Molecular Evolution of Sex-Biased Genes in *Drosophila ananassae*

Sonja Grath

München 2010

Molecular Evolution of Sex-Biased Genes in *Drosophila ananassae*

Sonja Grath

Dissertation
an der Fakultät für Biologie
der Ludwig–Maximilians–Universität
München

vorgelegt von
Sonja Grath
aus Starnberg

München, den 12.05.2010

Erklärung

Diese Dissertation wurde im Sinne von §12 der Promotionsordnung von Herrn Prof. Dr. John Parsch betreut. Ich erkläre hiermit, dass die Dissertation nicht einer anderen Prüfungskommission vorgelegt worden ist und dass ich mich nicht anderweitig einer Doktorprüfung ohne Erfolg unterzogen habe.

Ehrenwörtliche Versicherung

Ich versichere hiermit ehrenwörtlich, dass die vorgelegte Dissertation von mir selbstständig und ohne unerlaubte Hilfe angefertigt wurde.

München, 12.05.2010

Erstgutachter: Prof. Dr. John Parsch

Zweitgutachter: Prof. Dr. Wolfgang Stephan

Tag der Abgabe: 12.05.2010

Tag der mündlichen Prüfung: 18.06.2010

Contents

List of Figures

List of Tables

Abstract

Genes with sex-biased expression often show rapid molecular evolution between species. Previous population genetic and comparative genomic studies of *Drosophila melanogaster* and *D. simulans* revealed that male-biased genes have especially high rates of adaptive evolution.

In this dissertation, I investigate the forces affecting the evolution of sex-biased genes in a *Drosophila* species other than the well-studied *D. melanogaster*. Using custom-made PCR-amplicon microarrays and published microarray data, I analyzed sex-biased gene expression in the cosmopolitan Drosophilid species *D. ananassae* which occurs in highly structured populations throughout the subtropical and tropical regions of the world, mainly in Southeast Asia.

I assessed sex-biased gene expression for 129 *D. ananassae* protein-coding genes whose *D. melanogaster* orthologs had been extensively studied at the expression and population genetic level. For 43 of these genes, I surveyed DNA sequence polymorphism in a natural population of *D. ananassae* and determined divergence to the sister species *D. atripex* and *D. phaeopleura*. Seven of these genes were further analyzed in twelve populations located throughout the geographical range of *D. ananassae*. Sex-biased gene expression is generally conserved between *D. melanogaster* and *D. ananassae*, but about one-third of the genes have either gained or lost sex-biased gene expression in one of the species and about 4% of the genes have undergone a reversal in sex-bias. In contrast to the *melanogaster* lineage, the signal of adaptive protein evolution for male-biased genes is not as strong in *D. ananassae* and is limited to genes with conserved male-biased expression in both species.

In addition, I made use of whole-genome data on sex-biased gene expression and the rate of evolution measured by the ratio of nonsynonymous substitution rate (d_N) to the synonymous substitution rate (d_S) to reveal that the degree of sex-bias is positively correlated with the rate of evolution over 30 million years of Drosophilid evolution. There is a general pattern of faster evolution for highly male-biased genes compared to male-biased genes with a lower degree of sex-bias in *D. melanogaster*, *D. ananassae*, and *D. pseudoobscura*. Moreover, up to 50% of genes are not conserved in degree of sex-bias between species. Additionally, I could assess patterns of molecular evolution in relation to ancestral states of sex-biased gene expression for orthologous genes using *D. pseudoobscura* as an outgroup. In most cases, genes that show conserved male-biased gene expression over longer evolutionary time scales have significantly higher rates of molecular evolution.

Investigating mating discrimination and genetic differentiation at mitochondrial loci shed light on possible patterns of speciation within the species occurring in this geographical range. Finally, I found evidence for cold adaptation in *D. ananassae* strains collected from the more temperate regions of the species range and, in two populations, differences in cold tolerance between males and females.

Zusammenfassung

Gene mit geschlechtsspezifischer Expression weisen häufig rasche molekulare Evolution zwischen verschiedenen Arten auf. Vorausgegangene populationsgenetische und vergleichende genomische Studien von *Drosophila melanogaster* und *D. simulans* haben gezeigt, dass männlich-spezifische Gene besonders hohe Raten adaptiver Evolution aufweisen.

In dieser Dissertation untersuche ich Faktoren, welche die Evolution geschlechtsspezifischer Gene beeinflussen. Als Modellorganismus dient die Fruchtfliegenart *Drosophila ananassae*. Diese Spezies lebt in stark strukturierten Populationen in den Tropen und Subtropen, hauptsächlich in Asien und dem Südpazifik. Zur Verfügung standen zwölf Fliegenpopulationen des Verbreitungsgebietes. Zur Analyse der geschlechtsspezifischen Genexpression verwende ich individuell gefertigte PCR-Amplikon-Microarrays und publizierte Micorarraydaten. Für 129 protein-codierende Gene wurde die geschlechtsspezifische Expression in *D. ananassae* bestimmt. Orthologe dieser Gene in *D. melanogaster* wurden zuvor intensiv auf populationsgenetischer und expressionstechnischer Ebene untersucht. Für 43 dieser Gene habe ich den DNA-Sequenzpolymorphismus in einer natürlichen Population aus Bangkok, Thailand, untersucht und die Divergenz zu *D. atripex* bzw. *D. phaeopleura* bestimmt. Sieben dieser Gene wurden zusätzlich in elf weiteren Populationen untersucht. Die geschlechtsspezifische Expression erweist sich als teilweise konserviert zwischen *D. melanogaster* und *D. ananassae*. Dennoch verloren oder erwarben ungefähr ein Drittel der Gene eine geschlechtsspezifische Expression in einer der beiden Arten. Im Gegensatz zur *Melanogaster*-Linie ist das Signal adaptiver Proteinevolution für männlich-spezifische Gene in *D. ananassae* niedriger und begrenzt auf Gene mit konservierter männlich-spezifischer Expression in beiden Arten. Die Analyse genomweiter Daten zeigte, dass der Grad der geschlechtsspezifischen Expression über 30 Millionen Jahre *Drosophila*-Evolution hinweg mit der Evolutionsrate korreliert ist. Männlich-spezifische Gene mit hoher Expressionsspezifität weisen eine schnellere Evolution auf als Gene mit geringerem Grad an Geschlechtsspezifität. Zudem sind 50% der Gene nicht in der Stärke ihrer Geschlechtsspezifität zwischen den Arten konserviert. Mit Hilfe von Experimenten zum Paarungsverhalten innerhalb der Populationen und Arten dieser Region und der Untersuchung genetischer Differenzierung mitochondrialer DNA trägt diese Studie zur Aufklärung möglicher Artenbildung bei. Weiterhin konnte eine Kälteanpassung verschiedener *D. ananassae* Stränge gezeigt werden. Die Anpassung weist teilweise Unterschiede zwischen Männchen und Weibchen auf.

Note

This thesis was written using the software package LaTeX and is based on the LaTeX template for dissertations developed at the Ludwig-Maximilians-Universität (LMU) by Robert Dahlke and Sigmund Stintzing, 2002 (`http://edoc.ub.uni-muenchen.de/hinweise/LaTeXVorlage.zip`).
In this dissertation I present my doctoral research from February 2007 unil May 2010. It is organized in five chapters, some of which are the result of collaboration with other scientists.

For the work presented in Chapter 1, I generated part of the molecular data and performed some of the analyses. The study was carried out under supervision of Malcolm Schug, John F. Baines and Wolfgang Stephan. A paper based on the findings described in this chapter has been published under the following title:

> Schug, M. D., Baines, J. F., Killon-Atwood, A., Mohanty, S., Das, A., **Grath, S.**, Smith, S. G., Zargham, S., McEvey, S. F., and Stephan, W. (2008)
> Evolution of mating isolation between populations of *Drosophila ananassae*.
> Mol. Ecol. **17**(11):2706-21.

For Chapter 2, I carried out the experiments, analyzed the data and wrote the manuscript. John F. Baines participated in the design of the study and helped with sequence analysis and microarray preparation. John Parsch conceived of the study, participated in its design and coordination, and performed statistical analysis of microarray data. The study was carried out under supervision of John F. Baines and John Parsch. A paper based on the findings described in this chapter has been published under the following title:

> **Grath, S.**, Baines, J. F., and Parsch, J. (2009)
> Molecular evolution of sex-biased genes in the *Drosophila ananassae* subgroup.
> BMC Evol. Biol. **9**(1):291.

For Chapter 3, the data was generated by myself. I analyzed the data and did the writing. The study was carried out under supervision of John F. Baines and John Parsch.
For Chapter 4, I designed the study, analyzed the data and did the writing. The study was supervised by John Parsch.
For Chapter 5, I designed and carried out the experiments. I analyzed the data and did the writing. The study was supervised by John Parsch.

General Introduction

Phenotypic differences between males and females of the same species, collectively referred to as sexual dimorphism, are quite common in all kinds of higher eukaryotes. They are thought to result from the differential action of natural (or sexual) selection on individuals of the two sexes. Darwin (1871) found that extravagant secondary sexual characteristics are very frequent in species of many taxa, including crustaceans, insects, fish, birds, and mammals. He further observed that these traits often show large morphological differences between closely related species and that they are mostly limited to the male of the species. One famous example is the peacock with its long and colorful tail feathers which are lacking in the peahen. This phenomenon was explained using his theory of sexual selection, which consists of two components: direct male-male competition (intrasexual selection), in which males from one species use these phenotypic traits to compete to each other for territory, food, females, or other limiting resources that increase their reproductive success, and female choice (intersexual selection), in which females preferentially mate with males that they find "attractive" based on their phenotype. In both cases, we would expect strong selective pressure on male traits that leads to their rapid evolution.

Sexual selection may not be limited to secondary sexual traits. Morphological studies suggest that sexual selection also acts on primary sexual characteristics. For example, it was found that sperm size and morphology differ greatly among different insect species (Jamieson, 1987) as does the external morphology of male genitalia (Eberhard, 1985; Hosken and Stockley, 2004). These are sometimes the only reliable traits available for classifying different species.

Analyses of interspecific hybridization also suggest the rapid evolution of male reproductive characters. This observation is related to Haldane's rule, which states that when hybrid offspring of only one sex are either inviable or sterile, it is most often the heterogametic sex (Haldane, 1922). In mammals and *Drosophila*, the males are heterogametic, and so hybrid male offspring are more often inviable or infertile. There are two major hypothe-

ses to explain this observation: the "dominance" hypothesis and "faster male evolution". The first hypothesis claims that hybrid incompatibilities are often recessive, and thus are only observed in the sex with hemizygous sex chromosomes (Turelli and Orr, 1995). The "faster male evolution" hypothesis states that genes involved in male reproduction evolve faster than genes involved in female reproduction or genes having no reproductive function (Wu and Davis, 1993). This leads to more genetic incompatibilities related to male reproduction. Faster male evolution can explain the great preponderance of male sterility factors relative to inviability factors that have been identified from *Drosophila* hybridizations (Tao *et al.*, 2003; Wu and Davis, 1993). However, this theory can only explain Haldane's rule for taxa in which the males are the heterogametic sex, and the theory is expected to explain primarily hybrid sterility, not inviability. Moreover, the two hypotheses are not mutually exclusive and it is likely that both play a role in hybrid breakdown (Presgraves and Orr, 1998).

At the molecular level, sex-related genes show increased rates of evolution in their DNA/protein sequences. One possible reason for this is sexual selection affecting the evolution of genes with reproductive functions (Civetta and Singh, 1999; Singh and Kulathinal, 2000; Swanson and Vacquier, 2002). In *Drosophila*, a number of male-specific genes show evidence for rapid evolution due to positive selection, for example *Acp26Aa* (Aguadé, 1998; Tsaur *et al.*, 1998), *OdsH* (Ting *et al.*, 1998), *Sdic* (Nurminsky *et al.*, 1998), *Dntf-2r* (Betrán and Long, 2003), and *jan-ocn* (Parsch *et al.*, 2001).

Females and males are almost identical genetically. In most species, the two sexes differ by only a few genes located on sex-specific chromosomes, such as the Y-chromosome in mammals or the W-chromosome in birds. This leads to the conclusion that most sexual dimorphism results from different expression of genes present in both males and females. Genes with different gene expression in both sexes are referred to as sex-biased genes. Subsequently, these genes are divided into male- and female-biased genes according to the sex which shows higher expression. In contrast, genes that show equal expression in both sexes, are called unbiased. It has been shown that sex-biased gene expression is common in a range of taxa, including insects, nematodes, birds and mammals (Ellegren and Parsch, 2007).

Previous studies used microarray and comparative genomic data from *Drosophila* to compare molecular evolutionary rates of genes belonging to different sex-biased expression classes. In general, genes with male-biased expression tend to show the fastest rates of molecular evolution. Using *D. yakuba* as species for divergence, male-biased *D.*

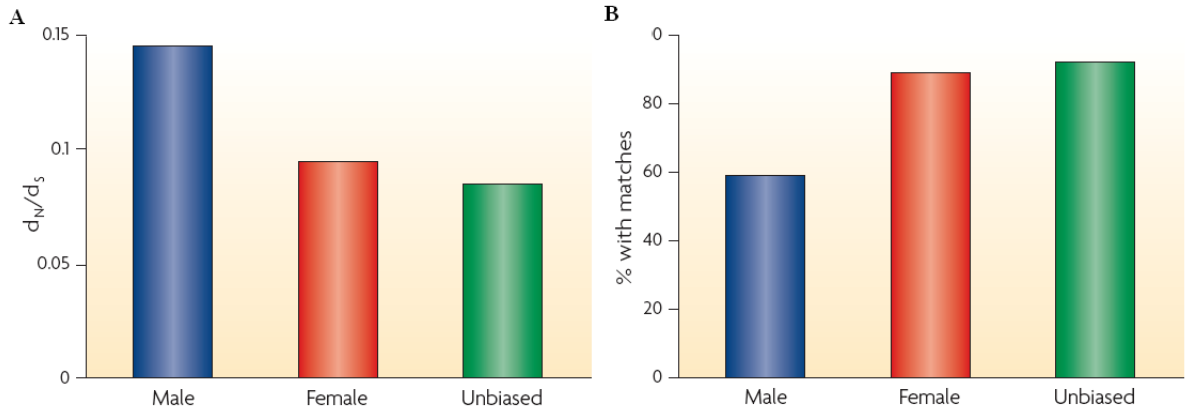

Figure 1: Divergence of sex-biased genes between *Drosophila* species

When orthologs genes are aligned between *D. melanogaster* and *D. simulans*, which are closely related species, male-biased genes show the greatest divergence between species. Shown is the ratio of the nonsynonymous and synonymous substitution rates (d_N/d_S) for a whole-genome comparison (A). For more distantly related species, it is more difficult to identify orthologs. Shown is the percentage of significant BLAST matches (E$<10^9$) when all *D. melanogaster* genes are aligned against the *D. pseudoobscura* genome. Male-biased genes show the fewest matches, indicating that they are the least conserved (B) (taken from Ellegren and Parsch (2007)).

melanogaster genes show significantly faster rates of evolution measured as the ratio of the nonsynonymous to the synonymous substitution rate (d_N/d_S or ω) than both the female- and unbiased genes. This difference was mainly due to an increased d_N in the male-biased genes, indicating rapid evolution of male genes at the amino acid level (Figure 1A). Additionally, male-biased genes have the lowest fraction of identifiable orthologs between more distantly related species as *D. melanogaster* and *D. pseudoobscura* (Figure 1B). Observations of faster male evolution have also been extended to the level of gene expression. Genes with male-biased expression have greater expression differences both within and between species than either female-biased or unbiased genes (Hutter *et al.*, 2008; Meiklejohn *et al.*, 2003; Ranz *et al.*, 2003).

These results clearly indicate that male-biased genes evolve faster than female-biased or unbiased genes. From interspecific comparisons described above alone it is not possible to determine the evolutionary forces leading to the differences in evolutionary rates between sex-biased expression classes. Indeed, the accelerated evolutionary rate observed for male-biased genes could have two different explanations. The first possibility is that male-biased genes are subject to less selective constraint than female-biased or unbiased genes, allowing them to accumulate more neutral (or slightly deleterious) amino acid changes between

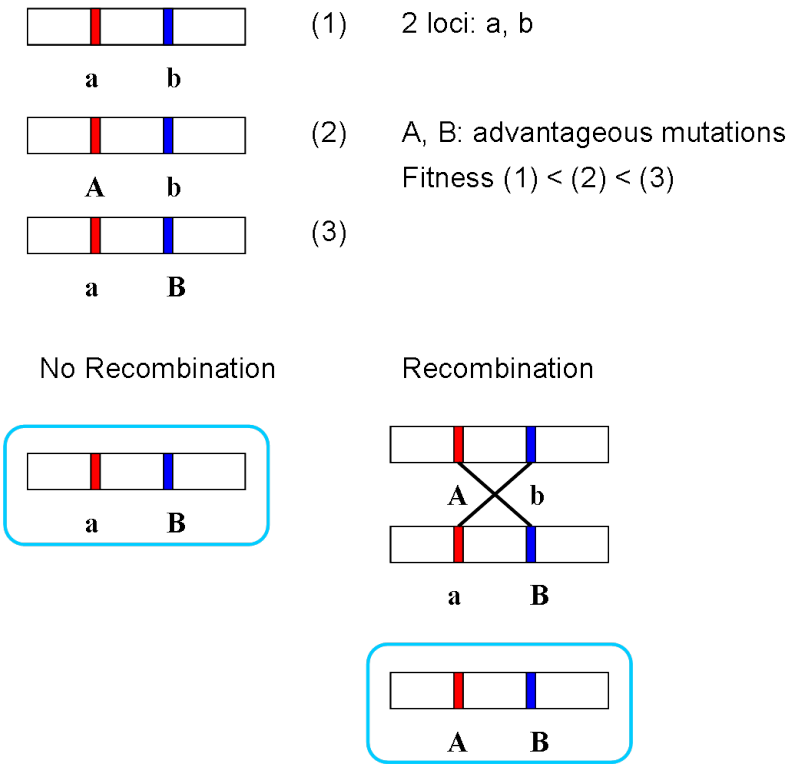

Figure 2: Hill-Robertson interference

Two loci a and b are linked on one chromosome. A and B are advantageous alleles which arise due to mutation. The different haplotypes have different fitness with ab (1) < Ab (2) < aB (3). With no recombination, aB haplotype gets fixed (blue frame), whereas with recombination the haplotype AB can arise in the population by crossover events and will be favored by selection.

species with no effect on fitness. Secondly, male-biased genes could be subject to more positive (or sexual) selection, and thus accumulate more adaptive amino acid changes between species. We can test these hypotheses by examining the correlation between evolutionary rate and recombination rate. Little recombination between sites which leads to increased linkage will lead to reduced levels of adaptation as positive selection is less efficient. Further, fixation of deleterious mutations is increased because of less efficient purifying selection. This effect of reduced efficiency of selection as a result of linkage between sites is known as Hill-Robertson interference.

Figure 2 describes one possible mechanism which may lead to such interference. In this case, two loci a and b are linked on one chromosome. Due to mutations, two more haplotypes Ab and aB arise in the population. Both alleles A and B are advantageous with haplotype fitness ab < Ab < aB. With no recombination, the aB haplotype gets fixed

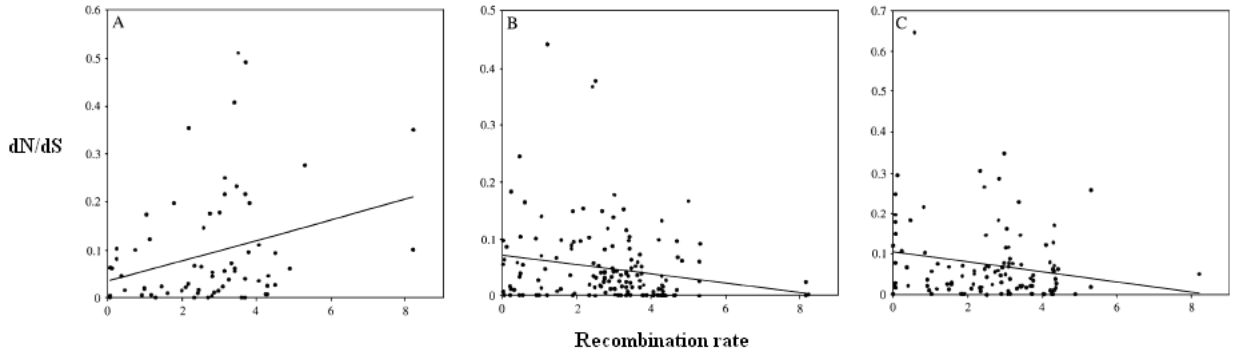

Figure 3: Correlation between d_N/d_S and recombination rate

Recombination rate is estimated by HKw (Hey and Kliman, 2002). Sex-biased gene expression is determined using a 1.5-cutoff. Genes with male-biased expression (A) show a positive correlation (P-value = 0.02, R = 0.30), both genes with female-biased expression (B) and unbiased expression (C) show a negative correlation (P-value = 0.02, R = -0.19 and P-value = 0.04, R = -0.19, respectively). Displayed are least-squares linear regression lines (taken from Zhang and Parsch (2005)).

as selection favors B. The advantageous allele A will disappear. When recombination occurs, haplotype AB can appear in the population due to crossing-over. Its fixation will be favored by selection. If most amino acid replacements are adaptive, then a positive correlation between d_N and recombination rate is expected. This is because positive selection is more effective in regions of higher recombination due to a relaxation of Hill-Robertson interference among selected sites. Purifying selection is less effective in regions of lower recombination. Subsequently, if most amino acid replacements are neutral (or slightly deleterious), then there should be no correlation or a negative correlation between evolutionary rate (d_N) and recombination rate. It has been found that for male-biased there is a positive correlation between d_N/d_S and recombination rate, suggesting that these genes are often targets of positive selection. In contrast, there is a negative correlation between d_N/d_S and recombination rate for female-biased and unbiased genes, suggesting that most amino acid changes in these genes are deleterious and are removed by purifying selection (Zhang and Parsch (2005), see Figure 3).

Moreover, male-biased genes should have more amino acid polymorphism within species than other genes and this should be proportional to the amount of divergence between species when there is less selective constraint on these genes. This can directly be tested by analyzing McDonald-Kreitman (MK) table data (McDonald and Kreitman, 1991). These are two-by-two contingency tables that partition molecular variation into polymorphism

and divergence, and into synonymous and nonsynonymous sites. The MK test compares the ratio of divergence at nonsynonymous and synonymous sites (D_N/D_S) to polymorphism at nonsynonymous and synonymous sites (P_N/P_S). Under a neutral model of molecular evolution, the ratio P_N/P_S is expected to equal the ratio D_N/D_S. An excess of nonsynonymous divergence ($D_N/D_S > P_N/P_S$) is indicative of positive selection driving amino acid replacements between species and the magnitude of this difference gives an estimate of the strength of selection and the fraction of selected amino acid replacements. An excess of nonsynonymous polymorphism ($P_N/P_S > D_N/D_S$) can be indicative of either balancing selection maintaining amino acid polymorphism in the population, or weak purifying selection, which allows slightly deleterious mutations to persist in a population but prevents them from becoming fixed between populations. Male-biased genes tend to show an excess of nonsynonymous divergence using *D. melanogaster* as species for polymorphism data and divergence to its sister species *D. simulans* (Baines *et al.*, 2008; Pröschel *et al.*, 2006). The results explained above indicate that male-biased genes are subject to more positive selection and accumulate more adaptive amino acid replacements.

To date, many microarray surveys of gene expression in *Drosophila* have focussed on some *D. melanogaster* strains. Thus, our classification of sex-biased genes mainly comes from just a small number of laboratory strains of one species. However, it is known that levels of sex-biased gene expression vary between closely related species (Ranz *et al.*, 2003) and even among strains of *D. melanogaster*. For this reason, we performed microarray experiments to determine sex-biased gene expression in another *Drosophila* species. As focal species served *D. ananassae*. This species is of particular interest because of its phylogenetic position as an outgroup to all species within the *melanogaster* lineage (Figure 4).

Moreover, *D. ananassae* is interesting for several other reasons. Its demographic history has been well studied, with the ancestral range of the species defined as a region of Southeast Asia that existed as a single landmass ("Sundaland") during the late Pleistocene around 18,000 years ago (Das *et al.*, 2004). Further, in contrast to *D. melanogaster*, it is a species showing significant population structure. This can be used for investigating selective forces operating on candidate loci by comparing the level and pattern of gene flow of these genes to putatively neutral loci (Das *et al.*, 2004). Finally, experimental tests of mate-choice have found levels of pre-mating isolation to be highest between several peripheral populations (Schug *et al.* (2008), see CHAPTER 1) suggesting that sexual selection occurs within local population, which leads to reproductive isolation among subpopulations.

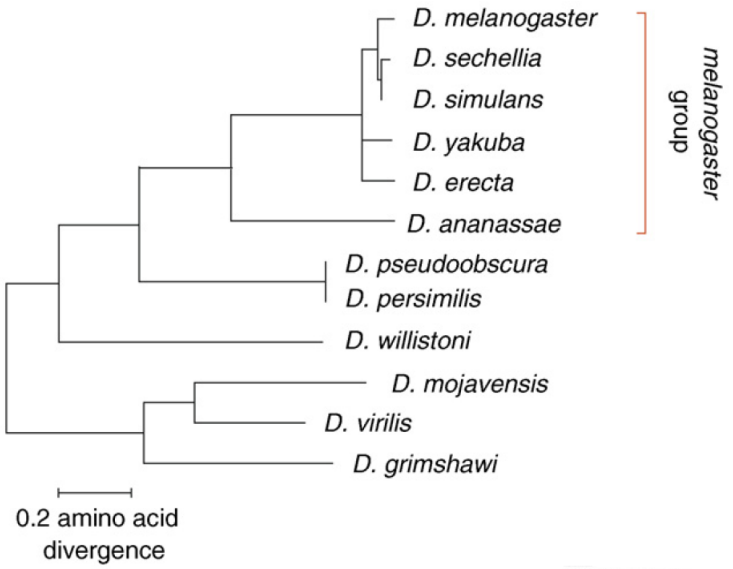

Figure 4: Drosophilid phylogeny

Shown are species sequenced in the *Drosophila* 12 Genomes Project (*Drosophila* 12 Genomes Consortium, 2007). Mean amino acid divergence for each branch was estimated in PAML (taken from Larracuente *et al.* (2008)).

We constructed microarrays from PCR (Polymerase Chain Reaction) products of genes whose *D. melanogaster* orthologs had been intensively studied previously at the population genetic and expression level (Baines *et al.*, 2008; Parsch *et al.*, 2009; Pröschel *et al.*, 2006). For these genes, sex-biased gene expression was known for *D. melanogaster*. We designed species-specific primers on the basis of the *D. ananassae* genome. For this type of microarrays, mRNA is isolated from two separate samples. This might be control or cancer cells or, in case of sex-biased gene expression, male and female tissues, respectively. Genes of interest are produced by PCR and robotically spotted onto glass slides. The two mRNA samples are reverse transcribed into cDNA, which is then labelled with two different fluorophores, *e.g.* green and red dyes. The probes are mixed and competitively hybridized to the microarray slides. Fluorophores are excited by laser and emission is scanned. Finally, relative fluorescence intensities of labelled spots are compared. The general principle of PCR-amplicon microarrays and competitive microarray hybridization is shown in Figure 5. The above approach allowed us to investigate the expression states for our candidate genes with high replication. Moreover, we made use of published whole-genome oligonucleotide array data (Zhang *et al.*, 2007).

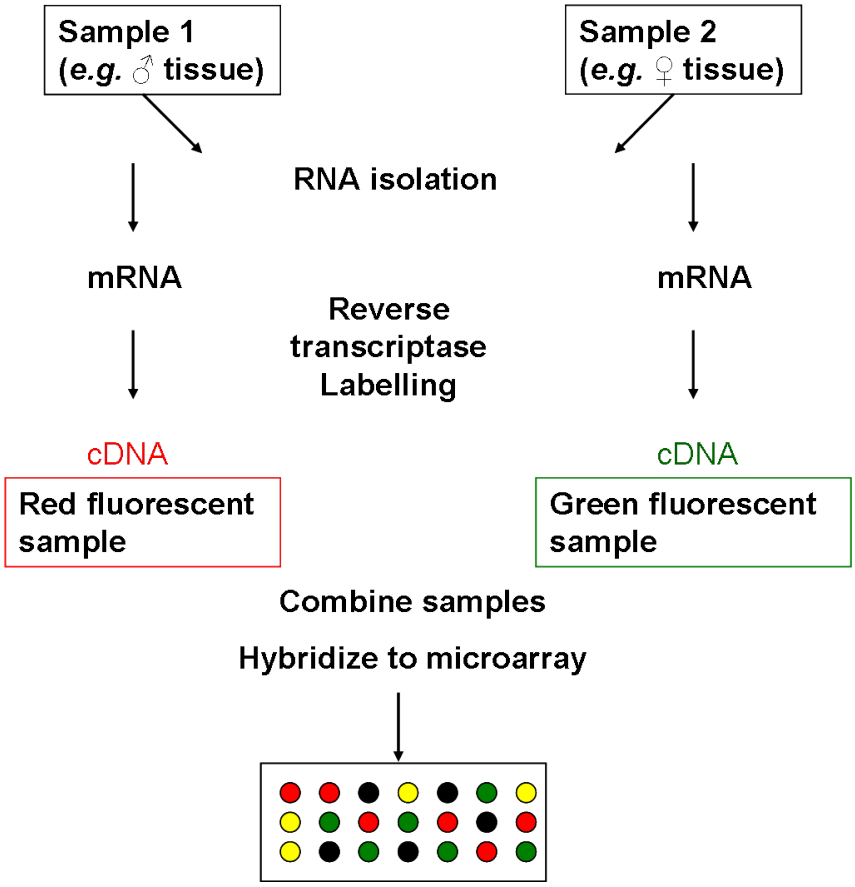

Figure 5: Principle of competitive microarray hybridization

RNA is isolated from two different samples, for instance, male and female tissue. The RNA is reverse transcribed into cDNA which is labelled with two fluorescent dyes, *e.g* red and green. The two labelled probes are competitively hybridized on the microarray slide which are scanned afterwards.

The overall goal of my doctoral project is to determine the forces affecting the evolution of sex-biased genes in a *Drosophila* species other than the well-studied *D. melanogaster*. My project has three major objectives: (1) analyzing sex-biased gene expression in *D. ananassae* and comparing this expression with *D. melanogaster*, (2) determining the type and strength of selection affecting the evolution of sex-biased genes in *D. ananassae*, and (3) identifying candidate genes possibly leading to population differentiation and reproductive isolation in *D. ananassae*. Moreover, I contributed to a study of mating discrimation in relation to genetic differentiation within the geographical range of *D. ananassae* and studied cold adaptation in this species within the course of my dissertation research.

Scope of this dissertation:

In CHAPTER 1, I contribute to a study measuring the level of genetic differentiation and mate discrimination among 18 populations of *D. ananassae* from throughout its geographical range and its sister species *D. pallidosa* located in the South Pacific. I collect mitochondrial DNA polymorphism data to measure genetic differentiation between all populations. Mate discrimination varies considerably throughout the species range, being higher among populations outside the ancestral Indonesian range, and highest in the South Pacific. Further, genetic differentiation appears to contribute to mate discrimination. Molecular variation differs among the ancestral, peripheral, and South Pacific populations, which is consistent with previous studies of intron polymorphism and microsatellites showing substantial genetic structure over the sampled geographical range. Subsequently, this study contributes to our understanding of demography in the highly structured species of *D. ananassae* and facilitates my further studies of sex-biased gene expression throughout the geographical range of this species.

In CHAPTER 2, I look at the type and strength of selection influencing the evolution of sex-biased genes in the *ananassae* lineage. I assess sex-biased gene expression in *D. ananassae* for 129 protein-coding genes using custom-made PCR-amplicon microarrays and published whole-genome microarray data. About 60% of these genes show conserved sex-biased gene expression between *D. melanogaster* and *D. ananassae*, whereas around one-third of the genes have either gained or lost sex-biased gene expression between these species and about 4% show a reversal of sex-bias between species, meaning that these genes show male-biased gene expression in *D. melanogaster* and female-biased gene expression in *D. ananassae* or vice versa. For 43 out of these 129 genes, I survey DNA sequence polymorphism in a *D. ananassae* population sample from Bangkok, Thailand, and determine divergence to one strain of the sister species *D. atripex* and/or *D. phaeopleura*. I find evidence for adaptive

evolution for male-biased genes in *D. ananassae* as previously observed for the *melanogaster* lineage. However, the rate of adaptive evolution is not greater for male-biased than for female-biased or unbiased genes in this species, suggesting that there are differences in sex-biased gene evolution between the two lineages.

In CHAPTER 3, I investigate the strength and pattern of selection in several *D. ananassae* populations for seven protein-coding genes which were previously analyzed for sex-biased gene expression and adaptive evolution in CHAPTER 2. I survey DNA sequence polymorphism throughout the geographical range of *D. ananassae*, including populations that were previously investigated for mate discrimination and genetic differentiation in CHAPTER 1. I determine divergence to a single sequence of either the sister species *D. atripex* or *D. phaeopleura*. Three out of the seven genes display patterns of positive selection that vary among genes and populations. The remaining four genes reveal unbiased gene expression both in *D. melanogaster* and in *D. ananassae* and are mainly consistent with neutral evolution in all populations.

In CHAPTER 4, I further investigate the influence of degree of sex-bias on the rate of molecular evolution for different Drosophilid species. Using published whole-genome data, I find a significant positive correlation between the degree of sex-bias measured by the male/female expression ratio and the rate of evolution measured by the ratio of the nonsynonymous substition rate to the synonymous substitution rate (d_N/d_S) for *D. melanogaster*, *D. ananassae*, and *D. pseudoobscura*. The higher the degree of sex-bias for both male- and female-biased genes, the higher is the rate of molecular evolution. Moreover, X-linked sex-biased genes generally show higher rates of evolution compared to autosomal genes. In addition, I investigate the influence of ancestral expression state on the evolution of orthologous genes between *D. melanogaster* and *D. ananassae* using *D. pseudoobscura* as an outgroup. Comparisons between orthologs reveal only about 50% conservation in degree of sex-bias between the *melanogaster* and the *ananassae* lineages. I suggest taking these findings into account when comparing sex-biased gene evolution between different species in future studies.

In CHAPTER 5, I investigate the possibility that *D. ananassae* populations have adapted to different environmental conditions. I concentrate on the ability of fly strains to recover from a cold-induced "chill coma". While this species is endemic to tropical and subtropical areas of the world, it still faces different climate conditions in nature, depending on factors such as latitude and elevation. I survey two different populations of *D. ananassae*, one from Bangkok, Thailand, and one from Kathmandu, Nepal.

Whereas Bangkok (latitude = 13N, elevation = 2m) has an average summer temperature of 29°C and an average winter temperature of 26°C, Kathmandu (latitude = 27N, elevation = 1,400m) has average summer and winter temperatures of 19°C and 3°C, respectively. Using three fly strains from each population and a standard test for determining cold resistance in insects (chill coma recovery test), I show that flies from Kathmandu are significantly more tolerant to colder temperatures than flies from Bangkok. In addition, I conduct an experiment using flies from Bangkok, Kathmandu, and Kumejima, Japan and observe that the Japanese flies have higher cold resistance compared to flies from Bangkok. Both the Japanese and the Bangkok populations show significant differences between male and female chill coma recovery time. This suggests that males and females have adapted differently to cold temperatures in this population and that sex-biased genes may be involved in environmental adaptation.

Chapter 1

Evolution of mating isolation between populations of *D. ananassae*

1.1 Introduction

Sexual isolation before fertilization may be one of the most important isolating mechanisms leading to speciation. Theory predicts that reproductive isolation is most likely to arise in allopatry as a by-product of one or more of three mechanisms: (i) adaptive divergence during exposure of isolated populations to different environments, (ii) gradual divergence and the accumulation of genetic incompatibilities by drift while populations are separated, or (iii) rapid genetic divergence and accumulation of incompatibilities by genetic drift as a consequence of founding events (Coyne and Orr, 2004; Howard and Berlocher, 1998; Otte and Endler, 1989; Tregenza *et al.*, 2000). Perhaps the only systematic attempt to distinguish between these three mechanisms was performed by Tregenza *et al.* (2000) in which a likelihood-based analysis revealed founder events during colonization as the most likely explanation for assortative mating patterns observed between populations of the European meadow grasshopper. They were, however, not able to rule out the additional effects of long-term isolation and sympatry with closely related species. Other studies have examined the relationship between genetic divergence and assortative mating in a number of species including salamanders (Tilley *et al.*, 1990), *Drosophila willistoni* (Gleason and Ritchie, 1998), and brown grasshoppers (Claridge *et al.*, 1985, 1988) to test for assortative mating associated with gradual genetic divergence, but have found no relationship. At present, the degree to which population isolation and genetic differentiation among populations influence the evolution of mate discrimination remains largely unclear.

One approach to identifying the evolutionary mechanisms that generate mate discrimination is to assay a species distributed across a broad geographical range that includes continuous and isolated populations, and additionally displays various levels of divergence in mating behaviors. *Drosophila ananassae* is distributed throughout the subtropical and tropical regions of the world, has been studied extensively by geneticists (Tobari, 1993), exists in highly structured populations in Asia and the South Pacific (Das *et al.*, 2004; Schug *et al.*, 2007; Tomimura *et al.*, 1993; Vogl *et al.*, 2003) and its biogeographical history is well characterized. Ancestral populations are from Indonesia and peripheral populations in Asia and the South Pacific have colonized these geographical regions since the last glaciation and human migration to Oceania (Das *et al.*, 2004; Schug *et al.*, 2007). The young age of the populations make it unlikely that natural selection has had a substantial affect on DNA sequence variation in most of the populations (Das *et al.*, 2004), but evidence for adaptive evolution was inferred from patterns of DNA sequence variation in northern vs. southern populations in Asia (Baines *et al.*, 2004; Chen *et al.*, 2000; Kim and Stephan, 2000; Stephan *et al.*, 1998). The strong population structure, potential local adaptation and observations of mate discrimination among populations (Spieth, 1952, 1966) suggest that prezygotic sexual isolation may be a salient feature of geographically local populations.

In this study, we assay levels and patterns of mate discrimination between 18 populations of *D. ananassae* spanning their geographical range and two populations of *Drosophila pallidosa* that occur in sympatry with *D. ananassae* in Fiji and Samoa. Previous studies provide strong evidence that the geographical origin of *D. ananassae* is in Southeast Asia, an area called the Sunda Shelf, and peripheral populations in Asia and the South Pacific represent migration since the time that sea levels rose approximately 20,000 years ago (Das *et al.*, 2004). We focus our sampling on populations that represent the ancestral range in Indonesia (Das *et al.*, 2004; Vogl *et al.*, 2003), the peripheral range extending to Nepal, Japan, Northern Australia, and Brazil, and to island populations in the South Pacific. Our sampling strategy intended to identify patterns of mate discrimination that may be associated with varying levels of population isolation and demography and the presence of a sister species in sympatry. Although the populations included in our survey are well-characterized by either DNA sequence polymorphism at introns (Das *et al.*, 2004), microsatellites (Schug *et al.*, 2007), or both, an analysis of mating behavior with respect to genetic differentiation using a common genetic data set would not be possible without the collection of additional data. Furthermore, several populations for which previously only

microsatellite data were available display striking patterns of differentiation at these loci, suggesting they may be morphologically indistinguishable nascent species (Schug *et al.*, 2007). Thus, to obtain a genetic data set common to all populations and confirm previous observations at other loci, we also obtained mitochondrial DNA (mtDNA) polymorphism data for all populations.

1.2 Materials and Methods

1.2.1 Population samples

We performed multiple choice mating experiments on 18 populations of *Drosophila ananassae* from throughout their worldwide geographical range (Figure 1.1). We chose populations that represented the ancestral range in Indonesia (A), the peripheral range in Asia and northern Australia (P), and island populations in the South Pacific (SP). For the purposes of our analysis, we include two populations from northeast Australia within the "South Pacific" group (SP). This is justified by our phylogenetic analysis (see below). We also assayed populations of *Drosophila pallidosa*: one from Nadi, Fiji, and another from Malololelei, Samoa (Figure 1.1). *Drosophila pallidosa* is recognizable in mixed samples because it is paler than *D. ananassae* and there is a reduced expression of metatarsal sex combs in males (Bock and Wheeler, 1972; McEvey *et al.*, 1987). Population samples are previously described (Das *et al.*, 2004; Schug *et al.*, 2007), except for the specimens from Nadi, Fiji, and Moorea, French Polynesia, which were collected in 2004 by M.D.S., A.K-A., S.G.S, and S.F.M.

1.2.2 Molecular analysis

We determined 1315 bp of mtDNA sequence including 645 bp of the cytochrome *b* gene and 670 bp of the control (AT-rich) region for 163 individuals previously assayed for mate discrimination behavior. I analyzed and sequenced samples from nine different populations for 64 individuals in total. One out of these populations (Nadi, Fiji) was *D. pallidosa*. Table 1.1 gives details on the all population samples. For each strain I sequenced, DNA was extracted using one single female fly following the Puregene DNA isolation kit protocol or the Phenol-Chloroform DNA extraction protocol. All protocol details are provided in Appendix H.

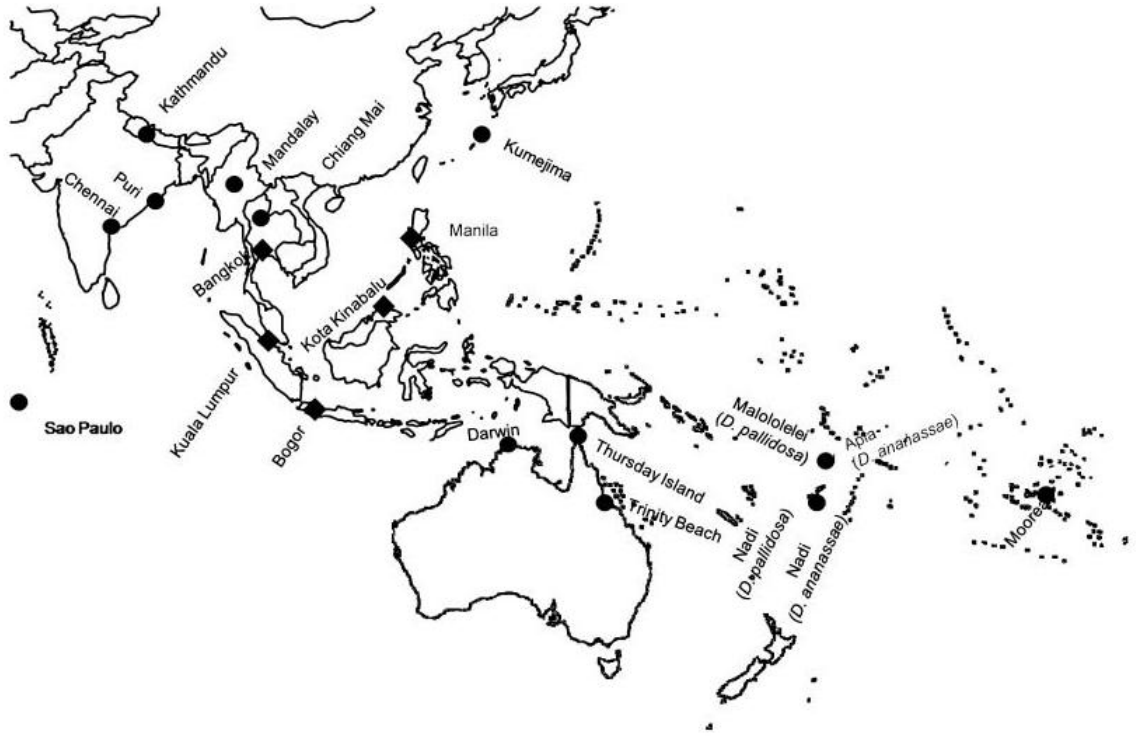

Figure 1.1: Map of populations

Map of *Drosophila ananassae* ancestral (A, closed diamonds), peripheral (P) and South Pacific (SP, closed dots) and *Drosophila pallidosa* (open dots) populations. Ancestral populations are inferred from multilocus DNA sequence variation (Das *et al.*, 2004; Vogl *et al.*, 2003).

Sampling location	Country of origin	Abbreviation	No. of lines
Apia	Samoa	APIA	10
Malololelei[#]	Samoa	MALP	6
Bangkok	Thailand	BKK	9
Chiang Mai[*]	Thailand	CNX	3
Bogor	Java	BOG	9
Chennai	India	CH	11
Puri[*]	India	PURI	14
Darwin	Australia	DAR	4
Kathmandu[*]	Nepal	KATH	6
Kota Kinabalu	Malaysia	KK	10
Kuala Lumpur[*]	Malaysia	KL	8
Kumejima[*], Okinawa	Japan	KMJ	8
Mandalay[*]	Myanmar	MAN	8
Manila[*]	Philippines	MNL	5
Moorea	French Polynesia	MFP	11
Nadi	Fiji	NADI	12
Nadi[*#]	Fiji	NADIP	4
Sao Paolo[*]	Brazil	SP	5
Trinity Beach	Australia	TB	10
Thursday Island	Australia	TI	10

[*] Populations where I collected sequence data.
[#] Populations of *D. pallidosa*.

Table 1.1: Population samples of *D. ananassae* and *D. pallidosa* used in this study

The primers used for cytochrome *b* amplification were AR: 5' GTA GCT CAA ACT ATT TCT TAT GAA G 3' obtained from Kastanis *et al.* (2003) and 10596F: 5' TTT TGG ATC ATT ACT TGG ATT ATG T 3'. Cycling conditions for this region were as follows: 95°C for two minutes followed 34 cycles of 95°C for 30 seconds, 50°C for one minute, 72°C for three minutes, and a final extension at 72 C for ten minutes. The control region was amplified with universal primers from $tRNA^{Ile}$ TI-N-8: 5' CTA TCA AGG TAA CCC TTT TTA TCA GGC A 3' and 12s rRNA SRJ-14612: 5' AGG GTA TCT AAT CCT AGT TT 3' described in Simon *et al.* (1994). Control region cycling conditions were initial denaturing at 95 C for two minutes, 30 cycles of 95 C for 30 seconds, 55°C for one minute, and 68°C for three minutes with a final extension at 68°C for ten minutes.

PCR products were verified using a 1% agarose gel with ethidium bromide staining. Verified PCR products were then purified with ExoSAP-IT. Sequencing of PCR products was carried out using BigDye chemistry and a 3730 automated sequencer (Applied Biosystems, Foster City, CA). For cytochrome *b* the PCR primers were also used as sequencing primers. Additional internal primers - two for cytochrome *b*, one for the control region - were used to get complete coverage. Internal primers for sequencing cytochrome *b* were 10630: 5' CAA ATT TTA ACC GGA TTA TTT TTA 3' and Cytobnew: 5' TCA ACT GGT CGA GCT CCA AT 3'. Primers for sequencing the control region were the end primer SRJ-14612 mentioned above and an internal primer CR-int: 5' TGC TGG CAC CAA TTT AGT CA 3'. Sequencing of strains not sequenced by myself was done on an Amersham MegaBACE or LI-COR IR2.

Sequences were concatenated and aligned using CLUSTAL W in MEGA 4.0 (Tamura *et al.*, 2007) and analyzed using DNASP version 4.10.9 (Rozas *et al.*, 2003). The evolutionary history of haplotypes was inferred using the minimum evolution (ME) method in MEGA 4. The evolutionary distances were computed using the maximum composite Likelihood method and are in the units of the number of base substitutions per site. The ME tree was searched using the close-neighbourinterchange (CNI) algorithm at a search level of 3. The neighbour-joining algorithm was used to generate the initial tree. All positions containing gaps and missing data were eliminated from the data set (complete deletion option). There were a total of 1,233 positions in the final data set. GenBank Accession numbers are EU528677-EU528841, and EU601555-EU601719.

1.2.3 Multiple choice experiments

We designed our experiments to reduce or eliminate potential effects of using isofemale lines that have been maintained in the laboratory for many generations. For populations from Indonesia, Australasia, and the South Pacific, virgin females and males from three different isolines were tested for mate discrimination for each pairwise combination of populations. These experiments were performed at University of North Carolina Greensboro (UNCG) in North Carolina. Three replicate experiments using three different isofemale lines per population were performed for each comparison, using 25 males and females from each line per replicate. Crosses between *D. ananassae* and *D. pallidosa* were performed with two replicates because of the number of isolines available. We maintained the flies for at least 3 days to ensure that females were virgins and both sexes were fully mature. In our experience, males court vigorously and both sexes consistently mate successfully at 3 days of age in *D. ananassae* and *D. pallidosa* and are thus fully mature. We also reared the flies at a constant temperature and humidity, collected flies for each experiment on the same day, and at the same time in the life cycle to reduce any potential effects of environmental conditions on variance in mating behavior among replicates. Replicate experiments between laboratories showed similar results, confirming the low variance among replicates.

Red or green food colouring was added to the food 12-24 h before each experiment to identify the flies after mating. Colour-swaps were performed such that each population received the same type of food colouring twice. The combination of three lines (replicates) per population and reversal of food colouring between replicates eliminates potential line-by-food colouring effects. Males were separated into individual vials 12-24 h before the experiment began to increase courtship vigour. Ultimately, approximately 300 males and virgin females (25 flies per sex per isofemale line x three lines per population), were released into a 25 x 37 x 25-cm mating cage made of Plexiglass. This is the same mating cage used by (Wu *et al.*, 1995), kindly provided to us on loan from C. Aquadro. The mating experiments took place soon after the beginning of the light cycle. Flies were observed for up to 2 h but the experiment was terminated earlier if approximately 50% of the flies had mated. Copulating pairs were aspirated from the cage, frozen at -20 C, and later scored for population origin based on the colour of their abdomens.

For populations from Indonesia, Asia, and Brazil, representative outcrossed populations were created by combining 10 isofemale lines collected from the same location and maintaining them for 10 generations. This method was used to eliminate individual line effects and reduce the workload associated with testing multiple independent lines. These

experiments were performed at the University of Munich, Germany. Virgin females from the outcrossed populations were collected and tested for mate discrimination as described above, using food colouring to identify the population origin of individuals in copula.

10 males and 10 virgin females from each population were introduced into the mating cage (40 flies total) for each replicate, and five replicates were performed for each pair of populations. Thus, a total of 200 flies were assayed per comparison using mating chambers constructed according to (Elens and Wattiaux, 1964). Mating experiments took place in the morning and lasted for 1 h.

We replicated these experiments in the UNCG, North Carolina laboratory using the independent line method for a subset of these same population samples including Kathmandu vs. Mandalay, Chennai vs. Mandalay, and Kathmandu vs. Chennai. The results were similar to those obtained in the Munich laboratory using the mass-mating method. We thus believe the two different experimental designs had little or no effect on the results.

1.2.4 Statistical analysis of behavioral data

We used a discrimination index described by (Rolan-Alvarez and Caballero, 2000) to quantify the relative mating preferences between each cross and overall discrimination for each population pair. This index measures the combined effects of sexual isolation and sexual selection and is described in detail by Coyne *et al.* (2005). The index is designed to disentangle the effects of mating propensity (sexual selection or asymmetric mating between species) from mate discrimination. The index is as follows:

PSI (pair sexual isolation) x PSS (pair sexual selection) = PTI (pair total index).
PSS measures the sexual selection differences between observed mating pairs and expected noncopulating pairs for each pair type. PSI measures sexual isolation by dividing the observed mating pairs for every pair type by the expected noncopulating pairs. Expected pair types are calculated assuming random mating. I_{PSI} measures a total isolation index based on discrimination and sexual isolation, and ranges from -1 (complete disassortative mating) to 0 (random mating), to 1 (complete assortative mating; Coyne *et al.* (2005)). Although there will always be some effect of mating propensity inherent in the total isolation index, the PTI approach outperforms other isolation indices at eliminating such effects and is thus the best index available for quantifying sexual isolation (Rolan-Alvarez and Caballero, 2000).

Because variance among replicates was low, we pooled the replicates and constructed contingency tables to calculate I_{PSI}, the significance of which we determined by 10,000

permutations of each contingency table using the software PTI version 1.0 (Rolan-Alvarez and Caballero, 2000; Takada and Rolan-Alvarez, 2000). N_{AB} represents a mating pair between population A and B, where the female is from population A and the male from population B, while N_{BA} indicates that a female is from population B and a male from population A. The calculation of I_{PSI} values requires at least one mating in each cell of the contingency table. In eight comparisons between *D. ananassae* and *D. pallidosa*, one cell in the contingency table had a zero value representing no mating pairs. To calculate I_{PSI} for these comparisons, we added a single mating for each empty cell. The construction of contingency tables to analyze mate discrimination also provides a means to quantify one-way sexual isolation (IA_{PSI}; Rolan-Alvarez (2004)) by examining patterns of asymmetry that may result from stronger mate discrimination in one population relative to another. We estimated IA_{PSI} and its statistical significance using JMATING (Carvajal-Rodriguez and Rolan-Alvarez, 2006), an updated version of PTI version 1.0 which evaluates statistical significance of asymmetry using bootstrapping.

1.2.5 Statistical analysis of relationships between behavioral and molecular data

To evaluate the correlation between genetic distance, geographical distance, and I_{PSI} (mate discrimination), we performed permutation analyses. To evaluate the statistical significance of the correlations between genetic distance (F_{ST}) and geographical distance (kilometres), we performed Mantel tests (Mantel, 1967) by permuting the matrices 1000 times and plotting the observed r value on the distribution of r values to arrive at a level of significance. For the relationship between genetic distance (F_{ST}) and mate discrimination (I_{PSI}), the matrix was not complete because we did not perform all of the possible pairwise comparisons of mate discrimination between populations. We re-sampled the I_{PSI} values without replacement 1000 times to obtain the probability of the observed r value, ignoring the missing data points in the I_{PSI} matrix, the equivalent of a Mantel test, but with an incomplete matrix.

1.3 Results

1.3.1 Levels and patterns of mate discrimination

Mate discrimination varied considerably among populations throughout the geographical range. In general, populations in the ancestral and peripheral range showed low and insignificant levels of discrimination (Table 1.3). In the two cases where mate discrimination was statistically significant after Bonferroni correction, one was between peripheral populations (Kathmandu and Puri) and the other between ancestral and peripheral populations (Bangkok and Chennai; Table 1.3). In contrast, populations from the South Pacific showed high levels of differentiation (Table 1.4). Thursday Island, Apia, and Trinity Beach display the highest levels of mate discrimination against all populations tested, including the ancestral Bogor population. Interestingly, populations from Nadi and Moorea show much lower levels of mate discrimination with one another and with the ancestral population from Bogor. Both populations did, however, show statistically significant discrimination with Apia, Trinity Beach, and Thursday Island.

Of the 54 pairwise mate discrimination assays, six showed significant asymmetric discrimination. Two of the six cases were between ancestral and South Pacific populations (Table 1.4) and the remaining was between peripheral and/ or South Pacific populations (Tables 1.3 and 1.4). Both cases of asymmetry between South Pacific and ancestral populations were in the opposite direction than predicted by the Kaneshiro hypothesis (Kaneshiro (1983, 1980); see Discussion). Thursday Island and Apia females were choosier than the ancestral Bogor females.

1.3.2 Distinctions between ancestral, Asia-peripheral, and South Pacific populations

A comparison of pairwise genetic differentiation and mate discrimination estimates for the ancestral, Asia-peripheral, and South Pacific populations shows a gradual increase in both estimates from the ancestral to South Pacific range (Figure 1.2). While there is a significant difference in the average levels of mate discrimination among the groups (Kruskal-Wallis $H = 18.29$, d.f. $= 3$, $P = 0.0004$), the means within each category may be biased by multiple comparisons because of the pairwise nature of the design. The results clearly do show, however, that mate discrimination among comparisons in the South Pacific range is substantially higher than those within either the ancestral or Asia-peripheral ranges,

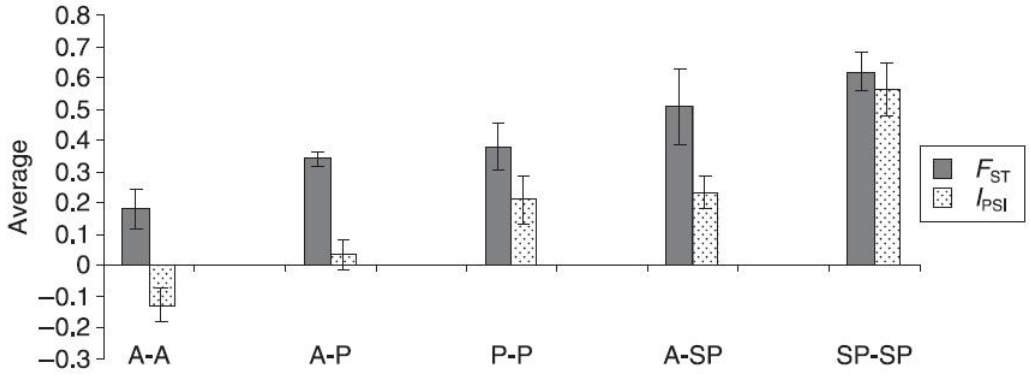

Figure 1.2: Comparison of pairwise genetic differentiation (F_{ST}) and mate discrimination (I_{PSI})

Mean and standard error of sexual discrimination between ancestral vs. ancestral (A-A), ancestral vs. Asia-peripheral (A-P), Asia-peripheral vs. Asia-peripheral (P-P), ancestral vs. South Pacific (A-SP), and South Pacific vs. South Pacific (SP-SP) populations of *Drosophila ananassae*. I_{PSI} is a measure describing a total isolation index based on discrimination and sexual isolation, and ranges from -1 (complete disassortative mating) to 0 (random mating), to 1 (complete assortative mating). F_{ST} is calculated by pairwise comparisons of mtDNA sequences between populations within each category.

and that there appears to be a trend towards higher levels of discrimination between Asia-peripheral populations than with ancestral populations.

1.3.3 Levels and patterns of mtDNA sequence polymorphism

Molecular variation (Table 1.2) varied considerably among the ancestral, peripheral, and South Pacific populations, consistent with previous studies of intron polymorphism (Das *et al.*, 2004; Schug *et al.*, 2007; Vogl *et al.*, 2003) and microsatellites (Schug *et al.*, 2007) that show substantial genetic structure throughout the geographical range we sampled. We found a total of 139 segregating sites among populations, 56 of which were singletons and 83 of which were shared among at least two individuals. There were a total of 111 haplotypes with an average haplotype diversity (H_d) of 0.99 and average nucleotide diversity (π) of 0.0086. Average F_{ST} between all *Drosophila ananassae* populations was 0.52 and ranged from 0 between Mandalay and Puri, and Mandalay and Chennai to 0.94 between Apia and Chiang Mai. The correlation between genetic distance and geographical distance between all *D. ananassae* populations was weak, negative, and not significant (Spearman's r = -0.031, P = 0.56). When we restricted our analysis of geographical distance and genetic differentiation to the 11 populations in the ancestral region, north, and northwest in Asia,

after taking the existence of Sundaland into account (Bogor, Bangkok, Chennai, Chiang Mai, Kathmandu, Kota Kinabalu, Kuala Lumpur, Kumejima, Mandalay, Manila, Puri), we found a statistically significant pattern of isolation-by-distance at mtDNA (Spearman's $r = 0.34$, $P = 0.005$), as for X-linked loci (Das *et al.*, 2004).

The estimate of F_{ST} reported here for mtDNA is higher than that found for the X-linked loci (Das *et al.*, 2004). However, this does not indicate that the mtDNA and the intron data sets are inconsistent. When only those 13 populations are analyzed for which both types of data are available, we find average pairwise estimates of $F_{ST} = 0.22$ and 0.32 for the X-linked and mtDNA data, respectively, while average nucleotide diversity θ is 0.0089 and 0.0046. We currently lack divergence data from an appropriate outgroup to estimate differences in the neutral mutation rate for these loci, but published estimates suggest that mtDNA sequences mutate on average twice as fast as nuclear ones in the subgenus Sophophora (Powell, 1997). Then, taking into account that the X is hemizygous and that the inheritance of mitochondria is maternal, we find that the ratio of N_e (effective population size estimated from the X-linked data) to that estimated from mtDNA is approximately $\frac{4}{3}$. This difference in N_e is qualitatively in agreement with our observation that F_{ST} at the mtDNA loci is larger than that at the X-linked introns. A possible cause for the difference may be a more pronounced effect of purifying selection at the mtDNA loci. This may also partly explain the observed generally negative Tajima's D values (Table 1.2), which are more strongly negative than at the introns (Das *et al.*, 2004). Finally, for our analysis of the mate discrimination data (below), it is important to note that the correlation between the observed values of F_{ST} at the mtDNA and X loci is positive and significant ($r = 0.45$, $P = 0.01$).

In general, populations from the ancestral range in Indonesia and the peripheral region in Asia and Australia showed less genetic differentiation than populations from the South Pacific Islands (Table 1.3, Table 1.4, and Figure 1.2). Among the populations for which we assayed mate discrimination, Apia and Chiang Mai showed the highest average pairwise F_{ST} values. A phylogenetic tree using mtDNA haplotypes clearly distinguishes Thursday Island, Trinity Beach and Apia from the remaining populations, placing them ancestral to all other populations which themselves were poorly resolved (Figure 1.3). A similar pattern is seen in a neighbour-joining tree based on pairwise F_{ST} between populations (Figure 1.4). Both populations of *Drosophila pallidosa* (Nadi and Malololelei) showed levels of F_{ST} within the range of the ancestral and peripheral populations and appeared within the poorly resolved area of the haplotype tree (Figures 1.3, 1.4).

Population	N	S	h	H_d	K	π	θ	Tajima's D
Apia, W. Samoa (SP)	10	1	5	0.67	1.2	0.000	0.0014	-1.74
Bangkok, Thailand (A)	9	2	9	1	7.67	0.005	0.0068	-1.28
Bogor, Indonesia (A)	9	2	8	0.97	6.17	0.003	0.0045	-0.87
Chennai, India (P)	12	2	11	0.99	4.18	0.003	0.0038	-0.68
Chiang Mai, Indonesia (P)	3	1	3	1	3.33	0.000	0.0005	n/a
Darwin, Australia (P)	4	1	3	1	3.33	0.005	0.0042	1.68
Kathmandu, Nepal (P)	6	1	6	1	8.27	0.004	0.0041	0.013
Kota Kinabalu, Indonesia (A)	10	2	9	0.98	5.82	0.004	0.006	-1.55
Kuala Lumpur, Indonesia (A)	8	3	8	1	9.11	0.003	0.0036	-0.139
Kumejima, Japan (P)	8	2	6	0.89	4.75	0.001	0.0018	-1.64
Moorea, French Polynesia (SP)	11	3	10	0.98	6.15	0.004	0.0063	-1.63
Mandalay, Myanmar (P)	8	2	8	1	8.57	0.003	0.006	0.0005
Manila, Philippines (A)	5	3	5	1	17.2	0.008	0.0075	1.23
Nadi, Fiji (SP)	12	5	11	0.98	12.68	0.005	0.0081	-1.23
Puri, India (P)	14	5	14	1	11.23	0.004	0.0072	-1.49
Sao Paulo, Brazil (P)	5	1	5	1	6.2	0.003	0.0034	1.12
Trinity Beach, Australia (SP)	10	2	9	0.97	7.93	0.005	0.006	-0.47
Thursday Island, Australia (SP)	10	2	10	1	15.09	0.011	0.0134	-0.73
D. pallidosa (Nadi, Fiji)	4	3	4	1	15.33	0.009	0.0097	-0.17
D. pallidosa (Malololelei, W. Samoa)	8	2	6	1	8.53	0.004	0.0058	-1.15

Biogeographical location of each population is designated at ancestral (A), peripheral (P), or South Pacific (SP).
N, number of individuals; S, segregating sites; h, number of haplotypes; H_d, haplotype diversity; K, average number of differences between individuals; π, average nucleotide diversity; θ, per site nucleotide diversity.

Table 1.2: Mitochondrial DNA variation within 18 populations of *D. ananassae* and two populations of *D. pallidosa*

Figure 1.3: Evolutionary relationships of 163 individuals based on mtDNA haplotypes

Individuals from Thursday Island are shown with closed circles, from Trinity Beach with open squares, and from Apia with open circles. *Drosophila pallidosa* individuals are shown with open diamonds (Fiji) and closed diamonds (Samoa). The optimal tree with the sum of branch length = 0.2084 is shown. The tree is rooted using Thursday Island which shows strongest similarity to the most closely related species, *D. atripex* and *D. phaeopleura* (see Figure 1.6). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

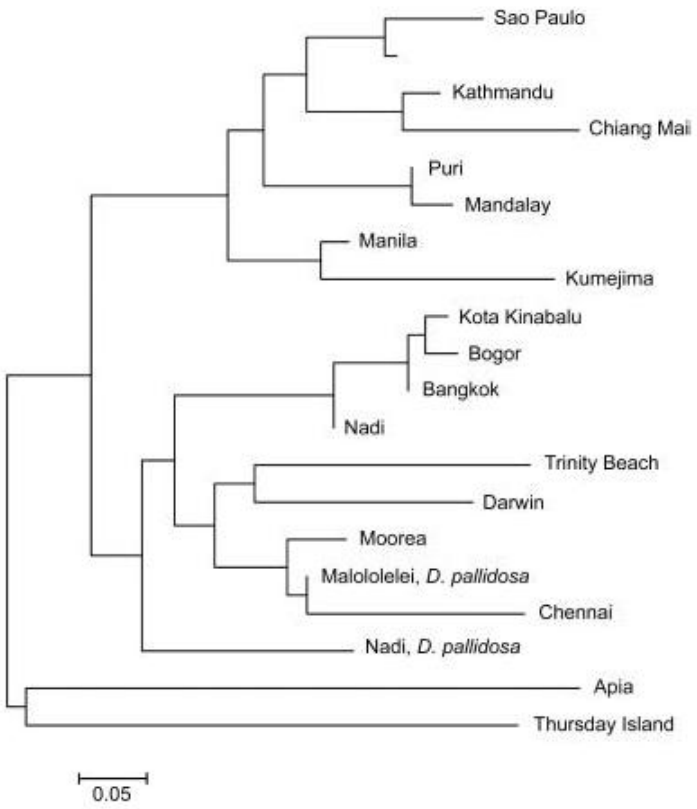

Figure 1.4: Evolutionary relationships of 18 populations of *D. ananassae* and two populations of *D. pallidosa*

The tree is based on pairwise F_{ST} (see Materials and Methods). The evolutionary history was inferred using the neighbour-joining method. The optimal tree with the sum of branch length = 2.810 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree.

As a reference to a well-known *Drosophila* species group, Figure 1.6 shows a phylogeny of 808 bp of the cytochome *b* gene for which we were able to obtain high-quality sequences from GenBank for *Drosophila melanogaster* and its sister species and the same 808 bp from *D. ananassae*, *D. pallidosa*, *D. atripex*, and *D. phaeopleura*. The tree confirms the ancestral position of Thursday Island and Trinity Beach and the relatively young age of the lineages relative to members of the *D. melanogaster* species group.

The correlation between pairwise genetic diversity (F_{ST}) and mate discrimination between all populations (Figure 1.5) was positive and significant using both the nonparametric Spearman's rank order correlation ($r = 0.54$, $P < 0.0001$) and Pearson's r ($r = 0.47$, $P < 0.0001$). The correlation between genetic distance (F_{ST}) and mate discrimination (I_{PSI}) including only the ancestral and Asia-peripheral populations was also positive and significant (Spearman's $r = 0.22$, $P = 0.04$). Thus, 22% (0.472) and 5% (0.222) of the variance in mate discrimination appears to be a function of genetic distance across the species range and within the ancestral and Asia-peripheral range, respectively.

1.3.4 Mate discrimination between species

Drosophila pallidosa and *D. ananassae* showed the highest levels of discrimination, but isolation was complete in only three of the 12 pairwise comparisons (Table 1.5). Mating isolation was high within islands where the species occur in sympatry; complete in Samoa (Malololelei *D. pallidosa* vs. Apia *D. ananassae*) and nearly complete in Fiji (where Nadi *D. ananassae* females discriminated completely against *D. pallidosa* and *D. pallidosa* females mated with *D. ananassae* at a low rate). In contrast, *D. pallidosa* from Fiji and Samoa did not discriminate.

Population A	Population B	Comp.	N_{AA}	N_{BB}	N_{AB}	N_{BA}	I_{PSI}	*sd*	P	F_{ST}
Bangkok	Chennai	A-P	16	15	11	2	0.516	0.124	<0.0001	0.393
Bangkok	Darwin	A-P	12	5	20	15	-0.38	0.133	0.004	0.424
Bangkok	Kumejima	A-P	17	2	10	2	0.113	0.312	0.676	0.379
Bangkok	Mandalay	A-P	11	15	22	7	0.018	0.153	0.928	0.405
Bangkok	Sao Paulo	A-P	15	21	17	9	0.173	0.133	0.18	0.175
Bogor	Chiang Mai	A-P	9	15	12	8	0.086	0.165	0.622	0.13
Bogor	Kathmandu	A-P	21	11	3	24	0.282	0.148	0.043	0.302
Bogor	Kota Kinabalu	A-A	7	8	19	7	-0.206	0.173	0.242	0.027
Bogor	Kuala Lumpur	A-A	19	14	19	13	0.019	0.126	0.902	0.113
Bogor	Chiang Mai	A-P	9	9	4	3	0.453	0.192	0.036	0.479
Bogor	Manila	A-P	3	14	5	18	-0.195	0.205	0.31	0.52
Bogor	Sao Paulo	A-P	12	27	7	15	0.276	0.141	0.06	0.25
Chennai	Kota Kinabalu	P-A	7	18	27	8	-0.142	0.151	0.332	0.042
Chennai	Kuala Lumpur	P-P	14	21	14	11	0.163	0.13	0.226	0.639
Chennai	Manila	P-A	12	13	12	5	0.23	0.156	0.144	0.386
Chennai	Sao Paulo	P-P	15	9	5	12	0.202	0.165	0.214	0.379
Darwin	Kota Kinabalu	P-A	12	17	14	8	0.154	0.15	0.288	0.459
Darwin	Kuala Lumpur	P-A	3	24	16	8	-0.142	0.181	0.412	0.432
Darwin	Manila	P-A	14	18	17	7	0.187	0.136	0.176	0.417
Kathmandu	Kota Kinabalu	P-A	7	19	29	8	-0.149	0.156	0.324	0.213
Kathmandu	Kuala Lumpur	P-A	22	1	5	12	-0.227	0.224	0.312	0.168
Kathmandu	Manila	P-A	8	15	23	8	-0.103	0.16	0.484	0.238
Kathmandu	Puri	P-P	14	13	10	4	0.604	0.143	<0.0001	0.109
Kathmandu	Sao Paulo	P-P	9	20	11	4	0.343	0.159	0.05	0.042
Kota Kinabalu	Kumejima	A-P	16	3	4	13	-0.026	0.226	0.87	0.418
Kota Kinabalu	Mandalay	A-P	19	17	6	26	0.183	0.136	0.172	0.466
Kota Kinabalu	Manila	A-A	10	11	14	17	-0.207	0.137	0.134	0.272
Kota Kinabalu	Sao Paulo	A-P	4	11	4	7	0.106	0.229	0.714	0.174
Kuala Lumpur	Kumejima	A-P	24	10	10	15	0.119	0.146	0.44	0.431
Kuala Lumpur	Mandalay	A-P	24	18	8	15	0.312	0.122	0.01	0.049
Kuala Lumpur	Manila	A-A	19	14	19	23	-0.122	0.118	0.302	0.288
Kuala Lumpur	Puri	A-P	19	12	9	18	0.084	0.144	0.57	0.273
Kuala Lumpur	Sao Paulo	A-P	3	7	4	20	-0.335	0.224	0.142	0.017
Kumejima	Mandalay	P-P	15	15	23	13	-0.072	0.132	0.586	0.603
Kumejima	Manila	P-A	4	7	12	6	-0.244	0.204	0.266	0.238
Kumejima	Puri	P-P	16	4	3	4	0.409	0.24	0.104	0.495
Kumejima	Sao Paulo	P-P	11	22	14	13	0.072	0.14	0.598	0.356
Mandalay	Manila	P-A	26	14	17	13	0.121	0.122	0.332	0.395
Mandalay	Puri	P-P	16	2	4	5	0.09	0.305	0.76	0
Manila	Sao Paulo	A-P	8	12	4	20	-0.489	0.183	0.782	0.27

Table 1.3: Multiple choice experiments between Indonesian and Asian populations of *D. ananassae*

I_{PSI} (Rolan-Alvarez and Caballero 2000)is the total mate discrimination index and P is based on 10,000 permutations. Genetic distance between populations was inferred between populations based on F_{ST} estimated from mtDNA polymorphism between individuals in each population. Biogeographical location of each population is designated as ancestral (A), or peripheral (P).
*Aysmmetry statistically significant; $P < 0.05$.
NS, not statistically significant after sequential Bonferroni correction (Rice, 1989).

Population A	Population B	Comp.	N_{AA}	N_{BB}	N_{AB}	N_{BA}	I_{PSI}	*sd*	P	F_{ST}
Apia	Bogor	SP-A	100	55	50	46	0.217*	0.064	<0.0001	0.816
Apia	Moorea	SP-SP	33	26	7	19	0.433*	0.097	<0.0001	0.0795
Apia	Thursday Island	SP-SP	107	111	8	5	0.889	0.03	<0.0001	0.859
Apia	Trinity Beach	SP-SP	60	74	24	25	0.465	0.066	<0.0001	0.8
Bogor	Moorea	A-SP	20	27	12	21	0.19	0.115	0.114	0.405
Bogor	Nadi	A-SP	18	21	14	16	0.134	0.125	0.306	0.131
Bogor	Thursday Island	A-SP	75	69	29	27	0.442	0.064	<0.0001	0.786
Bogor	Trinity Beach	A-SP	46	92	31	64	0.177*	0.069	0.011 (NS)	0.522
Moorea	Nadi	SP-SP	15	15	14	13	0.054	0.139	0.69	0.323
Moorea	Thursday Island	SP-SP	27	43	3	3	0.846*	0.061	<0.0001	0.789
Moorea	Trinity Beach	SP-SP	27	25	9	12	0.433	0.107	<0.0001	0.522
Nadi	Thursday Island	SP-SP	35	34	4	11	0.665	0.078	<0.0001	0.76
Nadi	Trinity Beach	SP-SP	28	37	5	19	0.517	0.087	<0.0001	0.168
Thursday Island	Trinity Beach	SP-SP	122	80	21	7	0.768*	0.041	<0.0001	0.764

I_{PSI} (Rolan-Alvarez and Caballero, 2000) is the total mate discrimination index and P is based on 10,000 permutations. Genetic distance between populations was inferred between populations based on F_{ST} estimated from mtDNA polymorphism between individuals in each population. Biogeographical location of each population is designated as ancestral (A), or South Pacific (SP).
*Aysmmetry statistically significant; $P < 0.05$.
NS, not statistically significant after sequential Bonferroni correction (Rice 1989).

Table 1.4: Multiple choice experiments between Indonesia, Australia, and Oceania populations of *D. ananassae*

Population A (*D. pall.*)	Population B (*D. ana.*)	N_{AA}	N_{BB}	N_{AB}	N_{BA}	I_{PSI}	*sd*	P	F_{ST}
Malololelei	Apia	56	75	0	0	0.929	0.031	<0.0001	0.623
Malololelei	Bogor	45	13	4	0	0.864	0.068	<0.0001	0.049
Malololelei	Moorea	20	31	21	0	0.461*	0.997	<0.0001	0.313
Malololelei	Nadi	32	15	16	17	0.146	0.113	0.19	0.385
Malololelei	Thursday Island	50	32	26	3	0.585	0.069	<0.0001	0.323
Malololelei	Trinity Beach	38	62	19	4	0.666	0.064	<0.0001	0.434
Nadi	Apia	35	24	0	0	0.937	0.045	<0.0001	0.806
Nadi	Bogor	24	36	0	1	0.923	0.054	<0.0001	0.524
Nadi	Moorea	67	51	2	0	0.967	0.023	<0.0001	0.486
Nadi	Nadi	61	44	6	0	0.883*	0.041	<0.0001	0.191
Nadi	Thursday Island	43	28	8	1	0.808	0.06	<0.0001	0.232
Nadi	Trinity Beach	41	15	0	0	0.797*	0.072	<0.0001	0.285
Nadi	Malololelei (*D. pall.*)	22	23	22	23	0.001	0.106	0.998	0.363

D. pallidosa populations from Malololelei, Samoa and Nadi, Fiji. All other populations are *D. ananassae*.
I_{PSI} (Rolan-Alvarez and Caballero, 2000) is the total mate discrimination index and P is based on 10,000 permutations. Genetic distance between populations was inferred between populations based on F_{ST} estimated from mtDNA polymorphism between individuals in each population. Biogeographical location of each population is designated as ancestral (A), or peripheral (P).
*Aysmmetry statistically significant; $P < 0.05$.
NS, not statistically significant after sequential Bonferroni correction (Rice, 1989).
I_{PSI} calculated for contingency tables with zero cells by adding 1 to each cell.

Table 1.5: Multiple choice experiments between *D. pallidosa* and *D. ananassae*

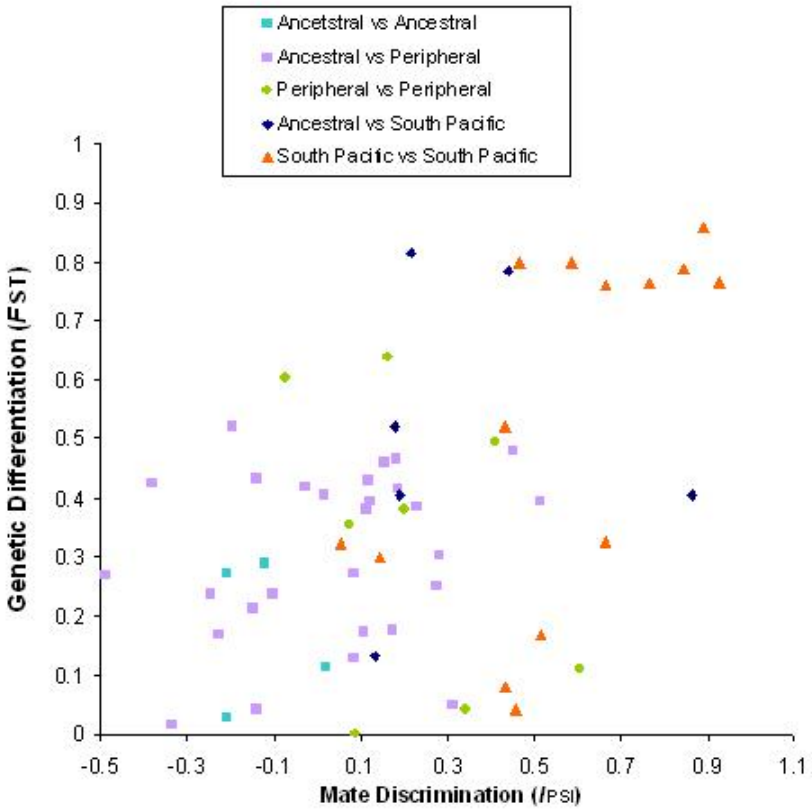

Figure 1.5: Genetic differentiation (F_{ST}) and mate discrimination (I_{PSI})

Scatterplot of genetic differentiation (F_{ST}) based on mtDNA polymorphism and mate discrimination (I_{PSI}) among 18 populations of *D. ananassae* (r = 0.54, $P < 0.0001$).

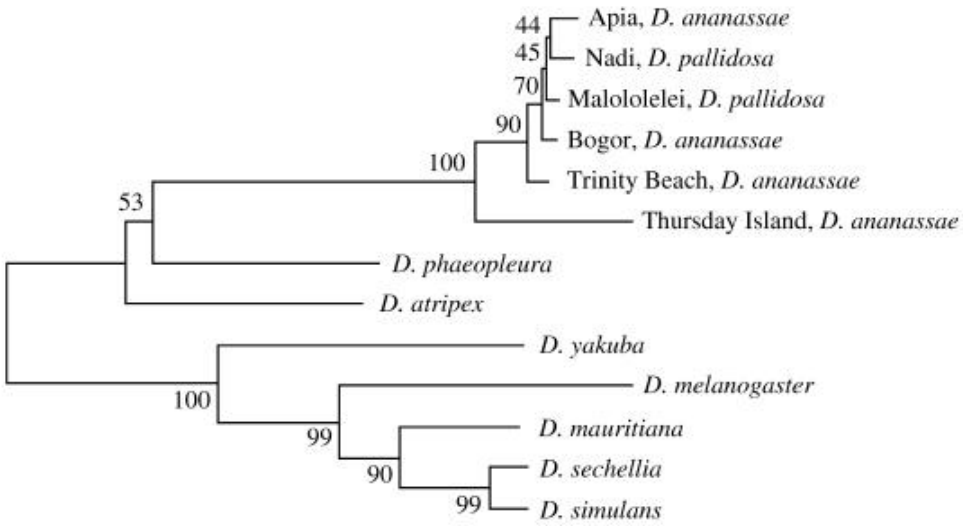

Figure 1.6: Evolutionary relationships of six *D. ananassae* and *D. pallidosa* populations

For these populations phylogenetic signal was resolvable (see Figure 1.3). Included are the most closely related species available (*D. atripex* and *D. phaeopleura*) and members of the *D. melanogaster* subgroup for comparison of evolutionary distance. The evolutionary history was inferred using the neighbourjoining method. The optimal tree with the sum of branch length = 0.350 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1025 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the maximum composite likelihood method and are in the units of the number of base substitutions per site. There were a total of 808 positions of the cytochrome *b* in the final data set.

1.4 Discussion

We have conducted one of the largest geographical surveys of mate discrimination in any species, performing 219 assortative mating experiments for 67 pairwise population comparisons and recording 5105 copulations between 18 *Drosophila ananassae* and two *Drosophila pallidosa* populations from throughout their worldwide geographical range. Our results show that (i) in contrast to a similar study in *Drosophila melanogaster* (Henderson and Lambert, 1982), substantial variation in mate discrimination exists, the strongest of which appears in island populations in the South Pacific; (ii) mate discrimination is significantly correlated with genetic differentiation between populations, but other evolutionary forces must also be involved, and at least two of the populations we assayed (Apia and Thursday Island) appear to be incipient species.

1.4.1 Geographical patterns of mate discrimination

Our results show that mate discrimination varies considerably across the geographical range of *D. ananassae* and is strongest between island populations in the South Pacific and in northeast Australia. Furthermore, mate discrimination occurs between some of the populations in the continuous geographical range in Asia (Puri vs. Kathmandu, and Chennai vs. Bangkok). These results contrast with the lack of mate discrimination between *D. melanogaster* populations in this same region (Henderson and Lambert, 1982), and are similar in magnitude to discrimination between cosmopolitan and Zimbabwe *D. melanogaster* populations (Hollocher *et al.*, 1997; Wu *et al.*, 1995). Thus, the patterns of marked geographical population structure that are a distinguishing feature of *D. ananassae* populations (Das *et al.*, 2004; Schug *et al.*, 2007; Tobari, 1993; Vogl *et al.*, 2003) appear to be accompanied by structure in the patterns of mate discrimination as well. Despite the fact that *D. ananassae* is one of the most common species throughout the tropical and subtropical regions of the world (Tobari, 1993), the patterns of mate discrimination are more characteristic of Drosophilids that exist in restricted habitats, such as *Drosophila immigrans* in the USA and Australia (Ehrman and Parsons, 1980), *Drosophila willistoni* in Central and South America (Gleason et al. 1998), and *Drosophila paulistorum* in Central America (Anderson and Ehrman, 1969; Dobzhansky *et al.*, 1964).

Although the populations of *D. ananassae* in isolated habitats on islands show the highest levels of mate discrimination, we also find significant levels of mate discrimination within the continuous geographical range in Asia, in addition to some apparently isolated

populations with relatively low levels of mate discrimination. These inconsistencies imply that there may be multiple reasons underlying mate discrimination. In one respect, our results are consistent with the study of the European meadow grasshopper (Tregenza *et al.*, 2000), which concluded that some aspect of the colonization process appears to be associated with the origin of prezygotic sexual isolation. However, in another respect, they are inconsistent because we find that mate discrimination has evolved in the continuous geographical region from the Sundaland islands, northwest where gene flow has had an important impact on genetic polymorphism. Indeed, observations of mate discrimination between populations with patterns of isolation-by-distance are uncommon (Tilley *et al.*, 1990; Tregenza, 2002). One possibility is that divergent mating behaviour may have evolved during a period of allopatry and has been maintained subsequently even in the presence of low levels of gene flow. There is growing evidence that mate discrimination may evolve rapidly as a by-product of diverse ecological conditions (*e.g.* Langerhans *et al.* (2007)), suggesting that prolonged periods of allopatry may not be required for divergent mating behaviour to evolve.

The most striking patterns of mate discrimination we observe are between *D. ananassae* populations from Apia, Trinity Beach, and Thursday Island. These populations may have become established during the last 5000 years following human migration (Schug *et al.*, 2007). Colonization is a process that is likely accompanied by population bottlenecks. However, evidence for recent population bottlenecks is not apparent in the patterns of mtDNA polymorphism or microsatellites (Schug *et al.*, 2007) for these populations. Tajima's D values at mtDNA were generally negative for the mtDNA sequences, consistent with population expansion (besides purifying selection mentioned above). The Apia and Thursday Island populations show high levels of mate discrimination against each other and all other populations, and phylogenetic analysis suggests that they are independent lineages. Since the negative Tajima's D values indicate that it is unlikely that recent bottlenecks are the cause of these high levels of mate discrimination, these behaviours may have evolved previous to the colonization of the islands in a different geographical region, or these populations are much more ancient, existing on the islands before human migration like species of other Drosophilid genera that are not peridomestic (*e.g. Mycodrosophila spp.*, McEvey and Polak (2005) Samoaia spp., Malloch (1934)). Interestingly, two haplotypes from the Manila population, which is in the ancestral species range, cluster with Apia haplotypes, suggesting that the Apia population may have diverged within the ancestral range before colonization. Alternatively, recent gene flow between Apia and Manila populations may

be reflected in some individuals. Identifying the ancestral source of the populations by further sampling and molecular analysis of populations within the ancestral Indonesian geographical region may help resolve this question.

Trinity Beach also shows high levels of mate discrimination, but lower than Apia and Thursday Island. While this population is currently highly isolated from surrounding geographical regions due to a strict agricultural quarantine in place since the 1970s, we observe mtDNA haplotypes that indicate the population is a mixture of at least two independent ancestral lineages, which is supported by the microsatellite analysis of Schug *et al.* (2007). Many colonization/migration scenarios are thus possible, but the relatively high pairwise F_{ST} values we observe indicate that population isolation may play a role in generating or maintaining behaviours underlying mate discrimination in this population. This interpretation is supported by the lower levels of genetic differentiation we observe in Nadi, Fiji, and Moorea, French Polynesia, which both show only modest and insignificant mate discrimination against each other and the ancestral population. The mtDNA haplotypes of these two populations appear scattered throughout the tree and pairwise F_{ST} values are relatively low, indicating that they are either of very recent origin or experience gene flow with populations in the ancestral geographical range. Ongoing gene flow seems possible because of the international tourist and agricultural trade in Fiji and French Polynesia.

1.4.2 Genetic differentiation and mate discrimination

Our results suggest that genetic differentiation may play a role in the divergence of mate discrimination behaviours. In the entire data set, genetic differentiation explains 22% of the variance in mate discrimination, but is highly influenced by the strong mate discrimination between Apia, Thursday Island, and Trinity Beach, which may be nascent species (see below) and ancestral to all other populations (Figure 1.3). Within the ancestral Indonesian and Asian geographical range, genetic differentiation explains only 5% of the variation in mate discrimination. This is perhaps not surprising because the range in genetic differentiation is relatively small in these regions, presumably representing current and historic gene flow among the populations which show a pattern of isolation-by-distance. However, this is the only study of which we are aware that has observed any relationship between genetic distance and mate discrimination (Claridge *et al.*, 1985, 1988; Gleason and Ritchie, 1998; Panhuis *et al.*, 2001; Tilley *et al.*, 1990; Tregenza, 2002; Tregenza *et al.*, 2000). This may reflect the breadth of the geographical sampling we performed relative to other studies

and suggests that broad geographical surveys may reveal insight into patterns of genetic differentiation that may underlie behavioural divergence.

We cannot, however, conclude that genetic differentiation is a prerequisite to the evolution of mate discrimination. In fact, our results show that some populations with high levels of mate discrimination do not always display high levels of genetic differentiation. For example, both the Trinity Beach vs. Nadi ($I_{PSI} = 0.517$, $F_{ST} = 0.168$) and Puri vs. Kathmandu ($I_{PSI} = 0.604$, $F_{ST} = 0.109$) assays show statistically significant mate discrimination and levels of genetic differentiation that are low relative to most of the other comparisons between populations, including those that show no evidence of mate discrimination. Thus, on a restricted geographical scale, our results are consistent with studies of various other organisms that found that genetic differentiation and mate discrimination were not correlated (Claridge *et al.*, 1985, 1988; Gleason and Ritchie, 1998; Panhuis *et al.*, 2001; Tilley *et al.*, 1990; Tregenza, 2002; Tregenza *et al.*, 2000). However, on a broader geographical scale, and in populations where strong mate discrimination is observed, there appears to be some relationship between the level of genetic differentiation and the evolution of mate discrimination.

1.4.3 Mate discrimination and adaptive evolution

There is a growing body of evidence that mate discrimination may evolve as a by-product of adaptation to environmental conditions while populations are in allopatry (*e.g.* Funk (1998); Langerhans *et al.* (2007); Vines and Schluter (2006)). Tests for such an effect require measurement of environmental variables that influence fitness which may be multifarious and are often difficult to identify in natural populations. In the case of the Asian *D. ananassae* populations for which we observe significant mate discrimination (Chennai, Kathmandu, Puri and Bangkok), putative environmental variables have not yet been identified, but may be related to temperature and elevation in particular geographical regions. We know, for example, that *D. ananassae* is a stenotherm and is reactive to differences in temperature (*e.g.* Das *et al.* (1995); Joshi (1999); Sisoda and Singh (2002)). In the South Pacific populations, the only potential environmental variable we can examine at this time is the presence of a sister species. The effect of a sympatric sister species on the evolution of mate discrimination was proposed by Zouros and d'Entremont (1980) based on their observation that *Drosophila mojavensis* in sympatry with *Drosophila arizonensis* showed higher levels of mate discrimination with other *D. mojavensis* populations than populations that were not in sympatry. This same phenomenon was also recently reported for

populations of *Drosophila subquinaria* in sympatry with *Drosophila recens* (Jaenike *et al.*, 2006).

We assayed two *D. ananassae* populations which show complete (Apia) or nearly complete (Nadi) discrimination with a sister species, *D. pallidosa*, in sympatry. In Apia, sympatry with a sister species is consistent with the hypothesis because it discriminates strongly against all other *D. ananassae* populations. However, in Nadi, sympatry with a sister species is not consistent with the hypothesis because it does not discriminate against another South Pacific-peripheral population, Moorea, or the ancestral population from Indonesia (Bogor). In fact, it is difficult to distinguish the potential influence of a sister species from the effects of demography and biogeographical history that may underlie mate discrimination in Apia. As we discussed previously, mate discrimination in Apia may have evolved before or after colonization and it is not possible to determine the extent to which the demographic events that accompanied and followed colonization vs. the presence of a sister species affected mate discrimination. Thus, while our results are consistent with the Zouros and d'Entremont (1980) hypothesis in Apia, we cannot rule out other scenarios that may cause a similar pattern of mate discrimination. Furthermore, the lack of mate discrimination of Nadi with other populations may reflect gene flow which could prevent the effects of divergence in mate discrimination behaviours.

Another clue to the effects of adaptation on the evolution of mate discrimination may lie in patterns of asymmetric choice of females in ancestral and derived populations. It was proposed (Kaneshiro, 1983, 1980) that asymmetry should evolve in founder populations, presumably as previously co-adapted gene complexes break up during the founder event, followed by adaptation to new niches. In such a case, he argued that females in derived populations should become less choosy than those in the ancestral population as they seek new mates in previously uninhabited niches. Such a pattern of asymmetry was observed in Hawaiian Drosophilids (Kaneshiro, 1983, 1980) and in the ancestral African and derived (Cosmopolitan) strains of *D. melanogaster* (Hollocher *et al.*, 1997). Although the mechanisms underlying the asymmetry have been debated (Hollocher *et al.*, 1997; Iwasa and Pomiankowski, 1995) it either appears to be irrelevant to the evolution of mate discrimination in *D. ananassae* populations or ancestral patterns of asymmetry that may have previously existed are not apparent in present-day populations. We found only six cases of asymmetry in mate discrimination and in two cases in the South Pacific where founder events are most likely to have occurred (Apia vs. Bogor, and Trinity Beach vs. Bogor). The asymmetry is in the opposite direction than Kaneshiro predicted and consis-

tent with Watanabe and Kawanishi (1979) who observed similar patterns across a wider range of *Drosophila* in the *melanogaster* and *virilis* subgroups and proposed that a failure of females of a newly derived species to mate with ancestral males may drive asymmetry.

1.4.4 Influence of biogeographical history

Our sampling regime was designed to include populations from the ancestral, Asia-peripheral, and South Pacific ranges with the intent of examining the data for patterns of mate discrimination that may reflect both the biogeographical migration of the populations out of Indonesia and into surrounding habitats during the past 20,000 years (Das *et al.*, 2004; Schug *et al.*, 2007). We were particularly interested in these comparisons because of the evidence that natural selection may have a strong influence on broad expanses of the genome in populations from the regions in Northern vs. Southern Asia (Baines *et al.*, 2004; Chen *et al.*, 2000; Kim and Stephan, 2000; Stephan *et al.*, 1998) and because of the obvious genetic drift that may accompany population isolation on the South Pacific islands and potentially some of the peripheral populations in Southeast Asia that surround the ancestral geographical range in Indonesia. The trend towards higher levels of mate discrimination in peripheral populations (Figure 1.2) is intriguing, particularly since it does not appear to correlate with levels of genetic differentiation or patterns of natural selection that were previously detected between the north and south populations. If this pattern is borne out in a more rigorous sampling and statistical analysis, it would suggest that some aspect of the biogeographical migration patterns has an important effect on the evolution of mate discrimination, even within a continuous geographical range.

1.4.5 Potential incipient speciation

Our results suggest that we may have identified three incipient species that show strong behavioural and molecular divergence from other *D. ananassae* populations and from each other. We are not the first to suggest this. Three other studies reached similar conclusions: (Futch, 1966) presented cytological evidence of a cryptic species in Papua New Guinea, (Tobari *et al.*, 1993) found a second cytological variant in Papua New Guinea, and (Klinken *et al.*, 2002) noted morphological variance in *D. ananassae* from Cairns and from Darwin. In our study, Thursday Island and Apia populations show strong mate discrimination with each other and all other populations including the ancestral Bogor population, and are also highly divergent at both mtDNA (this study) and microsatellites (Schug *et al.*, 2007).

Trinity Beach, a population that also shows high levels of genetic differentiation, shows strong mate discrimination from most other populations, and females from this population discriminate against the ancestral Bogor population (reflected by the statistically significant asymmetry). Both mtDNA differentiation (this study) and microsatellites (Schug *et al.*, 2007) show higher levels of genetic differentiation between these populations. Furthermore, all three of the populations are more clearly distinguished in a distance and haplotype tree than *D. pallidosa* from both Fiji and Samoa.

Cleary the level of divergence between the *D. ananassae* samples is much younger than between any of the *D. melanogaster* species group. The phylogeny supports the hypothesis that Thursday Island and Trinity Beach populations are ancestral and diverged before the colonization of Asia and the South Pacific, perhaps in the Indonesian geographical area (Sundaland). The deeper ancestry of these populations may underlie the strong divergence in mating behaviours. However among the many of the remaining populations, the extent of behavioural divergence is substantial and has likely occurred within the past 20,000 years.

The behavioural and phylogeographical results raise the possibility that Thursday Island, Trinity Beach, and Apia are true species, more divergent even that *D. pallidosa*. Body colour varies considerably in the South Pacific region (Tobari, 1993). Cosmopolitan *D. ananassae* in the ancestral and peripheral ranges have a typically light body colour similar to *D. melanogaster*. Thursday Island and Apia flies have a very dark body colour, and may be the undescribed species reported by Tobari (1993) called *D. pallidosa-like* or *Drosophila papuensis-like*. The Nadi population of *D. ananassae* also displays a slightly darker body colour than the cosmopolitan populations, and thus may additionally be a member of Tobari's undescribed species. This scenario is supported by our data, suggesting that in fact these populations may be currently in the very early stages of speciation. Trinity Beach populations are similar in colour to the cosmopolitan populations, but may also be an incipient species.

Prezygotic mating isolation between *D. ananassae* and *D. pallidosa* has been well documented previously (reviewed in Tobari (1993)) and appears to involve both differences in the male courtship song (Yamada *et al.*, 2002), female discrimination (Doi *et al.*, 2001), and potentially body size (Sisoda and Singh, 2004). The female discrimination behaviour has been mapped to a region of the sex chromosome that carries an inversion (Doi *et al.*, 2001; Sawamura *et al.*, 2006, 2007), which is fixed in the Samoa population (Futch, 1973). Thus, the recent speciation, yet close overall genetic relatedness of the two species may involve

the fixation of a polymorphic inversion in the *D. pallidosa* lineage. The identification of the underlying genetic factors leading to rapid behavioural divergence between *D. ananassae* and *D. pallidosa*, and studies of chromosomal inversion frequencies throughout the ancestral, peripheral, and South Pacific range should provide insight into the evolutionary mechanisms involved.

We have performed crosses among these South Pacific populations, *D. pallidosa*, and the ancestral Bogor population in the laboratory and have noticed no evidence of postzygotic sexual isolation (hybrid viability or sterility). Crosses between *D. ananassae* and *D. pallidosa* stocks from Fiji show strong behavioural isolation but no obvious postzygotic sexual isolation (M. Schug, C. Pantazis, unpublished results; M. Matsuda personal communication). Furthermore, relative to the *D. melanogaster* subgroup, the divergence among populations is clearly very young, which is consistent with previous microsatellite (Schug *et al.*, 2007) and intron (Das *et al.*, 2004) studies. The results suggest that Thursday Island, Trinity Beach, and Apia are nascent species and that the exuberant expression of tarsal sex combs that distinguishes *D. pallidosa* from *D. ananassae* may not be a reliable indicator of a "good" species. Furthermore, they suggest that behavioural divergence may evolve very rapidly.

1.5 Conclusions

Prezygotic mating isolation is likely to represent one of the first stages in the transition from populations to species. For this reason, it has been a major interest of evolutionary biologists during the last decades. Mate discrimination is one of the most commonly measured forms of prezygotic isolation. It appears to be relatively common among closely related species and has previously been used as a measure to distinguish populations from subspecies, races, and sister species. However, the influences of various evolutionary mechanisms that may generate mate discrimination are largely unknown.

In this study, we measured the level and pattern of mate discrimination among 18 populations of the cosmopolitan Drosophilid species *D. ananassae* and its sister species *D. pallidosa*. Additionally, we measured genetic differentiation between all 18 populations using mitochondrial DNA polymorphism data. I mainly contributed to the study by collecting and analyzing sequence data.

Mate discrimination varies considerably throughout the species range, being higher among populations outside the ancestral Indonesian range, and highest in the South Pa-

cific. Our results suggest that sexual selection occurs within local populations, leading to reproductive isolation among subpopulations. Colonization and genetic differentiation may have influenced the evolutionary origin of mate discrimination. Our phylogeographical approach clarifies the ancestral relationships of several populations from the South Pacific that show particularly strong mate discrimination and suggests that they may be in the early stages of speciation. Furthermore, both the genetic and behavioral results cast doubt on the status of *D. pallidosa* as a "good" species.

Chapter 2

Molecular evolution of sex-biased genes in the *D. ananassae* subgroup

2.1 Introduction

Sex-biased genes, *i.e.* those that differ in expression level between males and females, may be subject to differing selective constraints depending on the sex in which they are expressed or they may experience conflicting selective pressures in males and females (reviewed in Ellegren and Parsch (2007)). Previous studies of *Drosophila melanogaster* have shown that male-biased genes, especially those expressed in reproductive tissues, consistently exhibit high levels of adaptive protein evolution (Baines *et al.*, 2008; Pröschel *et al.*, 2006). Genome-wide comparisons of the ratio of the nonsynonymous substitution rate to the synonymous substitution rate (d_N/d_S) also indicate that male-biased genes are more functionally divergent between closely-related *Drosophila* species and are less likely to have identifiable orthologs between distantly-related species than genes with female-biased or unbiased expression (Baines *et al.*, 2008; Zhang and Parsch, 2005). A limitation of these studies is that they rely on gene expression data from *D. melanogaster* for sex-bias classifications and, thus, are not informative with respect to differences in sex-biased gene expression or sex-biased gene evolution between lineages.

In contrast to the studies that focused primarily on *D. melanogaster*, a recent SAGE (serial analysis of gene expression) study found no accelerated rate of protein evolution for male-biased genes in *D. pseudoobscura* (Metta *et al.*, 2006). This study measured the rate of evolution by the proportion of nonsynonymous substitutions (d_N) between species

and confirmed a higher rate of protein evolution in genes that had male-biased expression in both *D. melanogaster* and *D. pseudoobscura*, but found no evidence for an increased rate of evolution of genes that had male-biased expression only in *D. pseudoobscura*. The latter genes were only about half as divergent as the former and showed evolutionary rates similar to those of female-biased and unbiased genes. These results suggest that patterns of sex-biased gene evolution may have changed since the split of the *D. melanogaster* and *D. pseudoobscura* lineages. To further investigate this possibility, we analyzed sex-biased gene expression and DNA sequence polymorphism in *D. ananassae*, a species within the *melanogaster* group that serves as an outgroup to all species within the *melanogaster* subgroup, but is more closely related to *D. melanogaster* than *D. pseudoobscura* (*Drosophila* 12 Genomes Consortium, 2007; Larracuente *et al.*, 2008).

D. ananassae is distributed throughout the subtropical and tropical regions of the world. In contrast to *D. melanogaster*, *D. ananassae* is a species displaying significant population structure (Das *et al.*, 2004; Schug *et al.*, 2007). The demographic history has been investigated previously, with the ancestral range of the species being defined as a region of Southeast Asia that existed as a single landmass (Sundaland) during the late Pleistocene around 18,000 years ago (Das, 2005; Das *et al.*, 2004; Schug *et al.*, 2007). Ancestral populations are expected to have colonized Asia and the South Pacific since the last glaciation and human migration to Oceania.

We use species-specific microarrays to investigate sex-biased gene expression in *D. ananassae* for a set of genes previously investigated in *D. melanogaster*. We find that 60% of these genes show conserved sex-bias, while 40% differ in their sex-bias classification between the two species. Using multilocus statistical tests that compare ratios of polymorphism and divergence at synonymous and nonsynonymous sites, we detect a general signal of adaptive protein evolution in *D. ananassae*. However, this signal is not stronger for male-biased genes than for female-biased or unbiased genes, which is consistent with there being differences in sex-biased gene evolution on the *D. melanogaster* and *D. ananassae* lineages.

2.2 Materials and Methods

2.2.1 Microarray analysis

To analyze sex-biased gene expression in *D. ananassae*, we designed a species-specific PCR-amplicon microarray. We began with a set of 148 genes that had previously been studied in *D. melanogaster* (Baines *et al.*, 2008; Pröschel *et al.*, 2006) and used the available *D. ananassae* genome sequence (Assembly August 2005; `http://genome.ucsc.edu/`) to identify their orthologs and design PCR primers that amplify exonic sequences (mean length = 458 bp). A complete list of genes and PCR primers is provided in Appendix A. PCR products were purified using genPURE PCR 96 well filter plates (Genetix) and spotted on UltraGAPS slides (Corning) using a GeneMachines OmniGrid Accent microarrayer. *D. ananassae* genomic DNA (gDNA) was also spotted as a control. Eight replicates of each gene probe were spotted per array, with each replicate in a different subarray. 12-14 control gDNA probes were spotted per subarray.

For hybridization, we extracted total RNA from four- to five-day old males and females using Trizol reagent (Invitrogen) and the manufacturer's protocol. Two inbred strains of *D. ananassae* from Kota Kinabalu, Borneo were used for RNA extraction (Das *et al.*, 2004). Reverse transcription was performed using 25g of total RNA and an anchored oligo(dT) primer. cDNA was labeled with fluorescent dyes (Alexa Fluor 555 and 647) using the amino-allyl labeling system (Invitrogen). Labeled male and female cDNA was competitively hybridized to the arrays for 20 hours at 42°C. Arrays were scanned using a Genetix aQuire 2-laser microarray scanner and the Genetix Qscan software. In total, we performed 12 replicate hybridizations. Six of these were biological replicates, *i.e.* from different RNA extractions performed at different times, while the other six were dye-swap replicates where we exchanged the fluorescent dyes used to label male and female cDNA within each biological replicate. Each microarray consisted of two identical halfs. Each half contained eight replicate spots for 162 samples (candidate genes and control probes). The 162 samples were arranged in 9x9 subarrays. With this design, standard and dye-swap reaction could be performed in parallel on each glass slide (Figure 2.1).

Prior to statistical analysis, normalization of the two dye channels was performed. Since our arrays contained only a subset of the genome with a non-random distribution of sex-biased genes, we used the gDNA spots as controls for normalization. The signal for each spot was calculated by subtracting the median local background from the mean spot intensity.

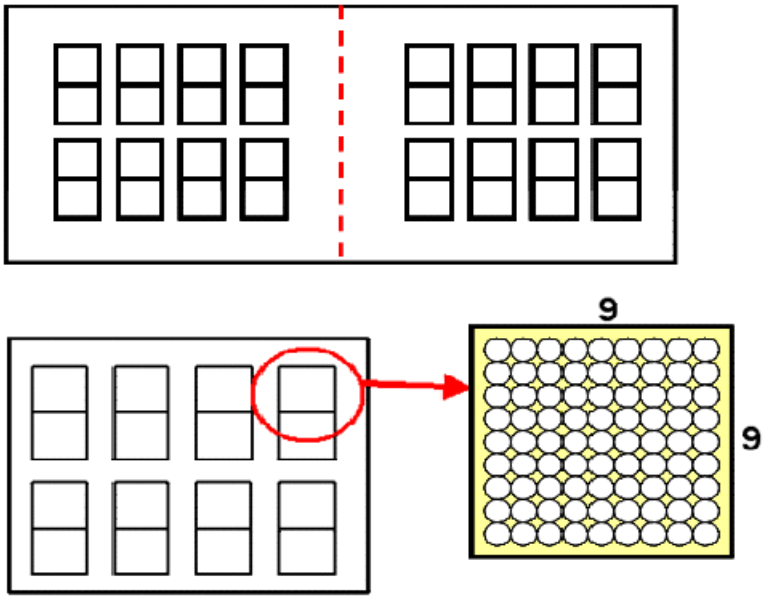

Figure 2.1: Array design for custom-made PCR-amplicon microarrays

Each array consisted of two identical halfs each containing eight fields of 162 probes. The 162 probes were arranged in 9x9 subarrays.

If this result was negative for a given channel, that channel was assigned a value of 0.5 according to the MINIMUM approach of LIMMA (Smyth and Speed, 2003). For each replicate subarray, the red/green ratio was calculated for each of the gDNA control spots and the mean of these ratios was taken as the normalization factor. The raw red value of each spot was then multiplied by this factor to obtain the normalized red value used to calculate the corrected red/green ratio. Note that this is a local normalization, with a different normalization factor for each replicate subarray. As a quality control measure, only spots with a mean intensity at least 20% above the mean local background in at least one of the channels were used. For any one array, we required that at least half of the replicate spots per gene (in this case four) displayed adequate signal as defined above. The median red/green ratio (after normalization) of these replicate spots was used for statistical analysis using the BAGEL software (Townsend, 2004; Townsend and Hartl, 2002). To determine the P-value cut-off used to define significant sex-biased genes, we performed randomizations of the BAGEL input file to estimate the false-discovery rate (FDR) for the dataset. For classification of sex-biased genes, we used a P-value cut-off of 0.01, which corresponds to a FDR of 10%.

2.2.2 Fly strains, polymerase chain reaction and DNA sequencing

For the polymorphism survey, we used 12 inbred strains of *D. ananassae* from Bangkok, Thailand. Like the Kota Kinabalu strains used for the microarray experiments, these strains were included in a previous population genetic survey of intronic loci and found to be in the ancestral range of *D. ananassae* (Das *et al.*, 2004). Thus, we expect little differentiation between these populations. We chose the Bangkok population for the polymorphism survey because a larger number of strains were available.
One strain each of *D. atripex* and *D. phaeopleura* (kindly provided by M. Schug) were used as outgroups. We used the *D. ananassae* genome (Assembly August 2005; `http://genome.ucsc.edu/`) to design PCR primers flanking the coding sequence of 43 target genes. These genes were chosen to represent different categories of sex-biased gene expression on the basis of the microarray data described above. Following PCR, the amplified products were purified with ExoSAP-IT (USB, Cleveland, OH) and sequenced on both strands using BigDye version 1.1 chemistry and a 3730 automated sequencer (Applied Biosystems, Foster City, CA). The PCR primers were also used as sequencing primers in the polymorphism survey. A complete list of primers is provided in Appendix B. For some genes, we were unable to get successful PCR or DNA sequence from all 12 *D. ananassae* strains (Appendix C). The average number of strains sequenced per gene was 11 and we required at least eight sequences for a gene to be included in our analysis. All new DNA sequences have been submitted to the GenBank/EMBL databases under accession numbers FN546265-FN546780.

2.2.3 Analysis of polymorphism and divergence

Sequences were edited using DNAstar (Madison, WI) and multiple alignments were calculated using MUSCLE (Edgar, 2004) (`http://www.ebi.ac.uk/Tools/muscle/`). Polymorphism and divergence statistics were calculated using DnaSP 4.5 (Rozas *et al.*, 2003). For McDonald-Kreitman (MK) table data we used the number of segregating mutations instead of the number of segregating sites as some genes had sites with three segregating variants. For divergence, we considered only sites with fixed differences between all *D. ananassae* lines and a single *D. atripex* (or *D. phaeopleura*) sequence. Multilocus Tajima's D tests were performed using the HKA program (`http://lifesci.rutgers.edu/~heylab/heylabsoftware.htm`). To calculate the fraction of positively selected amino acid substitutions, α, with the method of (Bierne and Eyre-

Walker, 2004), we used the DoFE program (`http://www.lifesci.sussex.ac.uk/home/Adam_Eyre-Walker/Software.html`). For MK tests and α calculations, the *D. atripex* sequence was used as the outgroup whenever available. When the *D. atripex* sequence was not available, *D. phaeopleura* was used as the outgroup.

We used the method of (Bustamante *et al.*, 2002) for estimating the posterior distribution of the selection parameter, γ, its 95% confidence intervals, and the proportion of the distribution falling below zero. The method is implemented in the MKPRF web server (`http://cbsuapps.tc.cornell.edu/mkprf.aspx`).

2.2.4 Phylogenetic analysis

For 13 genes we were able to get sequences from both *D. atripex* and *D. phaeopleura* (Appendix C). In addition, we downloaded the amino acid sequences of *D. melanogaster*, *D. simulans*, *D. ananassae*, and *D. pseudoobscura* for the entire set of genes analyzed from Flybase (`http://flybase.org/`). For one gene (*CG18418*), we were unable to get the orthologous sequence from *D. pseudoobscura*. This left us with 12 genes that could be aligned across all species (*CG2577, CG3004, CG3024, CG4593, CG6459, CG7840, CG8277, CG9135, CG10853, CG11379, CG15336,* and *CG15717*). Amino acid sequences of these 12 genes were concatenated for phylogenetic reconstruction. Bayesian inference (BI), as implemented in MrBayes version 3.1.1 (Huelsenbeck and Ronquist, 2001), was used to reconstruct the phylogeny with *D. pseudoobscura* set as the outgroup. For the BI analysis, two distinct 10,000 generation runs were conducted (three incrementally heated chains with model jumping between fixed-rate amino acid models were used and trees were saved to a file every 10 generations). Identical topologies were recovered from both runs. A burn-in period of 2,500 generations was determined graphically by plotting likelihood values for each sample. The results were presented in the form of a 50% majority-rule consensus tree in which trees corresponding to the burn-in period were discarded. Support for the nodes was given by posterior probability estimates of clades. Using model jumping, only the Jones (JTT) model (Jones *et al.*, 1992) contributed to the final result with 100% posterior probability. This was also supported by the program ProtTest version 2.0 (Abascal *et al.*, 2005), which determines the best-fit substitution model for amino acid data under a likelihood framework. Within this framework, the Akaike information criterion (Akaike, 1973) selected the JTT (Jones *et al.*, 1992) model including the proportion of invariable sites and the gamma distribution of rate variation among sites (JTT + I +

G) as the best-fitting model with an Akaike weight of 0.54. The second-best model was JTT+G with an Akaike weight of 0.46. Trees were visualized with TreeView (Page, 1996).

2.3 Results

2.3.1 Sex-biased gene expression in *D. ananassase*

To investigate sex-biased gene expression in *D. ananassae*, we designed a species-specific microarray of PCR-amplified exon sequences from 148 genes (Appendix A) whose orthologs had been previously classified as male-biased, female-biased or unbiased in *D. melanogaster* (see Methods). This set of genes was of particular interest because the vast majority (136 genes) had been studied at the population-genetic level in *D. melanogaster* and estimates of the rate of adaptive evolution in the *melanogaster* subgroup were already available (Baines *et al.*, 2008; Pröschel *et al.*, 2006). After quality control, we obtained sufficient hybridization signal to reliably classify 60 genes (22 male-biased, 13 female-biased, and 25 unbiased). Of the remaining genes, we were able to classify 69 (21 male-biased, 10 female-biased, and 38 unbiased) using the *D. ananassae* whole-genome microarray data of Zhang *et al.* (Zhang *et al.*, 2007). In total, 129 of the 148 genes (87%) were classified, with 43 male-biased, 23 female-biased, and 63 unbiased (Appendix D).

To examine the conservation of sex-biased gene expression between species, we compared the above *D. ananassae* classifications to those previously determined for *D. melanogaster*. The majority of genes (59%) showed a conserved expression pattern, with the same bias in both species (Figure 2.2A). However, 37% of the genes were classified as sex-biased in only one species, and 4% showed the opposite sex-bias in the two species (Figure 2.2A). Overall, male-biased genes showed the greatest conservation: 72% of the genes with male-biased expression in *D. ananassae* also had male-biased expression in *D. melanogaster*, 65% of the genes with female-biased expression in *D. ananassae* also had female-biased expression in *D. melanogaster*, and 48% of the genes with unbiased expression in *D. ananassae* also had unbiased expression in *D. melanogaster*. Of the genes that were sex-biased in only one species, most (69%) were sex-biased in *D. melanogaster*, but unbiased in *D. ananassae* (Figure 2.2B). Three genes (*CG13690, CG3024, CG4593*) were male-biased in *D. ananassae*, but female-biased in *D. melanogaster*, while two genes (*CG12684, CG7387*) were female-biased in *D. ananassae*, but male-biased in *D. melanogaster* (Figure 2.2B). For *CG13690*, the molecular function is described as ribonuclease H activity and RNA binding. It is involved in the RNA metabolic process. The molecular function of *CG3024*

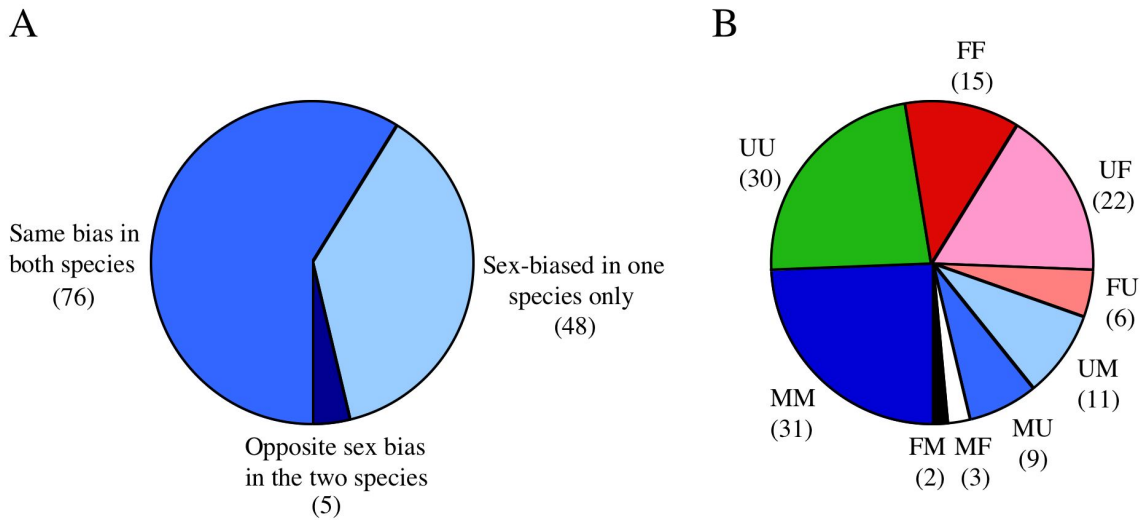

Figure 2.2: Conservation of sex-biased gene expression between *D. melanogaster* and *D. ananassae*

Panel A shows the overall conservation of sex-biased expression between the two species. Panel B shows the conservation of the specific sex-bias classes, with the first letter indicating the bias in *D. ananassae* and the second letter indicating the bias in *D. melanogaster*. "M" indicates male-biased, "F" indicates female-biased, and "U" indicates unbiased expression. The area of the chart taken up by each category indicates the percentage of genes falling into that category. The number of genes in each category is given in parentheses.

(*torp4a*) is described as ATP binding and unfolded protein binding. The gene is involved in protein folding. For *CG7387*, the function is given as unfolded protein binding and heat shock protein binding. It is also involved in protein folding. For the two remaining genes, no functions or biological processes are described. All described functions refer to what is known for *D. melanogaster*.

2.3.2 Phylogenetic relationship of the focal species

For analysis of evolutionary rates and tests for adaptive evolution, it is critical to have an appropriate outgroup species. Two recent molecular phylogenetic studies suggested that *D. atripex* and/or *D. phaeopleura* might serve as an appropriate outgroup to *D. ananassae* for these purposes (DaLage *et al.*, 2007; Matsuda *et al.*, 2009). However, these studies used a small number of loci that sometimes gave conflicting results. To further investigate the phylogenetic relationship of these species, we used the concatenated amino acid sequences of 12 genes that we sequenced in both *D. atripex* and *D. phaeopleura* (Appendix C) and

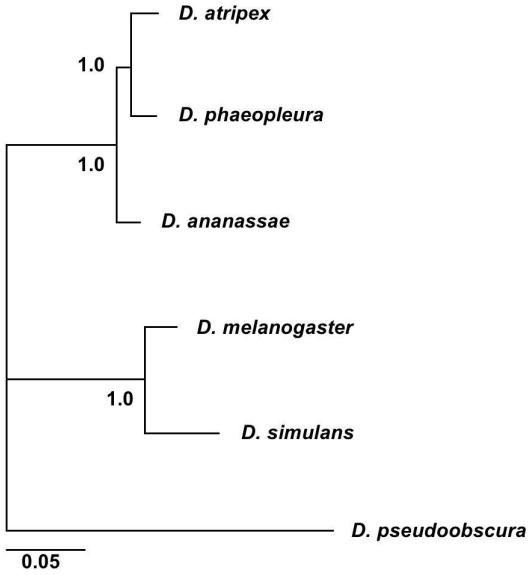

Figure 2.3: Phylogeny

50% majority-rule consensus tree generated from concatenated amino acid sequences of 12 genes (3,446 sites) and Bayesian inference analysis (10,000 generations). Identical topologies were recovered in two independent runs of MrBayes (Huelsenbeck and Ronquist, 2001). Numbers at the nodes indicate posterior clade probabilities.

for which sequence data were available from *D. ananassae*, *D. melanogaster*, *D. simulans*, and *D. pseudoobscura*. The concatenated alignment of 3,446 amino acid positions was used to generate a phylogenetic tree by Bayesian inference (Figure 2.3). The tree topology was strongly supported (100% posterior clade probabilities for each node) and indicated that *D. atripex* and *D. phaeopleura* are more closely related to each other than either is to *D. ananassae*. Thus, both of these species can be used as an outgroup to *D. ananassae*. Furthermore, the divergence between *D. ananassae* and *D. atripex*/*D. phaeopleura* is similar to the divergence between *D. melanogaster* and *D. simulans*, which facilitates the comparison of evolutionary patterns between the *melanogaster* and *ananassae* subgroups (see below).

Species	Bias	Genes	d_S	d_N	d_N/d_S
Dmel vs. *Dsim*	M	17	0.151	0.028	0.192
	F	16	0.128	0.019	0.154
	U	10	0.113	0.016	0.126
Dana vs. *Dat*	M	11	0.194	0.020	0.106
	F	8	0.222	0.017	0.071
	U	9	0.171	0.016	0.131
Dana vs. *Dph*	M	11	0.196	0.018	0.084
	F	8	0.160	0.019	0.118
	U	9	0.222	0.022	0.105

[a]For the *Dmel* vs. *Dsim* comparison, sex-bias classifications come from *D. melanogaster* microarray data. For the other comparisons, sex-bias classifications come from *D. ananassae* microarray data. M, male-biased; F, female-biased; U, unbiased.

Table 2.1: Evolutionary rates of genes with male-, female-, and unbiased expression

2.3.3 Levels of polymorphism and divergence

We surveyed DNA sequence polymorphism in a sample of 12 isofemale lines from Bangkok, Thailand for a subset of 43 genes used in our microarray analysis (Appendix C). This set included 17 male-biased genes, 12 female-biased genes, and 14 unbiased genes. Fourteen of these genes differed in their sex-bias classification between *D. ananassae* and *D. melanogaster*. In addition, we sequenced all genes in a single strain of either *D. atripex* or *D. phaeopleura* to use as an outgroup. For 13 genes we were able to get sequences from both *D. atripex* and *D. phaeopleura*, revealing that both exhibit similar levels of divergence to *D. ananassae* (on average, 20% at synonymous sites and 2% at nonsynonymous sites) and that the choice of outgroup does not bias estimates of interspecific divergence (Table 2.1).

Comparison of the divergence between *D. ananassae* and *D. atripex*/*D. phaeopleura* with the divergence between *D. melanogaster* and *D. simulans* for the different sex-bias classes of genes suggests that the accelerated rate of evolution of male-biased genes is limited to the *melanogaster* subgroup (Table 2.1). When comparing *D. melanogaster* and *D. simulans*, values of d_N and d_N/d_S are consistently higher for male-biased genes than for female-biased or unbiased genes, which is not the case for comparisons of *D. ananassae* to either *D. atripex* or *D. phaeopleura* (Table 2.1). However, there is no significant difference in d_N/d_S between male-biased genes on the different lineages (Mann-Whitney

Bias	Genes	π_s[a]	π_n[b]	π_n/π_s	θ_s[c]	θ_n[d]
		D. melanogaster (Zimbabwe, Africa)				
Male	17	0.0215	0.0015	0.0680	0.0234	0.0018
Female	16	0.0127	0.0017	0.1340	0.0135	0.0021
Unbiased	10	0.0149	0.0017	0.1115	0.0172	0.0020
All autosomal	29	0.0165	0.0013	0.0813	0.0175	0.0017
All X-linked	14	0.0176	0.0021	0.1198	0.0204	0.0025
All	43	0.0168	0.0016	0.0942	0.0185	0.0019
		D. ananassae (Bangkok, Thailand)				
Male	17	0.0197	0.0014	0.0723	0.0208	0.0017
Female	12	0.0234	0.0015	0.0628	0.0235	0.0018
Unbiased	14	0.0240	0.0017	0.0702	0.0249	0.0020
All autosomal	29	0.0236	0.0014	0.0590	0.0241	0.0017
All X-linked	14	0.0191	0.0018	0.0927	0.0203	0.0020
All	43	0.0221	0.0015	0.0687	0.0228	0.0018

[a]The average number of nucleotide differences per synonymous site.
[b]The average number of nucleotide differences per nonsynonymous site.
[c]Mean nucleotide diversity (per site) at synonymous sites.
[d]Mean nucleotide diversity (per site) at nonsynonymous sites.

Table 2.2: Intraspecific polymorphism in *D. melanogaster* and *D. ananassae*

test, $P > 0.05$). It should be noted, however, that the number and identity of the genes are not constant among the three comparisons.

Levels of DNA sequence polymorphism in the Bangkok population of *D. ananassae* are comparable to levels in an ancestral African population of *D. melanogaster* (Table 2.2). On average, synonymous polymorphism is slightly higher in *D. ananassae* than in *D. melanogaster*, suggesting that the former has a slightly larger effective population size (N_e). This is consistent with the Bangkok population representing an ancestral population of *D. ananassae* (Das *et al.*, 2004). Taken together, levels of polymorphism and divergence suggest that our *D. ananassae* population and outgroup species are appropriate for comparison to the African *D. melanogaster* population with *D. simulans* as an outgroup.

We also investigated levels of synonymous polymorphism on the X chromosome relative to the autosomes in both *D. ananassae* and *D. melanogaster*. This provides an estimate of the effective population size of males and females in each species. The ratio is expected to be $\frac{3}{4}$ if there is an equal number of breeding males and females in the population. If

	D. ananassae		*D. melanogaster*	
	X	Autosomal	X	Autosomal
θ_S	0.0203	0.0233	0.0204	0.0175
d_S	0.1860	0.1992	0.1418	0.1294
θ_S/d_S	0.1091	0.1210	0.1437	0.1352
$X:A^a$		0.91		1.06

[a]The ratio of X chromosomal to autosomal effective population size estimated from θ_S/d_S.

Table 2.3: Synonymous polymorphism and divergence at X-linked and autosomal loci

there is sexual selection acting on males, the male effective population size will be reduced and the X:autosome polymorphism ratio will be increased above $\frac{3}{4}$. After standardizing polymorphism by divergence to correct for possible differences in mutation rate between the X and the autosomes, we find that the X:autosome polymorphism ratio is greater than $\frac{3}{4}$ for both species (Table 2.3). To test the significance of this, we multiplied the X chromosomal values by $\frac{4}{3}$ and compared them to the autosomal values with a Mann-Whitney test. For *D. melanogaster*, the difference was significant ($P = 0.04$), indicating that the original ratio was significantly greater than $\frac{3}{4}$. For *D. ananassae*, an X:autosome polymorphism ratio of $\frac{3}{4}$ could not be rejected ($P = 0.13$). There was no significant difference in the X:autosome polymorphism ratio between *D. ananassae* and *D. melanogaster* ($P = 0.42$).

2.3.4 McDonald-Kreitman tests

To evaluate the type of selection operating on individual genes, we applied single-locus McDonald-Kreitman (MK) tests (McDonald and Kreitman, 1991) to the 43 genes for which we had polymorphism data from *D. ananassae* and divergence data from either *D. atripex* or *D. phaeopleura*. In total, six genes (14%) gave a significant MK test result (Appendix C). Half of the significant genes departed from neutrality in the direction of positive selection (*i.e.* a relative excess of nonsynonymous divergence) (Table 2.4), while the other half departed from neutrality in a pattern consistent with either balancing or weak purifying selection (*i.e.* a relative excess of nonsynonymous polymorphism). These are likely to be cases of weak purifying selection, as average values of Tajima's D (Tajima, 1989) are significantly negative at nonsynonymous sites (Table 2.5), which suggests that many segregating nonsynonymous polymorphisms are slightly deleterious.

Gene	Bias	D_S	P_S	D_N	P_N	P-value	D_S	P_S	D_N	P_N	P-value
		D. ananassae					*D. melanogaster*				
CG6980	MM	24	10	16	1	0.036	16	5	13	4	0.984
CG14717	UM	43	28	40	5	0.001	23	11	7	6	0.38
CG10750	UM	28	17	10	1	0.047	21	20	10	0	0.004
CG3085	MM	44	33	3	1	0.468	25	41	5	1	0.028
CG18341	UM	1	9	2	4	0.254	26	22	18	2	0.003
CG1314	MM	43	24	13	5	0.518	33	12	80	4	0.0004
CG9723	UU	64	10	16	7	0.076	36	18	61	5	0.0003

[a]First letter indicates expression in *D. ananassae*, second letter expression in *D. melanogaster* ("M"= male-biased, "F"= female-biased, "U"= unbiased). The last three rows show all male-, female-, and unbiased genes according to the classification in *D. ananassae*.
[b]The total number of synonymous fixed differences.
[c]The total number of nonsynonymous fixed differences.
[d]The total number of synonymous polymorphisms.
[e]The total number of nonsynonymous polymorphisms.
[f] P-value was determined by G-test when applicable, otherwise by Fisher's exact test.

Table 2.4: Genes with significant McDonald-Kreitman tests for positive selection

Bias	Synonymous	Nonsynonymous
M	-0.21 (0.270)	-0.47 (0.046)
F	-0.08 (0.453)	-0.76 (0.040)
U	-0.19 (0.291)	-0.56 (0.030)

P-values (in parantheses were determined as the proportion of 1000 simulations giving a Tajima's D value less than or equal to the observed.

Table 2.5: Average values of Tajima's D

A comparison of our MK test results to those from *D. melanogaster*/*D. simulans* (Baines *et al.*, 2008; Pröschel *et al.*, 2006) revealed that, in all but one case (*CG10750*), the individual genes showing significant evidence for positive selection (*i.e.* those with MK test P-values less than 0.05) differed between the two lineages (Table 2.4). Significance was assumed with a P-value smaller than 0.05. Most genes significant for positive selection were male-biased in at least one of the species. However, the proportion of positively-selected male-biased genes in *D. ananassae* ($\frac{1}{17} = 6\%$) was not as high as in previous studies of *D. melanogaster* where $\frac{7}{33}$ (21%) of autosomal male-biased (Pröschel *et al.*, 2006), and $\frac{7}{17}$ (41%) of X-linked male-biased genes (Baines *et al.*, 2008) gave significant MK tests for positive selection. Of the three genes with significant MK tests for positive selection in *D. ananassae*, one (*CG6980*) showed conserved male-biased expression between *D. ananassae* and *D. melanogaster*, while two (*CG14717* and *CG10750*) were male-biased in *D. melanogaster*, but unbiased in *D. ananassae* (Table 2.4).

To further investigate expression class-dependent patterns of evolution, we constructed 12 groups of genes (Table 2.6). The first nine classes were partitioned with regard to sex-bias in *D. ananassae* and *D. melanogaster*. This allowed us to compare genes with conserved sex-bias between the two species with genes showing sex-bias in only one species or genes showing a reversal of sex-bias between species. The final three classes were partitioned solely on expression in *D. ananassae*, without regard to expression in *D. melanogaster*. Application of the MK test to the summed values of polymorphism and divergence within each group (Fay *et al.*, 2001) revealed a significant departure from neutrality in the direction of positive selection for genes with conserved male-biased expression and for genes with male-biased expression private to *D. melanogaster*, whereas the female-biased and unbiased genes did not differ significantly from the neutral expectation (Table 2.6). The genes with male-biased gene expression private to *D. ananassae* also did not differ significantly from the neutral expectation.

2.3.5 Estimation of the proportion of adaptive amino acid replacements

We also used a multi-locus version of the MK test to estimate α , the fraction of amino acid replacements between species that were fixed by positive selection (Bierne and Eyre-Walker, 2004). For this, we divided the genes into six different groups. First, we considered male-, female-, and unbiased genes that had conserved sex-biased expression between *D.*

$Bias^a$	Genes	D_S	D_N	P_S	P_N	α	P_S	P_N	α
MM	13	448	164	203	43	0.42**	122	19	0.57***
MU	2	44	6	5	6	-7.80**	4	2	-2.67
MF	2	77	39	14	8	-0.13	4	3	-0.48
FF	9	296	106	134	33	0.31	90	15	0.53**
FU	2	51	25	19	3	0.68	16	0	na(**)
FM	1	49	16	16	5	0.04	7	1	0.56
UU	6	163	34	97	19	0.06	60	4	0.68*
UM	3	72	52	54	10	0.74***	20	5	0.65*
UF	5	154	49	41	22	-0.69	37	10	0.15
M-	17	569	209	222	57	0.30*	130	24	0.50**
F-	12	396	147	162	41	0.32	123	16	0.65***
U-	14	389	135	192	51	0.23	117	19	0.53**

Symbols are the same as in blabla-table.
[a]$\alpha = 1 - [(D_S P_N)/(D_N P_S)]$. Asterisks indicate the significance of the summed data as determined by a G-test ($^* P < 0.05$,$^{**} P < 0.01$,$^{***} P < 0.001$).
[b]Excluding low frequence ($\leq 15\%$) polymorphisms.

Table 2.6: Summary of polymorphic and divergent sites

melanogaster and *D. ananassae*. Second, we considered the genes according only to their bias in *D. ananassae*. We could not consider all 12 groups of genes mentioned above, because some of the groups contained too few genes to allow maximum likelihood analysis (*e.g.* there were only two genes with male-biased expression private to *D. ananassae*). Because the segregation of slightly deleterious nonsynonymous mutations can lead to a downward bias in the estimate of α, we excluded all polymorphisms (both synonymous and nonsynonymous) segregating at frequency $\leq 15\%$ from our analysis (Charlesworth and Eyre-Walker, 2008).

For the genes with conserved sex-biased expression between species, we observed a significant signal of positive selection only for those with conserved male-biased expression, although in all cases the mean estimate of α was greater than zero (Figure 2.5). However, when we considered the genes only by their sex-bias classification in *D. ananassae* we found evidence for adaptive evolution in all three classes of genes, with higher α estimates and more significant departures from neutrality detected for female-biased and unbiased genes than for male-biased genes (Figure 2.5). Thus, the increased rate of adaptive evolution seen for male-biased genes appears to be limited to those with conserved male-biased expression

between the two species. For *D. ananassae*, estimates of α were 43%, 60%, and 53% for male-biased, female-biased, and unbiased genes, respectively.

2.3.6 Estimation of the selection parameter

We used the Bayesian analysis method of (Bustamante *et al.*, 2002) to estimate the selection parameter $\gamma = 2N_e s$, where N_e is the effective population size and s the selection coefficient for amino acid replacements in a group of genes. The method assumes that γ is normally distributed among genes. In the case of neutral evolution, γ is expected to be zero. Values of γ above zero indicate positive selection, whereas negative values either indicate balancing or weak purifying selection. Again, we considered our data in six different groups (see above). The estimated selection parameters were greater than zero for all groups of genes except the group of all female-biased genes which had a mean value of -0.25, indicating weak purifying selection against nonsynonymous mutations (Figure 2.4A). In both classes of male-biased genes, the proportion of the distribution falling below zero was the lowest ($P_{(y\leq 0)} = 0.0096$ and $P_{(y\leq 0)} = 0.018$, for conserved and all male-biased genes, respectively), indicating positive selection favoring amino acid replacements. In contrast, the greatest difference could be seen between conserved female-biased and all female-biased genes ($P_{(y\leq 0)} = 0.237$ and $P_{(y\leq 0)} = 0.733$, respectively). After repeating the analysis with singleton polymorphisms excluded as above, the estimates for conserved male-, female-, and unbiased genes were 2.6, 1.24, and 1.18 ($P_{(y\leq 0)} = 0.001$, $P_{(y\leq 0)} = 0.044$, and $P_{(y\leq 0)} = 0.105$, respectively), The estimates for all male-, female-, and unbiased genes were 1.94, 0.8, and 2.7 ($P_{(y\leq 0)} = 0.0004$, $P_{(y\leq 0)} = 0.0644$, and $P_{(y\leq 0)} = 0.0074$, respectively; Figure 2.4B). Overall, the groups including unbiased genes showed the highest standard deviations. This can be explained by the fact, that unbiased genes had the lowest proportion of conserved genes between *D. melanogaster* and *D. ananassae* (six out of 14 genes) in our dataset. Moreover, as the group of conserved unbiased genes only included six genes, a higher standard deviation compared to conserved male-biased (13 genes) or female-biased (nine genes) genes is explicable. Generally, it can be seen that for the estimation of the selection parameter γ there is low variation between groups of genes.

As calculations might not be correct when using X and autosomal loci together (C. Bustamante, personal communication), we also performed the analyses using autosomal genes only which basically show the same results.

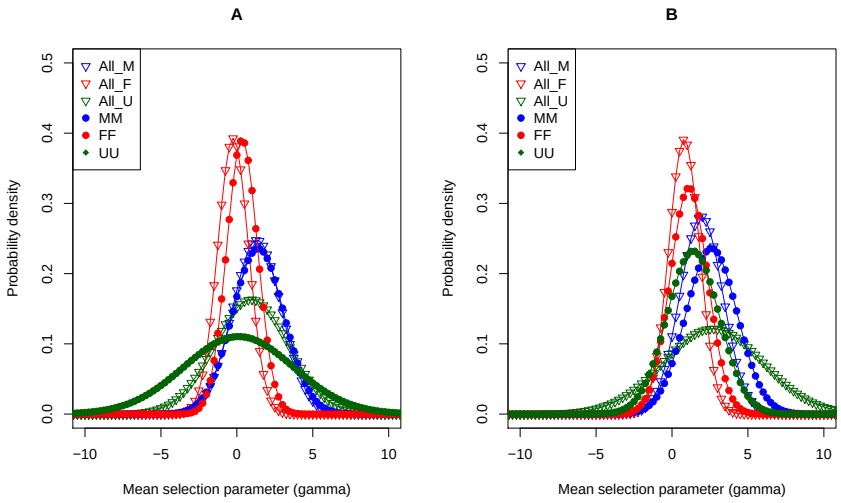

Figure 2.4: Estimation of the selection parameter γ

Bayesian posterior distribution of the mean selection parameter (γ) for male-biased (blue), female-biased (red), and unbiased (green) genes. Panel A displays the distribution of mean γ using all polymorphic sites. Panel B displays the distribution of mean γ after excluding all low-frequency polymorphisms (singletons).

2.4 Discussion

Although sex-biased expression was conserved between *D. ananassae* and *D. melanogaster* for the majority of the genes we analyzed, a large minority (40%) showed a gain, loss, or reversal of sex-biased expression between species (Figure 2.2). In a previous microarray study, Ranz *et al.* (Ranz *et al.*, 2003) found that 20% of genes showed a gain, loss, or reversal of sex-biased expression between *D. melanogaster* and *D. simulans*. The higher percentage observed in the present study is likely attributable to the greater evolutionary distance separating *D. melanogaster* and *D. ananassae* (Figure 2), which provides more opportunity for expression changes. Consistent with this, a previous SAGE study found that 34% of genes changed their sex-biased expression pattern between *D. melanogaster* and the more distantly-related *D. pseudoobscura* (Metta *et al.*, 2006).

Several aspects of our experimental design might also contribute to the gene expression patterns described above. First, the genes on our array were not a random set, but instead were enriched for those showing strong sex-biased expression in multiple, independent experiments in *D. melanogaster*. Thus, we might expect to see an overrepresentation of

genes that are sex-biased only in *D. melanogaster*, and an underrepresentation of genes that are sex-biased only in *D. ananassae*. Indeed, this is what we find: of the 48 genes that are sex-biased in only one species, 33 (69%) are sex-biased in *D. melanogaster*, while only 15 (31%) are sex-biased in *D. ananassae* (Figure 2.2B). A second factor that may influence our results is that our *D. ananassae* classifications come from a combination of our own experimental data and those from a published whole-genome microarray study (Zhang *et al.*, 2007) and the two studies differ in their design and replication. Our own microarray experiments examined only a small set of genes, which allowed for many replicate probes per gene to be present on each array. We also enforced strict quality control to exclude genes with weak hybridization signal (which would otherwise be classified as unbiased). The use of the published whole-genome arrays allowed us to classify more genes, but the classification may be less reliable due to the lower replication and the increased problem of multiple testing that weakens the statistical analysis. Indeed, the whole-genome data classify a higher fraction of unbiased genes (68% vs. 33% for our custom arrays), which would be expected since the null hypothesis of these comparisons is that there is no difference in expression level between males and females. Also, it should be noted that the whole-genome arrays of Zhang *et al.* (2007) only identified around 12% of the *D. ananassae* transcriptome as sex-biased, while the previous studies that our *D. melanogaster* classifications are based on identified 20-70% of the transcriptome as sex-biased (Gibson *et al.*, 2004; Gnad and Parsch, 2006; Parisi *et al.*, 2003; Ranz *et al.*, 2003). This difference is most likely the result of differences in experimental design and statistical power between the studies, not an underlying difference in the amount of sex-biased expression between the two species, as Zhang *et al.* (2007) also reported similarly low percentages of sex-biased genes in six other *Drosophila* species, including *D. melanogaster*.

Despite the presumably weaker power to detect sex-biased expression in *D. ananassae*, we observed 15 genes that were sex-biased in *D. ananassae*, but unbiased in *D. melanogaster* (Figure 2.2B). For 12 of these genes, the ancestral expression state could be inferred using microarray data from the outgroup species *D. pseudoobscura* (Zhang *et al.*, 2007) (Appendix E). Interestingly, 11 of these 12 genes showed a match between the *D. melanogaster* and the *D. pseudoobscura* classification, suggesting that sex-biased expression was gained on the *D. ananassae* lineage in the vast majority of cases. The ancestral expression state could also be inferred for four of the five genes that showed a reversal of sex-bias between *D. melanogaster* and *D. ananassae*. Three of these genes were

female-biased in *D. melanogaster* and *D. pseudoobscura*, but male-biased in *D. ananassae*. One gene was male-biased in *D. pseudoobscura* and *D. melanogaster*, but female-biased in *D. ananassae*. However, this gene (*CG7387*) also differed in its sex-bias classification between our custom microarrays (female-biased) and the whole-genome *D. ananassae* microarrays (male-biased) (Zhang *et al.*, 2007). Thus, it is possible that the sex-biased expression of this gene is strain- or condition-dependent. This was the only such conflict between the two *D. ananassae* expression datasets, although there were 26 cases where a gene was classified as sex-biased in one dataset and unbiased in the other.

Our survey of DNA sequence polymorphism in *D. ananassae* is the largest performed to date in terms of number of loci investigated and the first to examine genes with sex-biased expression. Overall, the level of polymorphism in the *D. ananassae* population from Bangkok, Thailand is similar to that in an ancestral African *D. melanogaster* population, which is consistent with Bangkok being within the ancestral species range of *D. ananassae* (Das *et al.*, 2004). When considering all loci we find that *D. ananassae* has slightly more synonymous polymorphism than *D. melanogaster* (Table 2.2), which suggest that the former has a larger N_e. Consistent with this interpretation, nonsynonymous polymorphism and the ratio of nonsynonymous to synonymous polymorphism are lower in *D. ananassae* (Table 2.2). This is expected if most segregating nonsynonymous mutations are slightly deleterious, as is suggested by their negative values of Tajima's D (Table 2.4), because purifying selection is more effective at removing deleterious mutations when N_e is large.

Multi-locus analyses of polymorphism and divergence indicate that adaptive protein evolution is prevalent in *D. ananassae*, with estimates of α in the range of 50-60% (Table 2.3 and Figure 2.5). These values are remarkably similar to those estimated for other *Drosophila* species (reviewed in Sella *et al.* (2009)), which suggests that there is a consistently high rate of adaptive protein evolution throughout the genus. However, despite the strong overall signal of positive selection acting on *D. ananassae* proteins, we do not see clear differences in the prevalence of adaptive protein evolution among male-, female-, and unbiased genes. This contrasts with previous results from *D. melanogaster* that indicated an increased rate of adaptive evolution in male-biased genes (Baines *et al.*, 2008; Pröschel *et al.*, 2006), but is consistent with the results of Metta *et al.* (2006), who found accelerated rates of evolution (as measured by d_N) for *D. melanogaster* male-biased genes, but not *D. pseudoobscura* male-biased genes. We see the same general trend for d_N (and d_N/d_S)

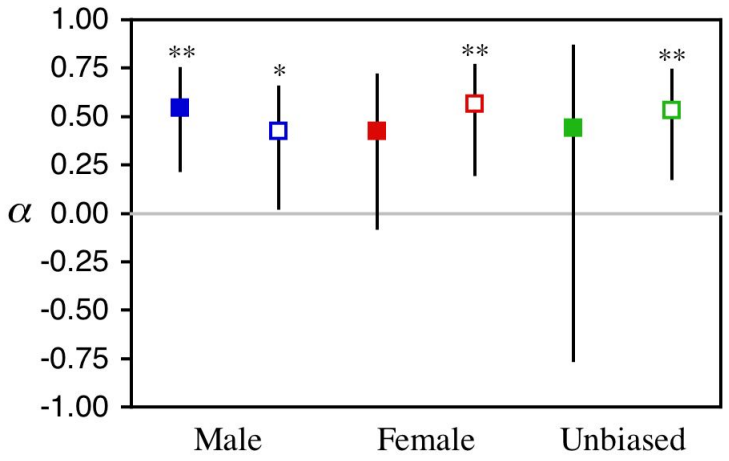

Figure 2.5: Maximum likelihood estimates of the fraction of positively selected amino acid replacements (α)

Values of for genes with male-, female-, and unbiased expression were calculated using a maximum likelihood method (Bierne and Eyre-Walker, 2004). Genes with conserved bias between *D. ananassae* and *D. melanogaster* are indicated with solid boxes, genes with bias according to classification in *D. ananassae* (*i.e.* including conserved genes and genes with bias private to *D. ananassae*) are indicated with open boxes. Low frequency polymorphisms ($\leq 15\%$) were excluded. Error bars indicate 95% CI. Asterisks indicate genes with a significant signal of positive selection ($^{*}P < 0.05, ^{**}P < 0.01$).

values in our data, which are highest for the male-biased genes of *D. melanogaster*, but lower for the male-biased genes of *D. ananassae* (Table 2.1). Taken together, these results suggest that male-biased genes have experienced an increase in the rate of adaptive protein evolution since the divergence of the *ananassae* and *melanogaster* subgroups (Figure 2.3).

The above results could be explained if sexual selection drives the evolution of many male-biased genes and the prevalence of sexual selection differs between the two subgroups. In ancestral *D. melanogaster* populations, X-linked synonymous polymorphism is significantly greater than $\frac{3}{4}$ of autosomal synonymous polymorphism (Parsch *et al.*, 2009), which is expected if sexual selection acts on males (reviewed in Ellegren (2009)). We observe a similar pattern for *D. ananassae* (Table 2.3), suggesting that sexual selection does not differ greatly between the species. However, due to the relatively small number of *D. ananassae* loci, we cannot detect a significant difference between the observed X:autosome diversity ratio and the expected ratio of $\frac{3}{4}$, nor can we detect a significant difference in this ratio between *D. ananassae* and *D. melanogaster*. Thus, we cannot exclude the possibility that sexual selection is weaker in the *D. ananassae* lineage. Mate-choice experiments,

however, indicate that there is significant female mate preference in *D. ananassae* (Schug *et al.*, 2008) and that male courtship song plays a major role in female mate discrimination (Doi *et al.*, 2001). Furthermore, it has been shown that *D. ananassae* males are subject to stronger intrasexual selection than females and that male mating success is correlated with morphological traits, such as body size and sternopleural bristle number (Singh and Singh, 2003). These findings suggest that sexual selection plays an important role in the species' evolution.

The observed differences in male-biased gene evolution between *D. ananassae* and *D. melanogaster* could also be influenced by the particular genes that were investigated. Our initial gene set was enriched for genes that showed strong and consistent sex-biased expression across multiple *D. melanogaster* microarray experiments. On average, the subset of male-biased genes showed a 12-fold male bias in *D. melanogaster*, but only a 5-fold male bias in *D. ananassae*. While it is difficult to directly compare male/female expression ratios across different experiments and microarray platforms, it is likely that most of the genes investigated had a stronger male bias in *D. melanogaster* than in *D. ananassae*. If the degree of male bias is correlated with the rate of adaptive evolution across the genus, as it is in the *melanogaster* subgroup (Baines *et al.*, 2008), it could explain differences in male-biased gene evolution between the lineages. This possibility could be addressed in future studies focusing on genes with exceptionally strong male-biased expression in *D. ananassae*. However, there is not a significant correlation between d_N/d_S and the male/female expression ratio within our *D. ananassae* dataset (Pearson's R=0.10, $P = 0.52$). In general, the female-biased genes included in our study showed weaker sex-biased expression than the male-biased genes in *D. melanogaster* (4-fold vs. 12-fold). This might explain why expression conservation between *D. melanogaster* and *D. ananassae* was greater for male-biased genes than female-biased genes (Figure 2.2), as highly biased genes are more likely to be detected as significant, while weakly biased genes are more likely to be non-significant and classified as unbiased.

2.5 Conclusions

Although sex-biased gene expression is abundant in *Drosophila* species, the sex-biased expression pattern of many genes differs between species. Species-specific microarray data indicate that 26% of genes with a strong sex-bias in *D. melanogaster* show no detectable sex-bias in *D. ananassae*, while 12% of genes with unbiased expression in *D. melanogaster* show significantly sex-biased expression in *D. ananassae*. The accelerated rate of adaptive evolution seen for male-biased genes in *D. melanogaster* is not observed for male-biased genes in *D. ananassae*, which suggests that there are differences in sex-biased gene evolution between the two lineages. This is in agreement with a previous study on rates of protein evolution in *D. pseudoobscura* (Metta *et al.*, 2006) and suggests that the rapid adaptive evolution of male-biased genes is unique to the *melanogaster* subgroup and not a general pattern in Drosophilids. Despite these differences, the overall signal of adaptive protein evolution is strong in *D. ananassae* ($\alpha \approx 50\%$) and is consistent with previous estimates throughout the genus.

Chapter 3

Population structure and adaptive evolution of protein-coding genes in *Drosophila ananassae*

3.1 Introduction

During the last decade, studies of population structure and genetic differentiation in *Drosophila ananassae* mainly concentrated on non-coding sequences. Work was done on intronic loci (Das *et al.*, 2004), microsatellite repeat-length polymorphisms (Schug *et al.*, 2007), and mitochondrial loci (Schug *et al.*, 2008). Studies on gene loci were limited to a few genes (Baines *et al.*, 2004; Chen *et al.*, 2000; Stephan *et al.*, 1998; Vogl *et al.*, 2003). *D. ananassae* is a highly-structured cosmopolitan and human commensal species. It is situated in the tropical and subtropical regions of the world, mainly in Asia and the Pacific, and its demographic history is well understood. Using a multilocus approach with ten neutral intronic loci, 16 population samples across the biogeographic species range, and a Bayesian approach (Vogl *et al.*, 2003), five populations from Southeast Asia were found to be ancestral (Das *et al.*, 2004). These five populations had been sampled from localities that belonged to one single landmass, called "Sundaland", that existed during the late Pleistocene around 18,000 years ago. The authors inferred migration routes out of Sundaland and found that these routes appear to parallel those of humans in this region. Moreover, *D. ananassae* can only be found close to human settlements in the wild (A. Kopp, personal communication). Population size expansion is especially evident in the ancestral populations. Strong evidence for two recent independent selective sweeps in

the northern and southern populations could be inferred at *furrowed* (*fw*), a gene that is located in a chromosomal region of very low recombination (Baines *et al.*, 2004; Chen *et al.*, 2000). This locus was compared to a gene (*Om(1D)*) located in a region of normal recombination. *Om(1D)* shows an isolation-by-distance effect and no deviation from neutral expectation. In addition, there is evidence for a recent selective sweep in another gene, *vermilion* (*v*) (Stephan *et al.*, 1998). Similar to *fw*, this gene is located in a chromosomal region of low recombination on the other side of the centromere on the X chromosome.

The above are the only protein-coding genes that have been used for population genetic analyses over the worldwide biogeographic range of *D. ananassae*. Microsatellite repeat-length variation analysis among several populations revealed high levels of genetic structure among all populations, particularly in Australasia and the South Pacific, and population expansion within all populations (Schug *et al.*, 2007). Isolation-by-distance among populations was evident at microsatellites, mitochondrial DNA and intronic loci (Das *et al.*, 2004; Schug *et al.*, 2007, 2008). Experimental tests of mate-choice found levels of pre-mating isolation to be highest between various peripheral populations (Schug *et al.* (2008), see CHAPTER 1). This suggests that sexual selection occurs within local populations, leading to reproductive isolation among subpopulations in this species.

To date, there is a lack of studies investigating genes that may lead to population differentiation and reproductive isolation. One possible starting point would be to survey genetic variation of sex-biased genes in *D. ananassae*. Recently, we investigated sex-biased gene expression in this species (Grath *et al.* (2009), see CHAPTER 2). We surveyed polymorphism and divergence levels of 43 protein-coding genes in one ancestral population from Bangkok, Thailand. For three autosomal genes, we found evidence for positive selection. These genes show male-biased gene expression either in *D. melanogaster* or both in *D. melanogaster* and *D. ananassae*. In this chapter, we survey DNA sequence polymorphism at these genes in twelve populations to determine the relative contribution of male-biased genes to genetic differentiation, adaptation, and reproductive isolation. Four autosomal genes with unbiased gene expression were previously found not to deviate from neutral evolution in the Bangkok population (Grath *et al.* (2009), see CHAPTER 2). We use these as reference genes and compare their variation in all populations to ten neutral X-linked genes analyzed by Das *et al.* (2004).

Previous studies of *D. ananassae* mainly used *D. pallidosa* for interspecific comparisons. However, *D. pallidosa* has found to be more closely related to some of the *D. ananassae* populations than many of the *D. ananassae* populations are to one another, at least when

measured by microsatellite repeat-length variation (Schug *et al.*, 2007). *D. pallidosa* shows incomplete mating discrimination against *D. ananassae* populations and no post-mating isolation (Schug *et al.*, 2008). Further, genetic divergence between these species at intronic and coding loci is very low (Baines *et al.*, 2004; Chen *et al.*, 2000; Stephan *et al.*, 1998; Vogl *et al.*, 2003). For this study, we use *D. atripex* and *D. phaeopleura* for determining interspecific divergence. Previously, we investigated DNA sequences of 13 protein-coding genes in these two species and found that both exhibit similar levels of divergence to *D. ananassae*, on average 20% at synonymous sites and 2% at nonsynonymous sites (Grath *et al.* (2009), see CHAPTER 2).

3.2 Materials and Methods

3.2.1 Population samples

A total of 107 *D. ananassae* isofemale lines from 12 locations in India, South-East Asia, Australia, and Japan were used for generating polymorphism data. The location, abbreviation, number of lines, and date of collection are listed for each population in Table 3.1. Most samples were collected by Aparup Das. Samples from Thursday Island, Australia and Trinity Beach, Australia, were provided by Malcolm Schug. One strain of *D. atripex* or *D. phaeopleura* (kindly provided by M. Schug) was used as an outgroup.

3.2.2 Candidate loci

For this study, we used three autsomal genes (*CG10750*, *CG14717*, *CG6980*) which were found to be under positive selection according to the results described in CHAPTER 2 (Grath *et al.*, 2009). Additionally, we included four autosomal unbiased genes (*CG10853*, *CG13189*, *CG6981*, *CG7508*) that showed no evidence of positive selection. Further information on all loci sequenced for this study is given in Table 3.2.

3.2.3 DNA extraction, polymerase chain reaction and DNA sequencing

We extracted genomic DNA from individual male flies using the Puregene DNA isolation kit (Gentra Systems, Minneapolis, MN) or phenol-chloroform extraction. Primers for amplification or sequencing are given in Table 3.3. Following PCR, the amplified products

Sampling location	Coordinates (latitude, longitude)	Country of origin	Abbr.	No. of isofemale lines	Collection date
Chennai	13:00 N, 80:10 E	India	CH	9	2000
Bhubaneswar	20:15 N, 85:52 E	India	BBS	9	2000
Puri	19:48 N, 85:52 E	India	PUR	16	2000
Bangkok	13:50 N, 100:29 E	Thailand	BKK	12	2002
Bogor, Java	06:09 S, 106:51 E	Indonesia	BOG	6	2001
Mandalay	22:00 N, 96:08 E	Myanmar	MAN	8	2001
Kathmandu	27:49 N, 85:21 E	Nepal	KATH	6	2000
Kota Kinabalu, Borneo	05:56 N, 116:03 E	Malaysia	KK	8	2002
Kuala Lumpur	03:08 N, 101:42 E	Malaysia	KL	9	2002
Kumejima, Okinawa	26:21 N, 127:46 E	Japan	KMJ	8	2000
Trinity Beach	16:47 S, 145:41 E	Australia	TB	7	2003
Thursday Island	10:35 S, 142:13 E	Australia	TI	9	2003

Table 3.1: Population samples of *D. ananassae* used in this study

Gene	Bias[a]	Length	Intron no.	Intron sites	D_S[b]	D_N[c]	P_S[d]	P_N[e]	P-value[f]
CG6980	MM	843	2	99	24	16	10	1	0.04
CG10750	UM	1171	3	176	33	15	26	2	0.009
CG14717	UM	894	0	0	43	40	28	5	0.0006
CG10853	UU	708	3	213	9	3	4	0	0.53
CG13189	UU	1020	0	0	42	0	36	3	0.11
CG6981	UU	824	2	335	17	1	8	0	1
CG7508	UU	933	1	159	12	3	33	7	0.83

Values refer to the population of Bangkok, Thailand, which was previously studied for sex-biased gene expression in *D. ananassae* (Grath *et al.*, 2009).
[a]First letter indicates expression in *D. ananassae*, second letter expression in *D. melanogaster* (M = male-biased, U = unbiased).
[b]The total number of synonymous fixed differences between Bangkok population and *D. atripex*/*D. phaeopleura*.
[c]The total number of nonsynonymous fixed differences between Bangkok population and *D. atripex*/*D. phaeopleura*.
[d]The total number of synonymous polymorphisms in Bangkok population.
[e]The total number of nonsynonymous polymorphisms in Bangkok population.
[f]*P*-value was determined by *G*-test when applicable, otherwise by Fisher's exact test.

Table 3.2: Overview of sequenced loci

Gene	Forward primer (5'-3') (F)	Reverse primer (5'-3') (R)	Internal Primer
CG14717	aaagccatgttttcttaccctt	gctgaaatttcaggaaactccc	
CG6980	cttcggttgctatagcatcc	acagatttgggcagtgttcacc	
CG10750	ctatcataataattgacacca	aagctacgaagcggagagc	atgcgccagcgtatggtgc (Fint) atgcgatcaatctcatggct (Rint)
CG10853	acacctgtgcgaatcagatg	ttctccctgtgtgtgagcc	
CG6981	tgatatagtgctatccagtg	ttgtttattcggcgattgc	
CG7508	tgcctgattgcctgccatcg	ctactgggctcctagttacagg	
CG13189	agagctcctcagttgaaagc	acagatgccaccacatcagc	

Table 3.3: PCR and sequencing primers

were purified with ExoSAP-IT (USB, Cleveland, OH) and sequenced on both strands using BigDye version 1.1 chemistry and a 3730 automated sequencer (Applied Biosystems, Foster City, CA). All protocol details are provided in Appendix H.

3.2.4 Sequence analysis

Sequences were edited using DNAstar (Madison, WI) and multiple alignments were calculated using MUSCLE (Edgar, 2004) (`http://www.ebi.ac.uk/Tools/muscle/`). We used the DnaSP 4.5 program (Librado and Rozas, 2009) for most intraspecific analyses. Nucleotide diversity, θ, was estimated according to Watterson (1975) and π according to Nei (1987). For divergence, we considered only sites with fixed differences between all *D. ananassae* lines and a single *D. atripex* (or *D. phaeopleura*) sequence. We considered the number of segregating mutations (instead of the number of segregating sites), as some genes had sites with three segregating variants. For MK tests and α calculations, the *D. atripex* sequence was used as the outgroup whenever available. When the *D. atripex* sequence was not available (for *CG14717* and *CG13189*), *D. phaeopleura* was used as the outgroup. MK tests were performed on each gene for all populations separately and for all strains pooled together.

3.2.5 Neutrality tests

To test if the pattern of polymorphism at any of the seven loci considered in this study was inconsistent with a neutral model of molecular evolution, we calculated Tajima's D-statistic (Tajima, 1989) and Fu and Li's D-statistic (Fu and Li, 1993) for each locus and population separately. We calculated Fu and Li's D using an outgroup sequence as proposed in the

original paper (Fu and Li, 1993). Tajima's D-statistic considers the normalized difference between the estimates of π and θ and is expected to be around zero under the standard neutral model. Fu and Li's D-statistic is based on the difference between the total number of polymorphic sites and the number of derived singletons.

In addition, we used a multilocus version of the Hudson-Kreitman-Aguadé (HKA) test (Hudson *et al.*, 1987) to compare the ratio of polymorphism to divergence across the seven genes for all 12 populations separately. Multilocus HKA tests were performed using the program HKA, kindly provided by Jody Hey (`http://lifesci.rutgers.edu/~heylab/heylabsoftware.htm`). This program was also used for multilocus Tajima's D and Fu and Li's D calculations. Here, the test statistics are compared with the distributions generated from 10,000 coalescent simulations (Kliman *et al.*, 2000).

Further, we performed the HKA test for all pairwise comparisons between loci (seven genes $\rightarrow$ 21 comparisons), for each of the 12 populations separately. For this, we used a program which was kindly provided by Lino Ometto. The program calculates the χ^2 values for each comparison. The critical value for a significant test (one degree of freedom) at significance level $P \leq 0.05$ would be 3.84. For each population, the probability of observing at least i significant tests at one specific locus given that n paired tests are performed and k are significant between the l loci, can be calculated as

$$P = \sum_{j=i}^{min(l-1,k)} \frac{\binom{k}{j}\binom{n-k}{l-1-j}}{\binom{n}{l-1}}$$

(Baines *et al.*, 2004).

We determined straight-line geographic distances between populations using geographic coordinates for each collecting site (Table 3.1) and a web-based calculator (`http://www.go.ednet.ns.ca/~larry/bsc/jslatlng.html`) and compared these with values for genetic differentiation measured by F_{ST} (Hudson *et al.*, 1992) to test for isolation-by-distance (Wright, 1943).

3.3 Results and Discussion

3.3.1 DNA polymorphism and divergence

All loci

Combining all seven loci, we sequenced 6,393 bp per individual and ≈643 kb in total. For divergence, we used outgroup sequences (*D. atripex*/*D. phaeopleura*) from Grath *et al.* (2009). We used 107 isofemale lines (Table 3.1) for our analyses. However, we could not get sequence data for each strain at each locus. The average number of strains sequenced per locus was 101. The total length of individual loci ranges from 824 to 1,171 (Table 3.1).

The number of haplotypes at individual loci varies from 36 (*CG10853*) to 76 (*CG6981*, *CG7508*). The number of segregating sites per locus is lowest in Kathmandu (12) and highest in Thursday Island (34). *CG10853* has the lowest number of segregating sites on average (11), *CG10750* the highest (34).

Mean nucleotide diversity, π, varies from 0.0128 (Kathmandu) to 0.0375 (Thursday Island). Nucleotide diversity measured by θ shows a similar pattern with the lowest value of 0.0141 in Kathmandu and highest value of 0.0391 in Thursday Island. The means of π and θ (at silent sites) over all populations are 0.0279 and 0.0276, respectively.

Over all seven loci, there were 235 fixed nucleotide differences (163 synonymous, 72 nonsynonymous). Per-site divergence at silent sites averaged over all loci was found to be 15.61%.

Mean nucleotide diversity at our genes is consistently higher compared to ten neutral intronic loci (Das *et al.*, 2004). Taking into account only populations which were analyzed in both studies, mean nucleotide diversity, π, at introns ranges from 0.0059 (KMJ) to 0.0124. Nucleotide diversity measured by θ shows the highest value (0.0135) in KK and the lowest value (0.0046) in KATH.

We divided our set of genes into three genes that were previously found to be under positive selection in one population from Bangkok, Thailand (*CG14717*, *CG6980*, and *CG10750*) and four genes that are presumably neutral (*CG10853*, *CG6981*, *CG7508*, and *CG13189*) (Grath *et al.*, 2009). DNA sequence polymorphism for all loci and populations is given in Appendix F.

Gene	BBS	BKK	BOG	MAN	CH	KATH	KK	KL	KMJ	PURI	TB	TI
CG14717												
π	0.0036	0.0466	0.0427	0.0312	0.0263	0.0015	0.0452	0.0381	0.0384	0.0344	0.0360	0.0346
θ	0.0034	0.0437	0.0438	0.0298	0.0235	0.0020	0.0457	0.0419	0.0352	0.0350	0.0438	0.0333
Taj. D	0.1959	0.0880	-0.3158	0.1473	0.4591	-1.132	-0.1733	-0.4417	0.0523	-0.1585	-1.2747	0.1175
K	0.2116	0.2198	0.2203	0.2176	0.2134	0.2086	0.2163	0.2139	0.2148	0.2134	0.2086	0.2067
CG6980												
π	0.0315	0.0222	0.0240	0.0226	0.0242	0.0063	0.0291	0.0191	0.0162	0.0390	0.0409	0.0351
θ	0.0244	0.0213	0.0222	0.0225	0.0215	0.0051	0.0271	0.0172	0.0191	0.0364	0.0351	0.0402
Taj. D	1.1768	0.2494	0.5086	-0.0900	0.3163	0.7077	0.1187	0.6664	-0.4509	-0.0332	0.7580	-0.7572
K	0.1997	0.1924	0.1916	0.1899	0.1936	0.1970	0.1945	0.2044	0.1946	0.2001	0.1957	0.1994
CG10750												
π	0.0363	0.0361	0.0189	0.0485	0.0372	0.0261	0.0463	0.0187	0.0140	0.0376	0.0429	0.0387
θ	0.0418	0.0410	0.0221	0.0429	0.0392	0.0288	0.0390	0.0195	0.0145	0.0330	0.0427	0.0409
Taj. D	-0.0219	0.4432	-0.1894	-0.3624	-1.2835	-1.3673	-0.2261	-0.7803	-0.0161	-1.0669	-1.1896	-0.2819
K	0.1986	0.1947	0.1887	0.1971	0.1954	0.1908	0.1983	0.1909	0.1914	0.1934	0.1951	0.1957

Values for divergence K, π and θ are for silent sites only. Tajima's D is given for all sites.

Table 3.4: Summary of polymorphism at loci under positive selection

Selected loci

Summary statistics for genes that were previously found to be under selection in Bangkok, Thailand, are given in Table 3.4. Mean nucleotide diversity, π, varies from 0.0113 (KATH) to 0.0402 (KK) with an average value of 0.0303. Nucleotide diversity measured by θ shows the same tendency with the lowest value (0.0120) in KATH and the highest (0.0405) in TB with an average value of 0.0300. Average divergence at silent sites over all loci and populations was found to be 20.13%. There is no remarkable variation in mean nucleotide diversity and divergence between genes.

Presumably neutral loci

Summary statistics for genes that were previously found to be under no selection in Bangkok, Thailand, are given in Table 3.5. All these genes show unbiased gene expression both in *D. melanogaster* and in *D. ananassae*. Mean nucleotide diversity, π, varies from 0.0140 (KATH) to 0.0386 (TI) with an average value over all populations of 0.0304. Nucleotide diversity measured by θ is lowest in KATH (0.0157) and highest in TI (0.0398) with an average value of 0.0258. *CG10853* shows the lowest values of π, θ, and divergence. Average divergence at silent sites over all populations is only 5.34% for this gene, whereas average divergence at *CG7508*, *CG6981*, and *CG13189* are 10.75%, 13.22%, and 19.6%. Average divergence over all populations and all four genes is lower compared to se-

Gene	BBS	BKK	BOG	MAN	CH	KATH	KK	KL	KMJ	PURI	TB	TI
CG10853												
π	0.0064	0.0078	0.0133	0.0091	0.0048	0.0039	0.0073	0.0072	0.0098	0.0086	0.0180	0.0244
θ	0.0067	0.0070	0.0136	0.0098	0.0064	0.0051	0.0076	0.0083	0.0094	0.0115	0.0228	0.0245
Taj. D	-0.0219	0.4432	-0.1894	-0.3624	-1.2835	-1.3673	-0.2261	-0.7803	-0.0161	-1.0669	-1.1896	-0.2819
K	0.0548	0.0512	0.0537	0.0523	0.0532	0.0538	0.0502	0.0515	0.0521	0.0542	0.0555	0.0582
CG6981												
π	0.0129	0.0148	0.0213	0.0139	0.0134	0.0038	0.0229	0.0153	0.0206	0.0143	0.0197	0.0292
θ	0.0138	0.0149	0.0219	0.0144	0.0138	0.0033	0.0241	0.0165	0.0175	0.0181	0.0216	0.0336
Taj. D	-0.3776	-0.4620	-0.2949	-0.6503	-0.2417	0.5081	-0.5067	-03882	0.7376	-1.0969	-0.6161	-0.9651
K	0.1325	0.1308	0.1326	0.1304	0.1272	0.1301	0.1338	0.1337	0.1348	0.1321	0.1342	0.1336
CG7508												
π	0.0427	0.0401	0.0381	0.0346	0.0238	0.0252	0.0262	0.0321	0.0272	0.0313	0.0439	0.0481
θ	0.0408	0.0458	0.0368	0.0324	0.0221	0.0276	0.0312	0.0331	0.0281	0.0359	0.0438	0.0496
Taj. D	0.1674	-0.7948	0.1745	0.1140	0.0870	-0.6944	-0.6867	-0.1718	-0.0589	-0.5211	-0.2460	-0.3065
K	0.1102	0.1105	0.1099	0.1038	0.0981	0.1064	0.1094	0.1092	0.1046	0.1078	0.1075	0.1128
CG13189												
π	0.0372	0.0499	0.0473	0.0482	0.0435	0.0230	0.0339	0.0379	0.0414	0.0441	0.0578	0.0526
θ	0.0368	0.0455	0.0418	0.0413	0.0351	0.0267	0.0281	0.0405	0.0295	0.0398	0.0494	0.0515
Taj. D	-0.1459	0.0835	0.6048	0.6048	0.7870	-0.9819	0.9358	-0.5509	1.8464	0.1805	0.9437	-0.1010
K	0.1961	0.1964	0.2004	0.1996	0.1974	0.1876	0.1932	0.2013	0.1982	0.1944	0.1922	0.1955

Values for divergence K, π and θ are for silent sites only. Tajima's D is given for all sites.

Table 3.5: Summary of polymorphism at presumably neutral loci

lected genes (12.23%). *CG7508* and *CG13189* show on average higher values for nucleotide diversity (π as well as θ) than the three genes under selecion.

3.3.2 Neutrality tests

Under the neutral model of molecular evolution we would expect an isolation-by-distance effect (Wright, 1943). Previous studies found isolation-by-distance at *Om(1D)*, where there is no evidence that the observed patterns of variation deviate from neutrality (Stephan *et al.*, 1998). Further, isolation-by-distance could be confirmed for microsatellite repeat-length variation at 23 loci (Schug *et al.*, 2007), for neutral intronic loci (Das *et al.*, 2004) and for mitochondrial DNA (Schug *et al.* (2008), see CHAPTER 1). We determined geographic distance between all our populations and measured genetic differentiation. There is no significant isolation-by-distance effect for most genes. Only one gene, *CG7508*, shows a statistically significant association between the geographic distance between populations and F_{ST} (Pearson's product-moment correlation, P-value $= 0.04$, $R = 0.25$). Table 3.6 provides the geographical distances between all populations, F_{ST} values are given in Appendix G.

The average values of divergence between *D. ananassae* and *D. atripex*/*D. phaeopleura* at all loci are given in Table 3.7. Divergence between *D. ananassae* and the outgroup is

Population	BBS	BKK	BOG	CH	KATH	KK	KL	KMJ	MAN	PURI	TB	TI
BBS												
BKK	1709											
BOG	3715	2319										
CH	1006	2202	3632									
KATH	840	2204	4412	1727								
KK	3626	1916	1682	4013	4047							
KL	2557	1191	1175	2609	3237	1622						
KMJ	4320	3162	4246	5174	4189	2580	3787					
MAN	1084	1014	3326	1964	1265	2784	2172	3242				
PURI	50	1691	3677	986	889	3607	2522	4335	1095			
TB	7050	5341	4388	7335	7453	3379	4739	4375	6189	7030		
TI	7713	6013	3924	7922	8162	4061	5312	5151	6898	7690	782	

Table 3.6: Geographical distances between populations

Gene	Outgroup	K all	K silent	K syn	K nonsyn
CG14717	Dph	0.0995	0.2137	0.2167	0.0625
CG6980	Dat	0.0802	0.1961	0.1883	0.0295
CG10750	Dat	0.0799	0.1942	0.1880	0.0219
CG10853	Dat	0.0354	0.0534	0.0944	0.0144
CG6981	Dat	0.0741	0.1321	0.1878	0.0027
CG7508	Dat	0.0418	0.1075	0.1071	0.0067
CG13189	Dat	0.0606	0.1960	0.1983	0.0003

Table 3.7: Divergence between *D. ananassae* and *D. atripex*/*D. phaeopleura*

highest for *CG14717*. There is no remarkable variation in divergence for the three genes that experienced positive selection in the Bangkok population. However, divergence at presumably neutral genes ranges from 3.54% (*CG10853*) to 7.41% (*CG6981*). Divergence at nonsynonymous sites is highest for *CG10853* (1.44%).

We would expect levels of polymorphism and divergence to be correlated under a constant-rate, neutral model of molecular evolution. We used a multilocus HKA test (Hudson *et al.*, 1987) to test this hypothesis. The test was applied to all seven loci taken together and revealed no significant departure from the equilibrium model in any of the populations (Table 3.8).

	Multilocus HKA test		Multilocus Tajima's D test						Multilocus Fu and Li's D test					
			Mean D			Variance D			Mean D			Variance D		
Pop	X^2	P	Obs.	Sim.	%higher	Obs.	Sim.	% higher	Obs.	Sim.	%higher	Obs.	Sim.	%higher
BBS	11.288	0.080	0.023	-0.069	39.44	0.384	0.815	86.85	0.367	-0.105	10.51	0.612	0.955	75.68
BKK	7.385	0.287	-0.160	-0.079	59.59	0.244	0.808	94.92	-0.117	-0.109	51.40	0.198	0.959	98.22*
BOG	3.895	0.691	-0.069	-0.059	50.58	0.297	0.780	94.01	-0.141	-0.114	52.47	0.527	0.929	80.49
MAN	3.023	0.806	0.041	-0.069	37.43	0.204	0.805	96.97*	0.423	-0.112	7.27	0.080	0.942	99.86*
CH	3.923	0.687	-0.037	-0.067	46.10	0.463	0.812	79.54	0.111	-0.103	28.37	1.097	0.959	34.32
KATH	12.477	0.052	-0.522	-0.047	91.36	0.655	0.836	66.49	-0.042	-0.093	44.92	1.251	0.957	24.62
KK	1.990	0.921	0.023	-0.069	39.28	0.363	0.808	88.31	0.210	-0.109	19.30	0.337	0.942	93.85
KL	5.716	0.456	-0.276	-0.070	72.05	0.212	0.813	96.72*	0.160	-0.106	24.31	0.485	0.956	84.59
KMJ	3.994	0.678	0.272	-0.062	16.24	0.615	0.820	64.91	0.388	-0.099	9.20	0.659	0.952	71.45
PURI	4.422	0.620	-0.321	-0.082	75.85	0.358	0.780	86.85	-0.110	-0.109	51.56	0.680	0.959	68.05
TB	2.028	0.917	-0.271	-0.055	73.54	0.751	0.791	49.99	-0.033	-0.120	40.15	1.217	0.926	24.70
TI	3.879	0.693	-0.402	-0.072	83.49	0.140	0.790	98.83*	-0.230	-0.123	61.93	0.152	0.937	99.35*

Pop. = population
X^2 = HKA test statistic
P = Probability
Obs. = Observed
Sim. = Simulated
%higher = percentage of simulated values higher than the observed ones
* = statistically significant ($P < 0.05$)

Table 3.8: Multilocus neutrality tests

In addition, we performed the HKA test for all pairwise comparisons between genes (seven genes $\rightarrow$ 21 comparisons), for each of the 12 populations separately. We calculated the probability of observing at least i significant tests at one specific gene given that n paired tests were performed and k were significant between the l loci using the equation given in the Materials and Methods section for each population. In total, only 16 pairwise comparisons out 252 possible (12 populations, 21 comparisons each) were significant at a significance level of $P \leq 0.05$. In all comparisons, one of the selected genes (*CG14717, CG6980, CG10750*) were involved and in 12 out of these 16 comparisons, the gene *CG14717* was involved. For this gene, the number of comparisons deviating from the neutral expectation was significantly higher for two northernmost populations (BBS and KATH). This could be consistent with adaptation to a new environment. If a gene is associated with adaptation to a new environment, gene flow is expected to increase between subpopulations sharing common environmental conditions. For the two northernmost populations from Kathmandu and Bhubaneswar, genetic differentiation is very low ($F_{ST} = 0.07$, see Appendix G) for this gene, which would be consistent with this hypothesis. The remaining four significant pairwise comparisons involve *CG6980* or *CG10750* against *CG7508*. Five populations do not show any significant comparisons (BOG, KK, MAN, TB, and TI). The results are summarized in Table 3.9.

3.3.3 McDonald-Kreitman tests

We performed McDonald-Kreitman (MK) tests for each gene and population separately. Additionally, MK tests were applied to each gene using the pooled sample of all strains over all populations. For *CG14717*, all populations except Bhubaneswar, India, and Kathmandu, Nepal, and the pooled sample gave significant results. Similarly, for *CG10750* almost all populations gave a significant result for positive selection. However, the populations that do not deviate from neutral expectation differ. For *CG10750*, populations from Kota Kinabalu, Kuala Lumpur, and Kumejima, Japan, do not give a significant MK test. In contrast, only the Bangkok population gives a significant result for *CG6980*, which has already been shown in Grath *et al.* (2009) (Table 3.10).

We find no evidence for selection between species using any population for *CG10853*, *CG6981*, *CG7508*, or *CG13189*. However, we find evidence for weak purifying selection acting on *CG13189* in the pooled sample of all strains (P-value = 0.03, Table 3.10).

Population	*CG14717*	*CG6980*	*CG10750*	Total	*P*
BBS	6	0	0	6	0
BKK	0	1	0	1	0.29
BOG	0	0	0	0	1
CH	1	0	0	1	0.29
KATH	4	1	0	5	0.01 (*CG14717*) 0.85 (*CG6980*)
KK	0	0	0	0	1
KL	0	0	1	1	0.29
KMJ	0	0	1	1	0.29
MAN	0	0	0	0	1
PUR	1	0	0	1	0.29
TB	0	0	0	0	1
TI	0	0	0	0	1

The HKA test was performed for all pairwise comparisons between genes, for each of the 12 populations. For each population, the probability of observing at least i significant tests at one specific gene given that n (21) paired tests were performed and k were significant between the l (7) genes was calculated using the equation given in the Materials and Methods section. In all significant comparisons, one of the selected genes (*CG14717, CG6980, CG10750*) were involved, *e.g.* for Kathmandu, four significant comparisons affect *CG14717* and one comparison affects *CG6980*.

Table 3.9: Pairwise HKA tests between all seven genes

Gene	Population	P_S[a]	P_N[b]	D_S[c]	D_N[d]	P-value[e]	NI[f]	α[g]
CG14717	BKK	29	3	43	40	0.00003	0.111	0.889
	BOG	22	3	43	40	0.00056	0.147	0.853
	MAN	17	2	42	40	0.00106	0.124	0.876
	CH	14	2	43	40	0.00481	0.154	0.846
	KK	26	1	42	40	0	0.04	0.96
	KL	25	2	41	40	0.00002	0.082	0.918
	KMJ	21	5	44	38	0.01071	0.276	0.724
	PURI	25	3	42	40	0.00014	0.126	0.874
	TB	22	3	37	40	0.00019	0.126	0.874
	TI	19	6	41	40	0.02181	0.324	0.676
	Pooled	63	18	37	38	0.00019	0.278	0.722
CG6980	BKK	10	1	24	16	0.03633	0.15	0.85
CG10750	BBS	27	2	34	15	0.00876	0.168	0.832
	BKK	26	2	33	15	0.00933	0.169	0.831
	BOG	13	0	36	16	0.02739	0	1
	MAN	27	2	34	15	0.00876	0.168	0.832
	CH	24	2	35	15	0.01791	0.194	0.806
	KATH	13	0	36	16	0.02723	0	1
	PURI	24	2	34	15	0.01588	0.189	0.811
	TB	22	1	35	16	0.00445	0.099	0.901
	TI	33	3	36	15	0.01712	0.232	0.768
	Pooled	51	8	29	14	0.02175	0.325	0.675
CG13189	Pooled	58	8	38	0	0.02572	-	-

[a]The total number of synonymous fixed differences.
[b]The total number of nonsynonymous fixed differences.
[c]The total number of synonymous polymorphisms.
[d]The total number of nonsynonymous polymorphisms.
[e]P-value was determined by G-test when applicable, otherwise by Fisher's exact test.
[f] Indicates the extent to which the levels of amino acid polymorphism depart from the expected in the neutral model (Rand and Kann, 1996).
[g] Indicates the proportion of amino acid substitutions driven by positive selection (Fay *et al.*, 2001).

Table 3.10: Significant McDonald-Kreitman tests

3.4 Conclusions and Outlook

To date, we have performed the largest population genetic study of sequence variation of autosomal genes in *D. ananassae*. We used populations sampled from almost the whole worldwide species range. We investigated DNA sequence variation of seven protein-coding genes in twelve populations. Using polymorphism data from the Bangkok, Thailand population, three of these genes showed evidence for adaptive evolution by the McDonald-Kreitman test (Grath *et al.* (2009), see CHAPTER 2). In the present study we performed McDonald-Kreitman tests on all of these genes in all populations. Whereas two genes (*CG14717* and *CG10750*) show evidence for positive selection between species using polymorphism data from almost all populations, one gene (*CG6980*) does not give a significant test for positive selection in any of the populations other than the one from Bangkok. The remaining four genes do not have patterns of variation deviating from neutrality in any of the twelve populations. However, levels of nucleotide diversity and divergence to the sister species *D. atripex* or *D. phaeopleura* vary considerably. Further, for only one gene (*CG7508*) could we confirm an isolation-by-distance effect that is expected under neutral evolution. Multilocus approaches involving all seven genes do not show strong evidence for deviation from neutrality. At the population level, the number of pairwise HKA comparisons deviating from the neutral expectation was significantly higher for only the two northernmost populations (BBS and KATH) and the gene *CG14717*. It is possible that we lack power to detect deviations from neutrality with the composition of our gene set. Previous studies (Baines *et al.*, 2004; Chen *et al.*, 2000) used one locus and compared this single locus to ten neutral introns to test for deviation from neutral expectations, whereas we have three genes under selection and only four presumably neutral genes. Moreover, variation in linkage or recombination rates between our genes might influence our results. We considered only autosomal genes. Some genes are located on the same chromosome arm (3R: *CG14717*, *CG6980*, *CG7508* and 3L: *CG10853*, *CG6981*), whereas the remaining two genes are not chromosomally linked to any of the other genes (*CG10750* is located on chromosome arm 2L and *CG13189* is located on chromosome arm 2R). However, all genes are several MB apart from each other, so that we might assume linkage equilibrium and independence between all loci.

Future studies might integrate more neutral loci for multilocus comparisons. We currently have sequence data for several intronic loci (Das *et al.*, 2004) for the outgroups (*D. atripex* and *D. phaeopleura*) used in this study. Moreover, we have polymorphism data for all seven genes from one population of *D. pallidosa* from Nadi, Fiji, on all seven genes

which allows for more analyses. Currently, there is a lack of information on molecular function or biological processes where the genes might be involved in. For *CG14717*, the molecular function is described as hydrolase activity. *CG6980* is involved in binding. For *CG10750*, no functions or biological processes are described. The molecular function of *CG7508* (*ato*) is described as DNA binding and transcription factor activity. The gene is involved in various biological processes, *e.g.* anatomical structure development, organ morphogenesis, sensory organ development, organ development, regulation of biological process, cell communication, cell recognition, reproductive process in a multicellular organism, oenocyte development, and sensory perception of mechanical stimulus. In *D. melanogaster*, 22 alleles are reported. For *CG13189*, the molecular function is described as metal ion transmembrane transporter activity. The gene is involved in transmembrane transport and metal ion transport. For *CG10853*, no functions or biological processes are described. All described functions refer to what is known from *D. melanogaster*. In the future, the question of molecular function and biological processes where our genes under selection might be involved in could be addressed further, using *e.g.* RNAi mechanisms or homologue recombination. Moreover, male-biased gene expression could be investigated further in *D. ananassae* for our genes of interest. Currently, only one gene (*CG6980*) shows conserved male-biased gene expression between *D. melanogaster* and *D. ananassase*, whereas all other genes show unbiased gene expression in *D. ananassae* (Grath *et al.*, 2009). However, it might be possible that some genes have tissue-dependent male-biased gene expression which could be investigated using quantitative real-time PCR experiments on different fly tissues, for example.

Chapter 4

The relationship between sex-biased gene expression and rate of molecular evolution in *Drosophila*

4.1 Introduction

Several factors affecting the evolution of sex-biased genes in *Drosophila* have been identified and studied previously. Genes with sex-biased gene expression often show rapid molecular evolution between species. Previous population genetic and comparative genomic studies of *D. melanogaster* and *D. simulans* revealed that male-biased genes have especially high rates of adaptive evolution (Baines *et al.*, 2008; Ellegren and Parsch, 2007; Pröschel *et al.*, 2006). However, for species other than those of the well-studied *melanogaster* subgroup, the situation is different and less clear. Studies on *D. ananassae* and *D. pseudoobscura* (Grath *et al.*, 2009; Metta *et al.*, 2006) could not confirm the pattern of higher rates of adaptive evolution for male-biased genes in these species.

Many factors might affect the variation in rates of evolution between different proteins in *Drosophila* by either influencing the rate of evolution itself or imposing evolutionary constraints. Larracuente *et al.* (2008) identified gene expression, intron and protein lengths, intron number, protein-protein interactions, recombination, and translational selection as possible affectors. In one recent study, protein secondary structure was found to influence positive selection in *Drosophila* (Ridout *et al.*, 2010). These authors found that amino acids forming disordered regions, *e.g.* random coils, are more likely to be under selection than amino acids situated in helices and β-structures. Here, we concentrate on the degree of sex-

bias and the ancestral expression state, both of which might contribute to the observed similarities and differences in sex-biased gene evolution among Drosophilid species. We compared male/female (M/F) expression ratios with the rate of evolution measured by the ratio of the nonsynonymous substitution rate to the synonymous substitution rate (d_N/d_S).

We used previously published expression data and d_N/d_S estimates from comparative genomic data for for several *Drosophila* species to compare the rate of functional evolution with the degree of sex-bias (male/female expression ratio). In general, we found that the rate of molecular evolution is correlated with the degree of sex-bias. There is a general pattern of faster evolution for highly male-biased genes compared to male-biased genes with a lower degree of sex-bias, female-biased or unbiased genes. For orthologous genes between *D. melanogaster* and *D. ananassae*, genes that show conserved male-biased expression between species evolve faster than those that have undergone a change in sex-biased expression.

4.2 Materials and Methods

4.2.1 Datasets

For *D. melanogaster* and *D. ananassae*, we used expression data from Zhang *et al.* (2007) and data on the rate of protein evolution (d_N/d_S) from Larracuente *et al.* (2008). In addition, we extracted expression and d_N/d_S data from Sebida (`http://www.sebida.de`) for *D. melanogaster*. Further, we used data from Jiang and Machado (2009) for *D. pseudoobscura*. M/F expression ratios were log_2-transformed. In total, we looked at four different datasets.

For *D. melanogaster*, we investigated a set of genes where we extracted d_N/d_S values from Larracuente *et al.* (2008) and expression data from Zhang *et al.* (2007) (dataset 1) and a set of genes where d_N/d_S and expression values were extracted from Sebida (dataset 2). These data come from a meta-analysis of sex-biased gene expression over several studies (Ayroles *et al.*, 2009; Gibson *et al.*, 2004; McIntyre *et al.*, 2006; Parisi *et al.*, 2003, 2004; Ranz *et al.*, 2003). For *D. ananassae*, the data (dataset 3) included M/F expression values extracted from Zhang *et al.* (2007) and d_N/d_S values from Larracuente *et al.* (2008). For *D. pseudoobscura*, we considered a set of genes with M/F expression levels and d_N/d_S values extracted from Jiang and Machado (2009) (dataset 4).

Dataset	MBG	FBG	UBG	MBG	FBG	UBG	MBG	FBG	UBG
				D. melanogaster					
		all			autosomal			X-linked	
1	947	341	6,659	865	284	5,648	82	57	1,011
2	3,384	4,990	4,061	2,985	4,054	3,400	395	929	659
				D. ananassae					
		all			autosomal			X-linked	
3	599	398	6,580	536	331	5,591	63	67	988
				D. pseudoobscura					
		all			autosomal			X-linked	
4	3,349	4,906	2,124	2,339	2,949	1,334	943	1,806	723

Table 4.1: Numbers of genes used in each analysis

4.2.2 Statistical analysis

We investigated the correlation of M/F expression ratios and d_N/d_S values for each species and different sets of genes. We looked at all available genes in each species and at autosomal and X-linked genes separately. The numbers of genes used for each analysis are given in Table 4.1.

For sex-biased genes, we looked in greater detail at the degree of sex-bias. For each species, we used unbiased genes and split the sets of male- and female-biased genes into three equally-sized groups each according to their degree of sex-bias (high, medium, and low) and compared these seven groups using Mann-Whitney U tests. We also tested for differences by splitting the sets of male- and female-biased genes into only two equally-sized groups (high and low degree of sex-bias). Including unbiased genes, this leads to a 7-group and a 5-group analysis, respectively.

Moreover, we investigated the ancestral state of expression for orthologous genes between *D. melanogaster* and *D. ananassae*. Two different studies were used to determine sex-biased gene expression in *D. pseudoobscura*. First, we used expression data from Zhang *et al.* (2007), second, we used expression data from Jiang and Machado (2009). *D. pseudoobscura* serves as an outgroup for both *D. melanogaster* and *D. ananassae* (Figure 4.1, see also Chapter 2) and was used to assess the ancestral sex-biased expression state for each gene.

Subsequently, we analyzed the conservation of degree of sex-bias between these orthologs. We divided sex-biased genes into different groups (high/medium/low or high/low, respectively) within each species and compared the overlap of genes between these groups. All statistical analyses were performed using R 2.9.0 (R Development Core Team, 2009).

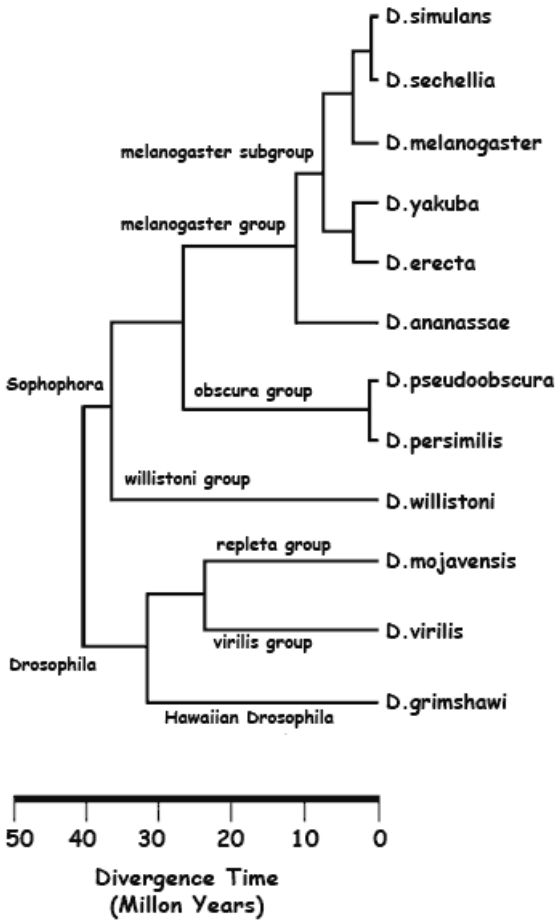

Figure 4.1: Drosopholid phylogeny (adapted from `www.flybase.org`)

4.3 Results and Discussion

4.3.1 Correlation between rate of evolution and degree of sex-bias

For all three species, there is a positive, significant correlation between the rate of protein evolution measured by d_N/d_S and male/female expression ratios (Figure 4.2). For each species, this holds for the complete dataset of all genes as well as the autosomal genes considered separately and for *D. melanogaster* and *D. ananassae* there are significant positive correlations for the X-linked genes. For male-biased genes, there are significant positive correlations between d_N/d_S and the degree of sex-bias. The higher a male-biased gene is expressed, the higher is the rate of protein evolution. This holds for the complete sets of all genes and the autosomal genes considered separately and for each species. In *D. melanogaster* (both datasets), the set of X-linked male-biased genes show significant positive correlations for the rate of molecular evolution with the degree of sex-bias. For *D. pseudoobscura*, there is no correlation for X-linked male-biased genes. All significance levels are given in Table 4.2.

4.3.2 Analysis for sex-biased and unbiased genes

For each species, we split the data into seven groups and ranked them from highest M/F expression ratio to lowest (Figure 4.3). There is a general pattern of faster evolution for highly expressed male-biased genes. For all datasets, this can be observed irrespectively of dividing sex-biased genes into three groups (high, medium, and low degree of sex-bias for female- and male-biased genes and unbiased genes, in total leading to a 7-group analysis of all genes) or considering five groups in total (high-medium, low-medium, unbiased, low-female, and high-female). (Figures 4.3, 4.4). Highly expressed male-biased genes show significantly higher d_N/d_S levels than lowly expressed male-biased genes in all species and taking all genes together or within only autosomal or X-linked genes. We give significance levels for all relevant comparisons in Table 4.3.

For female-biased genes, the pattern is less clear at first glance. However, there are quite a few significant or almost significant comparisons between female-biased genes with low degree of sex-bias and those with high degree of sex-bias for all three species. Interestingly, the genes with low degree of female-biased expression mostly reveal higher rates of molecular evolution. Note, that in the case of female-biased genes lower degree of sex-bias corresponds to higher values of M/F expression. Consequently, the pattern for

	All		Autosomal		X-linked	
Dataset	P-value	R	P-value	R	P-value	R
1						
all	<0.001	0.25	<0.001	0.25	<0.001	0.25
MBG	<0.001	0.25	<0.001	0.26	0.02	0.26
FBG	<0.001	-0.21	<0.001	-0.23	0.34	-0.13
UBG	<0.001	0.11	<0.001	0.10	<0.001	0.19
2						
all	<0.001	0.14	<0.001	0.14	<0.001	0.20
MBG	<0.001	0.17	<0.001	0.19	<0.001	0.21
FBG	0.21	-0.02	0.12	-0.02	0.85	0.006
UBG	0.54	0.01	0.54	0.01	0.70	0.01
3						
all	<0.001	0.14	<0.001	0.15	0.008	0.08
MBG	<0.001	0.14	<0.001	0.16	0.64	0.06
FBG	0.27	0.06	0.42	0.04	0.33	0.12
UBG	<0.001	-0.17	<0.001	-0.15	<0.001	-0.24
4						
all	<0.001	0.03	<0.001	0.08	0.61	0.009
MBG	0.08	0.03	0.001	0.07	0.59	-0.02
FBG	0.94	0.001	0.19	0.02	0.98	-0.001
UBG	0.03	0.05	0.02	0.07	0.79	-0.01

Table 4.2: Correlation between rate of evolution and degree of sex-bias

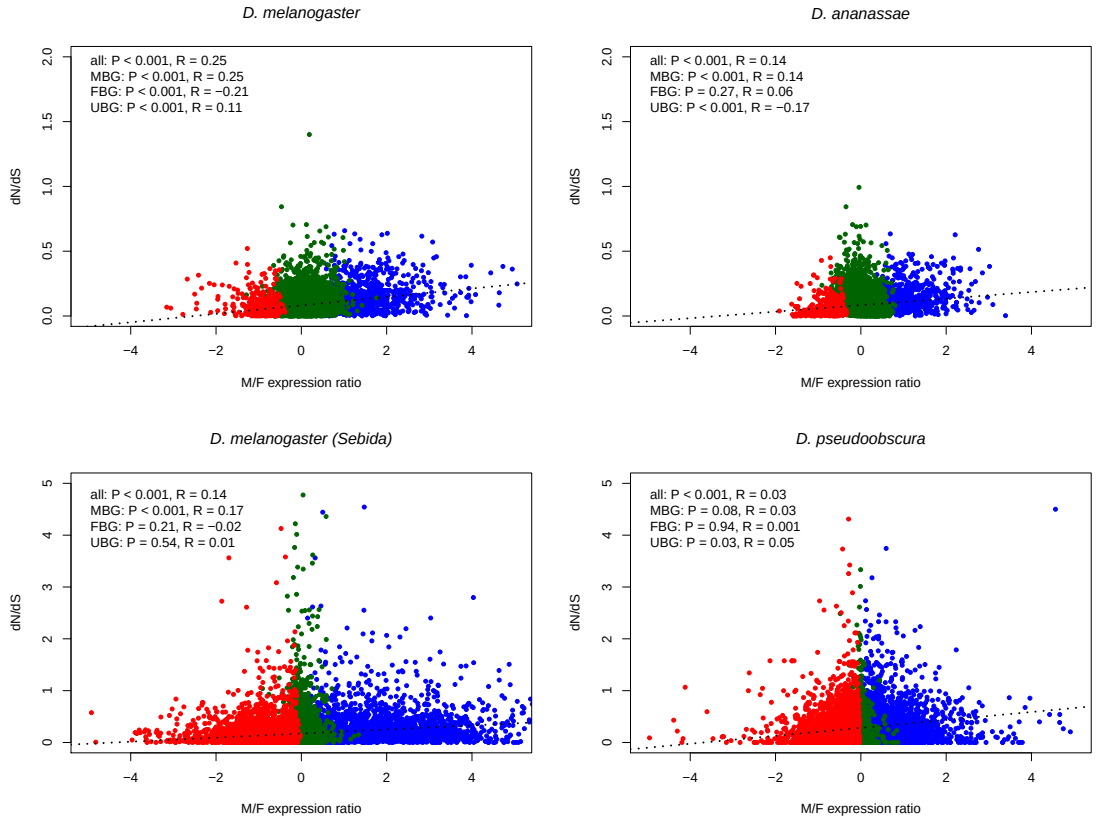

Figure 4.2: Correlation of M/F expression ratio and d_N/d_S (all genes)

Male-biased genes (MBG) are shown in blue, female-biased genes (FBG) in red and unbiased genes (UBG) in green. P-values are determined using Pearson's product-moment correlation. The correlation coefficients for each group are indicated as R. The regression line corresponds to the best fit linear correlation as determined for all genes.

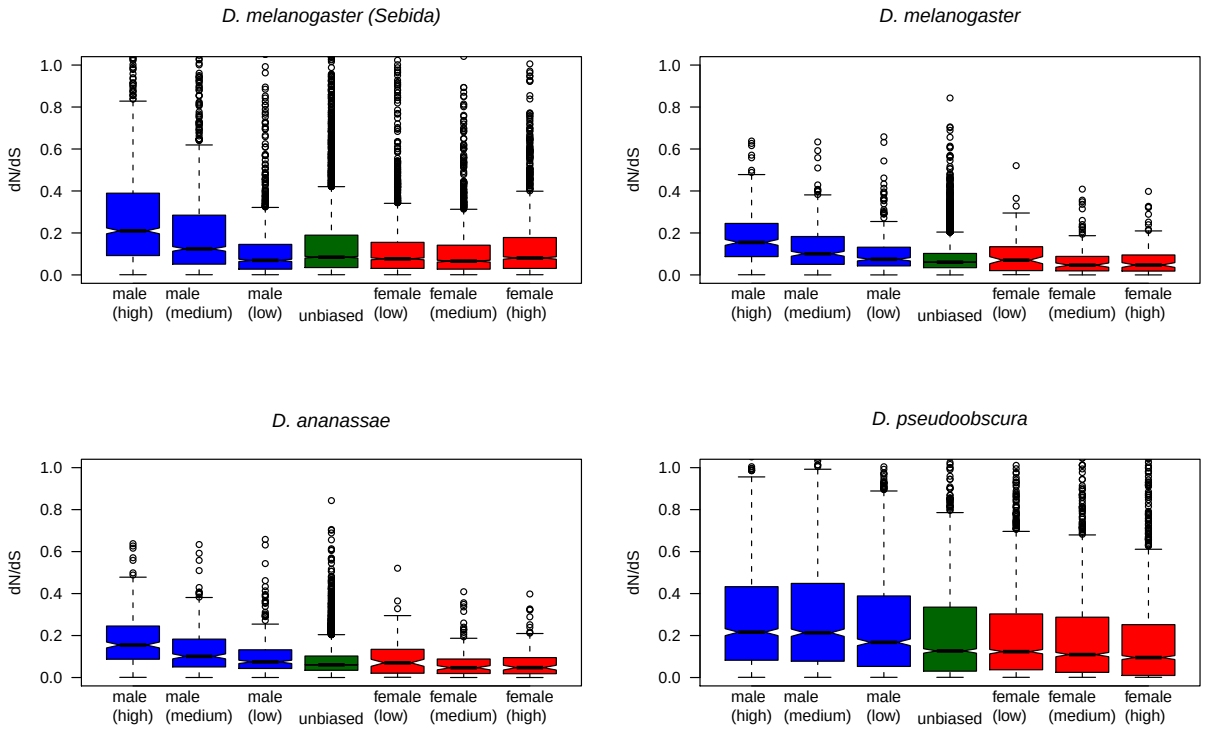

Figure 4.3: Group-based comparison of male-, female-, and unbiased genes (all genes)

Sex-biased genes are divided into three equally-sized groups for each sex according to degree of sex-bias (high, medium, and low). Male-biased genes are displayed in blue, female-biased genes in red, and unbiased genes in green.

female-biased genes is in accordance with genes displaying higher M/F expression ratios having higher rates of molecular evolution on average.

In general, for sex-biased genes, X-linked genes show higher d_N/d_S levels than autosomal genes (Figure 4.5). The difference is significant for male-biased genes in all three species (for *D. melanogaster* only in dataset 2). Female-biased X-linked genes reveal higher d_N/d_S compared to autosomal genes in *D. melanogaster* and *D. pseudoobscura*. Note, that for *D. pseudoobscura* even X-linked unbiased genes have an increased level for rate of evolution. In this species, X-linked genes generally show significantly higher d_N/d_S levels than autosomal genes. Significance levels for all comparisons are given in Table 4.4.

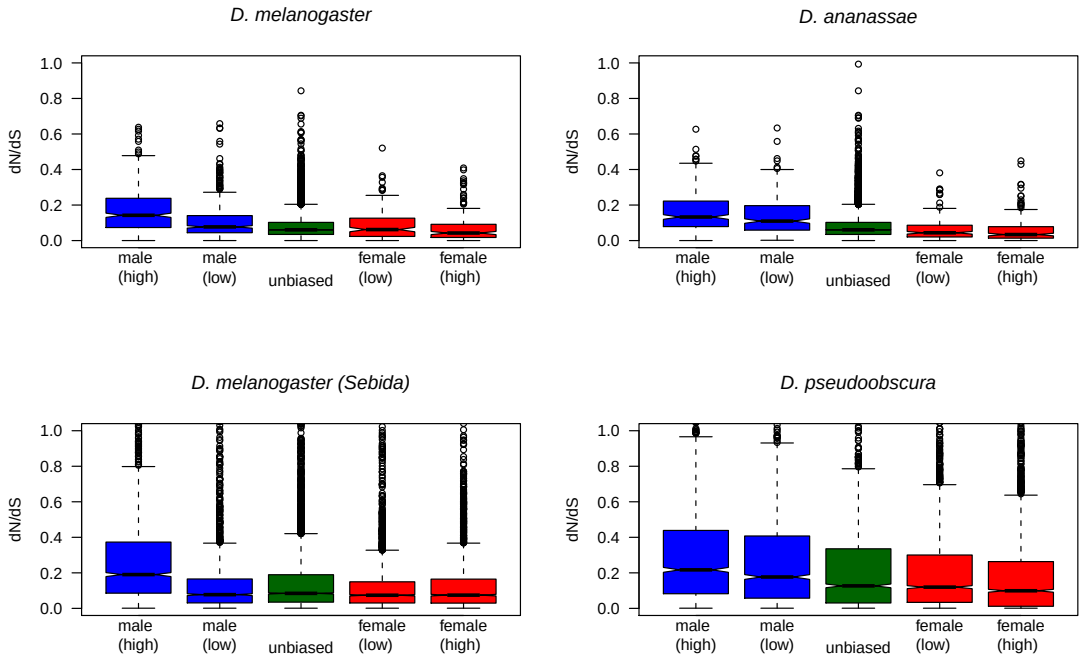

Figure 4.4: Group-based comparison of male-, female-, and unbiased genes with sex-biased genes split in two classes (all genes)

Sex-biased genes are divided into two equally-sized groups for each sex according to degree of sex-bias (high and low). Male-biased genes are displayed in blue, female-biased genes in red, and unbiased genes in green.

Comparison	All genes	Autosomal	X-linked
D. melanogaster, dataset 1			
7-group comparison			
high-male vs. low-male	< 0.001	< 0.001	< 0.001
high-female vs. low-female	0.122	< 0.001	0.006
5-group comparison			
high-male vs. low-male	< 0.001	< 0.001	0.01
high-female vs. low-female	0.038	< 0.001	0.011
D. melanogaster, dataset 2			
7-group comparison			
high-male vs. low-male	< 0.001	< 0.001	< 0.001
high-female vs. low-female	0.297	0.561	0.511
5-group comparison			
high-male vs. low-male	< 0.001	< 0.001	< 0.001
high-female vs. low-female	0.355	0.277	0.87
D. ananassae, dataset 3			
7-group comparison			
high-male vs. low-male	0.004	0.001	0.23
high-female vs. low-female	0.104	0.199	0.813
5-group comparison			
high-male vs. low-male	0.010	0.024	0.561
high-female vs. low-female	0.071	0.08	0.562
D. pseudoobscura, dataset 4			
7-group comparison			
high-male vs. low-male	< 0.001	< 0.001	0.002
high-female vs. low-female	< 0.001	< 0.001	0.005
5-group comparison			
high-male vs. low-male	< 0.001	0.029	0.006
high-female vs. low-female	< 0.001	0.013	0.001

Table 4.3: Significance levels of Mann-Whitney U tests for 5-group and 7-group comparisons

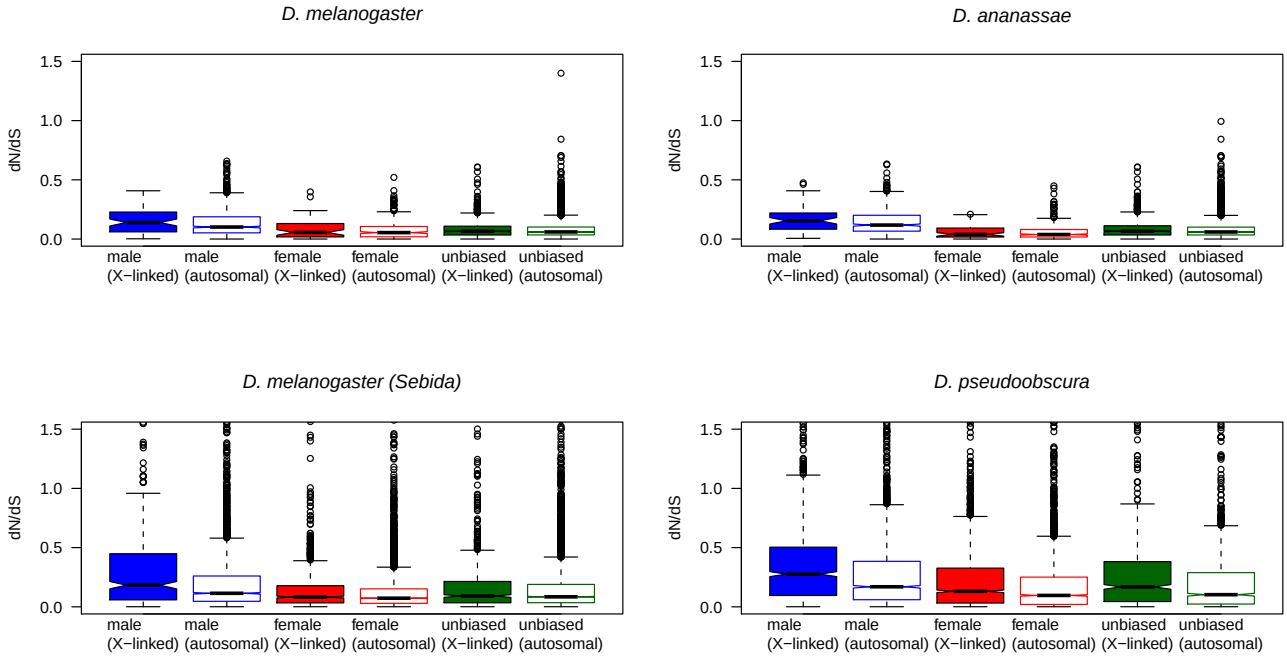

Figure 4.5: Comparison of X-linked and autosomal genes

X-linked genes are shown in full color (blue for male-biased genes, red for female-biased genes, and green for unbiased genes), autosomal genes are shown with colored borders.

Bias	autosomal vs. X-linked
D. melanogaster, dataset 1	
M	0.116
F	0.913
U	0.192
D. melanogaster, dataset 2	
M	< 0.001
F	0.016
U	0.236
D. ananassae, dataset 3	
M	0.042
F	0.604
U	0.036
D. pseudoobscura, dataset 4	
M	< 0.001
F	< 0.001
U	< 0.001

"M" indicates male-biased, "F" indicates female-biased, and "U" indicates unbiased expression.

Table 4.4: Significance levels of Mann-Whitney U tests for X-linked compared to autosomal genes

Expression bias	No. of genes
MM	425
FF	112
UU	5,598
MU	155
UM	372
FU	241
UF	190
MF	0
FM	32
Total	7,125

"M"= male-biased, "F"= female-biased, "U"= unbiased; the first letter indicates expression in *D. ananassae*, the second letter indicates expression in *D. melanogaster*.

Table 4.5: Expression between *D. melanogaster* and *D. ananassae*

4.3.3 Orthologs between species

For 7,125 orthologs, we could consider d_N/d_S and expression data in both *D. melanogaster* and *D. ananassae* using the dataset of Larracuente *et al.* (2008) and Zhang *et al.* (2007). We divided the set of genes into nine categories according to their expression conservation (Table 4.5). The first letter of each category indicates expression in *D. ananassae* , the second letter indicates expression in *D. melanogaster* ("M"= male-biased, "F"= female-biased, and "U"= unbiased, respectively).

Genes with conserved male-biased expression (MM) show the highest d_N/d_S values (Figure 4.6). Interestingly, the difference between MM and MU is not significant (Mann-Whitney U test, P-value = 0.0758), whereas the groups MM/UM and MU/UM differ significantly (P-value = 5.79e-10 and P-value = 0.0008125, respectively).

In addition, we determined orthologous genes between the two species using expression and d_N/d_S data from Zhang *et al.* (2007) and Larracuente *et al.* (2008), respectively, for *D. ananassae* and data taken from Sebida for *D. melanogaster*. Note, that for this comparison d_N/d_S values for *D. ananassae* are calculated over the *melanogaster* group (Larracuente *et al.*, 2008), whereas d_N/d_S for *D. melanogaster* are estimated between *D. melanogaster* and its sister species *D. simulans*. Using these data, we found 6,025 orthologs between the species. The lower number of genes can be explained by the fact that we used flybase identifiers (FBgn) for determining the overlap of the two species. However, the identification of genes in different datasets is not consistent and some genes are described

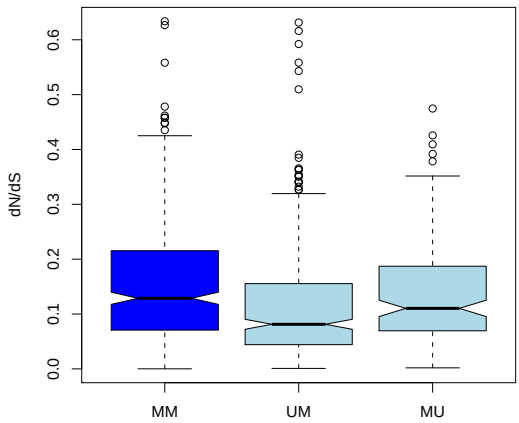

Figure 4.6: d_N/d_S between *D. melanogaster* and *D. ananassae*

"M"= male-biased, "F"= female-biased, "U"= unbiased; the first letter indicates expression in *D. ananassae*, the second letter indicates expression in *D. melanogaster*. Conserved male-biased expression are displayed in blue, genes with male-biased expression in only one species are displayed in lightblue.

via different identifiers in different datasets and did not show up in comparisons taking only FBgn. Taking into account all possible identifiers would increase the number of orthologous genes. Again, we divided the set of genes into nine categories according to their expression conservation (Table 4.6).

In general, the Sebida dataset contains a higher proportion of sex-biased genes compared to *D. melanogaster* genes taken from Zhang *et al.* (2007). Accounting for this, we would expect the number of conserved unbiased genes (UU) to decrease, whereas the numbers of genes which are unbiased in *D. ananassae* but sex-biased in *D. melanogaster* (UM and UF, respectively) should increase. This is exactly what we can observe.

Again, we compared male-biased genes with conserved expression between species (MM) to genes with male-biased expression either in *D. melanogaster* or *D. ananassae* (UM or MU, respectively). For these genes, the question is which d_N/d_S ratios should be taken for group-based comparisons. We considered three different methods for determining the rate of evolution for these genes. First, we compared the groups using d_N/d_S calculated over the complete *melanogaster* group (Larracuente *et al.*, 2008). This comparison only differs from the dataset before (Table 4.5) in the number and identity of genes. Like before, genes with conserved male-biased expression show the highest d_N/d_S ratios. In this

Expression bias	No. of genes
MM	469
FF	239
UU	1,563
MU	28
UM	964
FU	51
UF	2,667
MF	13
FM	31
Total	6,025

"M"= male-biased, "F"= female-biased, "U"= unbiased; the first letter indicates expression in *D. ananassae*, the second letter indicates expression in *D. melanogaster*. We extracted values for *D. melanogaster* from Sebida.

Table 4.6: Expression between *D. melanogaster* and *D. ananassae*

comparison, the rate of evolution is significantly higher compared to both cases where only one species reveal male-biased expression (MM-UM, P-value < 0.001; MM-MU, P-value < 0.001). However, the comparison between groups of non-conserved male-biased genes is not significant (MU-UM, P-value $= 0.81$). Using d_N/d_S ratios as determined from the Sebida database (*i.e.* between *D. melanogaster* and *D. simulans*), the comparison between conserved male-biased genes (MM) and genes with male-biased gene expression private to *D. melanogaster* stays significant with conserved male-biased genes still revealing the highest rate of evolution on average. Finally, we used the averages between the two d_N/d_S ratios described above. The results are in accordance with the first comparison (MM-UM, P-value < 0.001; MM-MU, P-value $= 0.04$; MU-UM, P-value $= 0.22$). Generally, our results show that the group of conserved male-biased genes reveal higher rates of evolution. The results are slightly dependent on the origin and determination of d_N/d_S ratios and the gene set considered. However, the d_N/d_S ratios as determined from Sebida (between *D. melanogaster* and *D. simulans*) are highly positively correlated with the d_N/d_S ratios calculated over the *melanogaster* group including *D. ananassae* (Spearman's rank correlation, P-value < 0.001, rho $= 0.789$).

Differences between groups of genes might be further influenced by the ancestral state of gene expression. Genes showing male-biased expression in *D. melanogaster* only (UM), for example, might either lost male-biased expression in *D. ananassae* or recently gained male-biased expression in *D. melanogaster* compared to the expression state in a common

ancestor of both species. Not taking this into account, might mislead conclusions for ortholog comparisons of only two species. This problem could be encountered using an outgroup species for determining ancestral expression states.

4.3.4 Influence of ancestral state of sex-biased gene expression

We wanted to investigate if the rate of evolution for sex-biased genes is dependent on the ancestral state of expression. We used expression states of *D. pseudoobscura* to determine the ancestral states for orthologous genes in *D. melanogaster* and *D. ananassae*. *D. pseudoobscura* serves as outgroup for both species (Figure 4.1). Each category from Table 4.4 showing gene expression conservation between *D. ananassae* and *D. melanogaster* can be divided into three groups according to the ancestral state in *D. pseudoobscura*. The ancestral state might be male-, female-, or unbiased expression in each case.

First, we used expression data from Zhang *et al.* (2007) for *D. pseudoobscura* which resulted in a dataset of 6,558 genes which are orthologous between all three species. According to expression state in different species, we split the data into 27 groups (Table 4.7). Note, that for this dataset d_N/d_S ratios are determined over the *melanogaster* group (Larracuente *et al.*, 2008).

Second, we used expression data and d_N/d_S information from Jiang and Machado (2009) for *D. pseudoobscura*. For this set, we determined the overlap with the orthologs between *D. melanogaster* and *D. ananassae* extracted from Zhang *et al.* (2007) (7,125 genes). This resulted in a dataset of 5,573 genes (Table 4.8). Again, the lower number of genes results from using only flybase identifiers (FBgn) for determining the overlap of the two species as well as the composition of the *D. pseudoobscura* dataset where expression and d_N/d_S were given with different identifiers. The dataset containing *D. pseudoobscura* genes as used by (Jiang and Machado, 2009) contains a higher proportion of sex-biased genes compared to *D. pseudoobscura* genes taken from (Zhang *et al.*, 2007). Taking this into account, we would expect the number of conserved unbiased genes (UUU) to decrease, whereas the numbers of genes which are unbiased in *D. ananassae* and/or *D. melanogaster* but sex-biased in *D. pseudoobscura* should increase. Again, this is exactly what we can observe (Table 4.7 and 4.8).

In general, the first comparison shows that genes which show conserved male-biased gene expression over longer evolutionary timescale reveal significantly higher rates of molecular evolution. Figure 4.7 displays comparisons of groups where at least one of the species exhibits male-biased gene expression. These groups are compared to conserved unbiased

Expression bias	No. of genes
FFF	95
FFM	0
FFU	8
FMF	3
FMM	7
FMU	18
FUF	126
FUM	8
FUU	93
MFF	0
MFM	0
MFU	0
MMF	0
MMM	313
MMU	16
MUF	5
MUM	88
MUU	30
UFF	132
UFM	3
UFU	42
UMF	6
UMM	151
UMU	162
UUF	946
UUM	413
UUU	3893
Total	6,558

"M"= male-biased, "F"= female-biased, "U"= unbiased; the first letter indicates expression in *D. ananassae*, the second letter indicates expression in *D. melanogaster*, and the third letter indicates expression in *D. pseudoobscura*.

Table 4.7: Expression between *D. melanogaster*, *D. ananassae*, and *D. pseudoobscura*

Expression bias	No. of genes
FFF	103
FFM	0
FFU	4
FMF	2
FMM	18
FMU	5
FUF	143
FUM	40
FUU	45
MFF	0
MFM	0
MFU	0
MMF	4
MMM	318
MMU	16
MUF	23
MUM	84
MUU	16
UFF	179
UFM	0
UFU	4
UMF	35
UMM	207
UMU	66
UUF	2,621
UUM	724
UUU	916
Total	5,573

"M"= male-biased, "F"= female-biased, "U"= unbiased; the first letter indicates expression in *D. ananassae*, the second letter indicates expression in *D. melanogaster*, and the third letter indicates expression in *D. pseudoobscura*.

Table 4.8: Expression between *D. melanogaster*, *D. ananassae*, and *D. pseudoobscura*

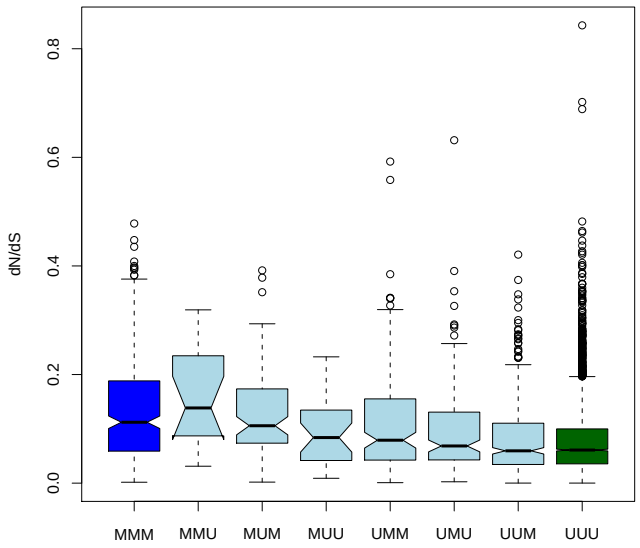

Figure 4.7: d_N/d_S between *D. melanogaster*, *D. ananassae*, and *D. pseudoobscura*

"M"= male-biased, "U"= unbiased; the first letter indicates expression in *D. ananassae*, the second letter indicates expression in *D. melanogaster*, and the third letter indicates expression in *D. pseudoobscura* Groups of conserved male-biased and unbiased genes are displayed in blue and green, respectively, groups containing male-biased genes in at least one species are displayed in lightblue.

genes (UUU). Genes showing reversal of sex-bias in one or more of the species comparisons are not considered because of the consistently low number of genes in each of these categories (Table 4.7).

Genes with conserved male-biased gene expression (MMM) reveal significantly higher rates of molecular evolution compared to genes which lost male-biased gene expression in *D. ananassae* (UMM, P-value $= 0.0012$) and genes which show male-biased gene expression in only one species (UMU, P-value < 0.001; MUU, P-value $= 0.056$; UMM, P-value < 0.001). Genes with conserved male-biased expression between *D. ananassae* and *D. melanogaster* (MMU) which correspond to expression conservation over a timescale of about 10 million years have significantly higher d_N/d_S ratios compared to genes with male-biased gene expression in only one of the two species (MUU, P-value $= 0.026$; UMU, P-value $= 0.004$) which is in accordance with the results described above. Moreover, rates of molecular evolution in this group are significantly higher compared to the group of genes where male-biased gene expression had been lost in *D. ananassae* (UMM, P-value $= 0.0012$). Note,

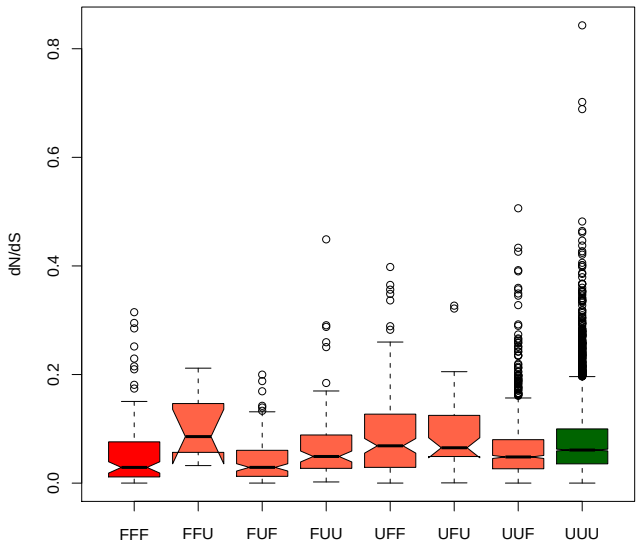

Figure 4.8: d_N/d_S between *D. melanogaster*, *D. ananassae*, and *D. pseudoobscura*

"F"= female-biased, "U"= unbiased; the first letter indicates expression in *D. ananassae*, the second letter indicates expression in *D. melanogaster*, and the third letter indicates expression in *D. pseudoobscura* Groups of conserved male-biased and unbiased genes are displayed in red and green, respectively, groups containing female-biased genes in at least one species are displayed in lightred.

that this group does not contain genes which show conserved expression over 30 million years of Drosophilid evolution between *D. melanogaster* and *D. pseudoobscura* (Figure 4.1). Despite revealing same expression for *D. melanogaster* and *D. pseudoobscura*, male-biased gene expression got lost at least once. Considering expression exclusively between these two species does not take into account possible gains or losts of sex-biased expression within this time period of evolution. Further, these genes show significantly lower d_N/d_S compared to genes with male-biased expression conserved between *D. ananassae* and *D. pseudoobscura* which corresponds to a timescale of about 20 million years (MUM, P-value = 0.022). Genes with female-biased expression states does not reveal similar pattern compared to male-biased genes. If anything, conserved female-biased genes display lower rates of molecular evolution, even compared to conserved unbiased genes (Figure 4.8).

For the second dataset, the question is which rates of molecular evolution should be used for group-based comparisons. It is possible to take d_N/d_S ratios calculated for the *melanogaster* group (Larracuente *et al.*, 2008), for *D. pseudoobscura* compared with its

sister species *D. persimilis* (Jiang and Machado, 2009), or d_N/d_S ratios averaged over the first two cases. Unfortunately, saturation in silent site divergence outside the *melanogaster* species group precludes the determination of d_N/d_S including *D. pseudoobscura* (Larracuente *et al.*, 2008; Metta *et al.*, 2006) which would be most appropriate for these comparisons. However, d_N/d_S ratios calculated over the *melanogaster* group are highly correlated with d_N/d_S ratios determined for the *obscura* group (Spearman's rank correlation, P-value < 0.001, rho $= 0.45$). Compared to the results described above, not all relevant comparisons stay significant. Consistently, genes with conserved male-biased gene expression reveal higher rates of molecular evolution compared to genes with male-biased gene expression in only one species. Again, female-biased genes does not display similar pattern in all comparisons.

4.3.5 Conservation of degree of sex-bias

We investigated the conservation of degree of sex-bias for ortholog genes between *D. melanogaster* and *D. ananassae*. Out of 7,125 genes, 425 genes showed conserved male-biased gene expression between these two species (Table 4.4). We first divided these genes into two groups of high and low degree of sex-bias for both species (Figure 4.9 A and B). About two third of all male-biased genes show conserved degree of sex-bias between *D. ananassae* and *D. melanogaster* (Figure 4.9 A and B). Secondly, we divided all male-biased genes into three groups with high, medium, and low degree of sex-bias, respectively. About one half of all genes reveal the same degree in both species for a 3-group analysis. (Figure 4.9 C and D).

We performed the same analyses for female-biased genes. Results are given in Figure 4.10. This set contains only 112 genes (FF, see Table 4.6). Dividing the data into two groups (Figure 4.10 A and B), reveals a 50% conservation of degree of sex-bias, whereas the degree is only 32% conserved dividing the data into three groups (Figure 4.10 C and D).

As for the consideration of orthologous genes between species in general (see above), we further used a dataset consisting of orthologs between *D. ananassae* (dataset 3, Table 4.1) and *D. melanogaster* genes extracted from Sebida (dataset 2, Table 4.1). This leads to 469 genes with conserved male-biased expression (MM, see Table 4.6) and 239 genes with conserved female-biased expression (FF, see Table 4.6). The conservation of degree of sex-bias is very similar between male- and female-biased genes and comparable to previous results using the first dataset (Table 4.5). For male-biased genes, 68% of genes reveal

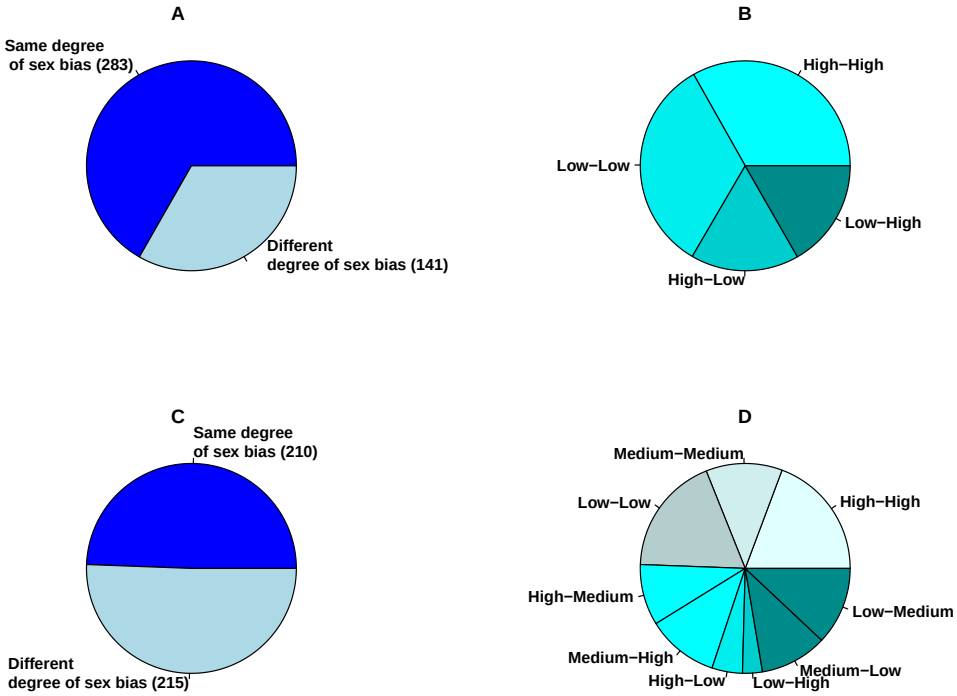

Figure 4.9: Conservation of degree of sex-bias between *D. melanogaster* and *D. ananassae* for male-biased genes with high and low degree of sex-bias

Panels A and B display results for dividing male-biased genes into two groups according to degree of sex-bias, high and low. Panels C and D display results for dividing male-biased genes into three groups; high, medium, and low degree of sex-bias. For panels A and C, the numbers of genes are given in parantheses.

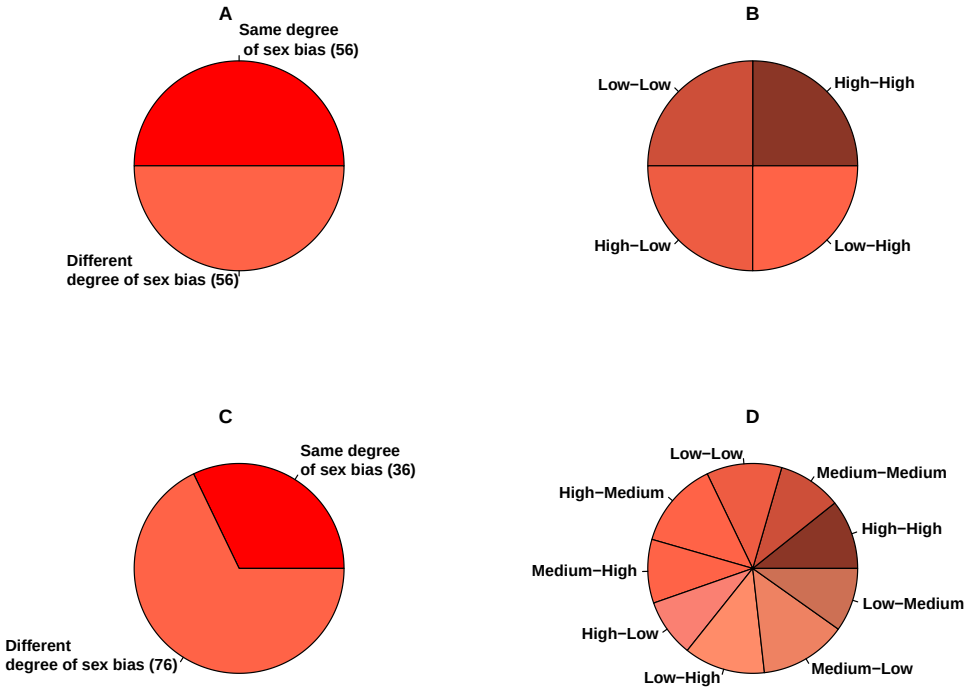

Figure 4.10: Conservation of degree of sex-bias between *D. melanogaster* and *D. ananassae* for female-biased genes with high, medium, and low degree of sex-bias

Panels A and B display results for dividing female-biased genes into two groups according to degree of sex-bias, high and low. Panels C and D display results for dividing female-biased genes into three groups; high, medium, and low degree of sex-bias. For panels A and C, the numbers of genes are given in parantheses.

a conserved degree of sex-bias using the 2-group analysis. For female-biased genes, this proportion is 61%. Dividing sex-biased genes into three groups (high, medium, and low degree of sex-bias), leads to a conservation of degree of sex-bias of 49% for male-biased genes and 40% for female-biased genes, respectively.

4.4 Conclusions

Using whole-genome data, we could show that for several *Drosophila* species the rate of molecular evolution measured by the ratio of the nonsynonymous substitution rate (d_N) to the synonymous substitution rate (d_S) is significantly positively correlated with the degree of sex-biased gene expression. As this is true for *D. melanogaster*, *D. ananassae*, and *D. pseudoobscura*, these findings correspond to a evolutionary timescale of about 30 million years.

Further, there is a general pattern of faster evolution for highly male-biased genes compared to male-biased genes with lower degree of sex-bias, female-biased, or unbiased genes in all three species. For orthologous genes between *D. melanogaster* and *D. ananassae*, genes with male-biased gene expression conserved between species reveal the highest rates of evolution.

We investigated whether differences between groups of orthologous genes are further influenced by the ancestral state of gene expression. We found evidence for the hypothesis that genes which show conserved male-biased gene expression over longer evolutionary time periods reveal significantly higher rates of molecular evolution. Moreover, we found the degree of sex-bias between *D. melanogaster* and *D. ananassae* is only about 50% conserved for both male- and female-biased genes.

We conclude that future studies comparing sex-biased gene expression between species should take into account both differences in degree of sex-bias and ancestral expression states.

Chapter 5

Phenotypic analysis of cold tolerance in *Drosophila ananassae* populations

5.1 Introduction

Adaptation to the surrounding environment is a key feature of Darwinian evolution. In *D. ananassae*, most populations are thought to be of a quite young age. This makes it unlikely that natural selection has had a considerable effect on DNA sequence variation in most of the populations (Das *et al.*, 2004). However, previous studies found evidence for adaptive evolution from patterns of DNA sequence variation in northern vs. southern populations in Asia (Baines *et al.*, 2004; Chen *et al.*, 2000; Kim and Stephan, 2000; Stephan *et al.*, 1998). In this study, we wanted to examine possible adaptation to temperature within *D. ananassae* populations. We investigated adaptation of different populations to their local environment by measuring the chill coma recovery time (CCRT) of the flies. CCRT is the time required for a fly to stand up after it has been placed in a "chill coma" by lowering its temperature (David *et al.*, 1998). We expect that flies from colder environments have shorter CCRT than those from warmer environments, if local adaptation has occurred. Drosophilids were previously classified into temperate or tropical species according to their CCRT (Gibert *et al.*, 2001) where single strains of different species were investigated. We wanted to take a further look into differences between populations of a tropical species.

5.2 Materials and Methods

5.2.1 Selection of populations

For our experiments, we used two *D. ananassae* populations. One population originates from Bangkok, Thailand (BKK), the other from Kathmandu, Nepal (KATH). Bangkok (latitude = 13N, elevation = 2m) has an average summer temperature of 29°C and an average winter temperature of 26°C. Kathmandu (latitude = 27N, elevation = 1,400m) has an average summer temperature of 19°C and an average winter temperature of 3°C. For each population, we randomly chose three different fly strains (BKK 9, 12, 17 and KATH 14, 19, 23, respectively). All flies were raised on standard cornmeal medium in our fly room at approximately 22°C, 40% humidity and constant lightning.

5.2.2 Chill Coma Recovery Test

We measured cold tolerance by using a modified protocol from the Chill Coma Recovery (CCR) test performed by David *et al.* (1998). The CCR test was shown to be tightly linked to cold tolerance ability and this trait follows latitudinal clines (Bubliy *et al.*, 2002; David *et al.*, 1998; Gibert *et al.*, 2001). Additionally, the protocol is simple and can be performed on a large number of individuals at the same time. To avoid influences of breeding conditions (David *et al.*, 1998), we kept all tested lines at low density. This limits competition between individual flies and allows them to reach their maximum body size.

After eclosion, young flies were sex-separated under light CO_2 anesthesia when around one day old and transfered into fresh 50ml vials containing 10ml of standard cornmeal food. At four to five days old, flies were separated without any anesthesia and placed into individual empty 8ml vials. As CO_2 anesthesia might affect the results of temperature experiments (Milton and Partridge, 2008; Nilson *et al.*, 2006), this procedure has the advantage of not exposing individuals to CO_2 immediately before the cold shock. These small vials were put into ice water (0°C ± 0.5°C) and stored in a refrigerator (4°C) for three hours. After this time, flies were put back to room temperature and CCRT was measured as the time (in minutes) necessary for a fly to stand on its legs (David *et al.*, 1998; Macdonald *et al.*, 2004; Morgan and Mackay, 2006). Males and females were scored in parallel. The experiment was stopped after 70 min of recording. Flies that were still alive but did not recover (they showed movements or waked up after slightly shaking the vials), were assigned a CCRT of 70 min. Dead individuals were not included into the analyses. They

Strain	BKK 9	BKK 12	BKK 17	KATH 14	KATH 19	KATH 23
No. of males	26	38	39	40	46	36
No. of females	26	47	45	44	52	36

Table 5.1: Number of flies assayed for chill coma recovery time (CCRT)

presented around 1% of the total flies tested and were randomly distributed among the lines tested and among males and females. In total, we performed eight measurements on four days. Each measurement comprised 60 flies and within each measurement each of the six fly strains were included. We were able to measure CCRT for 475 flies in total (five flies died during the experiment). Table 5.1 gives the total number of flies investigated for each strain.

5.2.3 Analysis

All data were analyzed using the software package R 2.9.0 (R Development Core Team, 2009). We analyzed differences in CCRT between different strains or between males and females of one strain using a Mann-Whitney U test. The correlation between male and female CCRT was determined with Spearman's rank correlation for all strains combined.

5.3 Results

We compared six different fly strains with respect to their cold resistance. In general, we found evidence for local adaptation from the CCRT assays. We found that flies from Kathmandu, Nepal, are more resistant to cold than flies from Bangkok, Thailand (Figure 5.1). The former flies started to wake up earlier, the maximum time needed for a fly of this population to wake up was lower and the strains had lower CCRT on average (Table 5.2 and Figure 5.3). There is a significant positive correlation (Spearman's rank correlation, P-value = 0.003, $\rho = 1$) between the mean CCRT of males and females for each strain (Figure 5.2).

However, there is one strain from Bangkok (BKK 12) which shows a similar pattern as the Kathmandu strains (Figures 5.3, 5.4). Flies from this strain were already awake after around 37 minutes on average (Table 5.2). There seem to be higher variability within the Bangkok population compared to the one originated from Kathmandu. Additional tests on different *D. ananassae* strains and populations could be performed to further confirm this pattern.

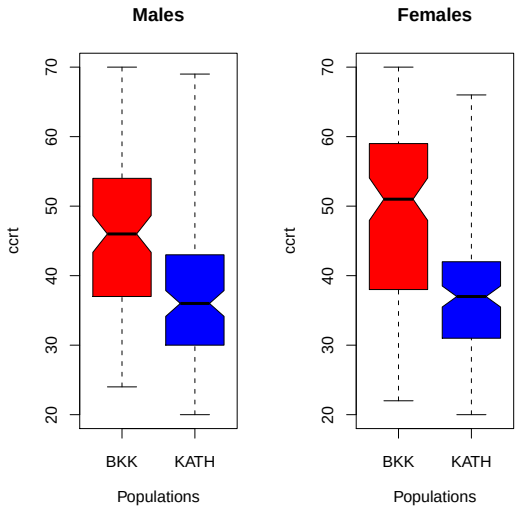

Figure 5.1: Comparison between Bangkok and Kathmandu flies

For flies from Bangkok, Thailand (red) the average chill coma recovery time (CCRT) is calculated over all three fly strains tested (BKK 9, 12, 17). For flies from Kathmandu, Nepal, CCRT is calculated over the strains KATH 14, 19, and 23. In both cases, male and female flies are considered separately. However, all experiments included both male and female flies from all strains.

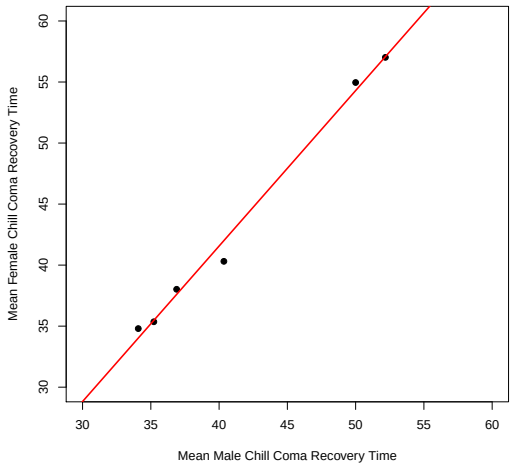

Figure 5.2: Correlation between male and female chill coma recovery time

For each of the six fly strains (BKK 9, 12, 17 and KATH 14, 19, 23, respectively) mean female and mean male chill coma recovery time was determined over all tested flies (on average around 40 flies per strain and sex).

Population	Mean	Median	Variance	*sd*	Min.	Max.
BKK 9						
Males	50.00	48.50	48.72	6.98	36	69
Females	54.96	58.00	79.48	8.92	38	70
BKK 12						
Males	36.89	34.50	19.72	10.26	24	70
Females	38.02	36.00	17.04	9.77	22	59
BKK 17						
Males	52.18	53.00	61.21	10.61	27	69
Females	57.02	58.00	43.22	7.62	33	70
KATH 14						
Males	34.08	32.00	10.71	8.77	20	69
Females	34.80	33.00	15.82	8.09	20	63
KATH 19						
Males	40.35	41.50	27.86	7.29	23	57
Females	40.31	39.50	19.66	9.11	25	66
KATH 23						
Males	35.22	35.00	25.99	7.04	21	49
Females	35.36	35.50	32.95	7.78	21	55

Table 5.2: General statistics for chill coma recovery time (CCRT) (in minutes) for Bangkok (BKK) and Kathmandu (KATH) flies

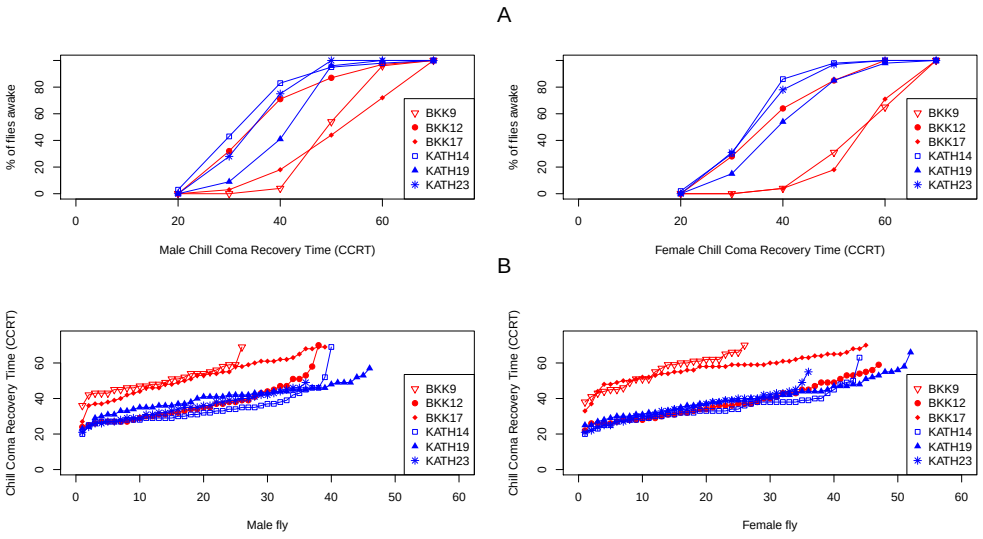

Figure 5.3: Chill coma recovery time of different populations

Panel A shows the percentage of flies awake according to observation time. Panel B shows the recovery times for single flies. All statistics are calculated for male and female flies separately.

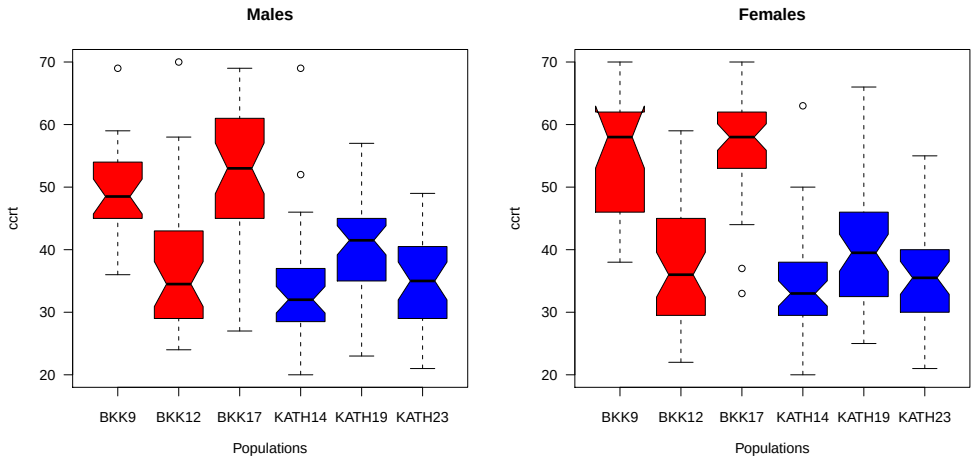

Figure 5.4: Comparison of populations

Shown are chill coma recovery times for each strain and population. Male (left) and female (right) flies are considered separately.

We compared CCRT of all strains using Mann-Whitney U tests. Significance values are given in Table 5.3. In general, we found no significant differences between males and females for flies from Kathmandu. For two strains from Bangkok (BKK9 and BKK17), male and female CCRT differ significantly (P-value = 0.03). Further, in the overall sample of Bangkok, males and females differ significantly in their CCRT (P-value = 0.04).

Comparison	P-value Males	P-value Females
Between populations comparison		
BKK-KATH	<0.001	<0.001
BKK 9-KATH 14	<0.001	<0.001
BKK 9-KATH 19	<0.001	<0.001
BKK 9-KATH 23	<0.001	<0.001
BKK 12-KATH 14	0.24	0.18
BKK 12-KATH 19	0.02	0.19
BKK 12-KATH 23	0.86	0.33
BKK 17-KATH 14	<0.001	<0.001
BKK 17-KATH 19	<0.001	<0.001
BKK 17-KATH 23	<0.001	<0.001
Within populations comparison		
BKK 9-BKK 12	<0.001	<0.001
BKK 9-BKK 17	0.28	0.48
BKK 12-KATH 17	<0.001	<0.001
KATH 14-KATH 19	<0.001	<0.001
KATH 14-KATH 23	0.24	0.47
KATH 19-KATH 23	<0.001	<0.001

Table 5.3: Significance of comparisons (Mann-Whitney U test)

5.4 Outlook

We showed that CCRT differed both between and within populations of *D. ananassae*. The ancestral Bangkok population seems to harbor higher variabilty among individuals. A preliminary study of three populations (20 male and female flies each) showed also differences in cold resistance for a Japanese *D. ananassae* population from Kumejima (latitude = 26N, elevation = 4m) compared to flies from Bangkok and Kathmandu. Kumejima has an average summer temperature of 27°C and and average winter temperature of 17°C. Japanese flies were more resistant to cold than flies from Thailand. In addition, in this population differences between males and females were significant (Figure 5.5). General statistics on these strains are given in Table 5.4.

Differences between males and females could also be found for strains and populations of *D. melanogaster*. Further studies could elucidate this pattern by extending the analysis to several strains and populations. Moreover, population genetic studies on candidate genes for cold resistance known from *D. melanogaster* might reveal if the same loci are responsible for cold resistance in *D. ananassae*.

Population	Mean	Median	Variance	*sd*	Min.	Max.
BKK						
Males	48.53	48.00	140.71	11.86	27	65
Females	52.05	55.00	72.37	8.51	33	60
KATH						
Males	36.75	37.00	77.36	8.80	22	52
Females	33.30	32.50	94.75	9.73	20	49
KMJ						
Males	33.60	35.00	52.67	7.26	23	49
Females	42.15	44.00	70.77	8.41	27	54

Table 5.4: General statistics for chill coma recovery time (CCRT) (in minutes) for Bangkok (BKK), Kathmandu (KATH) and Kumejima (KMJ) flies

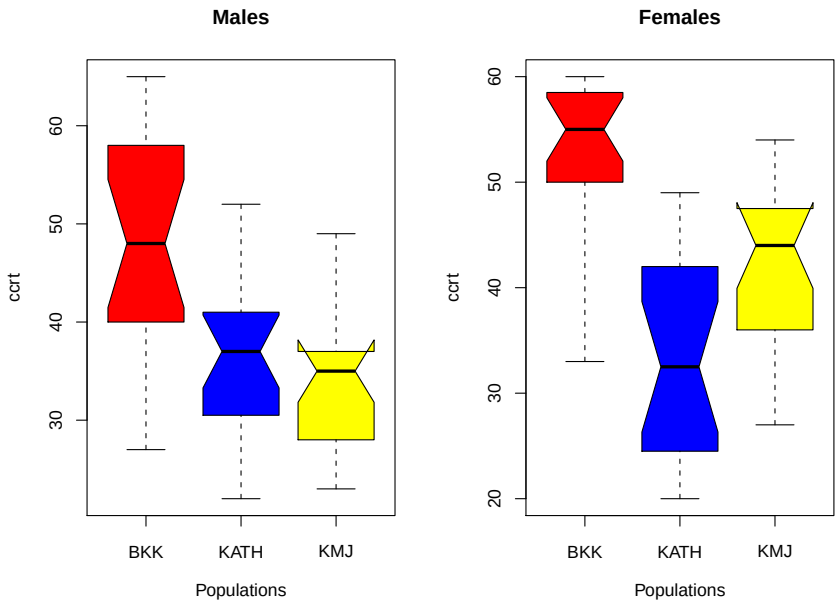

Figure 5.5: Comparison of flies from Bangkok, Kathmandu, and Kumejima

For each population and sex, on average 20 flies from two populations were analyzed. Flies from Bangkok, Thailand (BKK 9 and 17) are indicated in red, flies from Kathmandu, Nepal (KATH 14 and 23) in blue, and flies from Kumejima, Japan (KMJ 12 and 13) in yellow.

General Discussion

Sexual dimorphism is a very common phenomenon in nature. It is widespread among higher eukaryotes and is thought to result from the differential action of natural or sexual selection on male and female individuals of one species. Males and females are almost identical genetically. In our study organism *Drosophila*, the two sexes differ only in the sex-specific chromosomes X and Y, where the male forms the heterogametic sex (XY). Consequently, the vast majority of sexually dimorphic traits result from genes that are present in both sexes but differentially expressed in male and female individuals. These genes are commonly referred to as sex-biased genes and can be further grouped into male- and female-biased genes, according to the sex which shows higher expression. Previous studies of *Drosophila melanogaster* have shown that male-biased genes, particularly those expressed in reproductive tissues, consistently exhibit high levels of adaptive protein evolution (Baines *et al.*, 2008; Pröschel *et al.*, 2006). Moreover, male-biased genes are functionally divergent between closely-related *Drosophila* species and are less likely to have identifiable orthologs between distantly-related species than genes with female-biased or unbiased expression which has been indicated by genome-wide comparisons of the ratio of the nonsynonymous substitution rate (d_N) to the synonymous substiution rate (d_S) (Baines *et al.*, 2008; Zhang and Parsch, 2005). One limitation of these studies is that they rely on gene expression data from only one species, *D. melanogaster*. However, levels of sex-biased gene expression are known to vary between closely-related species (Jiang and Machado, 2009; Ranz *et al.*, 2003; Zhang *et al.*, 2007) and even between different strains of *D. melanogaster* (Meiklejohn *et al.*, 2003).

The overall goal of this dissertation was to determine the forces affecting the evolution of sex-biased genes in a *Drosophila* species other than the well-studied *D. melanogaster*. For this, we chose *D. ananassae* as study organism. This species is of particular interest because of its phylogenetic position as an outgroup to all species within the *melanogaster* species subgroup and its population structure (see INTRODUCTION). *D. ananassae* is

distributed throughout the subtropical and tropical regions of the world. The ancestral region of the species has been defined as a region of Southeast Asia that existed as a single landmass during the late Pleistocene around 18,000 years ago (Das, 2005; Das *et al.*, 2004; Schug *et al.*, 2007). One factor possibly influencing sex-related traits has been already proposed by Darwin (1871). He supposed that sexual selection, either through direct male-male competition or female mate choice, was responsible for the occurrence of extravagant secondary characteristics present in the males of many species.

In CHAPTER 1, we investigated mating discrimination and genetic differentiation in 18 populations of *D. ananassae*. Experimental tests of mate choice have found levels of pre-mating isolation to be highest between various peripheral populations. This suggests that sexual selection occurs within local populations of *D. ananassae*, leading to reproductive isolation among subpopulations. Several studies found evidence that mate discrimination may evolve as a by-product of adaptation to environmental conditions while populations are in allopatry (Funk, 1998; Langerhans *et al.*, 2007; Vines and Schluter, 2006). In the case of the Asian *D. ananassae* populations for which we found significant mate discrimination (Chennai, Kathmandu, Puri and Bangkok), putatative environmental variables that might influence fitness have not yet been identified. However, possible factors might be related to temperature and elevation in particular geographical regions.

In CHAPTER 5, we analyzed cold resistance in three Asian populations, two of which (Bangkok and Kathmandu) showing high mate discrimination and found evidence for local adaptation from chill coma recovery time (CCRT) assays. We could show that cold resistance measured by the CCRT of individual flies differ both between and within populations of *D. ananassae*. We found that flies from Kathmandu are more resistant to cold than flies from Bangkok. Moreover, there are differences in CCRT between males and females for two strains of Bangkok with the females taking more time to recover than the males. The same is true for females from another Asian population from Kumejima, Japan. This suggests that the two sexes have adapted differently to cold temperatures in these populations and that sex-biased genes may be involved in environmental adaptation.

Sex-biased genes may be subject to different selective constraints depending on the sex in which they are expressed. Even conflicting selective pressures in the two sexes are possible which is known as sexual antagonism. Here, the expression of one sex is beneficial to one sex but harmful to the other (Ellegren and Parsch, 2007). In *Drosophila*, sexually antagonistic genes occur frequently and there response to selection has been demonstrated. Previous studies even used sex-biased gene expression as a proxy for sexual antagonism.

It has been assumed, that differences in expression might reflect conflict at the genetic level. However, one recent study questions this assumption. Innocenti and Morrow (2010) investigated sex-biased gene expression in fly strains with different fitness in males and females. Whereas 91.5% of the transcriptome revealed sex-biased gene expression, only 8% of the transcripts experienced sexually antagonistic selection.

Previously, it has been found that male-biased genes show high levels of adaptive evolution compared to female-biased or unbiased genes in *D. ananassae* (Baines *et al.*, 2008; Pröschel *et al.*, 2006). In contrast, studies on *D. pseudoobscura* showed conflicting results. Whereas one SAGE (serial analysis of gene expression) study found no accelerated rate of protein evolution for male-biased genes in this species (Metta *et al.*, 2006), one recent study using whole-genome expression data could show that male-biased genes show faster rates of protein divergence than female-biased or unbiased genes (Jiang and Machado, 2009). This conflicting result might be explained by the fact that the first study is based on a sequencing method with relatively low coverage and on a lower number of genes compared to the second study. Metta *et al.* (2006) suggested that *D. pseudoobscura* revolved under a different selection regime and the patterns of sex-biased gene evolution may have changed since the split of the *D. melanogaster* and *D. pseudoobscura* lineages.

In CHAPTER 2, we further investigate this possibility by analyzing sex-biased gene expression and DNA sequence polymorphism in *D. ananassae*, which serves as an outgroup to *D. melanogaster* but is more closely related to *D. melanogaster* than *D. pseudoobscura* (*Drosophila* 12 Genomes Consortium, 2007; Larracuente *et al.*, 2008). We could show that sex-biased expression between *D. ananassae* and *D. melanogaster* is mainly conserved, but around 40% of the genes showed a gain, loss, or reversal of sex-biased expression between species. This is in accordance with other studies measuring sex-biased gene conservation between *Drosophila* species (Jiang and Machado, 2009; Metta *et al.*, 2006; Ranz *et al.*, 2003), but might be affected by our experimental design. As our set of candidate genes is enriched for genes showing strong sex-biased expression in multiple experiments in *D. melanogaster*, we see an overrepresentation of genes that show sex-biased gene expression private to *D. melanogaster*. Further, our classification of genes into sex-bias classes is based on two different microarray experiments as we combined results from our own PCR-amplicon microarrys with those from a published whole-genome microarray study (Zhang *et al.*, 2007). Our survey of DNA sequence polymorphism in *D. ananassae* and divergence to the sister species *D. atripex* and/or *D. phaeopleura* indicate that adaptive protein evolution exists in *D. ananassae*, with estimates of α in the range of 50-60%. However, we do

not observe an increased rate of adaptive evolution for male-biased genes. This contrasts to previous findings in *D. melanogaster* and is in accordance with the study of Metta *et al.* (2006), who found no accelerated rates of evolution for male-biased genes in *D. pseudoobscura*. Taken together, our results suggest that male-biased genes indeed have experienced an increase in the rate of adaptive protein evolution since the divergence of the *melanogaster* and *ananassae* subgroups. However, the observed differences might be influenced by the particular genes we investigated. Our gene set was enriched for genes showing strong and consistent sex-biased expression across multiple experiments in *D. melanogaster*. Previously, it has been found that the degree of male bias is correlated with rate of adaptive evolution across the *melanogaster* subgroup (Baines *et al.*, 2008). Such a correlation in the *ananassae* subgroup could explain differences in male-biased gene evolution between the lineages. Currently, we lack appropriate divergence data for outgroups to *D. ananassae* to investigate this further. Future studies might address this question by focusing on genes with exceptionally strong male bias in *D. ananassae*. Further, newly developed sequencing methods (reviewed in Mardis (2008, 2009); Morozova and Marra (2008); Morozova *et al.* (2009)) might be used to get whole-genome data for outgroups to *D. ananassae* or to infer sex-biased gene expression in these species.

We investigated the correlation between the degree of sex-bias and the rate of molecular evolution measured by d_N/d_S over the whole *melanogaster* group and compared this to the *obscura* group (CHAPTER 4).

We could show that the rate of molecular evolution measured by the ratio of the non-synonymous substitution rate (d_N) to the synonymous substitution rate (d_S) is significantly correlated with the degree of sex-bias measured by the M/F expression ratio. This is true for *D. melanogaster*, *D. ananassae*, and *D. pseudoobscura* and therefore corresponds to an evolutionary timescale of about 30 million years. Note, however, that this does not specifically address the question above if there are differences in the rates of adaptive protein evolution for male-biased genes between different lineages. This would be a correlation between the M/F expression ratio in the *melanogaster* lineage with the d_N/d_S between *D. melanogaster* and *D. simulans* (Baines *et al.*, 2008) and a correlation between the M/F expression ratio in the *ananassae* lineage and the d_N/d_S between *D. ananassae* and *D. atripex* (or *D. phaeopleura*). As we used d_N/d_S over the *melanogaster* group, only the expression data changes and we can just state that there is a general pattern of faster evolution of highly male-biased genes. In addition, we investigated the conservation of the degree of sex-bias between orthologous genes between *D. melanogaster* and *D. ananassae* and found

about 50% conservation for both male- and female-biased genes. Further, we found evidence for the hypothesis that genes showing conserved male-biased gene expression over longer evolutionary time scales show significantly higher rates of protein evolution. This might influence comparative studies of sex-biased genes.

For three protein-coding genes showing male-biased expression in *D. ananassae* and/or *D. melanogaster* and four unbiased genes, we investigated adaptive evolution in twelve populations of *D. ananassae* (CHAPTER 3). All these genes are located on the autsomes. Recent studies on gene loci were mainly limited to a few genes (Baines *et al.*, 2004; Chen *et al.*, 2000; Stephan *et al.*, 1998; Vogl *et al.*, 2003) or non-coding sequences (Das *et al.* (2004); Schug *et al.* (2007, 2008), see CHAPTER 1) on the X-chromosome. The first three genes show signatures of positive selection in one ancestral *D. ananassae* population from Bangkok, Thailand (see CHAPTER 2). The remaining four genes do not show any evidence for selection in the Bangkok population. We found patterns of adaptive evolution acting on gene (*CG14717*) particularly in the northern range of the population. Future studies could address molecular function of our candidate genes further and could integrate more data to improve the power of the statistical tests applied.

In my dissertation, I studied the evolutionary patterns of sex-biased genes in another Drosophilid species than the well-known *D. melanogaster*. Determing sex-biased gene expression in this species illustrated the importance of using genome-specific expression data. Around 40% of genes gained or lost sex-biased gene expression between *D. melanogaster* and *D. ananassae*. Consequently, concentrating on gene expression of only one species could result in misleading interpretations. By analyzing the DNA sequence polymorphism in *D. ananassae* and the divergence to the sister species *D. atripex* and *D. melanogaster* I could give further evidence on the hypothesis that selection patterns of male-biased genes have changed after the split of the *melanogaster* subgroup. However, there are similarities over a long time scale of Drosophilid evolution. To date, we have conducted the largest population genetic and gene expression study of sex-biased genes in *D. ananassae*. The major results presented in this dissertation are:

Mate discrimination varies considerably throughout the species range of *D. ananassae* and is partly influenced by genetic differentiation between different populations (Chapter 1).

Mate discrimination is one of the most commonly measured forms of prezygotic isolation and appears to be relatively common among closely related species. In this study, we measured the level and pattern of mate discrimination among 18 populations of *D. ananassae* from throughout its geographical range and its sister species *D. pallidosa*, which is restricted to the South Pacific Islands. Additionally, we measured genetic differentiation between all 18 populations using mitochondrial DNA polymorphism data. We could show that mate discrimination varies considerably throughout the species range, being higher among populations outside the ancestral Indonesian range, and highest in the South Pacific. Moreover, genetic differentiation seems to contribute to mating discrimination. Our results suggest that the colonization and genetic differentiation may have an influence on the evolutionary origin of mate discrimination. Several populations from the South Pacific show particularly strong mate discrimination suggesting that they might be in the early stages of speciation. Furthermore, both the genetic and behavioral results cast doubt on the status of *D. pallidosa* as a "good" species.

The signal of adaptive protein evolution for male-biased genes is lower in *D. ananassae* compared to *D. melanogaster* and restricted to genes with conserved male-biased gene expression (Chapter 2).

Previously, studies of the *melanogaster* subgroup revealed that male-biased genes have especially high rates of adaptive evolution compared to female-biased and unbiased genes.

We extended these studies to another species of the *melanogaster* group and investigated sex-biased gene evolution in *D. ananassae*. This species occurs in structured populations in tropical and subtropical regions. We used custom-made PCR-amplicon microarrays and published microarray data to characterize the sex-biased expression of 129 *D. ananassae* genes whose *D. melanogaster* orthologs had been classified previously as sex-biased or unbiased in their expression and had been studied extensively at the population genetic level. For 43 of these genes we surveyed DNA sequence polymorphism in a natural population of *D. ananassae* and determined divergence to the sister species *D. atripex* and *D. phaeopleura*. In general, we found sex-biased gene expression to be conserved between *D. ananassae* and *D. melanogaster*. However, about one-third of the genes have either gained or lost sex-biased expression in one of the species and about 4% showed a reversal of sex-biased expression between these species. The signal of adaptive protein evolution for male-biased genes is not as strong in *D. ananassae* as it is in *D. melanogaster* and is limited to genes with conserved male-biased expression. Our findings extend previous observations of widespread adaptive protein evolution to an independent *Drosophila* lineage, the *D. ananassae* subgroup. However, the rate of adaptive evolution is not greater for male-biased genes than for female-biased or unbiased genes, which suggests that there are differences in sex-biased gene evolution between the two lineages.

The strength and type of selection acting on individual genes differ among populations of *D. ananassae* (Chapter 3).

For seven protein-coding genes that were analyzed for sex-biased gene expression and adaptive evolution in *D. ananassae* in CHAPTER 2, we surveyed DNA sequence polymorphism in twelve populations throughout the geographical range of *D. ananassae* and analyzed divergence to *D. atripex* and *D. phaeopleura*. The same populations were investigated for mating discrimination and genetic differentiation in CHAPTER 1. We performed the largest population genetic study on sequence variation of autosomal genes in *D. ananassae* so far using populations sampled from almost the whole worldwide species range.

The rate of adaptive evolution is dependent on the degree of sex-bias over 30 million years of Drosophilid evolution (Chapter 4).

Previous studies provided evidence for differences in evolutionary patterns of sex-biased genes in various species of *Drosophila*. Whereas *D. melanogaster* shows a clear pattern of higher rates of adaptive protein evolution for male-biased genes, studies of *D. ananassae* and *D. pseudoobscura* suggested that this pattern developed after the split of the *melanogaster* subgroup from more basal *Drosophila* lineages. We used genome-wide ex-

pression information from *D. melanogaster*, *D. ananassae*, and *D. pseudoobscura* and data on the rate of evolution estimated from comparative genomic studies. We found that the rate of evolution measured by d_N/d_S is correlated with the degree of sex-bias. There is a general pattern of faster evolution for highly male-biased genes compared to male-biased genes with a lower degree of sex-bias for all three species. For orthologs between *D. melanogaster* and *D. ananassae*, male-biased genes show a different pattern if they are conserved in sex-bias between species or not, with the conserved genes showing faster rates of evolution. Moreover, up to 50% of genes are not conserved between these species with respect to their degree of sex-bias and there are differences depending on the ancestral state of sex-bias as inferred from expression in *D. pseudoobscura*. This study shows that more factors than previously known might influence evolutionary rates of sex-biased genes in different Drosophilid species.

***D. ananassae* populations show evidence for cold adaptation (Chapter 5).**
We performed a pilot study on cold resistance in several strains and populations of *D. ananassae*. Using a commonly used chill coma recovery test, we investigated chill coma recovery time of two *D. ananassae* populations. One of these populations (Bangkok, Thailand) experiences tropical conditions with little climate change over the year and little differences between day and night temperatures. The other population (Kathmandu, Nepal) is situated at higher altitude (1,400m) with lower average temperatures and higher differences between average summer and winter temperature. All flies were raised under standard fly room conditions over many generations at approximately 22°C. However, fly strains of different populations reveal differences in cold resistance. After a cold shock at 0°C for three hours, flies from Kathmandu, Nepal, recovered after approximately 35 minutes, whereas flies from Bangkok, Thailand, take around ten minutes longer to recover. Moreover, we see higher variability in chill coma resistance for flies from Bangkok, which is part of the ancestral range of this species. For Bangkok, we can observe differences between male and females, whereas for Kathmandu, we do not. In addition, we performed a preliminary study that included two strains from Kumejima, Japan. These flies showed higher cold tolerance compared to flies from Thailand and differences between males and females similar to those observed for *D. melanogaster*. These results suggest that sex-specific adaptation to environmental conditions has occurred in at least some populations of *D. ananassae*.

Bibliography

F. Abascal, R. Zardoya, and D. Posada. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*, 21:2104–2105, 2005.

M. Aguadé. Different forces drive the evolution of the *Acp26Aa* and *Acp26Ab* accessory gland genes in the *Drosophila melanogaster* species complex. *Genetics*, 150:1079–1089, 1998.

H. Akaike. Information theory and an extension of the maximum likelihood principle. *In: BN Petrov and F Csaki, eds. Second International Symposium on Information Theory, Budapest: Akademiai Kiado*, 8, 1973.

W. W. Anderson and L. Ehrman. Mating choice in crosses between geographic populations of *Drosophila psuedoobscura*. *American Midland Naturalist*, 81:47–53, 1969.

J. F. Ayroles, M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman, M. M. Magwire, S. M. Rollmann, L. H. Duncan, F. Lawrence, R. R. H. Anholt, and T. F. C. Mackay. Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genetics*, 41: 299–307, 2009.

J. F. Baines, A. Das, S. Mousset, and W. Stephan. The role of natural selection in genetic differentiation of worldwide populations of *Drosophila ananassae*. *Genetics*, 168:1987–1998, 2004.

J. F. Baines, S. A. Sawyer, D. L. Hartl, and J. Parsch. Effects of X-linkage and sex-biased gene expression on the rate of adaptive protein evolution in *Drosophila*. *Molecular Biology and Evolution*, 25:1639–1650, 2008.

E. Betrán and M. Long. *Dntf-2r*, a young *Drosophila* retroposed gene with specific male expression under positive Darwinian selection. *Genetics*, 164:977–988, 2003.

N. Bierne and A. Eyre-Walker. The genomic rate of adaptive amino acid substitution in *Drosophila*. *Molecular Biology and Evolution*, 21:1350–1360, 2004.

I. R. Bock and M. R. Wheeler. The *Drosophila melanogaster* species group. *University of Texas Publications*, 7213:1–102, 1972.

O. A. Bubliy, A. Riihimaa, F. M. Norry, and V. Loeschcke. Variation in resistance and acclimation to low-temperature stress among three geographical strains of *Drosophila melanogaster*. *Journal of Thermal Biology*, 27:337–344, 2002.

C. Bustamante, R. Nielsen, S. Sawyer, K. Olsen, M. Purugganan, and D. Hartl. The cost of inbreeding in *Arabidopsis*. *Nature*, 416:531–534, 2002.

A. Carvajal-Rodriguez and E. Rolan-Alvarez. JMATING: a software for the analysis of sexual selection and sexual isolation effects from mating frequency data. *BMC Evolutionary Biology*, 6:40, 2006.

J. Charlesworth and A. Eyre-Walker. The McDonald-Kreitman test and slightly deleterious mutations. *Molecular Biology and Evolution*, 25:1007–1015, 2008.

Y. Chen, B. J. Marsh, and W. Stephan. Joint effects of natural selection and recombination on gene flow between *Drosophila ananassae* populations. *Genetics*, 155:1185–1194, 2000.

A. Civetta and R. S. Singh. Broad-sense sexual selection, sex gene pool evolution, and speciation. *Genome*, 42:1033–1041, 1999.

M. F. Claridge, J. D. Hollander, and J. C. Morgan. Variation in courtship signals and hybridization between geographically definable populations of the rice brown planthopper, *Nilaparvata lugens* (Stal). *Biological Journal of the Linnean Society*, 24:35–49, 1985.

M. F. Claridge, J. Denhollander, and J. C. Morgan. Variation in hostplant relations and courtship signals of weed-associated populations of the brown planthopper, *Nilaparvata lugens* (Stal), from Australia and Asia: a test of the recognition species concept. *Biological Journal of the Linnean Society*, 35:79–93, 1988.

J. Coyne and H. A. Orr. *Speciation*. Sinauer & Associates, Sunderland, Massachusetts, 2004.

J. A. Coyne, S. Elwyn, and E. Rolan-Alvarez. Impact of experimental design on *Drosophila* sexual isolation studies: direct effects and comparison to field hybridization data. *Evolution*, 59:2588–2601, 2005.

J.-L. DaLage, G. J. Kergoat, F. Maczkowiak, J.-F. Silvain, M.-L. Cariou, and L. Daniel. A phylogeny of Drosophilidae using the *Amyrel* gene: questioning the *Drosophila melanogaster* species group boundaries. *Journal of Zoological Systematics and Evolutionary Research*, 45:47–63, 2007.

C. Darwin. *The Descent of Man*. John Murray, London, 1871.

A. Das. Population genomics and bioinformatic studies reveal evolutionary history of *Drosophila ananassae*. *Current Science*, 89, 2005.

A. Das, S. K. Jamiruddin, and B. B. Parida. Desiccation tolerance in an Indian natural population of *Drosophila ananassae*. *Journal of Environmental Biology*, 16:237–242, 1995.

A. Das, S. Mohanty, and W. Stephan. Inferring the population structure and demography of *Drosophila ananassae* from multilocus data. *Genetics*, 168:1975–1985, 2004.

R. J. David, P. Gibert, E. Pla, G. Petavy, D. Karan, and B. Moreteau. Cold stress tolerance in *Drosophila*: analysis of chill coma recovery in *D. melanogaster*. *Journal of Thermal Biology*, 23:291–299, 1998.

T. Dobzhansky, L. Ehrman, O. Pavlovsky, and B. Spassky. The superspecies *Drosophila paulistorum*. *Proceedings of the National Academy of Sciences*, 51:3–9, 1964.

M. Doi, M. Matsuda, T. Masatoshi, H. Matsubayashi, and O. Yuzuru. A locus for female discrimination behavior causing sexual isolation in *Drosophila*. *Genetics and Molecular Research*, 98:6714–6719, 2001.

Drosophila 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, 450:203–218, 2007.

W. G. Eberhard. *Sexual selection and animal genitalia*. Harvard University Press, Cambridge, MA, 1985.

R. C. Edgar. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5:113, 2004.

L. Ehrman and P. A. Parsons. Sexual isolation among widely distributed populations of *Drosophila immigrans*. *Behavior Genetics*, 10:401–407, 1980.

A. A. Elens and J. M. Wattiaux. Direct observation of sexual isolation. *Drosophila Information Services*, 39:118–119, 1964.

H. Ellegren. The different levels of genetic diversity in sex chromosomes and autosomes. *Trends in Genetics*, 25:278–284, 2009.

H. Ellegren and J. Parsch. The evolution of sex-biased genes and sex-biased gene expression. *Nature Reviews Genetics*, 8:689–698, 2007.

J. C. Fay, G. J. Wyckoff, and C.-I. Wu. Positive and negative selection on the human genome. *Genetics*, 158:1227–1234, 2001.

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

D. J. Funk. Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. *Evolution*, 52:1744–1759, 1998.

D. J. Futch. A study of speciation in South Pacific populations of *Drosophila ananassae*. *University of Texas Publications*, 6615:79–120, 1966.

D. J. Futch. On the ethological differentiation of *Drosohila ananassae* and *Drosophila pallidosa* in Samoa. *Evolution*, 420:456–467, 1973.

P. Gibert, B. Moreteau, G. Pétavy, D. Karan, and J. R. David. Chill-coma tolerance, a major climatic adaptation among *Drosophila* species. *Evolution*, 55:1063–1068, 2001.

G. Gibson, R. Riley-Berger, L. Harshman, A. Kopp, S. Vacha, S. Nuzhdin, and M. Wayne. Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. *Genetics*, 167:1791–1799, 2004.

J. M. Gleason and M. G. Ritchie. Evolution of courtship song and reproductive isolation in the *Drosophila willistoni* species complex: do sexual signals diverge the most quickly? *Evolution*, 52:1493–1500, 1998.

F. Gnad and J. Parsch. Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression. *Bioinformatics*, 22:2577–2579, 2006.

S. Grath, J. F. Baines, and J. Parsch. Molecular evolution of sex-biased genes in the *Drosophila ananassae* subgroup. *BMC Evolutionary Biology*, 9:219, 2009.

J. B. S. Haldane. Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, 12:101–109, 1922.

N. R. Henderson and D. M. Lambert. No significant deviation from random mating of worldwide populations of *Drosophila melanogaster*. *Nature*, 300:437–440, 1982.

J. Hey and R. M. Kliman. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics*, 160:595–608, 2002.

H. Hollocher, C. T. Ting, M. L. Wu, and C. I. Wu. Incipient speciation by sexual isolation in *Drosophila melanogaster*: extensive genetic divergence without reinforcement. *Genetics*, 147:1191–1201, 1997.

D. J. Hosken and P. Stockley. Sexual selection and genital evolution. *Trends in Ecology and Evolution*, 19:87–93, 2004.

D. J. Howard and S. H. Berlocher. *Endless Forms: Species and Speciation*. Oxford University Press, New York, 1998.

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

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

J. P. Huelsenbeck and F. Ronquist. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17:754–755, 2001.

S. Hutter, S. Saminadin-Peter, W. Stephan, and J. Parsch. Gene expression variation in African and European populations of *Drosophila melanogaster*. *Genome Biology*, 9:R12, 2008.

P. Innocenti and E. H. Morrow. The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biology*, 8:e1000335, 2010.

Y. Iwasa and A. Pomiankowski. Continual change in mate preferences. *Nature*, 377:420–422, 1995.

J. Jaenike, K. A. Dyer, C. Cornish, and M. S. Minhas. Asymmetrical reinforcement and Wolbachia infection in *Drosophila*. *PloS Biology*, 4:e325, 2006.

B. G. M. Jamieson. *The ultrastructure and phylogeny of insect spermatozoa*. Cambridge University Press, Cambridge, 1987.

Z.-F. Jiang and C. A. Machado. Evolution of sex-dependent gene expression in three recently diverged species of *Drosophila*. *Genetics*, 183:1175–1185, 2009.

D. T. Jones, W. R. Taylor, and J. M. Thornton. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, 8:275–282, 1992.

D. S. Joshi. Latitudinal variation in locomotor activity rhythm in adult *Drosophila ananassae*. *Canadian Journal of Zoology*, 77:865–870, 1999.

K. Kaneshiro. Sexual selection and direction of evolution in the biosystematics of Hawaiian Drosophilidae. *Annual Review of Entomology*, 28:161–178, 1983.

K. Y. Kaneshiro. Sexual isolation, speciation and the direction of evolution. *Evolution*, 34: 437–444, 1980.

P. Kastanis, E. Eliopoulos, G. N. Goulielmos, S. Tsakas, and M. Loukas. Macroevolutionary relationships of species of *Drosophila melanogaster* group based on mtDNA sequences. *Molecular Phylogenetics and Evolution*, 28:518–528, 2003.

Y. Kim and W. Stephan. Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics*, 155:1415–1427, 2000.

R. M. Kliman, P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman, A. J. Berry, J. McCarter, J. Wakeley, and J. Hey. The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics*, 156:1913–1931, 2000.

R. D. V. Klinken, G. H. Walter, and M. K. Ross. Drosophilidae (Diptera) of Australia's Northern Territory: ecology and biogeography. *Australian Journal of Entomology*, 41: 236–242, 2002.

R. B. Langerhans, M. E. Gifford, and E. O. Joseph. Ecological speciation in Gambusia fishes. *Evolution*, 61:2056–2074, 2007.

A. M. Larracuente, T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh, D. Sturgill, Y. Zhang, B. Oliver, and A. G. Clark. Evolution of protein-coding genes in *Drosophila*. *Trends in Genetics*, 24:114–123, 2008.

P. Librado and J. Rozas. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25:1451–1452, 2009.

S. S. Macdonald, L. Rako, P. Batterham, and A. A. Hoffmann. Dissecting chill coma recovery as a measure of cold resistance: evidence for a biphasic response in *Drosophila melanogaster*. *Journal of Insect Physiology*, 50:695–700, 2004.

J. R. Malloch. Insects of Samoa and other Samoan terrestrial Arthropoda, part VI. Diptera. *British Museum*, 8:267–328, 1934.

N. Mantel. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27:209–220, 1967.

E. R. Mardis. Next-generation DNA sequencing methods. *Annual Reviews of Genomics and Human Genetics*, 9:387–402, 2008.

E. R. Mardis. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 24:133–141, 2009.

M. Matsuda, C.-S. Ng, M. Doi, A. Kopp, and Y. N. Tobari. Evolution in the *Drosophila ananassae* species subgroup. *Fly*, 3:157–169, 2009.

J. H. McDonald and M. Kreitman. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, 351:652–654, 1991.

S. F. McEvey and M. Polak. *Mycodrosophila* (Diptera: Drosophilidae) of Fiji and Vanuatu with a description of nine new species. In N. L. Edenhuis and D. J. Bickel, editors, *Fiji Arthropods II*, pages 35–67. Bishop Museum Occassional Papers, Honolulu, Hawai, 2005.

S. F. McEvey, J. R. David, and L. Tsachs. The *Drosophila ananassae* complex with description of a new species from French Polynesia (Diptera: Drosophilidae). *Annales de la Société entomologique de France (N.S.)*, 23:377–385, 1987.

L. M. McIntyre, L. M. Bono, A. Genissel, R. Westerman, D. Junk, M. Telonis-Scott, L. Harshman, M. L. Wayne, A. Kopp, and S. V. Nuzhdin. Sex-specific expression of alternative transcripts in *Drosophila*. *Genome Biology*, 7:R79, 2006.

C. D. Meiklejohn, J. Parsch, J. M. Ranz, and D. L. Hartl. Rapid evolution of male-biased gene expression in *Drosophila*. *Proceedings of the National Academy of Sciences*, 100: 9894–9899, 2003.

M. Metta, R. Gudavalli, J.-M. Gibert, and C. Schlötterer. No accelerated rate of protein evolution in male-biased *Drosophila pseudoobscura* genes. *Genetics*, 174:411–420, 2006.

C. C. Milton and L. Partridge. Joint effects of natural selection and recombination on gene flow between *Drosophila ananassae* populations. *Journal of Insect Physiology*, 54:32–40, 2008.

T. J. Morgan and T. C. M. Mackay. Quantitative trait loci for thermotolerance phenotypes in Drosophila melanogaster. *Heredity*, 50:232–242, 2006.

O. Morozova and M. A. Marra. Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92:255–264, 2008.

O. Morozova, M. Hirst, and M. A. Marra. Applications of new-sequencing technologies for transcriptome analysis. *Annual Reviews of Genomics and Human Genetics*, 10:135–151, 2009.

M. Nei. *Molecular Evolutionary Genetics*. Columbia University Press, New York, 1987.

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

D. I. Nurminsky, M. V. Nurminskaya, D. D. A. D, and D. L. Hartl. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. *Nature*, 396:572–575, 1998.

D. Otte and J. A. Endler. *Speciation and its Consequences*. Sinauer & Associates, Sunderland, Massachusetts, 1989.

R. D. Page. TreeView: an application to display phylogenetic trees on personal computers. *Bioinformatics*, 12:357–358, 1996.

T. M. Panhuis, R. Butlin, M. Zuk, and T. Tregenza. Sexual selection and speciation. *Trends in Ecology and Evolution*, 16:364–371, 2001.

M. Parisi, R. Nuttall, D. Naiman, G. Bouffard, J. Malley, J. Andrews, S. Eastman, and B. Oliver. Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science*, 299:697–700, 2003.

M. Parisi, R. Nuttall, P. Edwards, J. Minor, D. Naiman, J. N. Lü, M. Doctolero, M. Vainer, C. Chau, J. Malley, S. Eastman, and B. Oliver. A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. *Genome Biology*, 5:R40, 2004.

J. Parsch, C. D. Meiklejohn, E. Hauschteck-Jungen, P. Hunziker, and D. L. Hartl. Molecular evolution of the ocnus and janus genes in the *Drosophila melanogaster* species subgroup. *Molecular Biology and Evolution*, 18:801–811, 2001.

J. Parsch, Z. Zhang, and J. F. Baines. The influence of demography and weak selection on the McDonald-Kreitman test: an empirical study in *Drosophila*. *Molecular Biology and Evolution*, 26:691–698, 2009.

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

D. C. Presgraves and H. A. Orr. Haldane's rule in taxa lacking a hemizygous X. *Science*, 282:952–954, 1998.

M. Pröschel, Z. Zhang, and J. Parsch. Widespread adaptive evolution of *Drosophila* genes with sex-biased expression. *Genetics*, 174:893–900, 2006.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

D. M. Rand and L. M. Kann. Excess amino acid polymorphism inmitochondrial dna: contrasts among genes from drosophila, mice, and humans. mol. biol. evol. 13: 735-748. *Molecular Biology and Evolution*, 13:735–748, 1996.

J. M. Ranz, C. I. Castillo-Davis, C. D. Meiklejohn, and D. L. Hartl. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Molecular Biology and Evolution*, 300:1742–1745, 2003.

W. Rice. Analyzing tables of statistical tests. *Evolution*, 43:223–225, 1989.

K. E. Ridout, C. J. Dixon, and D. A. Filatov. Positive selection differs between protein secondary structure elements in *Drosophila*. *Genome Biology and Evolution*, 2010:166–179, 2010.

E. Rolan-Alvarez. Evolution of asymmetry in sexual isolation: a criticism of a test case. *Evolutionary Ecology Research*, 6:1099–1106, 2004.

E. Rolan-Alvarez and A. Caballero. Estimating sexual selection and sexual isolation effects from mating frequencies. *Evolution*, 54:30–36, 2000.

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

K. Sawamura, Y. Tomimura, H. Sato, H. Yamada, M. Matsuda, and Y. Oguma. Establishing interspecific mosaic genome lines between *Drosophila ananassae* and *Drosophila pallidosa* by means of parthenogenesis. *Genetical Research*, 88:1–11, 2006.

K. Sawamura, H. Zhi, K. Setoguchi, H. Yamada, T. Miyo, and M. Matsuda. Genetic analysis of mating recognition between *Drosophila ananassae* and *Drosophila pallidosa*: application of interspecific mosaic genome lines. *Genetica*, 2007.

M. D. Schug, S. G. Smith, A. Tozier-Pearce, and S. F. McEvey. The genetic structure of *Drosophila ananassae* populations from Asia, Australia and Samoa. *Genetics*, 175: 1429–1440, 2007.

M. D. Schug, J. F. Baines, A. Killon-Atwood, S. Mohanty, A. Das, S. Grath, S. G. Smith, S. Zargham, S. F. McEvey, and W. Stephan. Evolution of mating isolation between populations of *Drosophila ananassae*. *Molecular Ecology*, 17:2706–2721, 2008.

G. Sella, D. A. Petrov, M. Przeworski, and P. Andolfatto. Pervasive natural selection in the *Drosophila* genome? *PloS Genetics*, 5:e1000495, 2009.

C. Simon, F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. Evolution, weigthing, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, 87: 651–701, 1994.

R. S. Singh and R. J. Kulathinal. Sex gene pool evolution and speciation: a new paradigm. *Genes & Genetic Systems*, 75:19–130, 2000.

S. R. Singh and B. Singh. Behavioral genetics of *Drosophila ananassae*. *Proceedings of the National Academy of Sciences*, 98:6714–6719, 2003.

S. Sisoda and B. N. Singh. Effect of temperature on longevity and productivity in *Drosophila ananassae*: evidence for adaptive plasticity and trade-off between longevity and productivity. *Genetica*, 114:95–102, 2002.

S. Sisoda and B. N. Singh. Size dependent sexual isolation in *Drosophila ananassae*. *Genetica*, 121:207–217, 2004.

G. K. Smyth and T. Speed. Normalization of cDNA microarray data. *Trends in Genetics*, 31:265–273, 2003.

H. T. Spieth. Genetic studies of natural populations of *Drosophila*. *Bulletin of the American Museum of Natural History*, 99:395–474, 1952.

H. T. Spieth. Genetic studies of natural populations of *Drosophila*. V. mating behavior of *D. ananassae* and *ananassae*-like flies from the Pacific. *University of Texas Publications*, 6615:133–145, 1966.

W. Stephan, L. Xing, D. A. Kirby, and J. M. Braverman. A test of the background selection hypothesis based on nucleotide data from *Drosophila ananassae*. *Proceedings of the National Academy of Sciences*, 95:5649–5654, 1998.

W. J. Swanson and V. D. Vacquier. The rapid evolution of reproductive proteins. *Nature Review Genetics*, 3:137–144, 2002.

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

Y. Takada and E. Rolan-Alvarez. Assortative mating between phenotypes of the intertidal snail *Littorina brevicula*: a putative case of incipient speciation? *Ophelia*, 52:1–8, 2000.

K. Tamura, J. Dudley, M. Nei, and S. Kumar. MEGA 4: molecular evolutionary genetic analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24:1596–1599, 2007.

Y. S. Tao, S. Chen, D. L. Hartl, and C. C. Laurie. Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. i. Differential accumulation of hybrid male sterility effects on the X and autosomes. *Genetics*, 164:1383–1397, 2003.

S. G. Tilley, P. A. Verrell, and S. J. Arnold. Correspondence between sexual isolation and allozyme differentiation: a test in the salamander *Desmognathus ochrophaeus*. *Proceedings of the National Academy of Sciences*, 87:2715–2719, 1990.

C. Ting, S. C. Tsaur, M. L. Wu, and C.-I. Wu. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science*, 282:1428–1429, 1998.

Y. N. Tobari. *Drosophila ananassae: Genetical and Biological Aspects*. Japan Scientific Societies Press/Karger, Tokyo/Basel, 1993.

Y. N. Tobari, B. Goni, Y. Tomimura, and M. Matsuda. Chromosomes. In Y. N. Tobari, editor, *Drosophila ananassae: Genetical and Biological Aspects*, pages 23–29. Japan Scientific Societies Press/Karger, Tokyo/Basel, 1993.

Y. Tomimura, M. Matsuda, Y. N. Tobari, M. L. Cariou, J. L. D. Lage, W. Stephan, and C. H. Langley. Population genetics. In Y. N. Tobari, editor, *Drosophila ananassae: Genetical and Biological Aspects*, pages 139–198. Japan Scientific Societies Press/Karger, Tokyo/Basel, 1993.

J. P. Townsend. Resolution of large and small differences in gene expression using models for the Bayesian analysis of gene expression levels and spotted DNA microarrays. *Trends in Genetics*, 5:54, 2004.

J. P. Townsend and D. L. Hartl. Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments. *Genome Biology*, 3:R71, 2002.

T. Tregenza. Divergence and reproductive isolation in the early stages of speciation. *Genetica*, 116:291–300, 2002.

T. Tregenza, V. L. Pritchard, and R. K. Butlin. The origins of premating reproductive isolation: testing hypotheses in the grasshopper *Chorthippus parallelus*. *Evolution*, 54: 1687–1698, 2000.

S. C. Tsaur, C. T. Ting, and C.-I. Wu. Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of *Drosophila*. II. Divergence versus polymorphism. *Molecular Biology and Evolution*, 15:1040–1046, 1998.

M. Turelli and H. A. Orr. The dominance theory of Haldane's rule. *Genetics*, 140:389–402, 1995.

T. H. Vines and D. Schluter. Strong assortative mating between allopatric sticklebacks as a by-product of adaptation to different environments. *Proceedings of the Royal Society B: Biological Sciences*, 273:911–916, 2006.

C. Vogl, A. Das, M. Beaumont, S. Mohanty, and W. Stephan. Population subdivision and molecular sequence variation: theory and analysis of *Drosophila ananassae* data. *Genetics*, 165:1385–1395, 2003.

T. K. Watanabe and M. Kawanishi. Mating preferences and the direction of evolution in *Drosophila*. *Science*, 205:906–907, 1979.

G. A. Watterson. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7:256–276, 1975.

S. Wright. Isolation by distance. *Genetics*, 28:114–138, 1943.

C.-I. Wu and A. W. Davis. Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. *The American Naturalist*, 142:187–212, 1993.

C. I. Wu, H. Hollocher, D. J. Begun, C. F. Aquadro, Y. Xu, and M. L. Wu. Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. *Proceedings of the National Academy of Sciences*, 92:2519–2523, 1995.

H. Yamada, M. Matsuda, and Y. Oguma. Genetics of sexual isolation based on courtship song between two sympatric species: *Drosophila ananassae* and *D. pallidosa*. *Genetica*, 116:225–237, 2002.

Y. Zhang, D. Sturgill, M. Parisi, S. Kumar, and B. Oliver. Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature*, 450:233–237, 2007.

Z. Zhang and J. Parsch. Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression. *Molecular Biology and Evolution*, 22:1945–1947, 2005.

E. Zouros and C. J. d'Entremont. Sexual isolation among populations of *Drosophila mojavensis*: response to pressure from a related species. *Evolution*, 34:421–430, 1980.

Appendix A

Primers for PCR-amplicon microarrays

Gene	Chro.[a]	Exp.[b]	Forward primer (5' - 3')	Reverse primer (5' - 3')	Length[c]
CG3481	2L	Control (Adh)	gagtgataatgtccgacg	gaacttacggctagggagc	653
CG3661	2R	Control (RpL23)	ccctcttctacccacaggtc	gcctagagaagatccacc	334
CG2867	3R	Female	acttgccttcgctggtggc	gatcaagccgcggtcacactgg	473
CG9383	3L	Female	tcggttgcaaatctgatccg	ccagaccattactgttaccg	527
CG9273	2L	Female	tgcatgccaaacatctcg	ccaacaggccacattaatgc	376
CG4973	3R	Female	gtatgtggcaatttaccgcc	ccactcacttggttaccac	784
CG1239	3R	Female	acggcgtaaaaagctatcgg	tttgcttccagtaggcttgg	249
CG6554	3R	Female	acctttcctacctttccc	gaatagctaacccagtcatccg	321
CG7840	2L	Female	caagctttccttgcactcgc	tggcccgtcaggttgtcg	739
CG9135	2L	Female	tgcagatcgttgactttgcc	gcagtattgtagttggcagcgg	699
CG5363	2L	Female	gctgcatttagtggcttggg	ctggaaactggaaagtctggc	292
CG3831	2R	Female	caccaagagcacacatagc	ccacgacacaaagtgttctcc	475
CG12276	3R	Female	gcttggttgatgctttcggg	ctggtttaagcggtcttggtgc	496
CG17950	2R	Female	atctctctttgcagctcg	agtgtgagtgggcgacttgg	524
CG5499	3R	Female	gaactttctccaggtgctcg	ctttcatattgaacctcgcc	538
CG10206	2L	Female	cgtacaactaatgccagcc	cgatgacgataaaccttggc	412
CG4299	3R	Female	accgcagtctctgctatcg	cgaaaccaagatggactacc	484
CG4236	3R	Female	agcagcggaatcctttgacg	cgcaatcgtaaatctggc	512
CG5272	3L	Female	gtgtggacgatgatgtaacg	agggtcagttggatgagg	576
CG5757	2R	Female	gagtctgtcaggtacagctc	ctattaagcgtggagcgc	472
CG12314	2L	Female	tgtcacatactcgtcgtcgg	ccatatatacgttcagctcg	355
CG4570	3R	Female	ctcagcatggagtcgaagcg	tgcacaatcgttagctgggc	405
CG12909	2R	Female	gtcacacttaactccacg	cttggtgagctgtgaaggc	746
CG6874	3L	Female	caggtgtactgcggattgg	ctcttacgcttccttctggc	463
CG13690	2L	Female	aaagcggcactcttcgcg	aacaccatcagcaccagc	572
CG3509	3R	Female	cccaatagccctccaagaaagc	ctttacgcatgcgctttccc	371
CG17361	3L	Female	acggtcacactaatcggagg	cgtcgtctaacacgtgtcgg	640
CG32409	3L	Female	aacagacgcgaagcggaagg	agtacctacatcgccacaccg	292
CG6459	2R	Female	gattccatttcgtcgccg	tagaagtctcgcggcgttgc	367

CG3975	2L	Female	gcttgtcgactccgttcactcc	atagctgagcttgcagctcggc	480
CG14434	X	Female	tcctccttgatgccaaacgcc	ttgtttaccatgtgcgaccc	526
CG15717	X	Female	ccttgacacttatcaccc	agcacctggcacctagttcg	702
CG9125	X	Female	aatcgccaccaagaacaaac	taccagtcgggaaggatctg	335
CG3024	X	Female	cgtaactgatcgagttcgcg	agttcgcaaagcagatcccg	660
CG8326	X	Female	gagtggattgcgcatgttgg	cacgagttcgcattgggaggc	146
CG3704	X	Female	aaccagaaagggcaggcg	caactggaggacatcaagc	560
CG12117	X	Female	cgcgaacatgaatctgagcc	ctcgaggacttcgatgaagc	738
CG3004	X	Female	tgccgccaaatagcgtccg	gaatgcgttggacagaacacc	426
CG4593	X	Female	cacgagcaacgtagtccagc	gcatcatagttggtacgc	587
CG8675	X	Female	cgaattccgagtcatcctcg	tccacaatgagcagccagc	645
CG11130	X	Female	tgtgtggactccagctacc	gatgttcgtgtatccgatcccc	317
CG2222	X	Female	gcgcagtgtctcaaagatga	cctgcccctctggtacataa	220
CG9915	X	Female	atagtgccattttgcgaccc	gacctaagtcaacagtcg	228
CG1749	X	Female	gtagatctcgtcctaagc	atccgccaccaactcaatcc	521
CG7931	3R	Male	ggttaggatttctgagggc	tgtagaatactcctcgtcc	391
CG7929	3R	Male	catagtacccaccgtgaagc	ttctaccatgtggtgcctcg	445
CG10252	3R	Male	caaatacgcgccaggtcc	ggtagagcctcattctatgc	215
CG14926	2L	Male	cagaacgaaagggaatcgaa	atcctcaagcgggatacctt	232
CG17956	3R	Male	cgaatcaatcatgtgttgcg	tctaacaacggaagcagc	217
CG17376	2L	Male	cggatctcaacttacatgg	ccaactctcggaccaatcg	279
CG1980	3R	Male	cacgactctggaggtagtac	cgcctgttcttcacatttcc	393
CG5565	2L	Male	attgcgatcgagaccttggc	ctctgtcatggatctcagc	569
CG18418	3L	Male	tgcaactgacctcgtactcg	gcggaagagcttgttgtagg	341
CG6980	3R	Male	tgcaggactcgaatctgcgg	cgctaatatccgctcttcatgc	297
CG6255	3R	Male	tggccatcagatagtcggc	ggacaccaaggtcatggtc	655
CG6332	3R	Male	gcgaacgattgaagaagctc	gcgaagagggcttgtacttg	321
CG3483	2R	Male	ggtcaccatgatcatgggc	cgcatcacatcgccgtaca	656
CG10307	2R	Male	gcagcgccagactaatcta	ccacagatttatgcgttctac	335
CG10750	2L	Male	gtcgaggtgaacggattcat	gtcggagaccttcaccacat	222
CG3085	2R	Male	tactctgcaggatatccg	attctccgccaggagcttcacg	550
CG8564	3L	Male	accttagtgcgagagtactc	gccctcgagatcaactggat	319
CG11475	2R	Male	ggataacggccttgctgcc	cgaagcttcatcctgtgcg	333
CG7387	3L	Male	ggtcgaactcgttgctcttc	acgcctacaacatcctcacc	249
CG18266	2L	Male	acaaagcaggaggaggagcg	ccggtcttgggaggatcaatg	184
CG9314	2L	Male	gcaggtgatgctcaacgagg	acttggagatcctggccg	679
CG6971	3R	Male	taaagctcacgccgcacagg	aggacataacaggatcgc	374
CG7251	2L	Male	agggtggatcgataggac	caccgtgagcctgatgctgcct	261
CG9531	2L	Male	gtgtcctcaataccagcg	attggaacgggcagctggtcg	222
CG6130	3R	Male	ctaataaggaaaccctggc	cacctgctcaatctttgcc	397
CG5045	2L	Male	cgcaacatcaacctgattcc	aaatccaacctgtgcgccgcc	401

CG5276	3R	Male	caggtacaacgaaaccaagg	gttgccacgggctttgtacg	568
CG15179	3R	Male	cgttggcttggccattga	agtggcagtcgcagtttgag	531
CG8277	3L	Male	cagccgtttggactgctaacg	gtacactctctgtattccg	191
CG14717	3R	Male	ccatcgtggtgatgcacg	cgatgtttgagctggtcacg	650
CG11037	3L	Male	cgtcgatgtgggaaagctgg	caacatccgtgtagacgcc	749
CG6036	3R	Male	gaaatacacatgggtggc	gtgataccagctgctcgcagc	621
CG13527	2R	Male	cgagttctgtcggtcacgg	cgtgtagacagatggcag	549
CG11379	X	Male	tcgccttacagctctgggcc	ttcgtctcggagttaggcgg	602
CG15208	X	Male	cgagatggaaagagagcgcg	gacgactgtccatcagacgg	488
CG15035	X	Male	tcgctcagcacctcgagcagc	caatctatactccgctcaccc	533
CG12684	X	Male	gccatgtccatgtcatgtcc	agcacaacacccagagttcc	597
CG2574	X	Male	ctggacggctggcgttaatgg	ctgttgctcatgagtctcgg	405
CG1314	X	Male	caagaaggagatcgacgagc	tcttgcccagacccatctcg	553
CG3708	X	Male	ctatttgagttcggagcagg	gctggatctggagagtaacg	562
CG9156	X	Male	gagaagagcgtaaccagc	ctgagggcgaaattcgaggc	711
CG12681	X	Male	catcgttcacgatcatcccg	ggctagttattgagtcagcc	535
CG12395	X	Male	cagcgtcacagcagaatcgc	ctgaagagctgcccgacgtgc	244
CG13759	X	Male	cggagtggtgcttaatggc	ccgcatccgtatacttgtcg	367
CG2577	X	Male	ggtaatggatctattggg	gtggaacttcatcatcagc	654
CG1950	X	Male	ctcgaagatgggcatcagc	cgtctgtacgagctggacgg	395
CG1503	X	Male	gggtctataaatcagggagg	gttgctgcagtcctcgagcg	278
CG6789	X	Male	gaacagagaatcaagcacgc	tgagcaacacatcgatggcc	727
CG11697	X	Male	cttctttcgttggtagtgggc	ttagccctagagtgctgtccc	576
CG5662	X	Male	ctatctgctgcagtgcgg	agcagaggctcacctcgtgc	448
CG6999	X	Male	aggcgctgaggaagaaagcg	ccttgcgagcctgtgacacttc	556
CG5334	X	Male	gtgagcattgcccataccgc	ctctgtagctttcgagtacgc	630
CG18341	X	Male	cttaccctccaccttgcagc	ccacgcagaggtcgatgttcg	425
CG10920	X	Male	ctgtgccctgtgttataccg	ggatcatgatcacgtggaagg	578
CG7860	X	Male	gcaatccccttaagacgg	ggatctactggatgaatggc	467
CG11981	3R	Unbiased	catcatctagtccatgcgg	aacggaggatgtgttgtggc	605
CG5919	3R	Unbiased	acttccgatacctcacccg	gcatattgagtatcgtgcgc	558
CG5915	3R	Unbiased	agtacatcgaaatggccagc	tgctccagggcattcttagc	603
CG9893	2R	Unbiased	cagcagtctgttaggcagg	ccagactgacatcattccc	673
CG13189	2R	Unbiased	agctctgtggtatggacagc	cagatgccaccacatcagc	404
CG10853	3L	Unbiased	tcatccgcctggatctgcg	tctccctgtgtgtgagccc	291
CG6913	3R	Unbiased	gcaagttgataaggactcgg	ccaaacagaacagacccaag	258
CG9437	2R	Unbiased	cgtcagtccagagttgctcg	gtcaatgccgacatcatcaacg	333
CG8392	2R	Unbiased	cagaaccgagtaactcttaccc	tggccgtcgaatttgatgg	483
CG7953	2L	Unbiased	gtgctatctctagttgcc	gatcacagtccacgacgatcgg	738
CG13419	3R	Unbiased	ccgtgatacatgtcctgc	ggagttatgtgactgtgacgg	435
CG7484	3L	Unbiased	attttggaggtctgcacctg	gtcggcatcctctaccacac	270

CG9283	3L	Unbiased	agtgttcagtcagagtcgcc	acggacatggacatggcgtggg	540
CG13934	3L	Unbiased	gtggtttgtaggcacttgcc	cggtcagcatggcatatgg	518
CG17404	3R	Unbiased	cagtaaaggttgacactcgg	tgcagtcgaattacgtgggc	411
CG10623	2L	Unbiased	gtaacagcctctgcctctagc	gtgagcaacagattcaggagc	541
CG3652	2L	Unbiased	tggaagaatgagctgaaggc	gggattgtacgtcatcac	318
CG9617	3R	Unbiased	tacatctgtgtctccggtcg	gagaaagacaaggcggtggc	323
CG6981	3L	Unbiased	gtttccaccactcagttccc	ctctactgcgtgtagtcgg	292
CG9822	2R	Unbiased	ctgctgtcaatcacgctatg	gatggtacagttcatgtggc	459
CG13845	3R	Unbiased	aacccctaccgatcctcacc	ctagtgcgtctcgccttcg	706
CG3683	2R	Unbiased	actcatttcagctgccgc	agctactgtggatgctgcc	453
CG3476	2L	Unbiased	gtgacctacgaattcctacagg	gctccacaccaaagaatacagc	366
CG6094	2L	Unbiased	ggcagcgacaataagttcaacg	agaagtccactactccactg	447
CG8844	2L	Unbiased	catcaagtgtatgtgccg	catcagctcaaacgaccg	244
CG7508	3R	Unbiased	gatttgatgctgttcggc	actgggctcctagttacagg	419
CG16985	3L	Unbiased	gctccagagcactgtaatcg	gtcgatagtcaagcaacg	370
CG11785	3R	Unbiased	tgaaagggagaatgctgg	aacgcgtccaataccaac	356
CG10035	3R	Unbiased	ccacttagtttacgagcc	gtcccattgtcgtccagacc	492
CG18553	3R	Unbiased	caagacccatggtgttgtgc	gatgaacagcatgacgaaggc	520
CG14629	X	Unbiased	gcaaggatgtagatgcctcagc	gacgaagaccatcaagcttcc	715
CG1885	X	Unbiased	cgtagctatcgctctctcc	cgaggagaacaagatgagc	564
CG1751	X	Unbiased	aacaagtgggacggatcgg	tccaggacgatgccgttctg	463
CG15247	X	Unbiased	tgggcgttcttctagcgtgc	acttccacgctcaactcg	605
CG14227	X	Unbiased	cttggatgctatttcgcag	gagaacctttatgtccgcc	399
CG9919	X	Unbiased	cgatctcataagaaccgg	gtccttagtagtctttggc	501
CG9538	X	Unbiased	tcaagtggaacgacgaactg	tggtagcagcgtaattgcag	372
CG2555	X	Unbiased	ctccatcagtcgagatccg	gtagttgccattggcatcgc	381
CG1397	X	Unbiased	ggactaatctaggatcgg	tgcgtcgctgctaccaatgc	390
CG9571	X	Unbiased	cagaacaacgacaagctcgacg	ggcaggattcatgcgctgg	608
CG14797	X	Unbiased	ctcttccgatcaccatccg	cgttggtggaggcggtgcta	199
CG3603	X	Unbiased	gattgccgctgatcgtaacc	cggtgacatcgatggcagc	631
CG14772	X	Unbiased	cggcaatgggaatggaagc	ctactgtgctttctcctg	376
CG11126	X	Unbiased	ctggcaggacgaactggac	ctggaggaatgggcactcg	285
CG15313	X	Unbiased	actttccacatccgctgaac	atcagtctagtcccggcaga	170
CG9723	X	Unbiased	accaaaccacgtagtccc	gcgaaagaagattccactgtcg	373
CG9164	X	Unbiased	gattactgtcaggtgaggc	atgacgatacgctggtgtcg	546
CG15336	X	Unbiased	cgtgcgtctaacaatagtgc	tgtcgcccgctgcctcttctgc	500

[a]Chromosomal location in *D. melanogaster*.

[b]Sex-biased expression category in *D. melanogaster*.

[c]Length of amplified fragment.

Appendix B

PCR and sequencing primers

Gene	Forward primer (5'-3') (F)	Reverse primer (5'-3') (R)	Internal Primer
CG10035	ggaaatcttcgacgaagacact	gtgaaatcggttttatggcgt	
CG10252	tctaatcaagtaggcctggc	aaccgcaaggcaaagtagc	
CG10750	ctatcataataattgacacca	atgcgatcaatctcatggct	
CG10853	acacctgtgcgaatcagatg	ttctccctgtgtgtgagcc	
CG10920	tgcgagtgtgagacagacatcg	caggctgctccgaattcagtc	
CG11379	tttctgtctctgccagaagc	agcggttggcattacttaaccc	
CG11697	acactgtcaacagagatgcc	agcaacaaactgccaccttgc	
CG11981	aagctgtcagttgccagac	tagtgggttcattggcggtgc	
CG12276	gaagcgttagtgctttttggc	gataagttcgggacattaaaaacg	
CG1239	tcaatggtcacctgtggttagt	taaacgcgcaaaacagcact	
CG1314	tttagtggtgtggctctaatcagt	ccgatccgtaaaaacctgagaa	ctcatcttcgacagtccg (F) ttctgctcctgctcctgc (R)
CG13189	agagctcctcagttgaaagc	acagatgccaccacatcagc	
CG14717	aaagccatgttttcttaccctt	gctgaaatttcaggaaactccc	
CG15336	agatttggcacacgcagtgc	acaccgtttggtcccactc	
CG15717	cagctatcgccgatcttcgc	aaggacctaccccaatcaccc	
CG1749	cgacggcagagaaattgaaa	cgaaaatagagcattcatttacag	
CG18266	tctcaaacacatcaggtcgc	actacctaggcacttcaaccg	
CG18341	acacattcaaaaccttctgc	tgtgcggtaaattagcatgc	
CG18418	acggttaacatagggactttgct	cattttgcatatacatccgg	
CG2222	tctataaccggtatttgaaccgg	tgatggacatggattgtt	
CG2577	ttagttccttgcgcccggttcg	gaatcgagaggccttctggg	
CG3004	tgaatgcagctggtcacact	ggttaccctataaagaccataacg	
CG3024	tcgcgattaggtcacactagc	taaccaaggatctgggaccg	ctacgtatccaagtacctcgg (F)
CG3085	cagttgggaggacaaattcagt	cgagggggattacatactttta	
CG3476	aggttggcaggtcttaaaccg	gcataattctggttgattcatgg	
CG3509	atcatccgccggtgcagtagtg	tcagaccaagcaactgccctggc	agaggcggcgaaggataagg (F) tagtgacctttgccctctc (R)

CG4593	tcgataactgtcagctggaaagc	aattgggaggacagctgagagg	
CG4973	ttaccacctctagcagtcgg	tttaggtagtaactgttgg	
CG5272	atgcctggagccactatatg	tcagatcgctctactttaacc	cacaatagaaccagtggcagca (F)
			agggctcatagtggactcct (R)
CG5499	aatggctccttctctgacgg	tgcgacagaatgacgttgcc	
CG5915	agtccagtcgtgttatctgc	agaggccattaaggagtctgc	
CG6036	tctagatctaattggcttcgg	ataacgcacagctgctgcc	
	gaggatcgacggcaatatgttg	aacccgctttatcatcacactgc	
CG6459	taagcgacaaccctagttgc	ttgaagctcctcctattccg	
CG6971	cagagccagaaactcatttgtgt	gccaagatggaggagcttaag	atctccaggagatgctggacac (F)
			gacatggcgatctcatcgcggat (R)
CG6980	cttcggttgctatagcatcc	acagatttgggcagtgttcacc	
CG6981	tgatatagtgctatccagtg	ttgtttattcggcgattgc	
CG7387	caatcggacaagttgtaggc	gaacaacaagatcctccagc	
CG7508	tgcctgattgcctgccatcg	ctactgggctcctagttacagg	
CG7840	tcgacattcaccgcgacaagc	atcgactgcagcctgatcc	
CG8277	tgtaagcaattcccactcgg	cgaagccaatgtcatcttcc	
CG9135	agccgcttgtgcgttacgctttgc	agcagcagatagccactttcggg	tggttcggtaacccacttggagc (R)
CG9383	acaaggcggagccataaacag	agggtacgaggaggactatg	
CG9723	tatgttccaaacaaaccgctg	tgcgtatttggcatttctgc	

Appendix C

Comparison *D. ananassae*/*D. melanogaster*

		D. ananassae									D. melanogaster							
Gene	Bias[a]	D_S[b]	P_S[c]	D_N[d]	P_N[e]	P^f	α^g	TD_{syn}[h]	TD_{non}[i]	Out.	D_S[b]	P_S[c]	D_N[d]	P_N[e]	P^f	α^g	TD_{syn}[h]	TD_{non}[i]
Autosomal																		
CG10035	FU	32	11	25	3	0.112	0.651	0.235	-1.642	Dph	25	5	9	4	0.31	-1.222	-0.99	0.5
CG10252	MM	28	13	6	3	0.925	-0.077	0.164	-0.804	Dat	19	3	0	0	na	na	-0.11	na
CG11981	FU	19	8	0	0	na	na	0.438	na	Dat	15	1	0	0	na	na	0.33	na
CG12276	FF	43	10	17	7	0.321	-0.771	0.451	0.066	Dat	41	3	8	5	0.0081	-7.542	0.02	-0.91
CG1239	UF	45	6	22	3	0.976	-0.023	-0.447	-1.187	Dat	23	5	18	2	0.4386	0.489	-0.99	-0.03
CG14717	UM	43	28	40	5	0.001	0.808	0.124	-1.014	Dph	23	11	7	6	0.38	-0.792	-0.18	-1.43
CG3509	FF	59	15	33	3	0.095	0.642	0.131	0.278	Dat	26	13	35	6	0.048	0.657	0.36	-0.24
CG4973	UF	13	5	4	8	0.034	-4.2	-0.325	0.354	Dat	43	8	9	9	0.0053	-4.375	-0.02	-1.81
CG5499	FF						na			Dph	7	3	0	3	0.0699	na	0.2	-1.45
CG5915	MU	13	4	0	2	0.088	na	0.585	-1.461	Dph	9	3	1	0	1	1	1.19	na
CG6036	MM	53	9	11	0	0.338	1	0.376	na	Dat	26	21	11	5	0.3406	0.437	0.32	-1.7
CG6971	MM	26	3	1	0	1	1	-1.111	na	Dph	26	7	0	0	na	na	-1.33	na
CG6980	MM	24	10	16	1	0.036	0.85	0.618	1.202	Dat	16	5	13	4	0.9839	0.015	0.43	0.18
CG7508	UU	12	33	3	7	0.832	0.152	-0.679	-1.118	Dat	23	4	9	3	0.4543	-0.917	-0.37	-0.42
CG10853	UU	9	4	3	0	0.529	1	0.998	na	Dat	6	7	2	4	0.6	-0.714	0	-0.78
		12	3	6	0	0.526	1			Dph								
CG18418	MM	33	9	11	4	0.682	-0.333	0.151	-0.893	Dat	27	33	13	5	0.0397	0.685	-0.01	0.25
		33	9	12	4	0.772	-0.222			Dph								
CG5272	FF	25	4	21	6	0.41	-0.786	-1.387	-0.505	Dph	17	4	14	3	0.911	0.089	-0.9	-0.42
CG6981	UU	17	8	1	0	1	1	0.139	na	Dat	13	4	0	0	na	na	0.1	na
CG7387	FM	49	16	16	5	0.94	0.043	-0.737	-0.908	Dat	45	12	18	6	0.7721	-0.25	-0.69	-1.07
CG8277	MM	23	12	8	10	0.138	-1.396	-1.45	-1.223	Dat	14	5	24	3	0.18	0.65	0.12	-0.38
		17	12	5	10	0.109	-1.833			Dph								
CG9383	UF	35	17	8	1	0.259	0.743	-0.424	0.409	Dph	14	1	2	0	1	1	-0.09	na
CG13189	UU	42	36	0	3	0.107	na	0.279	-1.627	Dph	26	22	3	0	0.2494	1	-0.22	na
CG3085	MM	44	33	3	1	0.468	0.556	-0.297	-1.122	Dph	25	41	5	1	0.0281	0.878	-0.45	-0.07
CG6459	FF	31	11	5	3	0.724	-0.691	-0.268	-0.51	Dat	21	7	5	3	0.465	-0.8	0.93	-1.75
		28	14	11	3	0.391	0.455			Dph								

CG10750	UM	28	17	10	1	0.047	0.835	0.518	-0.129	Dat	21	20	10	0	0.0039	1	-0.33	na
CG18266	MM	53	23	47	10	0.089	0.51	-1.073	0	Dph	35	11	55	9	0.1893	0.479	-0.83	-0.69
CG3476	MU	31	1	6	4	0.004	-19.667	-0.032	-0.162	Dph	17	16	0	6	0.0267	na	-0.99	-1.56
CG7840	FF	32	25	4	3	0.96	0.04	-0.116	-1.56	Dat	18	17	12	4	0.1057	0.647	-0.04	0.42
		31	25	2	3	0.653	-0.86			Dph								
CG9135	FF	51	27	2	3	0.264	-1.833	0.784	-1.55	Dat	28	19	3	0	0.2788	1	-0.78	na
		52	27	3	3	0.444	-0.926			Dph								
X-linked																		
CG10920	MM	61	18	23	3	0.193	0.558	0.518	-1.536	Dat	43	24	56	11	0.099	0.648	-0.06	-0.27
CG11697	MM	0	19	1	3	0.174	1	0.041	1.94	Dat	28	1	17	5	0.0733	-7.235	-1.13	-0.32
CG15336	UU	19	6	11	2	0.529	0.424	-0.032	-1.449	Dat	10	17	10	13	0.64	0.235	-1.47	-0.68
		21	6	8	2	0.883	0.125			Dph								
CG15717	FF	28	23	20	8	0.146	0.513	-0.816	-0.543	Dat	12	7	16	8	0.8107	0.143	-0.9	0.12
		25	23	9	8	0.951	0.034			Dph								
CG1749	UF	35	7	9	0	0.328	1	-0.48	na	Dph	26	5	4	9	0.0007	-10.7	-0.94	-0.62
CG18341	UM	1	9	2	4	0.254	0.778	-1.417	0.925	Dat	26	22	18	2	0.0026	0.869	-0.08	-1.42
CG2222	FF	21	12	4	0	0.282	1	0.437	na	Dph	10	11	2	4	0.531	-0.818	-0.31	-1.79
CG2577	MM	39	22	12	2	0.095	0.705	-0.091	-1.436	Dat	32	26	4	0	0.132	1	-0.98	na
		44	22	12	2	0.135	0.667			Dph								
CG3004	UF	26	6	6	10	0.003	-6.222	-0.307	-0.018	Dat	23	18	7	0	0.036	1	-0.43	na
		26	6	9	11	0.007	-4.296			Dph								
CG3024	MF	54	7	37	7	0.512	-0.459	-1.19	-1.189	Dat	35	5	51	6	0.7637	0.176	-0.68	-0.58
		63	7	44	7	0.529	-0.432			Dph								
CG4593	MF	23	7	2	1	0.709	-0.643	-0.692	0.543	Dat	12	19	3	0	0.076	1	0.48	na
		17	7	4	1	0.668	0.393			Dph								
CG9723	UU	64	10	16	7	0.076	-1.8	-0.588	-1.305	Dph	36	18	61	5	0.0003	0.836	-0.46	-1.01
CG11379	MM	21	8	12	1	0.119	0.781	0.028	-1.133	Dat	17	7	5	4	0.413	-0.943	-1.07	-1.39
		23	8	7	1	0.401	0.589			Dph								
CG1314	MM	43	24	13	5	0.518	0.311	-0.194	0.286	Dph	33	12	80	4	0.0004	0.863	-1.3	-0.82

[a]First letter indicates expression in *D. ananassae*, second letter expression in *D. melanogaster* ("M"=male-biased, "F"=female-biased, "U"=unbiased).

[b]The total number of synonymous fixed differences.

[c]The total number of nonsynonymous fixed differences.

[d]The total number of synonymous polymorphisms.

[e]The total number of nonsynonymous polymorphisms.

[f]*P*-value of McDonald-Kreitman test. Red indicates a relative excess of nonsynonymous divergence (positive selection), blue indicates a relative excess of nonsynonymous polymorphism.

[g]$\alpha = 1 - [(D_S * P_N)/(D_N * P_S)]$.

[h]Tajima's D at synonymous sites.

[i]Tajima's D at nonsynonymous sites.

Appendix D

Sex-biased gene expression in *D. ananassae* compared to *D. melanogaster* and *D. pseudoobscura*

Gene	Bias *D. mel*	M/F *D.mel*[a]	Bias *D. ana*	M/F *D. ana*	Bias *D. pse*	M/F *D.pse*[e]
CG10035	U	1.06	F	0.88d	U	1
CG10206	F	0.18	F	0.37c	/	/
CG10252	M	35.16	M	15.14d	M	8.69
CG10623	U	0.95	M	1.29d	U	0.97
CG10750	M	12.73	U	0.91c	M	2.71
CG10853	U	0.99	U	1.05b	/	/
CG10920	M	4.91	M	4.82d	M	2.5
CG11126	U	1.01	U	1.20b	/	/
CG11130	F	0.4	U	0.96b	F	0.61
CG11379	M	3.04	M	1.77d	M	1.98
CG11697	M	4.81	M	1.94b	/	/
CG11785	U	1.07	U	0.90b	/	/
CG11981	U	1.03	F	0.72d	U	1.02
CG12117	F	0.38	U	0.83b	F	0.8
CG12276	F	0.27	F	0.60b	U	0.99
CG1239	F	0.34	U	0.91b	F	0.68
CG12395	M	7.11	M	1.31d	M	2.65
CG12681	M	12.14	M	2.53b	/	/
CG12684	M	6.17	F	0.43c	/	/
CG12909	F	0.26	U	1.13d	F	0.58
CG1314	M	8.95	M	1.67b	M	1.56
CG13189	U	1.03	U	1.18d	U	1.01
CG13419	U	0.98	U	1.12b	U	1.04
CG13690	F	0.18	M	1.60d	F	0.67
CG13845	U	1.02	U	0.98b	/	/
CG13934	U	0.92	F	0.84d	U	1
CG1397	U	0.97	U	1.04b	U	1
CG14227	U	0.78	M	2.65d	/	/

CG14434	F	0.24	U	0.88b	U	0.96
CG14629	U	1.09	M	1.47d	U	1.02
CG14717	M	16.29	U	1.45b	/	/
CG14797	U	1.05	U	0.99b	U	1
CG14926	M	25.19	M	3.74b	M	4.74
CG1503	M	6.16	U	1.24b	M	1.49
CG15035	M	10.1	U	1.17d	/	/
CG15179	M	14.09	M	2.13d	M	5.9
CG15208	M	15.34	M	2.71b	M	9.22
CG15247	U	1.01	U	1.01b	U	1.01
CG15313	U	0.99	M	5.21c	/	/
CG15336	U	1.08	U	0.76d	U	1.01
CG15717	F	0.28	F	0.69d	F	0.62
CG16985	U	1.1	F	0.69b	U	1
CG17361	F	0.32	U	0.95b	/	/
CG17376	M	12.5	M	10.82d	/	/
CG17404	U	0.93	U	1.07b	U	0.99
CG1749	F	0.42	U	0.97d	U	0.98
CG1751	U	1.03	U	1.12d	M	1.34
CG18266	M	13.68	M	1.94d	M	4.81
CG18341	M	4.66	U	1.14b	M	2.52
CG18418	M	16.06	M	2.21b	/	/
CG18553	U	1.06	U	1.06b	U	1.02
CG1885	U	0.98	M	2.49d	U	0.98
CG1950	M	6.68	U	1.24b	/	/
CG2222	F	0.4	F	0.74d	U	0.94
CG2555	U	1	M	1.94d	U	0.99
CG2574	M	7.34	M	1.69b	U	1.03
CG2577	M	8.26	M	1.48d	M	1.71
CG2867	F	0.3	U	0.75d	F	0.67
CG3004	F	0.37	U	0.82b	F	0.67
CG3024	F	0.35	M	3.43d	F	0.63
CG3085	M	19.81	M	2.20d	M	5.2
CG32409	F	0.26	U	0.70b	F	0.5
CG3476	U	0.98	M	24.59d	U	0.97
CG3481	Control (Adh)	-	F	0.69b	M	2.09
CG3483	M	10.77	M	2.55b	F	0.16
CG3509	F	0.13	F	0.49b	U	0.97
CG3603	U	0.98	U	0.68b	U	1.02
CG3652	U	0.98	U	0.98b	F	0.26
CG3661	Control (RpL23)	0.81	F	0.55d	U	0.99

CG3683	U	1.04	U	0.74b	U	0.97
CG3704	F	0.33	U	1.04b	M	1.41
CG3708	M	5.52	M	1.52b	F	0.43
CG3831	F	0.32	F	0.65d	U	0.96
CG3975	F	0.37	U	0.85b	U	1
CG4236	F	0.32	U	0.71b	F	0.36
CG4299	F	0.24	F	0.46b	F	0.66
CG4570	F	0.1	U	0.89b	F	0.58
CG4593	F	0.37	M	4.38d	U	0.96
CG4973	F	0.32	U	0.95d	M	2.77
CG5045	M	19.81	M	1.58d	F	0.32
CG5272	F	0.08	F	0.42d	M	1.53
CG5276	M	5.67	U	1.19b	/	/
CG5334	M	3.93	M	1.29c	F	0.52
CG5363	F	0.21	F	0.58b	F	0.36
CG5499	F	0.24	F	0.50d	/	/
CG5565	M	20.83	M	4.39d	/	/
CG5662	M	4.06	U	1.08b	F	0.69
CG5757	F	0.26	F	0.63b	/	/
CG5915	U	1	M	1.90d	U	1
CG5919	U	0.98	U	1.21d	M	4.58
CG6036	M	13.66	M	2.19d	F	0.75
CG6094	U	1	U	0.92b	M	2.88
CG6130	M	8.22	M	1.95b	U	1.06
CG6255	M	17.96	U	1.40d	M	6.08
CG6332	M	23.88	M	4.61d	F	0.47
CG6459	F	0.27	F	0.52b	F	0.38
CG6554	F	0.24	U	1.11b	U	1.04
CG6789	M	7.4	M	1.88d	U	1.02
CG6913	U	1	U	1.16b	M	2.11
CG6971	M	9.72	M	1.89b	M	2.43
CG6980	M	8.08	M	2.96d	U	0.99
CG6981	U	1.03	U	1.06d	/	/
CG7251	M	11.14	U	1.36b	M	1.78
CG7387	M	19.63	F/M	0.95d	U	1
CG7484	U	0.95	F	0.63c	U	1
CG7508	U	1.11	U	0.83d	U	0.96
CG7840	F	0.18	F	0.44b	U	1.01
CG7860	M	3.11	U	1.08b	U	0.97
CG7953	U	1.07	U	0.79b	M	3.35
CG8277	M	19.91	M	4.78d	U	0.95

CG8326	F	0.34	U	0.82b	U	1.02
CG8392	U	0.95	U	0.88b	M	4.27
CG8564	M	35.21	M	2.35d	F	0.72
CG8675	F	0.42	U	0.93b	U	0.98
CG8844	U	1.04	M	1.23d	F	0.63
CG9125	F	0.31	U	0.89b	F	0.31
CG9135	F	0.16	F	0.35c	U	1.02
CG9164	U	1.02	U	1.01b	F	0.5
CG9273	F	0.22	U	0.74b	U	1
CG9283	U	0.98	U	1.12b	M	2.99
CG9314	M	24.06	M	2.64b	/	/
CG9383	F	0.28	U	0.98d	U	0.99
CG9437	U	0.95	U	1.05b	M	1.84
CG9531	M	3.48	M	1.74b	F	0.81
CG9538	U	0.82	F	0.86d	/	/
CG9617	U	0.98	U	1.08b	/	/
CG9723	U	0.95	U	0.86b	U	0.98
CG9893	U	0.95	U	0.96b	F	0.59
CG9915	F	0.43	U	0.83d	U	0.99
CG9919	U	1.02	U	1.00b	U	1

[a]Average male/female expression ratio from Parisi *et al.* (2003), Ranz *et al.* (2003), and Gibson *et al.* (2004).

[b]Male/female expression ratio from Zhang *et al.* (2007).

[c]Male/female expression ratio from PCR-amplicon microarrays (this study).

[d]Average value from Zhang *et al.* (2007) and PCR-amplicon microarrays.

[e]Male/female expression ratio from Zhang *et al.* (2007) ("/" indicates genes with either no ortholog in *D. pseudoobscura* or for which expression data are not available).

Appendix E

Inference of ancestral sex-biased expression state

Gene	Dana	Dmel	Dpse
CG10035	F	U	U
CG11981	F	U	U
CG13934	F	U	U
CG16985	F	U	U
CG7484	F	U	U
CG9538	F	U	F
CG10623	M	U	U
CG14629	M	U	U
CG1885	M	U	U
CG2555	M	U	U
CG3476	M	U	U
CG8844	M	U	U
CG7387	F/M	M	M
CG13690	M	F	F
CG3024	M	F	F
CG4593	M	F	F
CG11130	U	F	F
CG12117	U	F	F
CG1239	U	F	F
CG12909	U	F	F
CG14434	U	F	U
CG17361	U	F	U
CG1749	U	F	U
CG2867	U	F	F
CG3004	U	F	F
CG32409	U	F	F
CG3704	U	F	U
CG3975	U	F	U
CG4236	U	F	U

CG4570	U	F	F
CG4973	U	F	U
CG6554	U	F	F
CG8326	U	F	U
CG8675	U	F	F
CG9125	U	F	F
CG9273	U	F	F
CG9383	U	F	U
CG9915	U	F	F
CG10750	U	M	M
CG1503	U	M	M
CG15035	U	M	U
CG18341	U	M	M
CG5276	U	M	M
CG5662	U	M	U
CG6255	U	M	U
CG7251	U	M	U
CG7860	U	M	U

"M" indicates male-biased, "F" indicates female-biased, and "U" indicates unbiased expression.

Appendix F

DNA sequence polymorphism at seven loci from twelve populations

Gene	Length	Loc	Param	BBS	BKK	BOG	MAN	CH	KATH	KK	KL	KMJ	PURI	TB	TI
CG14717	894	3R	N	9	12	6	8	9	6	8	9	9	15	6	8
			Seg. sites	2	30	25	19	16	2	27	27	26	27	25	24
			Hap	3	9	6	7	4	2	6	6	7	7	4	7
			π	0.0009	0.0122	0.0117	0.0085	0.0072	0.0008	0.0114	0.0102	0.0109	0.0093	0.0098	0.0111
			θ_W	0.0008	0.0119	0.0123	0.0082	0.0066	0.001	0.0117	0.0111	0.0107	0.0096	0.0123	0.0108
			Fu D	1.0173	-0.7425	-0.5894	0.4381	1.7624	0.8827	0.1089	0.1033	0.5955	0.096	-0.0765	-0.3404
			Taj. D	0.1959	0.088	-0.3158	0.1473	0.4591	-1.132	-0.1733	-0.4417	0.0523	-0.1585	-1.2747	0.1175
CG6980	843	3R	N	9	11	6	8	9	6	8	9	7	16	7	9
			Seg. sites	22	16	19	18	20	5	21	17	16	40	33	38
			Hap	7	10	6	6	9	3	8	5	6	14	5	7
			π	0.0125	0.0073	0.0108	0.0082	0.0098	0.0029	0.0109	0.0085	0.0072	0.0147	0.0189	0.0143
			θ_W	0.0101	0.0069	0.0099	0.0082	0.0092	0.0026	0.0105	0.0074	0.0078	0.0147	0.0165	0.0166
			Fu D	1.22	0.4727	0.7935	0.1566	0.6597	1.4945	0.5582	1.4945	0.1523	-0.1037	1.6937	0.1858
			Taj. D	1.1768	0.2494	0.5086	-0.09	0.3163	0.7077	0.1187	0.6664	-0.4509	-0.0332	0.758	-0.7572
CG10750	1,171	2L	N	9	12	6	8	9	6	8	8	8	10	4	8
			Seg. sites	44	50	19	44	40	26	43	19	14	37	29	41
			Hap	9	12	6	7	9	5	8	6	6	8	4	7
			π	0.0125	0.0118	0.0064	0.0168	0.012	0.0087	0.017	0.0063	0.0048	0.0133	0.0147	0.0135
			θ_W	0.0148	0.0122	0.0075	0.0152	0.0129	0.0097	0.0148	0.0066	0.005	0.0121	0.0149	0.0148
			Fu D	-0.9674	-0.2964	-0.8789	0.8684	-0.4282	-0.0948	0.7409	0.3394	0.0049	1.5315	-0.1651	-0.3616
			Taj. D	-0.8303	-0.7257	-0.9694	0.5216	-0.3835	-0.6944	0.6963	-0.2646	-0.2083	0.4481	-0.2695	-0.5185
CG10853	708	3L	N	8	10	6	8	9	6	8	9	9	15	6	8
			Seg. sites	8	6	11	8	8	6	7	8	8	16	18	24
			Hap	3	8	5	7	3	2	6	6	5	8	4	7
			π	0.0044	0.0036	0.0062	0.0046	0.0027	0.0024	0.0037	0.0039	0.0049	0.0055	0.0097	0.0136
			θ_W	0.0044	0.003	0.0062	0.0049	0.0037	0.0031	0.0038	0.0047	0.0047	0.0075	0.012	0.0153
			Fu D	0.9837	-0.3183	-0.8346	0.2244	-1.4366	-1.9963	0.4049	-0.8255	-0.5942	-0.3806	-1.3874	-0.028
			Taj. D	-0.0219	0.4432	-0.1894	-0.3624	-1.2835	-1.3673	-0.2261	-0.7803	-0.0161	-1.0669	-1.1896	-0.2819
CG6981	824	3L	N	9	12	6	8	9	6	8	9	9	16	5	7
			Seg. sites	18	24	23	20	17	7	30	20	22	28	20	36
			Hap	5	12	6	7	8	3	7	6	8	13	5	7

			π	0.0063	0.0077	0.011	0.0072	0.007	0.002	0.0119	0.0079	0.0107	0.0075	0.0102	0.0151
			θ_W	0.006	0.0078	0.0114	0.0075	0.0072	0.0017	0.0126	0.0086	0.0091	0.0094	0.0113	0.0175
			Fu D	-0.0789	-0.0115	-0.4097	0.2372	-0.3972	0.3197	-0.5651	-0.1553	0.8829	-1.1673	-1.1713	-0.7977
			Taj. D	-0.3776	-0.462	-0.2949	-0.6503	-0.2417	0.5081	-0.5067	-0.3882	0.7376	-1.0969	-0.6161	-0.9651
CG7508	933	3R	N	9	10	6	8	9	6	8	9	8	15	6	9
			Seg. sites	35	45	30	25	23	23	26	32	24	38	26	37
			Hap	8	9	6	6	8	4	6	8	6	13	5	9
			π	0.0179	0.0164	0.0158	0.0142	0.0096	0.0107	0.011	0.0132	0.0117	0.0122	0.0168	0.0183
			θ_W	0.0176	0.0189	0.0152	0.0138	0.0094	0.0119	0.0126	0.0136	0.0118	0.0137	0.0174	0.0191
			Fu D	0.423	-0.3587	0.1142	0.3227	-0.2103	-0.4823	-0.5794	0.0067	-0.1882	-0.4522	-0.1646	-0.5388
			Taj. D	0.1674	-0.7948	0.1745	0.114	0.087	-0.6944	-0.6867	-0.1718	-0.0589	-0.5211	-0.246	-0.3065
CG13189	1,020	2R	N	8	12	6	8	9	6	7	7	8	8	5	8
			Seg. sites	25	36	24	27	26	17	20	26	21	29	27	37
			Hap	5	11	6	8	8	5	7	5	4	7	5	7
			π	0.0096	0.013	0.0119	0.0126	0.0114	0.0062	0.0094	0.0098	0.0108	0.0118	0.0145	0.0139
			θ_W	0.0098	0.0127	0.0107	0.011	0.0097	0.0073	0.008	0.0108	0.0079	0.0113	0.0127	0.014
			Fu D	-0.028	0.4354	0.8202	0.7794	0.8239	-0.4205	0.8002	0.1597	1.8597	-0.2968	1.0415	0.2691
			Taj. D	-0.1459	0.0835	0.6048	0.6048	0.787	-0.9819	0.9358	-0.5509	1.8464	0.1805	0.9437	-0.101

Length = total length of the fragment in nucleotide bases.
Loc = chromosomal location.
Param = different parameters.
N = number of individuals analyzed.
Seg. sites = number of segregating sites.
Hap = number of haplotypes.
π and θ_W = two different measures of nucleotide diversity.
Fu D = Fu and Li's D.
Taj D = Tajima's D.

Appendix G

Genetic differentiation between populations of *D. ananassae*

Population 1	Population 2	*CG10853*	*CG13189*	*CG6981*	*CG7508*	*CG14717*	*CG6980*	*CG10750*
BBS	BKK	0.30	0.04	0.00	0.00	0.32	0.07	0.05
BBS	BOG	0.30	0.00	0.00	0.00	0.16	0.09	0.16
BBS	BURMA	0.31	0.02	0.05	0.00	0.27	0.20	0.11
BBS	CH	0.00	0.06	0.05	0.09	0.14	0.00	0.09
BBS	KATH	0.00	0.29	0.15	0.05	0.07	0.49	0.19
BBS	KK	0.21	0.04	0.01	0.03	0.14	0.00	0.05
BBS	KL	0.12	0.00	0.00	0.00	0.12	0.11	0.10
BBS	KMJ	0.21	0.01	0.12	0.06	0.27	0.12	0.24
BBS	PURI	0.00	0.02	0.03	0.02	0.18	0.00	0.09
BBS	TB	0.03	0.09	0.00	0.03	0.01	0.01	0.00
BBS	TI	0.26	0.00	0.11	0.10	0.26	0.05	0.04
BKK	BOG	0.00	0.00	0.00	0.02	0.00	0.16	0.02
BKK	BURMA	0.00	0.02	0.00	0.04	0.00	0.17	0.05
BKK	CH	0.27	0.08	0.00	0.08	0.02	0.20	0.06
BKK	KATH	0.33	0.15	0.16	0.00	0.31	0.53	0.06
BKK	KK	0.10	0.04	0.00	0.02	0.00	0.00	0.00
BKK	KL	0.18	0.07	0.02	0.03	0.04	0.42	0.04
BKK	KMJ	0.00	0.06	0.07	0.07	0.00	0.20	0.11
BKK	PURI	0.13	0.06	0.00	0.03	0.01	0.08	0.00
BKK	TB	0.11	0.07	0.02	0.02	0.13	0.10	0.01
BKK	TI	0.22	0.02	0.13	0.09	0.18	0.18	0.08
BOG	BURMA	0.00	0.00	0.00	0.09	0.00	0.14	0.14
BOG	CH	0.27	0.03	0.00	0.02	0.00	0.07	0.15
BOG	KATH	0.33	0.20	0.12	0.05	0.16	0.49	0.13
BOG	KK	0.05	0.10	0.00	0.01	0.00	0.08	0.12
BOG	KL	0.15	0.00	0.00	0.00	0.00	0.34	0.10
BOG	KMJ	0.00	0.01	0.02	0.04	0.00	0.30	0.15
BOG	PURI	0.17	0.03	0.00	0.00	0.00	0.05	0.04
BOG	TB	0.13	0.09	0.01	0.00	0.02	0.15	0.09
BOG	TI	0.19	0.00	0.09	0.17	0.10	0.16	0.17
BURMA	CH	0.30	0.00	0.00	0.18	0.00	0.27	0.00
BURMA	KATH	0.35	0.21	0.21	0.13	0.27	0.48	0.07
BURMA	KK	0.12	0.05	0.03	0.12	0.00	0.10	0.00
BURMA	KL	0.18	0.00	0.03	0.09	0.02	0.44	0.16
BURMA	KMJ	0.00	0.00	0.05	0.11	0.00	0.39	0.22
BURMA	PURI	0.13	0.00	0.00	0.12	0.02	0.14	0.00
BURMA	TB	0.10	0.10	0.02	0.10	0.09	0.22	0.04

BURMA	TI	0.22	0.01	0.12	0.15	0.17	0.27	0.05
CH	KATH	0.00	0.26	0.18	0.11	0.11	0.57	0.05
CH	KK	0.16	0.07	0.00	0.04	0.00	0.08	0.05
CH	KL	0.03	0.12	0.05	0.04	0.00	0.18	0.09
CH	KMJ	0.18	0.00	0.08	0.01	0.00	0.22	0.16
CH	PURI	0.00	0.00	0.00	0.00	0.00	0.01	0.00
CH	TB	0.02	0.14	0.00	0.00	0.00	0.06	0.00
CH	TI	0.25	0.02	0.11	0.24	0.13	0.11	0.03
KATH	KK	0.24	0.26	0.14	0.06	0.14	0.49	0.12
KATH	KL	0.15	0.30	0.29	0.07	0.11	0.67	0.00
KATH	KMJ	0.23	0.28	0.22	0.07	0.25	0.69	0.04
KATH	PURI	0.00	0.25	0.22	0.06	0.15	0.41	0.01
KATH	TB	0.02	0.23	0.11	0.02	0.00	0.43	0.10
KATH	TI	0.28	0.26	0.22	0.16	0.24	0.50	0.20
KK	KL	0.02	0.10	0.03	0.00	0.00	0.27	0.15
KK	KMJ	0.09	0.03	0.06	0.02	0.00	0.06	0.19
KK	PURI	0.09	0.05	0.00	0.00	0.00	0.00	0.00
KK	TB	0.07	0.09	0.00	0.00	0.00	0.06	0.00
KK	TI	0.22	0.01	0.08	0.21	0.09	0.14	0.04
KL	KMJ	0.11	0.00	0.10	0.01	0.01	0.44	0.00
KL	PURI	0.04	0.03	0.04	0.00	0.00	0.21	0.08
KL	TB	0.03	0.17	0.03	0.00	0.00	0.19	0.04
KL	TI	0.21	0.03	0.11	0.16	0.09	0.16	0.18
KMJ	PURI	0.06	0.00	0.08	0.00	0.00	0.12	0.12
KMJ	TB	0.03	0.11	0.00	0.00	0.07	0.13	0.12
KMJ	TI	0.20	0.00	0.11	0.21	0.13	0.29	0.23
PURI	TB	0.00	0.11	0.00	0.00	0.02	0.04	0.00
PURI	TI	0.21	0.00	0.11	0.19	0.09	0.09	0.05
TB	TI	0.09	0.00	0.03	0.16	0.04	0.06	0.00

Genetic differentiation as determined by F_{ST} (Hudson *et al.*, 1992).

Appendix H

Protocols

DNA Extraction

Phenol-Chloroform small-scale preparation of *Drosophila* genomic DNA

1. Put 1 live, knocked-out fly in a 1.5ml tube and keep on ice.
2. Add 50µl cold Hydration Buffer (see below) and squish flies completely.
3. Add 95µl Hydration Buffer and 5µl Proteinase K stock solution (20mg/ml in hydration buffer).
4. Add 10µl 10% sarcosyl and mix by tapping the tube.
5. Incubate for 1-2 hours at 37°C.
6. Add 80µl phenol and 80µl chloroform, mix well.
7. Centrifuge at 14,000 rpm for 2 min.
8. Transfer aqueous top layer to a fresh tube.
9. Add 100µl chloroform, mix well.
10. Centrifuge at 14,000 rpm for 2 min.
11. Transfer aqueous top layer to a fresh tube.
12. Add 16µl 3M sodium acetate, pH 5.2, and 350µl 100% ethanol, mix well, incubate at room temperature for 5 min.
13. Centrifuge at 14,000 rpm for 15 min, carefully remove all ethanol.
14. Air dry the pellet at room temperature until it is completely dry.
15. Resuspend pellet in 50µl of distilled water.

Hydration buffer
0.1M NaCl
0.2M Sucrose
10mM EDTA, pH 8.0
30mM Tris-Cl, pH 8.0
0.5% Triton X-100

Protocol for Purgene DNA Isolation kit for one fly (Gentra Systems, Minneapolis, Michigan, USA)

1. Put 1 fly to the 50µl chilled Cell-Lysis-Solution in 1.5ml tube on ice and squish fly completely.
2. Add an additional 49.5µl Cell-Lysis-Solution and 0.5µl Proteinase K stock solution.
3. Homogenize thoroughly.
4. Incubate 1 hour at 55°C, then 10 minutes at 65°C.
5. Add 1.5µl RNase solution A (2mg/ml) to the tube
6. Mix the sample by inverting the tube and incubate 15 min at 37°C.
7. Cool sample at room temperature.
8. Add 33µl Protein-Precipitation-Solution and vortex.
9. Keep the tube 5 min on ice.
10. Centrifuge at 13,200 rpm for 5 min.
11. Transfer supernatant to fresh tube.
12. Add 150µl of 100% isopropanol and mix the sample by inverting gently.
13. Incubate 5 min at room temperature.
14. Centrifuge at 13,200 rpm for 5 min, remove supernatant.
15. Wash the pellet with 150µl of 70% ethanol.
16. Centrifuge 2min at 12,000 rpm, carefully remove the ethanol.
17. Dry the pellet at room temperature, resuspend the pellet in 50µl of distilled water.

Polymerase Chain Reaction (PCR)

25µl reaction mix 1x

Ingredient	Amount (µl)
ddH_2O	19.625
10x PCR buffer	2.5
dNTP (10mM)	0.625
Forward primer (10µM)	0.5
Reverse primer (10µM)	0.5
DNA	1.00
Taq Polymerase	0.25

PCR standard cycling conditions

1. Incubate at 94°C for 00:02:00
2. Incubate at 94°C for 00:00:45
3. Incubate at * °C for 00:00:45
4. Incubate at 72°C for **
5. Cycle to step 2 for 34 more times
6. Incubate at 72°C for 00:07:00
7. Incubate at 12°C forever

* Proper annealing temperature
** Proper extension time (ca. 1min/1kB)

PCR products are verified on 1% agarose gel

Purification

PCR product	20µl
ddH_2O	3.5µl
ExoSAP-IT (USB, Cleavland, OH, USA)	1µl
10x PCR buffer	0.5µl

1. Incubate at 37°C for 00:30:00
2. Incubate at 80°C for 00:15:00
3. Incubate at 12°C forever

DNA Sequencing

Setting up sequencing reaction

Big Dye sequencing mix	2.0 µl
5x Sequencing Buffer	1.0 µl
Sequencing Primer	1.0 µl
DNA Template	2-5 µl
ddH_2O	optional
Total	10.0 µl

Sequencing cycling conditions

1. Incubate at 96°C for 00:01:00
2. Incubate at 96°C for 00:00:10
3. Incubate at 50°C for 00:00:15
4. Incubate at 60°C for 00:04:00
5. Cycle to step 2 for 34 more times
6. Incubate at 12 C forever
7. Add 10µl of distilled water to each sample

Microarrays

PCR plate clean-up

Cleaning of PCR plates for microarray printing using Genetix GenPure Kit

1. Mix PCR product in Wash Plate with 5fold volume Binding Buffer, wait for at least 5 min.
2. Transfer to Filter Plate, put Wash Plate underneath.
3. Spin 2 min at 1,250g, discard buffer.
4. Spin again for 5 min.
5. Dry plate by putting on heatblock at 37°C for about 30 min.
6. Elute with 30µl RNase-free water (prewarmed to 42°C), wait for 5 min.
7. Spin 2 min at 1,250g.

8. Elute with another 25µl RNase-free water.

9. Spin 5 min at 1,250g.

The following protocols are designed for hybridization to PCR amplicon-based DNA microarrays and adapted mainly from those provided by the *Drosophila* Genomics Research Center (DGRC) (http://dgrc.cgb.indiana.edu/) and Corning/Promega (http://www.corning.com).

Note: Use Rnase-free filter/barrier tips for every step of the following protocols and wear only nitrile gloves, as latex may cause fluorescent background.

RNA Extraction

1. Begin culturing fly stocks several weeks in advance, flies should be kept under standard conditions at 25°C for at least two generations before collection.

2. For each unit of extraction (enough RNA for two hybridizations) collect three sets of 20-25 males or 8-10 females, aged 4-6 days.

3. When ready for extraction, knock flies out and transfer each set of flies to individual 1.5ml tubes on ice.

4. Add 200µl Trizol to each tube and grind flies completely. Combine the 3 tubes of males or females into 1 tube.

5. Add an additional 400µl Trizol to the combined homogenate for a total of 1ml, mix by inverting the tube and incubate at room temperature for 5 min.

6. Centrifuge 12,000 g at 4°C for 10 min, transfer supernatant to fresh tube.

7. Add 200µl chloroform, mix well by shaking the tubes vigorously for 15 sec by hand, incubate at room temperature for 3 min.

8. Centrifuge 12,000 g at 4°C for 10 min, transfer the aqueous upper phase to a clean tube.

9. Add 500µl isopropanol, mix, and incubate at room temperature for *exact* 10 min.

10. Centrifuge 12,000 g at 4°C for 10 min, remove supernatant, a clearly-visible white pellet should remain.

11. Wash the pellet with 1ml 75% ethanol prepared with Rnase-free water.

12. At this stage, the sample is stable in ethanol at -20°C. Store the sample until shortly before beginning with cDNA synthesis and labeling.

13. Remove the ethanol completely and air dry for 5-10 min. It may be helpful to first remove the ethanol, then briefly centrifuge and again remove any ethanol that accumulated from the spin. Do not over-dry or samples may be difficult to resuspend.

14. Resuspend the pellet in 30µl of Rnase-free water. Dissolving may be aided by several tapping and brief centrifugations and/or heating at 37-55°C until dissolved.

15. Quantify RNA by specifying 2µl sample + 98µl Rnase-free water. Get concentration by A_{260} and sample purity by A_{260}/A_{280}.

16. Run 1µl on a gel, rRNA band should be visible.

17. You should be left with around 27µl of sample at a concentration of 2-4µg/µl. This is more than enough for 2 hybridizations (20µg per hybridization required).

cDNA Synthesis and Amino Allyl Labeling

This protocol makes use of the following Invitrogen packages: SuperScriptTM Indirect cDNA Labeling System, Alexa Fluor® 555 and Alexa Fluor® 647 Reactive Dye Deca-Packs, and cDNA Labeling Purification Module ("'SNAP"' columns).

First-Strand cDNA Synthesis

1. Mix and centrifuge each component in 200µl Rnase-free tubes for individual reactions, or 1.5µl Rnase-free tubes for larger reactions.

Component	Volume
25µg total RNA	Xµl
Anchored Oligo(dT)B20B Primer (2.5µg/µl)	2µl
DEPC treated H_2O	To 18µl

2. Incubate at 70°C for 5 min, and then quick chill on ice for 1 min.

3. Add the following to each tube:

Component	Volume
5X First-Strand Buffer	6µl
0.1 DTT	1.5µl
dNTP mix	1.5µl
RNaseOUT^{TM} (40U/µl)	1µl
SuperScript^{TM} III RT (400U/µl)	2µl
Final Volume:	30µl

4. Mix gently and collect the contents of each tube by brief centrifugation. Incubate at 46°C for 3 hours.

5. Proceed directly to Hydrolysis and Neutralization.

Hydrolysis and Neutralization

This step is taken to degrade the original RNA.

1. Add 15µl of 1N NaOH to each reaction tube from the First Strand cDNA synthesis reaction. Mix thoroughly.
2. Incubate tube at 70°C for 10 min.
3. Add 15µl of 1N HCl immediately after the 10 min incubation to neutralize the pH and mix gently.
4. Proceed directly to Purifying First-Strand cDNA.

Purifying First-Strand cDNA

This step removes unincorporated dNTPs by ethanol precipitation.

1. Add 12µl 3M Sodium Acetate, pH 5.2 and 1µl Glycogen to the neutralized reaction and mix.
2. Add 180µl ice-cold 100% ethanol and mix by vortexing.
3. Place at -20°C for at least 1 hour. Samples can also be incubated overnight at -20°C, as this may increase yield, or stored for several days.
4. Centrifuge at 14,000g at 4°C for 20 min. Carefully remove and discard the supernatant.
5. Wash the pellet with 500µl 75% cold ethanol and centrifuge at 14,000g at 4°C for 2 min. Carefully remove and discard the supernatant.
6. Centrifuge tube and carefully remove any supernatant that accumulated from the spin.
7. Air dry the samples to evaporate any ethanol that may still be on the sample. Sample will turn from white to clear and viscous (glass-like) when ready. Avoid over-drying, as it will be harder to resuspend the samples.
8. After samples have been appropriately dried, resuspend each sample in 5µl of 2X Coupling Buffer.
9. Proceed directly to Labeling with Fluorescent Dye.

Labeling with Fluorescent Dye
Note: This step will label the amino-modified cDNA with the Alexa Fluor® dyes. While working with the dyes or the already labeled cDNA be sure not to expose them to direct sun or overhead light

1. Remove the appropriate Alexa Fluor® dye vials from -20°C storage.
2. Add 2µl of DMSO directly to each dye vial and mix thoroughly.
   Note: DMSO is hygroscopic and will absorb moisture from the air, which reacts with the dyes to reduce the coupling efficiency. So, warm DMSO to room temperature before use and keep the cap close on the vial when not in use
3. Centrifuge vials briefly.
4. Add the DMSO/dye solution to the tube from the Ethanol Precipitation step above. Add 3µl of DEPC-treated water to bring the final volume of the sample to 10µl.
5. Mix samples by vortexing, centrifuge briefly, and incubate at room temperature in the dark for 1-2 hours.
6. Go to Purification of Labeled cDNA after the dark incubation.

Purifying Labeled cDNA
Before starting:
Prepare **Binding Buffer** with **Isopropanol**:
Add 6.5ml Isopropanol to 18ml (entire bottle) of Binding Buffer.
Prepare **Wash Buffer** with **Ethanol**:
Add 20ml Ethanol to 5ml (entire bottle) of Wash Buffer.

1. Add 700µl of Binding Buffer to the reaction tube containing the labeled cDNA from Coupling. Vortex briefly to mix.
2. Each Low-Elution Volume Spin Cartridge is preinserted into a collection tube. For multiple reactions, clearly label each collection tube, and then load the cDNA/Binding Buffer solution directly onto the Spin Cartridge.
3. Centrifuge at 3,300g in a microcentrifuge for 1 min. Remove the collection tube and discard the flow-through.
4. Place the Spin Cartridge in the same collection tube and add 600µl of Wash Buffer to the column.
5. Centrifuge at maximum speed for 30 sec. Remove the collection tube and discard the flow-through.

6. Place the Spin Cartridge in the same collection tube and centrifuge at maximum speed for 30 sec to remove any residual Wash Buffer. Remove the collection tube and discard.

7. Place the Spin Cartridge onto a new amber collection tube (supplied in the kit).

8. Add 20µl of DEPC-treated water to the center of the Spin Cartridge and incubate at room temperature for 1 min.

9. Centrifuge at maximum speed for 1 min to collect the purified cDNA. The eluate contains your purified labeled cDNA.
   The sample can be stored at -20°C for up to one week prior to hybridization. Avoid freeze/thawing.

10. Next, it will be necessary to dry down the sample in a speed vac for around 30 min in order to resuspend the sample in 12 µl of Pronto!^{TM} Long Oligo/cDNA Hybridization Solution.
    The labeled cDNA resuspended in hybridization solution may be stored for up to one week at 4°C and protected from light.

Prehybridization and Hybridization

Preparation of Wash Solutions

The volumes of Universal Wash Reagents A and B provided in the Hybridization Kit are sufficient for processing 25 arrays, depending upon which system is being used.
Carefully follow the order of addition.

Wash Solution 1	Volume
deionized water	1,118.75ml
Universal Wash Reagent A	125ml
Universal Wash Reagent B	6.25ml
Wash Solution 2	**Volume**
deionized water	3,562.5ml
Universal Wash Reagent A	187.5ml
Wash Solution 3	**Volume**
Wash Solution 2	750ml
deionized water	3,000ml

Presoak and Prehybridization

1. Heat required volumes of both Pronto!™ Universal Pre-Soak Solution and Pronto!™ Universal Pre-Hybridization Solution to 42°C for at least 30 min.
2. Add 250µl Sodium Borohydride Solution to 24.75ml of 42°C Universal Pre-Soak Solution. Pipet several times to mix thoroughly. Do not add Sodium Borohydride to the Pre-Soak solution more than 15 min before use and do everything in the hood.
3. Immerse arrays in solution from Step 2 and incubate at 42°C for 20 min.
4. Transfer arrays to Wash Solution 2 and incubate at ambient temperature for 30 sec.
5. Transfer to a fresh container of Wash Solution 2 for 30 sec.

Hybridization

1. Wash the required number of pieces of cover glass (M-series lifter slips) with nuclease-free water, followed by ethanol. Dry cover glass by blowing high-purity compressed nitrogen gas or allow to air-dry in a dust-free environment.
2. Combine 11µl of each of the two samples to be competitively hybridized (labeled cDNA resuspended in 12-24µl hybridization solution) into a PCR tube and mix well.
3. Incubate the labeled cDNA solution at 95°C for 5 min, protecting samples from light.
4. Centrifuge the cDNA at 13,500g for 2 min to collect condensation. Do not place the solution on ice because this will cause precipitation of some of the components.
5. Place array in a Corning Hybridization Chamber (make sure to fill the two moisturizing wells with 10µl ddH_2O each). Pipet the labeled cDNA gently up and down and then transfer onto surface of the printed side of the slide. Carefully place the cover glass on the array. Avoid trapping air bubbles between the array and the cover glass. Assemble the chamber.
6. Incubate the chamber-array assembly at 42°C for 20 h using a water bath or a hybridization oven.

Post-Hybridization Washes *Note: Do not allow the arrays to dry out between washes, as this irreversibly increases background levels.*

1. Heat required volume of Wash Solution 1 to 42°C for at least 30 min.
2. Disassemble the hybridization chambers with the printed array side facing up.
3. Immerse arrays in Wash Solution 1 at 42°C for 1-2 min until the cover glass falls from the slide.

4. Transfer arrays to a fresh container of Wash Solution 1 at 42°C and incubate for 5 min.
5. Transfer arrays to Wash Solution 2 at ambient temperature (22-25°C) and incubate for 10 min.
6. Transfer arrays to Wash Solution 3 at ambient temperature and incubate for 2 min.
7. Repeat Step 6 twice.
8. Dry arrays by centrifugation at 1,600g for 2 min.
9. Store arrays in light-proof container until ready to scan.

Curriculum Vitae

Personal Details

Name	Sonja Grath
Date of birth	May 23rd, 1980
Place of birth	Starnberg, Germany
Nationality	German
Marital status	Married
Current work address	LMU Biozentrum, Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany
Telephone	++49 - 89 - 2180 74 109
Fax	++49 - 89 - 2180 74 104
E-Mail	grath@bio.lmu.de

Education

Feb 2007 - present	University of Munich (LMU), Munich, Germany PhD student in Biology Advisor: Prof. Dr. John Parsch
Aug 2005 - Dec 2006	University of Munich (LMU), Munich, Germany Technical University of Munich (TUM), Munich, Germany Master of Science (M. Sc.) in Bioinformatics Thesis: Comparison of distance- and kernel-based methods for structured data Advisor: Prof. Dr. Stefan Kramer (TUM)

Oct 2001 - Jul 2005	University of Munich (LMU), Munich, Germany Technical University of Munich (TUM), Weihenstephan, Germany Bachelor of Science (B. Sc.) in Bioinformatics Thesis: Influence of fusogene peptides on fusion of liposome-membranes (in german) Advisor: Prof. Dr. Dieter Langosch, Dr. Markus Guetlich (TUM)
Nov 1999 - Sep 2001	Technical University of Munich (TUM), Munich, Germany Studies of Computer Science

Working Experience

Mar 2006 - Jan 2007	University of Munich (LMU), Munich, Germany Student assistant Section of Evolutionary Biology, Evolutionary and Functional Genomics, Prof. Dr. John Parsch
Apr 2005 - Mar 2006	Erzbischoefliches Ordinariat Muenchen IT consulting/User help-desk
Apr 2004 - Aug 2004	Technical University of Munich (TUM), Weihenstephan, Germany Practical training, preparation of Bachelor's thesis Center for Integrated Protein Science Munich, Chemistry of Biopolymeres, Prof. Dr. Dieter Langosch

Skills

Languages	German (Native language) English (Fluent) French (Basic knowledge)

Computer/IT	Microsoft Word, Excel, Outlook, PowerPoint, Access Working experience with both Windows (XP/Vista) and Linux
Programming	Perl, Java, R, MySQL

Selected conferences and presentations

Grath, S., Baines, J. F., and Parsch, J. (Jan 2010)
Studying adaptive evolution of sex-biased genes in *Drosophila ananassae*.
Talk at the 43rd Population Genetics Group Meeting (Liverpool, United Kingdom).

Grath, S., Baines, J. F., and Parsch, J. (Aug 2009)
Molecular evolution of sex-biased genes in the *Drosophila ananassae* subgroup.
Poster presentation at the 12th congress of the European Society for Evolutionary Biology (Torino, Italy).

Grath, S., Baines, J. F., and Parsch, J. (Sep 2008)
Sex-biased gene expression and adaptive evolution in *Drosophila ananassae*.
Talk at the 101st Annual Meeting of the Deutsche Zoologische Gesellschaft (Jena, Germany).

Grath, S., Baines, J. F., and Parsch, J. (Sep 2007)
Identification and analysis of genes with sex-biased expression in *Drosophila ananassae*.
Poster presentation at the 20th European Drosophila Research Conference (Vienna, Austria).

Publications

Grath, S., Baines, J. F., and Parsch, J. (2009)
Molecular evolution of sex-biased genes in the *Drosophila ananassae* subgroup.
BMC Evol. Biol. **9**(1):291.

Schug, M. D., Baines, J. F., Killon-Atwood, A., Mohanty, S., Das, A., **Grath, S.**, Smith, S. G., Zargham, S., McEvey, S. F., and Stephan, W. (2008)
Evolution of mating isolation between populations of *Drosophila ananassae*.
Mol. Ecol. **17**(11):2706-21.

Acknowledgements

The first person I would like to thank is Prof. Dr. John Parsch for giving me the opportunity to conduct my PhD research under his supervision. He contributed largely to improve my scientific skills and writing. I owe a debt of gratitude for his patience and accuracy in correcting my papers and conference abstracts and for his confidence in me and my work.
I owe many thanks to Prof. Dr. John Baines for introducing me to experimental work and for always keeping me on track. Performing scientific work at an university always keeps a lot of distractions ready. He taught me not to loose my main goals and improved my scientific thinking and ability to work independently.
I would like to extend my gratitude to Prof. Dr. Wolfgang Stephan for giving me the opportunity to contribute to one of his studies and for being part in my committee. In addition, I want to thank Prof. Dr. Susanne Foitzik for her insightful comments during my committee meeting.
Thanks to Prof. Dr. Dirk Metzler and Dr. Martin Hutzenthaler for help with all matters concerning statistics and the R course. I thank Dr. Pleuni Pennings for sharing her great fascination in science and for her continous encouragement. Even in times of doubts and hard work, she never lost the conviction that being a scientist is one of the greatest professions.
The group of evolutionary biology is steadily growing and a lot of people crossed my paths during the four years I was part of the lab, first as practical, than as PhD student.
First, I want to thank all the current and former technicians in our lab. Ana, Anne, Beate, Carmen, Gisela, Hedwig, Hilde, Simone, and Yvonne - thank you for your help and advice in the lab and with keeping our flies happy, even under standard fly room conditions. Without Katrin's and Ingrid's support I would have had definitively more trouble in handling university bureaucracy.
I thank my "oldest" room mates, Lena Müller and Miri Linnenbrink for their patience and friendship, particularly during the last months of writing this thesis. When I am leaving,

I will miss you not only for your questions ("Sonjaaaa..."), but especially for your everlasting encouragement and confidence in my work. I am happy that I could support Miri at her first conference, the PopGroup meeting in Liverpool, and I am looking forward to climbing up more and more walls together with Lena. I want to thank Claus Kemkemer for his enduring happyness, frankness and suggestions on fly work. Further, I thank Winfried Hense for the advide on fly genetics and handling and the philosophical discussions in the fly room. Stefan Laurent, Rebecca Meredith and Jennifer Lohr were a great help in organizing the DZG PhD meeting 2009 in Munich. Pavlos Pavlidis helped me loosing fear of any questions concerning demography.
We *Drosophila* people would be really lonesome without the Tomato people in our group. I thank Iris Fischer for always lending me an ear and having good advice on population genetics. Anja Hörger and Aurélien Tellier - I really enjoyed our time at the ESEB meeting in Torino and many other discussions. Lisha Naduvilezath, Robert Piskol, and Sergej Nowoshilow - we built a small Bioinformatician island within all the biologists. You often helped me out when I was at the brink of desparation because of some troubles with R, LaTeX, C++, Perl... .
I want to thank my former colleague Nicolas Svetec for his advice for the chill coma recovery experiments and his friendship. The same is true for Annegret Werzner. Thanks for your insights on experimental settings and all the chats.
Special thanks go to all the practical students who joined me in the lab. Carla, Johanna, Ricardo, Pablo and Francisco - you not only contributed to my thesis with your efforts in the lab, but further helped me critically reflecting my work. I want to thank Monika Poxleitner for taking over the CCRT experiments in the course of her Bachelor thesis.
Thanks to all my other current and former colleagues for contributing to such a great working atmosphere. I had the feeling that I could always come and see someone to get some advice on experiments or just a small word of encouragement.
I want to thank all my friends and especially my parents for their support and for believing in me and my abilities. Last, but definitively not least, I owe a debt of gratitude to my husband Matthias without whom nothing would be possible. You not only helped me through all my studies including Bachelor and Master thesis, but finally through this dissertation. I could always rely on your love, patience, motivation, confidence in me and my work and constructive criticism. Thank you for all.