2004), with and without singleton polymorphism.

Figure 5.1 Estimated dN, dS and dN/dS values for each gene and for a concatenated dataset.



CHAPTER 6

CONCLUSIONS

The study of repeated evolution is rife with controversy over the proper categorization of parallel vs. convergent events (Williams & Ebach 2006; Arendt & Reznick 2007; Desutter-Grandcolas et al. 2007; Leander 2008). While that controversy is far from resolved, discovery of repeated evolution at the genetic level is becoming more commonplace. Reports of striking examples of repeated evolution raise questions about the frequency of and the factors influencing repeated evolution. Frequently, parallel genetic adaptation is found to underlie low complexity traits. With the exception of parallel changes in pigmentation genes, most instances are likely constrained such that only changes in a few specific genes could conceivably result in the adaptive trait (Gompel & Prud'homme 2009). It is unclear, however, if this phenomenon is a result of sampling bias or if it is a result of a requirement for strong selection and low complexity in parallel adaptation.

Our understanding of repeated evolution will benefit from study of a variety of systems in which the presence of either similar traits or similar selection pressures indicates the presence of parallel evolution, and the trait complexity varies. The male-killing *Wolbachia* strain characterized in Chapter 2 provides a component of just such a system. Genetic and phylogenetic analyses presented here show that the male-killing *Wolbachia* strain infecting *Drosophila borealis* is very closely related to a male-killing strain of *Wolbachia* infecting *Drosophila innubila*. This host-parasite interaction, in conjunction with the already-characterized male-killing infections in *D. innubila* and *Drosophila bifasciata,* provides a network within which comparisons of the genetic mechanisms of male-killing in multiple *Wolbachia* lineages and adaptation of multiple hosts to similar parasites can be investigated.

In Chapters 3-5, I directly investigated the role of parallel evolution in adaptation to a climatic gradient in *Drosophila americana* and *Drosophila melanogaster*. The two *Drosophila* species inhabit a similar North American range and both show variation in cold tolerance and diapause (Hoffman et al. 2001; Schmidt et al. 2005; unpublished data). However, *D. americana* is endemic to the region while *D. melanogaster* is a recent invader. Analysis of variation at five loci for which there is evidence of local adaptation in *D. melanogaster* revealed no evidence that variation in those genes plays a similar role in *D. americana.*

At the *Alcohol dehydrogenase* (*Adh*) locus investigated in Chapter 3, very low frequency sequence variation in *D. americana* indicates purifying selection across the *Adh* region, especially in introns. Comparison with closely-related species *Drosophila lummei* and *Drosophila virilis* indicates that the intron sequences are not only constrained within species, but also conserved among species. The *Adh* locus of *D. americana* segregates very little neutral variation, and there is no evidence for locally adaptive variants. These results contrast with the presumed role of *Adh* in local adaptation in *D. melanogaster*. If the two species are experiencing similar selection pressures on ethanol tolerance, they are responding differently at the genetic level. The lack of variation at the *Adh* locus of *D. americana* makes it a bad candidate for the target of selection that maintains the nearby chromosomal fusion polymorphism.

The *Phosphoglucomutase* (*Pgm*) locus investigated in Chapter 4 segregates more variation than the *Adh* locus, but that variation is likely neutral, rather than adaptive. Parallels between *D. americana* and *D. melanogaster* can be drawn at this locus: Both species segregate many amino acid variants, and both species segregate variants clinally. However, patterns of variation throughout the gene indicate that most polymorphism is neutral and that the primary selective force acting on amino acid variants is purifying selection. Furthermore, noncoding clinal variants have at most a slight effect on PGM activity and starvation resistance, and are likely neutral as well. The taxonomically

widespread variation observed at the *Pgm* locus is likely a result of excess metabolic capacity that makes many alleles selectively equivalent, resulting in relaxed selection on amino acid sequence.

In Chapter 5 I investigated the role of positive selection in the evolution of a group of genes involved in metabolism and growth. Each of the investigated genes shows signs of selection on locally adapted variants in *D. melanogaster*. In *D. americana*, these genes show little evidence of positive selection for adaptive variants or adaptive divergence. Estimates of the proportion of changes that have been fixed by positive selection are low for nonsynonymous and noncoding sites, indicating that most divergence from *D. virilis* is a result of fixation of neutral variants. Estimates of the proportion of changes driven by positive selection in *D. melanogaster* are much higher. The observed difference may be a result of demographic differences between the two species that influence estimates.

Together, these analyses show little evidence of parallel adaptation to climate between *D. melanogaster* and *D. americana* at the level of the gene. The two species inhabit a similar range, display similar phenotypic variation across that range, and their genomes likely encode homologs of most genes (*Drosphila* 12 Genomes Consortium 2007). But, while the substrate and selection are similar, the genes investigated here underlie a complex trait and the two species differ in demographic history.

Adaptation to climate is frequently attributed to variation in metabolism and growth, two processes that are the result of complex pathways. There may be many different evolutionary routes that allow adaptation under the same selective pressures. For example, Flybase lists 119 genes involved in glucose metabolism, and 256 with alcohol dehydrogenase activity (Tweedie et al. 2009). Because parallel genetic change is likely negatively correlated with trait complexity (Gompel & Prud'homme 2009), its absence in the investigated genes is not altogether unexpected.

The distinct demographic histories of the North American populations of two species may also play a role. Currently, they inhabit a similar range but historically they have very different origins. *D. melanogaster* in North America represent a recent colonization from an ancestral African population, and show signs of population bottlenecks (Baudry et al. 2004; Haddrill et al. 2005). This resulted in a reduction in polymorphism relative to African populations. So, even if selective pressures are the same for North American *D. melanogaster* and *D. americana*, the substrates for evolutionary change are fundamentally different in terms of divergence as well as standing variation. In the case of *D. americana* and *D. melanogaster*, it is likely that contingency has played a part in preventing parallel adaptation in metabolic pathways. Parallel adaptation can occur in species that are distantly-related, but the probability likely decreases with evolutionary distance.

Often, it is tempting to infer parallel genetic change for closely-related species with independently derived similar phenotypes, and to infer convergent genetic change when the evolutionary relationship is more distant. Instead, repeated adaptation should be investigated in many systems of varying evolutionary relatedness, and the results should be framed in terms of the level of homology at which a trait is parallel. While the results reported here do not indicate parallel change at the level of the gene, they highlight the importance of reporting negative as well as positive results in the study of repeated evolution. Without proper documentation in a multitude of systems, we will be unable to make inferences about the importance of repeated adaptation in evolution.

APPENDIX

Table A.1 GenBank accession numbers for *Wolbachia* sequences

Host/strain	*wsp*	*ftsZ*	*gatB*	*coxA*	*hcpA*	*fbpA*
A clade						
Glossina morsitans centralis ωCen	AF020078	--	--	--	--	--
Bactrocera dorsalis ωDor	DQ834379	--	--	--	--	--
Cadra cautella ωCauA	AF020075	--	--	--	--	--
D. bifasciata	AJ271121	--	--	--	--	--
Drosophila auraria ωRi	AF020062	--	--	--	--	--
Drosophila borealis	FJ415468	FJ415469	FJ415471	FJ415470	FJ415472	FJ415473
Drosophila innubila	AY552553	DQ842316	DQ842428	DQ842280	DQ842391	DQ842354
Drosophila melanogaster ωMel	AF020072	DQ842340	DQ842452	DQ842304	DQ842415	DQ842378
Drosophila melanogaster ωMelCS	AF020065	--	--	--	--	--
Drosophila melanogaster ωMelH	AF020066	--	--	--	--	--
Drosophila recens	AY154399	DQ842319	DQ842431	DQ842283	DQ842394	DQ842357
Drosophila sechellia ωHa	AF020073	--	--	--	--	--
Drosophila simulans ωAu	DQ235409	DQ235342	DQ842432	DQ842284	DQ842395	DQ842358
Drosophila simulans ωCof	AF020067	--	--	--	--	--
Drosophila simulans ωHa	AF020068	--	--	--	--	--
Drosophila simulans ωRi	AF020070	--	--	--	--	--
Glossina morsitans morsitans ωMors	AF020079	--	--	--	--	--
Muscidifurax uniraptor ωUni	AF020071	--	--	--	--	--
Nasonia vitripenis ωVit	AF020081	DQ842333	DQ842445	DQ842297	DQ842408	DQ842371
Pleistodontes imperialis	AY567684	--	--	--	--	--

Table A.1 continued

Host/strain	*wsp*	*ftsZ*	*gatB*	*coxA*	*hcpA*	*fbpA*
B clade						
Acraea encedon T	AJ271198	--	--	--	--	--
Acraea encedon U	AJ130716	--	--	--	--	--
Acraea eponina	AJ271194	--	--	--	--	--
Adalia bipunctata Z	AJ130715	--	--	--	--	--
*Adalia bipunctata*Y	AJ130714	--	--	--	--	--
Aedes albopictus ωAlbB	AF020059	--	--	--	--	--
Cadra cautella ωCauB	AF020076	--	--	--	--	--
Coleomegilla maculata fuscilabris	AF217724	--	--	--	--	--
Culex pipiens ωPip	AF020061	--	--	--	--	--
Culex quinquefasciatus ωPip	AF020060	--	--	--	--	--
Drosophila simulans ωMa	AF020069	--	--	--	--	--
Drosophila simulans ωNo	AF020074	--	--	--	--	--
Hypolimnas bolina	AJ307076	--	--	--	--	--
*Laodelphax striatellus*ωStri	AF020080	--	--	--	--	--
Polygonia c-aureum	AB094378	--	--	--	--	--
*Triblium confusum*ωCon	AF020083	--	--	--	--	--
*Trichogramma Deion*ωDei	AF020084	--	--	--	--	--
C clade						
Brugia pahangi	AY527208	--	--	--	--	--
Wuchereria bancrofti	AJ252180	--	--	--	--	--

Table A.2 GenBank accession numbers for *Drosophila* sequences.

Gene	Dinn	Drec	Dbor	Dvir	Dmoj	Dgri	Dwil	Dpse	Dana	Dyak	Dsim	Dmel
COI	AF519389	DQ851807	EU390720	ref	ref	ref	ref	--	ref	ref	ref	ref
COII	AF519325	AF147123	EU390740	ref	ref	ref	ref	ref	ref	ref	ref	ref
period*	AY541174	EF188978	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
vermillion*	AY541152	EF189062	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Tpi*	AY541236	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG2789*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG2794*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG2819*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG2848*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG3238*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG3883*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG4184*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG4644*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG9663*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG9883*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG9961*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG14381*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG17262*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref
Dmel\CG17660*	--	--	--	ref	ref	ref	ref	ref	ref	ref	ref	ref

Note: reference sequences obtained from the alignments of Rasmussen & Kellis (2007). (http://compbio.mit.edu/spidir/)

REFERENCES

Akashi H, and Schaeffer SW. (1997). Natural selection and the frequency distributions of "silent" DNA polymorphism in Drosophila. Genetics 146:295-307.

Andolfatto P (2007). Hitchhiking effects of recurrent beneficial amino acid substitutions in the Drosophila melanogaster genome. Genome Res. 17:1755-1762.

--- (2005). Adaptive evolution of non-coding DNA in Drosophila. Nature 437:1149-1152.

Arendt J, and Reznick D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends Ecol.Evol. 23:26-32.

Ayala FJ, Powell JR, Tracey ML, Mourao CA, and Perez-Salas S. (1972). Enzyme variability in the Drosophila willistoni group. IV. Genic variation in natural populations of Drosophila willistoni. Genetics 70:113-139.

Baldo L, and Werren JH. (2007). Revisiting Wolbachia supergroup typing based on WSP: spurious lineages and discordance with MLST. Curr.Microbiol. 55:81-87.

Baldo L, Lo N, and Werren JH. (2005). Mosaic nature of the wolbachia surface protein. J.Bacteriol. 187:5406-5418.

Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, et al. (2006). Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. Appl.Environ.Microbiol. 72:7098-7110.

Baudry E, Viginier B, and Veuille M. (2004). Non-African populations of Drosophila melanogaster have a unique origin. Mol.Biol.Evol. 21:1482-1491.

Berry A, and Kreitman M. (1993). Molecular analysis of an allozyme cline: alcohol dehydrogenase in Drosophila melanogaster on the east coast of North America. Genetics 134:869-893.

Bierne N, and Eyre-Walker A. (2004). The genomic rate of adaptive amino acid substitution in Drosophila. Mol.Biol.Evol. 21:1350-1360.

Burkhart BD, Montgomery E, Langley CH, and Voelker RA. (1984). Characterization of Allozyme Null and Low Activity Alleles from Two Natural Populations of Drosophila melanogaster. Genetics 107:295-306.

Burns JM, and Johnson FM. (1967). Esterase polymorphism in natural populations of a sulfur butterfly, Colias eurytheme. Science 156:93-96.

Bustamante CD, Nielsen R, Sawyer SA, Olsen KM, Purugganan MD, and Hartl DL. (2002). The cost of inbreeding in Arabidopsis. Nature 416:531-534.

Cai JJ, Macpherson JM, Sella G, and Petrov DA. (2009). Pervasive hitchhiking at coding and regulatory sites in humans. PLoS Genet. 5:e1000336.

Caletka BC, and McAllister BF. (2004). A genealogical view of chromosomal evolution and species delimitation in the Drosophila virilis species subgroup. Mol.Phylogenet.Evol. 33:664-670.

Carlini DB (2004). Experimental reduction of codon bias in the Drosophila alcohol dehydrogenase gene results in decreased ethanol tolerance of adult flies. J.Evol.Biol. 17:779-785.

Carlini DB, and Stephan W. (2003). In vivo introduction of unpreferred synonymous codons into the Drosophila Adh gene results in reduced levels of ADH protein. Genetics 163:239-243.

Carson HL (1956). A female-producing strain of D. borealis Patterson. D.I.S. 30:109-110.

Carter PA, and Watt WB. (1988). Adaptation at specific loci. V. Metabolically adjacent enzyme loci may have very distinct experiences of selective pressures. Genetics 119:913-924.

Charlat S, Hurst GD, and Mercot H. (2003). Evolutionary consequences of Wolbachia infections. Trends Genet. 19:217-223.

Charlesworth D (2006). Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. 2:e64.

Charlesworth J, and Eyre-Walker A. (2006). The rate of adaptive evolution in enteric bacteria. Mol.Biol.Evol. 23:1348-1356.

Chen Y, and Stephan W. (2003). Compensatory evolution of a precursor messenger RNA secondary structure in the Drosophila melanogaster Adh gene. Proc.Natl.Acad.Sci.U.S.A. 100:11499-11504.

Chimpanzee Sequencing and Analysis Consortium (CSAC) (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69-87.

Christin PA, Salamin N, Savolainen V, Duvall MR, and Besnard G. (2007). C4 Photosynthesis evolved in grasses via parallel adaptive genetic changes. Curr.Biol. 17:1241-1247.

Civetta A, and Singh RS. (1995). High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in Drosophila melanogaster and Drosophila virilis group species. J.Mol.Evol. 41:1085-1095.

Clark AG, Glanowski S, Nielsen R, Thomas P, Kejariwal A, Todd MJ, et al. (2003). Positive selection in the human genome inferred from human-chimp-mouse orthologous gene alignments. Cold Spring Harb.Symp.Quant.Biol. 68:471-477.

Desalle R (1992). The Origin and Possible Time of Divergence of the Hawaiian Drosophilidae - Evidence from Dna-Sequences. Mol.Biol.Evol. 9:905-916.

Desutter-Grandcolas L, Legendre F, Grandcolas P, and et al. (2007). Distinguishing between convergence and parallelism is central to comparative biology: a reply to Williams and Ebach. 23:90-94.

DiMichele L, and Powers DA. (1982). Physiological basis for swimming endurance differences between LDH-B genotypes of Fundulus heteroclitus. Science 216:1014-1016.

--- (1982). LDH-B genotype-specific hatching times of Fundulus heteroclitus embryos. Nature 296:563-564.

DiMichele L, Paynter KT, and Powers DA. (1991). Evidence of lactate dehydrogenase-B allozyme effects in the teleost, Fundulus heteroclitus. Science 253:898-900.

Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, Yang SP, et al. (2008). A catalog of neutral and deleterious polymorphism in yeast. PLoS Genet. 4:e1000183.

Drosophila 12 Genomes Consortium, Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, et al. (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203-218.

Duret L, and Mouchiroud D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, Arabidopsis. Proc.Natl.Acad.Sci.U.S.A. 96:4482-4487.

Dyer KA, and Jaenike J. (2004). Evolutionarily stable infection by a male-killing endosymbiont in Drosophila innubila: molecular evidence from the host and parasite genomes. Genetics 168:1443-1455.

Dyson EA, Kamath MK, and Hurst GD. (2002). Wolbachia infection associated with all-female broods in Hypolimnas bolina (Lepidoptera: Nymphalidae): evidence for horizontal transmission of a butterfly male killer. Heredity 88:166-171.

Eanes A (1999). Analysis of selection on enzyme polymorphisms. Annu Rev Ecol Syst 30:301-326.

Eanes WF, Merritt TJ, Flowers JM, Kumagai S, Sezgin E, and Zhu CT. (2006). Flux control and excess capacity in the enzymes of glycolysis and their relationship to flight metabolism in Drosophila melanogaster. Proc.Natl.Acad.Sci.U.S.A. 103:19413-19418.

Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792-1797.

Endler JA (1973). Gene flow and population differentiation. Science 179:243-250.

Evans AL, Mena PA, and McAllister BF. (2007). Positive selection near an inversion breakpoint on the neo-X chromosome of Drosophila americana. Genetics 177:1303-1319.

Eyre-Walker A (2006). The genomic rate of adaptive evolution. Trends Ecol.Evol. 21:569-575.

--- (2002). Changing effective population size and the McDonald-Kreitman test. Genetics 162:2017-2024.

Fay JC, and Wu CI. (2000). Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.

Feldman CR, Brodie ED,Jr, Brodie ED,3rd, and Pfrender ME. (2009). The evolutionary origins of beneficial alleles during the repeated adaptation of garter snakes to deadly prey. Proc.Natl.Acad.Sci.U.S.A. 106:13415-13420.

Freriksen A, Seykens D, Scharloo W, and Heinstra PW. (1991). Alcohol dehydrogenase controls the flux from ethanol into lipids in Drosophila larvae. A 13C NMR study. J.Biol.Chem. 266:21399-21403.

Fry JD, Donlon K, and Saweikis M. (2008). A worldwide polymorphism in aldehyde dehydrogenase in Drosophila melanogaster: evidence for selection mediated by dietary ethanol. Evolution Int.J.Org.Evolution 62:66-75.

Fu YX, and Li WH. (1993). Statistical tests of neutrality of mutations.Genetics 133:693-709.

Gibson JB, Cao A, Symonds J, and Reed D. (1991). Low activity sn-glycerol-3-phosphate dehydrogenase variants in natural populations of Drosophila melanogaster. Heredity 66 (Pt 1):75-82.

Gillespie JH. (2000) Genetic drift in an infinite population. The pseudohitchhiking model. Genetics 155: 909-919.

Gompel N, and Prud'homme B. (2009). The causes of repeated genetic evolution. Dev.Biol. 332:36-47.

Gould S. J. 1989. Wonderful Life: The Burgess Shale and the Nature of History. W. W. Norton & Co. Ltd., London.

Gross JB, Borowsky R, and Tabin CJ. (2009). A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish Astyanax mexicanus. PLoS Genet. 5:e1000326.

Haddrill PR, Charlesworth B, Halligan DL, and Andolfatto P. (2005). Patterns of intron sequence evolution in Drosophila are dependent upon length and GC content. Genome Biol. 6:R67.

Haddrill PR, Thornton KR, Charlesworth B, and Andolfatto P. (2005). Multilocus patterns of nucleotide variability and the demographic and selection history of Drosophila melanogaster populations. Genome Res. 15:790-799.

Haldane J. B. S. 1932. The Causes of Evolution. Longman, Green & Co. Limited, London.

Halligan DL, Oliver F, Eyre-Walker A, Harr B, and Keightley PD. (2010). Evidence for pervasive adaptive protein evolution in wild mice. PLoS Genet. 6:e1000825.

Heath BD, Butcher RD, Whitfield WG, and Hubbard SF. (1999). Horizontal transfer of Wolbachia between phylogenetically distant insect species by a naturally occurring mechanism. Curr.Biol. 9:313-316.

Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, and Werren JH. (2008). How many species are infected with Wolbachia?--A statistical analysis of current data. FEMS Microbiol.Lett. 281:215-220.

Hoekstra HE, and Nachman MW. (2003). Different genes underlie adaptive melanism in different populations of rock pocket mice. Mol.Ecol. 12:1185-1194.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, and Crossland JP. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313:101-104.

Hoffmann RJ (1985). Properties of allelic variants of phosphoglucomutase from the sea anemone Metridium senile. Biochem.Genet. 23:859-876.

--- (1981). Evolutionary genetics of Metridium senile. II. Geographic patterns of allozyme variation. Biochem.Genet. 19:145-154.

Hoffmann AA, Hallas R, Sinclair C, Mitrovski P (2001) Levels of variation in stress resistance in Drosophila among strains, local populations, and geographic regions: patterns for desiccation, starvation, cold resistance and associated traits. Evolution, 55, 1621–1630

Hornett EA, Duplouy AM, Davies N, Roderick GK, Wedell N, Hurst GD, et al. (2008). You can't keep a good parasite down: evolution of a male-killer suppressor uncovers cytoplasmic incompatibility. Evolution 62:1258-1263.

Hornett EA, Charlat S, Duplouy AM, Davies N, Roderick GK, Wedell N, et al. (2006). Evolution of male-killer suppression in a natural population. PLoS Biol. 4:e283.

Hudson RR, Kreitman M, and Aguade M. (1987). A test of neutral moleculat evolution based on nucleotide data. Genetics 116:153-159.

Huelsenbeck JP, and Ronquist F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Hurst LD (1991). The incidences and evolution of cytoplasmic male killers. Proc. R. Soc. Lond. B 244:91-99.

Jolley KA, Chan MS, and Maiden MC. (2004). mlstdbNet - distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics 5:86.

Jost MC, Hillis DM, Lu Y, Kyle JW, Fozzard HA, and Zakon HH. (2008). Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. Mol.Biol.Evol. 25:1016-1024.

Kamping A, and van Delden W. (1999). A long-term study on interactions between the Adh and alpha Gpdh allozyme polymorphisms and the chromosomal inversion In(2L)t in a seminatural population of D-melanogaster. J.Evol.Biol. 12:809-821.

Kaplan NL, Hudson, RR, Langley CH. The “hitchhiking effect” revisited. Genetics 123: 887-899.

Keightley PD, Lercher MJ, and Eyre-Walker A. (2005). Evidence for widespread degradation of gene control regions in hominid genomes. PLoS Biol. 3:e42.

Kim Y, Stephan W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160: 765-777.

Kirby DA, Muse SV, and Stephan W. (1995). Maintenance of pre-mRNA secondary structure by epistatic selection. Proc.Natl.Acad.Sci.U.S.A. 92:9047-9051.

Kreitman M, and Hudson RR. (1991). Inferring the Evolutionary Histories of the Adh and Adh-Dup Loci in Drosophila-Melanogaster from Patterns of Polymorphism and Divergence. Genetics 127:565-582.

Laurie CC, Bridgham JT, and Choudhary M. (1991). Associations between DNA sequence variation and variation in expression of the Adh gene in natural populations of Drosophila melanogaster. Genetics 129:489-499.

Leander BS (2008). Different modes of convergent evolution reflect phylogenetic distances: a reply to Arendt and Reznick. Trends Ecol.Evol. 23:481-2; author reply 483-4.

Leigh Brown AJ (1977). Physiological correlates of an enzyme polymorphism. Nature 269:803-804.

Lewontin RC, and Hubby JL. (1966). A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. Genetics 54:595-609.

Li H, and Stephan W. (2006). Inferring the demographic history and rate of adaptive substitution in Drosophila. PLoS Genet. 2:e166.

Librado P, and Rozas J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. (2009). Population genomics of domestic and wild yeasts. Nature 458:337-341.

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

Maside XL, Lee AWS, and Charlesworth B. (2004). Selection on codon usage in Drosophila americana. 14:150-154.

Mateos M, Castrezana SJ, Nankivell BJ, Estes AM, Markow TA, and Moran NA. (2006). Heritable endosymbionts of Drosophila. Genetics 174:363-376.

Matzkin LM (2004). Population genetics and geographic variation of alcohol dehydrogenase (Adh) paralogs and glucose-6-phosphate dehydrogenase (G6pd) in Drosophila mojavensis. Mol.Biol.Evol. 21:276-285.

Maynard Smith J, and Haigh J. (1974). The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23-35.

McAllister BF (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 BF, and Evans AL. (2006). Increased nucleotide diversity with transient Y linkage in Drosophila americana. PLoS ONE 1:e112.

McAllister BF, and Charlesworth B. (1999). Reduced sequence variability on the neo-Y chromosome of Drosophila americana americana. Genetics 153:221-233.

McAllister BF, Sheeley SL, Mena PA, Evans AL, and Schlotterer C. (2008). Clinal distribution of a chromosomal rearrangement - a precursor to chromosomal speciation? 62:1852-1865.

McCracken KG, Barger CP, Bulgarella M, Johnson KP, Sonsthagen SA, Trucco J, et al. (2009). Parallel evolution in the major haemoglobin genes of eight species of Andean waterfowl. Mol.Ecol. 18:3992-4005.

Mcdonald JH, and Kreitman M. (1991). Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652-654.

Mena P (2009). The role of chromosomal rearrangements in adaptation of Drosophila americana. University of Iowa.

Merçot H, Defaye D, Capy P, Pla E, and David JR. (1994). Alcohol tolerance, ADH activity, and ecological niche of Drosophila species. Evolution 48:746-757.

Montooth KL, Siebenthall KT, and Clark AG. (2006). Membrane lipid physiology and toxin catabolism underlie ethanol and acetic acid tolerance in Drosophila melanogaster. J.Exp.Biol. 209:3837-3850.

Mundy NI (2005). A window on the genetics of evolution: MC1R and plumage colouration in birds. Proc.Biol.Sci. 272:1633-1640.

Nachman MW (2005). The genetic basis of adaptation: lessons from concealing coloration in pocket mice. Genetica 123:125-136.

Nielsen R and Yang Z. (1998). Likelihood models for detecting positively selected amino acid sites and application to the HIV-1 envelope gene. Genetics 148:929-936.

Nielsen R and Yang Z. (2003). Estimating the distribution of selection coefficients from phylogenetic data with applications to mitochondrial and viral DNA. Mol. Biol. Evol. 20:1231-1239.

Nurminsky DI, Moriyama EN, Lozovskaya ER, and Hartl DL. (1996). Molecular phylogeny and genome evolution in the Drosophila virilis species group: Duplications of the Alcohol dehydrogenase gene. Mol.Biol.Evol. 13:132-149.

Oakeshott JG, Gibson JB, Anderson PR, Knibb WR, Anderson DG, and Chambers GK. (1982). Alcohol dehydrogenare and glycerol-3-phosphate dehydrogenase clines in Drosophila melanogaster on different continents. Evolution 36:86-96.

Orr HA (2005). The probability of parallel evolution. Evolution 59:216-220.

Pogson GH (1991). Expression of overdominance for specific activity at the phosphoglucomutase-2 locus in the Pacific oyster, Crassostrea gigas. Genetics 128:133-141.

Posada D, and Crandall KA. (1998). MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818.

Powell JR, and Desalle R. (1995). Drosophila Molecular Phylogenies and their Uses. 28:87-138.

Powers DA, and Place AR. (1978). Biochemical genetics of Fundulus heterolitus (L.). I. Temporal and spatial variation in gene frequencies of Ldh-B, Mdh-A, Gpi-B, and Pgm-A. Biochem.Genet. 16:593-607.

Prakash S, Lewontin RC, and Hubby JL. (1969). A molecular approach to the study of genic heterozygosity in natural populations. IV. Patterns of genic variation in central, marginal and isolated populations of Drosophila pseudoobscura. Genetics 61:841-858.

Rasmussen MD, and Kellis M. (2007). Accurate gene-tree reconstruction by learning gene- and species-specific substitution rates across multiple complete genomes. Genome Res. 17:1932-1942.

Rokas A, and Carroll SB. (2008). Frequent and widespread parallel evolution of protein sequences. Mol.Biol.Evol. 25:1943-1953.

Ronquist F, and Huelsenbeck JP. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Rosenblum EB, Hoekstra HE, and Nachman MW. (2004). Adaptive reptile color variation and the evolution of the Mc1r gene. Evolution 58:1794-1808.

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

Schaeffer SW, and Miller EL. (1992). Moleculat population genetics of an electrophoretically monomorphic protein in the alcohol dehydrogenase region of Drosophila pseudoobscura. Genetics 132:163-178.

Schaeffer SW, Goetting-Minesky MP, Kovacevic M, Peoples JR, Graybill JL, Miller JM, et al. (2003). Evolutionary genomics of inversions in Drosophila pseudoobscura: Evidence for epistasis. Proc.Natl.Acad.Sci.U.S.A. 100:8319-8324.

Schmidt PS, Paaby AB, Heschel MS (2005) Genetic variance for diapause expression and associated life histories in Drosophila melanogaster. Evolution, 59, 2616–2625

Schulte PM, Gomez-Chiarri M, and Powers DA. (1997). Structural and functional differences in the promoter and 5' flanking region of Ldh-B within and between populations of the teleost Fundulus heteroclitus. Genetics 145:759-769.

Sella G, Petrov DA, Przeworski M, and Andolfatto P. (2009). Pervasive natural selection in the Drosophila genome? PLoS Genet. 5:e1000495.

Sezgin E, Duvernell DD, Matzkin LM, Duan Y, Zhu CT, Verrelli BC, et al. (2004). Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in Drosophila melanogaster. Genetics 168:923-931.

Simpson GG. (1961) Principles of Animal Taxonomy. New York, Columbia University Press.

Smith NG, and Eyre-Walker A. (2002). Adaptive protein evolution in Drosophila. Nature 415:1022-1024.

Stam LF, and Laurie-Ahlberg CC. (1982). A semiautomated procedure for the assay of 23 enzymes from Drosophila melanogaster. Insect Biochem 12:537-544.

Stouthamer R, Breeuwer JA, and Hurst GD. (1999). Wolbachia pipientis: microbial manipulator of arthropod reproduction. Annu.Rev.Microbiol. 53:71-102.

Suzuki Y and Gojobori T. (1999). A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. 16: 1315-1328.

Swofford DL (2002). PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinaur Associates, Sunderland, Massachusetts.

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.

Tamura K, Subramanian S, and Kumar S. (2004). Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol.Biol.Evol. 21:36-44.

Tinsley MC, and Majerus ME. (2007). Small steps or giant leaps for male-killers? Phylogenetic constraints to male-killer host shifts. BMC Evol.Biol. 7:238.

Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, et al. (2009). FlyBase: enhancing Drosophila Gene Ontology annotations. Nucleic Acids Res. 37:D555-9.

Umina PA, Weeks AR, Kearney MR, McKechnie SW, and Hoffmann AA. (2005). A rapid shift in a classic clinal pattern in Drosophila reflecting climate change. Science 308:691-693.

Vasemagi A (2006). The adaptive hypothesis of clinal variation revisited: single-locus clines as a result of spatially restricted gene flow. Genetics 173:2411-2414.

Vavre F, Fleury F, Lepetit D, Fouillet P, and Bouletreau M. (1999). Phylogenetic evidence for horizontal transmission of Wolbachia in host-parasitoid associations. Mol.Biol.Evol. 16:1711-1723.

Veneti Z, Toda MJ, and Hurst GD. (2004). Host resistance does not explain variation in incidence of male-killing bacteria in Drosophila bifasciata. BMC Evol.Biol. 4:52.

Verrelli BC, and Eanes WF. (2001). The functional impact of Pgm amino acid polymorphism on glycogen content in Drosophila melanogaster. Genetics 159:201-210.

--- (2001). Clinal variation for amino acid polymorphisms at the Pgm locus in Drosophila melanogaster. Genetics 157:1649-1663.

--- (2000). Extensive amino acid polymorphism at the pgm locus is consistent with adaptive protein evolution in Drosophila melanogaster. Genetics 156:1737-1752.

Waterson GA (1975). On the number of segregating sites in genetical models without recombination. Theoret Popul Biol 7:256-276.

Weinreich DM, Delaney NF, Depristo MA, and Hartl DL. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312:111-114.

Weisburg WG, Barns SM, Pelletier DA, and Lane DJ. (1991). 16S ribosomal DNA amplification for phylogenetic study. J.Bacteriol. 173:697-703.

Williams DM, and Ebach MC. (2006). Heterology: the shadows of a shade. 23:84-89.

Wood TE, Burke JM, and Rieseberg LH. (2005). Parallel genotypic adaptation: when evolution repeats itself. Genetica 123:157-170.

Yang Z (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol.Biol.Evol. 24:1586-1591.

Zamer WE, and Hoffmann RJ. (1989). Allozymes of glucose-6-phosphate isomerase differentially modulate pentose-shunt metabolism in the sea anemone Metridium senile. Proc.Natl.Acad.Sci.U.S.A. 86:2737-2741.

Zhang J, and Kumar S. (1997). Detection of convergent and parallel evolution at the amino acid sequence level. Mol.Biol.Evol. 14:527-536.

Zhang L, and Li WH. (2005). Human SNPs reveal no evidence of frequent positive selection. Mol.Biol.Evol. 22:2504-2507.

Zhou W, Rousset F, and O'Neil S. (1998). Phylogeny and PCR-based classification of Wolbachia strains using wsp gene sequences. Proc.Biol.Sci. 265:509-515.