Dissertation zur Erlangung des Doktorgrades
der Naturwissenschaften an der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

Molecular evolution of sex-biased genes in *Drosophila*

Zhi Zhang
Aus
Jilin, China
2006

Erklärung

Diese Dissertation wurde im Sinne von §13 Abs. 3 bzw. 4 der Promotionsordnung von Prof. John Parsch betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbstständig, ohne unerlaubte Hilfe erarbeitet.

München, 24.10.2006

Zhi Zhang

Dissertation eingereicht am: 24.10.2006

1. Gutachter: Prof Dr. John Parsch
2. Gutachter: Prof Dr. Wolfgang Stephan

Mündliche Prüfung am: 20.12.2006

Table of contents

List of abbreviations

A	equal to $\frac{1}{2}\log_2(R*G)$ during microarray analysis
α	The fraction of positively selected amino acid replacements between species
Acp	Accessory gland protein
BAGEL	Bayesian Analysis of Gene Expression Levels
bp	Base pair
CARMA	Comprehensive R based Microarray Analysis web
d_N	Nonsynonymous substitution rate estimated from nucleotide divergence of nonsynonymous sites
d_S	Synonymous substitution rate estimated from nucleotide divergence of synonymous sites
d_N/d_S	The ratio of the nonsynonymous substitution rate to the synonymous substitution rate, estimated from nucleotide divergence
ENC	Effective number of codons
F_{op}	Frequency of optimal codons
M	equal to $\log_2(R/G)$ during microarray analysis
Mb	Megabases
MK	McDonald-Kreitman test

MKPRF	McDonald-Kreitman test with Poisson random field
N_e	Effective population size
PAML	Phylogenetic analysis by maximum likelihood
π_N	Nucleotide diversity of nonsynonymous sites based on the mean pairwise differences between sequences
π_S	Nucleotide diversity of synonymous sites based on the mean pairwise differences between sequences
R	Correlation coefficient
r	Local recombination rate
γ	Selection parameter, equal to $2N_es$
R/G	The intensity ratio of two channels (e.g. red and green) in a two color microarray experiment
s	Selection coefficient

Introduction

Charles Darwin first proposed the theory of sexual selection in his book *On The Origin of Species*, which was primarily devoted to natural selection (Darwin 1859). In his later book, *The Descent of Man and Selection in Relation to Sex*, Darwin considered the subject of sexual selection comprehensively, because he felt that natural selection alone was unable to explain certain types of apparently non-competitive adaptations, such as the tail of the peacock (Darwin 1871). The original meaning of sexual selection comes from the theory that competition for mates between individuals of the same sex (typically males) drives the evolution of certain traits. Within a species, one sex (typically females) acts as a limiting resource for the other (typically males). Sexual selection can be categorized into two major forms: intrasexual selection (e.g. male-male competition) in which members of the less limiting sex compete aggressively between themselves for mating with the more limiting sex, and intersexual selection (e.g. female choice) in which males compete with each other to be chosen by females (Panhuis et al. 2001). In most sexual species, the males and females behave differently when it comes to investment in producing offspring. Thus, sexual selection may affect males and females differently. Typically, the effects of sexual selection are more pronounced in males than in females. Differences in secondary sexual characteristics between males and females of a species are referred to as sexual dimorphisms and include traits that may be subtle, such as size differences (sexual size dimorphism), or more extreme, such as the presence of horns or sex-specific color patterns. The case of the peacock, with its extravagant and colorful tail feathers that are lacking in the peahen, is often cited as perhaps the most extraordinary example of sexual dimorphism. On account of their sometimes greatly ornamented nature, secondary sexual characteristics can prove to be an obstruction to the animal and lower its fitness. By the same token, sexual selection can produce individuals with such elaborate ornaments that they must be either energetically costly to generate, costly to conserve, or even impose a direct survival cost for the individual with the ornament traits. In this sense, sexual selection can lead to maladaptive traits. Because traits driven by sexual selection often conflict with the survival interests of the individual, the question then arises as to why such

apparent liabilities are permitted to persist while “survival of the fittest” is considered the capital rule in nature. Darwin's theory of sexual selection gave a plausible explanation for the origin and maintenance of many splendid, though potentially costly, ornaments.

Understanding how sexual selection acts upon genetic variation is central to understanding the evolution of mating behavior and speciation. On the one hand, this topic has been studied based on ecology and morphology, widely and deeply, particularly for male traits (Barraclough et al. 1995; MØller and Cuervo 1998). Studies of morphology and interspecific hybridization have suggested that sex-related characters (particularly those involved in male reproduction) evolve rapidly relative to nonsex-related characters. It has long been known that in multicellular eukaryotes, phenotypes associated with sex and reproduction accumulate interspecific differences more rapidly than other phenotypes (Darwin 1871). Such rapid evolution is often evident in male secondary sexual characteristics (e.g. feather coloration or song in male birds), and also in primary sexual characteristics (e.g. genital or sperm morphology in insects). Darwin considered sexual selection important enough to give it a separate name and the evolution of male novelty has been the subject of intensive and prolonged study in evolutionary biology. It is assumed that sexual selection and/or sexual antagonism drive the evolution of sexual traits. The rapid evolution of male sexual characters has recently been extended to the molecular level, and is suggested by 2D gel electrophoresis of reproductive proteins (Coulthart and Singh 1988), global analysis of sex-biased gene expression levels (Meiklejohn et al. 2003), and DNA-based studies of specific genes involved in male reproduction (Civetta and Singh 1995). For example, one group of genes that has been extensively studied at the nucleotide level in *Drosophila* is that encoding male accessory gland proteins (*Acps*) (Tsaur and Wu 1997; Begun et al. 2000; Swanson et al. 2001; Kern et al. 2004). *Acps* function in male reproduction, and some show a very strong signal of adaptive evolution in polymorphism and divergence data. Gene expression data also show that genes with male-biased expression change rapidly in terms of both gene expression over a 2.5 million year time scale (Ranz et al. 2003) and sequence over a 250 million year time scale (Parisi et al. 2003). It has also been found that male-biased genes have higher levels of expression polymorphism than female-biased genes and nonsex-biased genes (Meiklejohn et al. 2003). These results suggest that

sex-dependent selection may drive changes in expression of the most rapidly evolving genes in the *Drosophila* transcriptome. Accordingly, several lines of evidence suggest that the diversity of traits affecting male fitness presents a large target for sexual selection (Singh and Kulathinal 2005). A study of newly-evolved genes in *Drosophila melanogaster* supports this assumption: these novel genes, derived from ancestral noncoding sequence, show male-biased expression and likely invaded populations under selective pressure relating to male reproduction (Levine et al. 2006). Levine et al. (2006) suggested that novel male reproduction-related genes could be fixed by frequent adaptive evolution driven by their beneficial function in male reproduction. It has also been found that many new retrotransposed genes that originated from X-linked parental genes show expression in testes, and some of these genes show the molecular hallmarks of positive selection (Betran et al. 2002; Betran and Long 2003; Arguello et al. 2006).

Although much research has focused on the evolution of male reproductive traits, the effects of sexual selection are not limited only to the males of the species. Each sex is part of the environment of the other sex and females also play an important role in sexual reproduction. Perpetual co-evolution between the sexes can occur when adaptation by one sex reduces fitness of the other (Rice 1996). Such a situation is known as sexual conflict or sexual antagonism. Sexual conflict can impose direct selection on female mating preference that leads to increased female resistance to harmful male traits. The conflict between the sexes is becoming increasingly clear through a set of recent studies, and it now seems that sexual conflict between males and females is the rule rather than the exception. Experiments using *D. melanogaster* indicate that components of male seminal fluid not only increase male fitness, but also reduce the competitive ability of sperm from other males. For example, Rice (1996) showed that when coevolving with other males, males rapidly adapt to a static female phenotype, becoming "super fertile". This male adaptation leads to a reduction in female survivorship, which is mediated by an increased rate of remaining and increased toxicity of seminal fluid.

Males and females share a common genome, but they perform many different biological functions and experience different selective pressures (Rice and Chippindale 2002). For example, males are selected to reproduce via microgametes and females via megagametes. This differential selection can lead to numerous differences in phenotypic optima involving behavioral, anatomical, and

physiological/cell-biological traits. Although the evolution of male traits has been studied in detail, little is known about the female reproductive molecules that are involved in male-female interaction. The few cases studied so far suggest that adaptive evolution may also occur in female reproductive molecules. Positive selection on female reproductive molecules has been detected in mammals (Swanson et al. 2001; Jansa et al. 2003) and abalone (Galindo et al. 2003). Swason et al. (2004) present the first systematic attempt to identify genes encoding female reproductive proteins in *Drosophila* and to initiate evolutionary analyses of several such genes. Their results demonstrate that several of these genes have been subjected to positive selection. Their expression in female reproductive tracts, presence of signal sequences/transmembrane domains, and rapid adaptive evolution indicate that they are prime candidates to encode female reproductive molecules that interact with rapidly evolving male *Acps*. A recently-published follow-up to this work revealed additional genes expressed in female reproductive tracts that are rapidly evolving and subject to positive selection, presumably due to co-evolution between males and females (Panhuis and Swanson 2006). In addition, large-scale transcriptomic and proteomic experiments have described the transcriptional and translational response to mating that occurs within the female reproductive tract and identified many more genes that are likely to be involved in mating, sperm maintenance and utilization (Mack et al. 2006).

The molecular evolution of genes with known functional roles in male and female reproduction has also been compared to that of genes with non-reproductive functions (Musters et al. 2006). Here, the availability of complete genome sequences of *D. melanogaster* and *D. pseudoobscura* provided an opportunity to investigate factors involved in sequence divergence. Using orthologs to calculate rates of nonsynonymous substitution (d_N), it was found that genes having reproductive functions in either the male, the female, or both sexes had greater divergence than genes without reproductive function. Divergence was even higher for genes involved in male-specific functions. This rapid evolution of sex-related genes could be due to increased positive selection, although the authors could not exclude the possibility of relaxation of selection constraint on reproductive genes. In numerous studies (e.g. Civetta and Singh 1998), a select group of proteins involved in reproduction, particularly male reproduction, has been identified as having high divergence rates. Similarly, proteins expressed in both testis and ovaries were found to be twice as

divergent compared to proteins in other tissues (Civetta and Singh 1995). In summary, there is much evidence supporting that genes with reproductive function evolve faster than genes without reproductive function.

What is the larger biological significance of sex-related gene evolution? Speciation, which is the division of one species into two distinct species over time, plays a central role in evolutionary theory. However, the evolutionary forces that lead to speciation are not well understood. Darwin observed that elaborate secondary sexual characters tended to occur in groups that also had high species richness. This suggests that species divergence can be related to sexual selection. Consistent with this, differences in the proportions of sexually dichromatic and monochromatic species were found in 20 sister pairs of passerine bird tribes, where 12 out of 15 comparisons with marked differences in the frequency of dichromatism showed differences in species richness in the expected direction (Barraclough et al. 1995). There are many similar comparative studies to support speciation by sexual selection: by comparing the number of species in taxa with different mating systems and different degrees of feather ornamentation in birds (MØller and Cuervo 1998) and by comparing lengths of nectar spurts in plants (Hodges and Arnold 1995), it was found that ornamented species had more subspecies, which suggested ongoing differentiation. This differentiation is presumably caused by sexual selection and/or sexual conflict. During this sexual selection/conflict, a male tries to maximize the proportion of a female's reproductive effort invested into his offspring, even if the trait related to sexual conflict is deleterious to the female.

Theory predicts that incompatibility involving sex-related genes will lead to population differentiation, reproductive isolation and speciation (Orr and Presgraves 2000). The Biological Species Concept (BSC; Mayr 1963) is the most widely accepted view and defines species as groups of organisms that are reproductively isolated (Mayr 1963). Reproductive isolation can occur both prezygotically (e.g. behavioral, physical, spatial or temporal isolation) and postzygotically (e.g. hybrid infertility or inviability). The latter mechanism has been of great interest to geneticists because it is thought to occur at the earliest stages of reproductive isolation and can be investigated with interspecific crosses in model organisms. However, until now little is known about the genetic mechanisms of speciation. The central questions are: what forces drive speciation? and can the footprint of these forces be found at the DNA

sequence level? My hypothesis is that different classes of sex-biased (and nonsex-biased) genes are subject to different evolutionary forces, which shape their molecular evolution. By investigating the changes that occur in their DNA sequences, I hope to determine the type and strength of selection acting on sex-biased genes and, thus, shed light on the mechanisms responsible for genetic differentiation between species.

Scope of this dissertation:

In chapter one, I report a general comparison of evolutionary rates of sex-biased genes using data from microarray experiments and comparative genomic studies of *Drosophila.* Comparisons of nonsynonymous/synonymous substitution rates (d_N/d_S) between species of the *D. melanogaster* subgroup revealed that genes with male-biased expression had significantly faster rates of evolution than genes with female-biased or nonsex-biased expression. The difference was due primarily to a higher d_N in the male-biased genes. The same pattern was observed for comparisons among more distantly related species. In comparisons between *D. melanogaster* and *D. pseudoobscura*, genes with highly male-bised expression were significantly more divergent than genes with highly female-bised expression. In many cases, orthologs of *D. melanogaster* male-biased genes could not be identified in *D. pseudoobscura* through a BLAST search. In contrast to the male-biased genes, there was no clear evidence for accelerated rates of evolution in female-biased genes, and most comparisons indicated a reduced rate of evolution in female-biased genes relative to nonsex-biased genes. Male-biased genes did not show an increased ratio of nonsynonymous/synonymous polymorphism within *D. melanogaster*, and comparisons of polymorphism/divergence ratios suggest that the rapid evolution of male-biased genes is due to positive selection.

In chapter two, I take a different approach to test whether the rapid evolution of male-biased genes is the result of increased positive (or sexual) selection, or if it is due to a relaxation of selective constraint. To distinguish between these two possibilities, I analyzed the relationship between the nonsynonymous substitution rate (d_N) and local recombination rate for 343 *Drosophila* genes that were classified as male-, female-, or nonsex-biased in their expression. For the male-biased genes, a positive correlation between d_N and recombination rate was observed. This can be explained by an increased rate of adaptive evolution in regions of higher

recombination due to a reduction of Hill-Robertson interference. In contrast, the correlation between d_N and recombination rate was negative for both female-, and nonsex-biased genes, suggesting that these genes are primarily subject to purifying selection, which is expected to be less effective in regions of reduced recombination.

In chapter three, I further investigate the type and strength of selection influencing the evolution of sex-biased genes by surveying DNA sequence polymorphism in 91 protein-encoding genes in a *D. melanogaster* population sample from Zimbabwe, Africa. In addition, I use the sequence of a single *D. simulans* strain for interspecific comparisons. Of the 91 genes, 33 show male-biased expression, 28 female-biased expression, and 30 nonsex-biased expression. The combination of within-species polymorphism and between-species divergence data allows the application of powerful statistical methods to infer the selective history of each group of genes. In general, these methods are based on the McDonald and Kreitman (MK) test, which compares the ratio of polymorphism and divergence at synonymous sites to that at nonsynonymous sites. My analyses of polymorphism and divergence indicate that adaptive evolution occurs more frequently in sex-biased genes (both male and female) than in nonsex-biased genes. Male-biased genes, in particular, appear to be consistent targets of positive selection. Female-biased genes show more variance in the type of selection they experience, with positive selection affecting some genes and purifying selection affecting others. Nonsex-biased genes appear to evolve primarily under purifying selection and have undergone relatively little adaptive evolution since the split of *D. melanogaster* and *D. simulans*.

In chapter four, I use two types of microarrays to investigate sex-biased gene expression in four *D. melanogaster* strains (two European and two African) and in one strain of *D. simulans*. One type of microarray was custom made and had probes corresponding to the 91 genes investigated in chapter 3. The other type of microarray was commercially available and contained probes to approximately 75% of the genes in the *D. melanogaster* genome. The custom arrays allowed for high replication and, in many cases, could confirm the expected expression pattern in the different *D. melanogaster* strains. The whole genome arrays revealed general patterns of sex-biased gene expression in *D. melanogaster* and *D. simulans*. For the vast majority of genes, the ratio of male to female expression was similar in two *D. melanogaster* strains and in *D. simulans*. In addition, my results from *D. simulans* agreed well with previously published results that used a different microarray platform.

Chapter 1

1. Molecular Evolution of Sex-Biased Genes in *Drosophila*

1.1 Introduction

It has long been known that traits associated with sexual reproduction (particularly those related to male reproductive success) often show greater interspecific divergence than non-reproductive traits. Darwin (1871) documented the frequent occurrence of extravagant secondary sexual characteristics in species of many taxa (including crustaceans, insects, fish, birds, and mammals) and observed that these traits often show large morphological differences between closely related species. Furthermore, Darwin noted that such traits were, with few exceptions, limited to the male of the species. These observations were explained by the theory of sexual selection, which posited that male reproductive traits evolved in response to male-male competition for mating opportunities and the preferential mating of females to males with "attractive" phenotypes. Morphological evidence suggests that sexual selection also acts on primary sexual characteristics. For example, sperm size and morphology are known to differ greatly among insect species (Jamieson 1987), as is the external morphology of male genitalia (Eberhard 1985; Hosken and Stockley 2004).

Studies of interspecific hybridization have also suggested the rapid evolution of male reproductive characters. In 1922, Haldane noted a common pattern regarding the viability and fertility of species hybrids. His observation, known as Haldane's Rule, was that when hybrid offspring of only one sex are either inviable or infertile, it is most often the heterogametic sex (Haldane 1922). In many taxa, such as mammals and *Drosophila*, the males are heterogametic (XY), and thus hybrid male offspring are more prone to be inviable or sterile. Two major hypotheses have been proposed to explain Haldane's rule. The first hypothesis, known as the "dominance" hypothesis posits that hybrid incompatibilities are often recessive, and thus are only observed in the sex with hemizygous sex chromosomes (Turelli and Orr 1995). The second

hypothesis is known as "faster male evolution" and posits that genes involved in male reproduction evolve faster than genes involved in female reproduction or genes with non-reproductive function (Wu and Davis 1993). This hypothesis can only completely explain Haldane's rule for taxa in which the males are heterogametic, and is expected to apply primarily to hybrid sterility, not inviability. However, the two hypotheses are not mutually exclusive and it is likely that both faster male evolution and dominance play a role in hybrid breakdown (Presgraves and Orr 1998). Faster male evolution may also explain the overwhelming preponderance of male sterility factors relative to inviability factors that have been identified from *Drosophila* hybridizations (Wu and Davis 1993; True et al. 1996; Tao et al. 2003)

Recent studies have indicated that sex-related genes show increased rates of evolution in their DNA/protein sequences, and it has been suggested that sexual selection affects the evolution of a broad range of genes with reproductive functions (see reviews by Civetta and Singh 1999; Singh and Kulathinal 2000; Swanson and Vacquier 2002). In *Drosophila*, the rapid evolution of reproductive proteins is suggested by the relatively large interspecific differences in migration pattern observed for these proteins on two-dimensional electrophoresis gels (Coulthart and Singh 1988; Civetta and Singh 1995), and in the elevated rate of amino acid substitution between *D. melanogaster* and *D. simulans* observed for large number of male-specific accessory gland proteins (Swanson et al. 2001). In addition, a number of male-specific genes showing evidence for rapid evolution due to positive selection have been identified, including *Acp26Aa* (Tsaur and Wu 1997; Aguadé 1998; Tsaur et al. 1998), *OdsH* (Ting et al. 1998), *Sdic* (Nurminsky et al. 1998; Nurminsky et al. 2001), *Dntf-2r* (Betrán and Long 2003), and *jan-ocn* (Parsch et al. 2001a; Parsch et al. 2001b). Recently, observations of faster male evolution have been extended to the level of gene expression (Meiklejohn et al. 2003; Ranz et al. 2003). These studies used competitive cDNA microarray hybridizations to demonstrate that genes with male-biased expression show greater expression differences both within and between species than either female-biased or nonsex-biased genes.

In this chapter, I use data from published microarray experiments to identify *Drosophila* genes showing either male- or female-bias in their expression. I then investigate the rates of evolution of these genes (using the ratio of nonsynonymous/synonymous substitution rates, d_N/d_S) among *Drosophila* species and compare them to a collection of control genes that show no sex-bias in their

expression. My results indicate that male-biased genes have a significantly higher rate of evolution than both female-biased and nonsex-biased genes. A similar pattern is observed for the evolution of highly sex-biased genes between *D. melanogaster* and *D. pseudoobscura*, with male-biased genes showing significantly greater divergence between these two species than either female-biased or nonsex-biased genes. Polymorphism and divergence data suggest that these differences are due to increased positive selection on male-biased genes.

1.2 Materials and Methods

1.2.1 Identification of genes with sex-biased expression

Two independent data sets that compared male and female gene expression in *D. melanogaster* through competitive microarray hybridizations were used to classify genes as male-, female-, or nonsex-biased in their expression (Parisi et al. 2003; Ranz et al. 2003). Although these data sets were generated using similar experimental approaches, a direct comparison of the two is difficult due to methodological differences in microarray platform, gene number, RNA source material, replication scheme, and statistical analysis. For example, Parisi et al. (2003) used arrays of PCR amplicons (averaging 410 bp in length) corresponding to individual exons of approximately 75% of the predicted genes in the *D. melanogaster* genome, while Ranz et al. (2003) used arrays of full-length cDNAs corresponding to approximately 40% of all predicted genes (Rubin et al. 2000). In addition, Parisi et al. (2003) performed dissections to compare gene expression between testes and ovaries, while Ranz et al. (2003) performed comparisons of whole flies. Finally, Parisi et al. (2003) used an expression difference of twofold to classify genes as sex-biased, while Ranz et al. (2003) used a Bayesian approach (Townsend and Hartl 2002) that could detect significant expression differences of less than twofold. To be conservative, I used the twofold cut-off for both data sets. That is, genes with a male/female (or testes/ovaries) ratio greater than two were considered male-biased, genes with a ratio less than one-half were considered female-biased, and genes with a ratio between one-half and two were considered nonsex-biased (Table 1 and Figure 1). In rare cases (7% of genes) where the two data sets resulted in different classifications (e.g. greater than twofold male bias in one data set and less than twofold male bias in the other), the gene was considered sex-biased. However, exclusion of these genes from my analyses does not affect the results. A small number of genes (0.08%) with sex-bias conflict (e.g. male-biased in one data set and female-biased in the other) were excluded from further analyses. The twofold cutoff for expression bias was chosen as a conventional standard to allow comparison of microarray results generated using different array platforms and experimental designs. Re-analysis of the data using cutoff values ranging from 1.5-fold to 3-fold did not alter the qualitative pattern or statistical

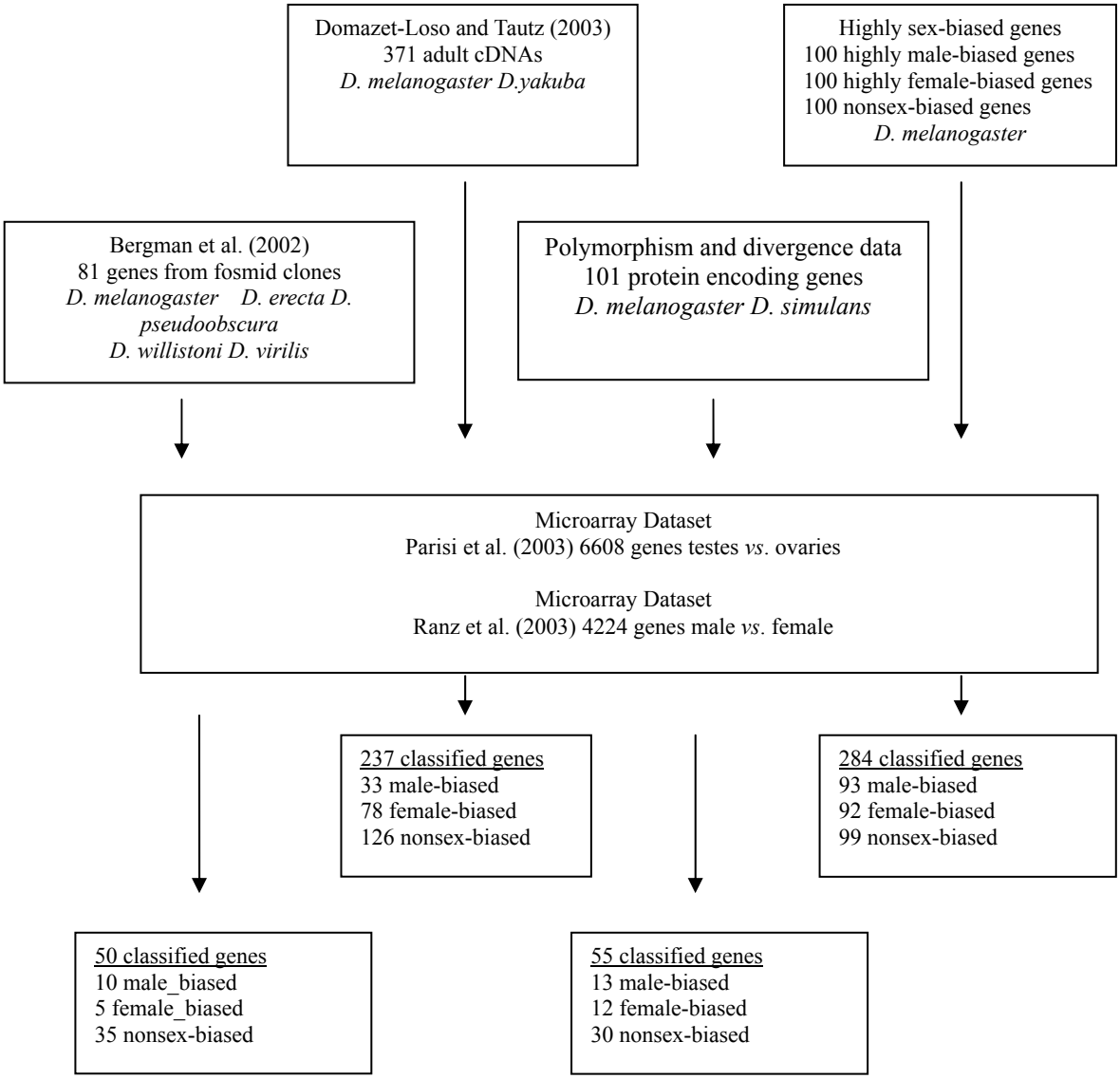

Figure 1. The construction of the data set analyzed in this study. Genes in the various comparative genomic data sets (top) were classified into three groups using expression data from two microarray experiments (center). Sex-biased genes were defined as those with greater than twofold expression difference between the sexes in either data set. Nonsex-biased genes were those with less than twofold difference in both data sets. The number of classified genes for each comparative genomic data set, after removing redundancies and genes with sex-bias conflicts, is shown at the bottom.

significance of my results. Finally, it should be emphasized that all sex-bias classifications are based on expression studies of *D. melanogaster*. In some cases, the expression bias could be confirmed in *D. simulans* (Ranz et al. 2003). However, expression levels in more distantly related species, such as *D. yakuba* or *D.*

pseudoobscura, have not been determined.

1.2.2 Analysis of comparative genomic data sets

Comparative sequence data for 371 adult cDNAs from *D. yakuba* (Domazet-Loso and Tautz 2003) were kindly provided by T. Domazet-Loso. Of the 371 cDNAs, 237 could be classified as male-, female-, or nonsex-biased based on microarray expression data (Table 1 and S1; and Figure 1). Evolutionary rates of these genes (d_N, d_S, and d_N/d_S) were calculated from pairwise comparisons of *D. yakuba* and *D. melanogaster* using the codeml program of the PAML software package (Yang 1997). Levels of codon bias were calculated as either effective number of codons (ENC; Wright 1990) or frequency of optimal codons (F_{op}; Ikemura 1981) based on the full-length transcript from *D. melanogaster* using the codonW program (http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html). In cases where multiple transcripts were predicted for a gene, the longest transcript was used. For calculating F_{op}, codon frequency data from *D. melanogaster* were used.

Comparative sequence data for 81 orthologous genes from *D. melanogaster*, *D. erecta*, *D. pseudoobscura*, *D. willistoni*, and *D. virilis* (Bergman et al. 2002) were downloaded from the Berkeley *Drosophila* Genome Project web site (http://www.fruitfly.org/comparative/index.html). These data come from sequenced fosmid clones (≈40 kb each) corresponding to the *D. melanogaster* genomic regions containing the *apterous*, *even-skipped*, *fushi tarazu*, *twist*, and *Rhodopsin 1*, *2*, *3*, and *4* genes. Of the 81 genes, 50 could be classified as male-, female-, or nonsex-biased based on the microarray data (Table 1 and S2; Figure 1). For these genes, values of d_N, d_S, and d_N/d_S were calculated for all pairwise comparisons of species using PAML (codeml runmode –2). For genes with sequences available from three or more species, d_N/d_S was also calculated for each gene using all available sequences and assuming a constant d_N/d_S over all branches of the phylogenetic tree (codeml runmode 0, model 0). In addition, I applied the "free-ratio" model (Yang 1998) that allows d_N/d_S to vary over all branches of the tree (codeml runmode 0, model 1) and compared the likelihood ratio of the two models using a χ^2 test with the degrees of freedom equal to the difference in parameter number (i.e., the number of branches in the tree minus one). For these analyses, the phylogenetic relationship of *Drosophila* species given in Bergman et al. (2002) was used.

Table 1. Summary of gene expression and comparative genomic data sets.

Data set	Total[a]	Male-biased[b]	Female-biased[b]	Nonsex-biased[b]
Gene expression				
Parisi et al. (2003)	6608	1102	1039	4467
Ranz et al. (2003)	4224	552	788	2884
Combined[c]	8484	1489	1567	5428
Comparative genomic				
Domazet-Loso and Tautz (2003)	237	33	78	126
Bergman et al. (2002)	50	10	5	35
Highly sex-biased	284	93	92	99
D. melanogaster polymorphism	55	13	12	30

[a] Genes with expression data in at least one microarray data set.
[b] Sex-biased genes were defined as those with greater than twofold expression difference between the sexes in either data set. Nonsex-biased genes were those with less than twofold difference in both data sets.
[c] Gene number after removing redundancies (28%) and genes with sex-bias conflicts (0.08%).

1.2.3 Comparison of highly sex-biased genes between *Drosophila* genomes

To investigate the evolutionary rates of genes showing the strongest sex-bias in expression, I selected the 50 genes with the highest and lowest male/female (or testes/ovaries) ratios from both the Parisi et al. (2003) and Ranz et al. (2003) data sets. As a control, I selected the 50 genes showing a male/female ratio closest to one from each data set. Only genes corresponding to predicted transcripts in the *D. melanogaster* genome release 3.0 (Celniker et al. 2002) were included. Of the 100 genes selected for each expression class, 93 male-, 92 female-, and 99 nonsex-biased genes remained after removing redundancies (Table 1 and S3; Figure 1). The paucity of overlapping genes between the two data sets was primarily due to differences in array composition (i.e., genes present in one data set, but absent in the other; 50% of the genes). For another 47% of the genes, the difference was only in the level of sex-bias (i.e., the gene was in the top 50 of its expression class in one data set, but not the other). For 3% of the genes, there was a conflict such that a gene was sex-biased in one data set, but nonsex-biased in the other. There were no cases of conflicting

male/female sex-bias classification among these genes. The coding sequences of these genes from *D. melanogaster* were used for a BLASTn (version 2.0; Altschul et al. 1997) search of the *D. pseudoobscura* genome sequence using the Baylor College of Medicine *Drosophila* Genome Project web site (http://www.hgsc.bcm.tmc.edu/projects/drosophila/). Because the coding sequences of many of the genes could not be aligned between the two species, I report divergence as either BLAST e-values or BLAST scores. For score calculation, the default values of -5 and -2 were used for the gap creation and gap extension penalties, respectively. Levels of codon bias were calculated as described above using the full-length coding sequences from *D. melanogaster*.

1.2.4 Comparison of polymorphism and divergence

Because the largest number of DNA sequence polymorphism surveys has been conducted in *D. melanogaster*, I chose this species to investigate levels of polymorphism in sex-biased genes. I began with a database of 101 protein-encoding genes extracted from the literature and GenBank (kindly provided by D. Presgraves) for which multiple (at least six) *D. melanogaster* alleles had been sequenced and at least one *D. simulans* allele was available for divergence analysis. A total of 55 genes could be classified as male-, female-, or nonsex-biased based on the microarray expression data (Table 1 and S4; Figure 1). One gene (*Dntf-2r*) was classified as nonsex-biased based on the Parisi et al. (2003) data set, although it had been shown to be testis-specific by RT-PCR (Betrán and Long 2003). In this case, I classified the gene as male-biased. The conflicting microarray result may be due to cross-hybridization between *Dntf-2r* and its close paralog *Dntf-2*, which is expressed in both sexes (Betrán and Long 2003). DnaSP (version 4; Rozas et al. 2003) was used to calculate levels of polymorphism and divergence and to perform the MK test (McDonald and Kreitman 1991). All available *D. melanogaster* sequences were used for polymorphism calculations, while a single *D. simulans* allele was used for divergence.

1.3 Results

1.3.1 Comparison of *D. yakuba* cDNAs to *D. melanogaster*

In an analysis of orphan gene evolution, Domazet-Loso and Tautz (2003) cloned and sequenced cDNAs from *D. yakuba* and estimated rates of evolution (as d_N, d_S, and d_N/d_S) by comparing the cDNA sequences to the *D. melanogaster* genome sequence (Celniker et al. 2002). I used the microarray results of Parisi et al. (2003) and Ranz et al. (2003) to classify the adult cDNAs analyzed by Domazet-Loso and Tautz (2003) as male-, female-, or nonsex-biased, then compared evolutionary rates among genes of the three expression classes (Table 2 and S1; Figure 2). These comparisons clearly indicated an increased rate of evolution in male-biased genes relative to female-biased genes. The average value of d_N/d_S was three-fold higher for male-biased genes than for female-biased genes, while that of d_N was over five-fold higher. Male-biased genes also showed faster evolutionary rates than nonsex-biased genes, with average values of d_N/d_S and d_N both differing by a factor of two between the two expression classes. In contrast, female-biased genes had lower values of d_N/d_S and d_N than nonsex-biased genes. The differences in d_N/d_S and d_N were significant for all comparisons and were highly significant for comparisons between male- and female-biased genes (Table 2). Using more and less stringent fold-change cutoffs (threefold and 1.5-fold, respectively) to classify genes as male-, female-, or nonsex-biased produced results consistent with the twofold analysis (Table 2 and Figure 2).

I also observed significant differences in synonymous substitution rates among the three classes of genes: Male-biased genes had significantly higher d_S than nonsex-biased genes, while female-biased genes had significantly lower d_S than nonsex-biased genes (Table 2 and S1; Figure 2). To investigate whether or not these differences could be explained by differing selective constraints on synonymous codon usage among genes of the three expression classes, I determined levels of codon bias for all genes by two measures (Table 2; see *Materials and Methods*). For both ENC (where lower values indicate greater bias) and F_{op} (where higher values indicate greater bias), female-biased genes showed the greatest codon usage bias, while male-biased genes showed the least. The differences in codon bias were significant for all comparisons (Table 2 and S1; Figure 2) and indicated an inverse

Table 2. Evolutionary rates and levels of codon bias for genes with sex-biased expression compared between *D. melanogaster* and *D. yakuba.*

		Male-biased	Female-biased	Nonsex-biased	P_{MF}[b]	P_{MN}[b]	P_{FN}[b]
1.5-fold[a]	Number	60	102	75			
	d_N	0.044	0.010	0.017	<0.0001	0.0007	0.0033
	d_S	0.366	0.221	0.276	<0.0001	0.0002	0.0008
	d_N/d_S	0.110	0.043	0.062	<0.0001	0.0078	0.0271
	ENC[c]	45.79	40.15	42.39	<0.0001	0.0114	0.0316
	F_{op}[d]	0.546	0.655	0.614	<0.0001	0.006	0.0102
2-fold[a]	Number	33	78	126			
	d_N	0.053	0.009	0.020	<0.001	0.002	<0.001
	d_S	0.359	0.193	0.297	<0.001	0.034	<0.001
	d_N/d_S	0.129	0.042	0.065	<0.001	0.006	<0.001
	ENC[c]	46.10	39.27	42.74	<0.001	0.038	<0.001
	F_{op}[d]	0.549	0.667	0.611	<0.001	0.032	<0.001
3-fold[a]	Number	22	59	156			
	d_N	0.063	0.007	0.020	<0.0001	0.0034	<0.0001
	d_S	0.374	0.189	0.294	<0.0001	0.0102	<0.0001
	d_N/d_S	0.151	0.038	0.065	<0.0001	0.0076	00.0001
	ENC[c]	47.45	38.65	42.94	0.0001	0.0128	0.0001
	F_{op}[d]	0.495	0.680	0.606	<0.0001	0.0037	<0.0001

[a] Sex-biased genes were defined as those with greater than 1.5, 2 and 3-fold expression difference between the sexes in either data set, respectively. Nonsex-biased genes were those with less than 1.5, 2 and 3-fold difference in both data sets, respectively..
[b] Two-tailed *P* from Mann-Whitney test of male-biased versus female-biased (P_{MF}), male-biased versus nonsex-biased (P_{MN}), and female-biased versus nonsex-biased (P_{FN}).
[c] Effective number of codons (Wright 1990).
[d] Frequency of optimal codons (Ikemura 1981).

relationship between level of codon bias and d_S for the three classes of genes (Figure 3). Furthermore, there was a significant negative correlation between F_{op} and d_S within the female-biased genes (Figure 3D), suggesting that selection for optimal codon usage might be responsible for the reduced synonymous substitution rate observed for this class of genes. It is also possible that the fixation of strongly-selected amino acid replacements results in the fixation of linked,

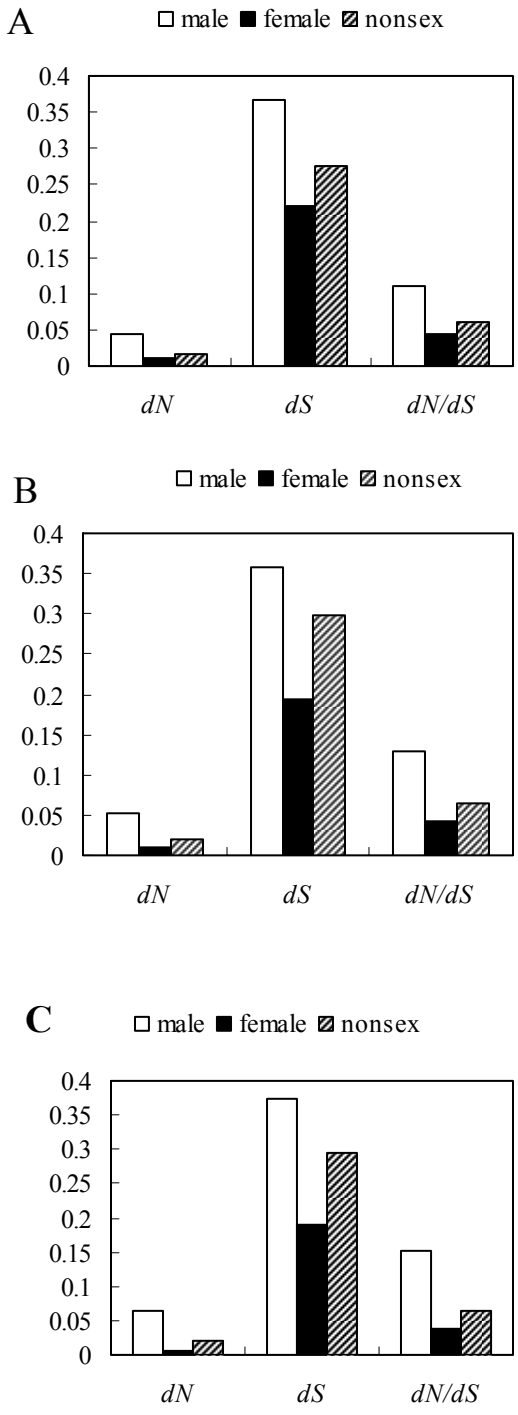

Figure 2. Nonsynonymous substitution rates (d_N), synonymous substitution rates (d_S) and evolutionary rates (d_N/d_S) estimated by divergence between *D. melanogaster* and *D. yakuba* (Domazet-Loso and Tautz 2003) for sex-biased genes with different sex-biased classification
(A) 1.5-fold
(B) 2-fold
(C) 3-fold

weakly-deleterious nonsynonymous substitutions (Betancourt and Presgraves 2002; Kim 2004). Consistent with this, I observed a significant negative correlation between F_{op} and d_N within the female-biased genes (Figure 3C) and the nonsex-biased genes. There was also a negative correlation between F_{op} and d_N within the male-biased genes (Figure 3A), although this correlation was not significant due to the small sample size and the absence of male-biased genes with high levels of codon bias. The

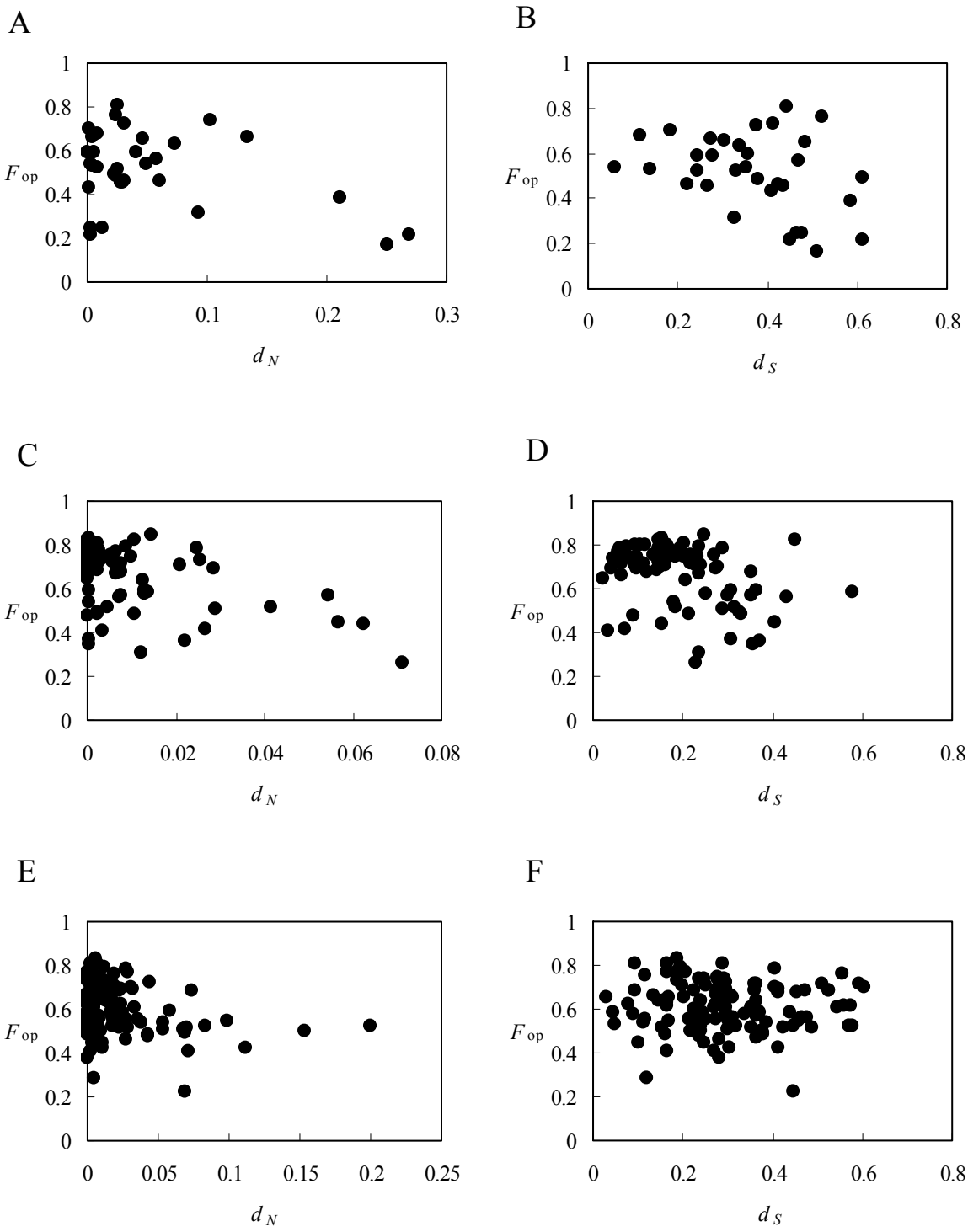

Figure 3. Correlation between substitution rates and levels of codon bias for male-, female- and nonsex-biased genes compared between *D. melanogaster* and *D. yakuba.* Plot of F_{op} versus d_N for (A) Male-biased genes (Spearman rank correlation test, R=-0.103, P = 0.56). (C) Female-biased genes (R=-0.36, P = 0.002). (E) Nonsex-biased genes (R=-0.015, P = 0.017), respectively. Plot of F_{op} versus d_S for (B) Male-biased genes (R=-0.316, P = 0.07). (D) Female-biased genes (R=-0.31, P = 0.007). (F) Nonsex-biased genes (R=-0.032, P = 0.72), respectively. F_{op} (frequency of optimal codons; Ikemura 1981) was calculated based on *D. melanogaster* codon usage.

use of different fold-change cutoffs for sex-biased classification does not change the result of codon bias rank among the three classes (Table 2).

Previous studies in *Drosophila* demonstrated a significant paucity of male-biased genes and a significant excess of female-biased genes on the *X* chromosome (Swanson et al. 2001; Parisi et al. 2003; Ranz et al. 2003). In my comparison, 6% (2/33) of the male-biased genes were on the *X* chromosome, while 15% (12/78) of the female-biased genes were on the *X*. Thus, a general tendency for slower evolution of *X* chromosomal genes (Orr and Betancourt 2001) could potentially explain my observations. Consistent with previous observations (Betancourt et al. 2002), I observed slightly lower (though not significantly so) evolutionary rates for *X*-linked genes (Figure 4). However, this pattern cannot explain the observed evolutionary rate differences between male- and female-biased genes. Considering only the autosomal genes, the average values of d_N/d_S, d_N, and d_S, for male-biased genes are 0.13, 0.05, and 0.35, while the corresponding values for female-biased genes are 0.05, 0.01, and 0.20. All of these differences are significant (Mann-Whitney test, $P < 0.001$). The small number of *X*-linked male-biased genes (two) precluded testing for evolutionary rate differences between male- and female-biased genes located on the *X* (Figure 5).

1.3.2 Comparison of orthologous genomic regions among *Drosophila* species

To assess the impact of comparative sequence data on genome annotation, Bergman et al. (2002) sequenced fosmid clones (≈40 kb each) corresponding to the *D. melanogaster* genomic regions containing the *apterous*, *even-skipped*, *fushi tarazu*, *twist*, and *Rhodopsin 1*, *2*, *3*, and *4* genes in four diverse *Drosophila* species: *D. erecta*, *D. pseudoobscura*, *D. willistoni*, and *D. virilis* (Table S2). These autosomal regions contain 81 known or predicted genes in *D. melanogaster*, however orthologous sequences from all of these genes were not obtained from every species due to either incomplete overlap of fosmid clones or genomic rearrangements between species. Based on microarray data, I was able to classify 50 of the genes as male-, female-, or nonsex-biased in their expression and compare evolutionary rates among the three classes (Figure 6A). Using all available sequences and assuming a constant d_N/d_S over all branches of the phylogenetic tree, I calculated average d_N/d_S values of 0.11, 0.02, and 0.07 for male-, female-, and nonsex-biased genes, respectively. The difference between male- and female-biased genes was significant (Mann-Whitney test, $P = 0.04$), though comparisons among other classes were not significant ($P > 0.05$) due to

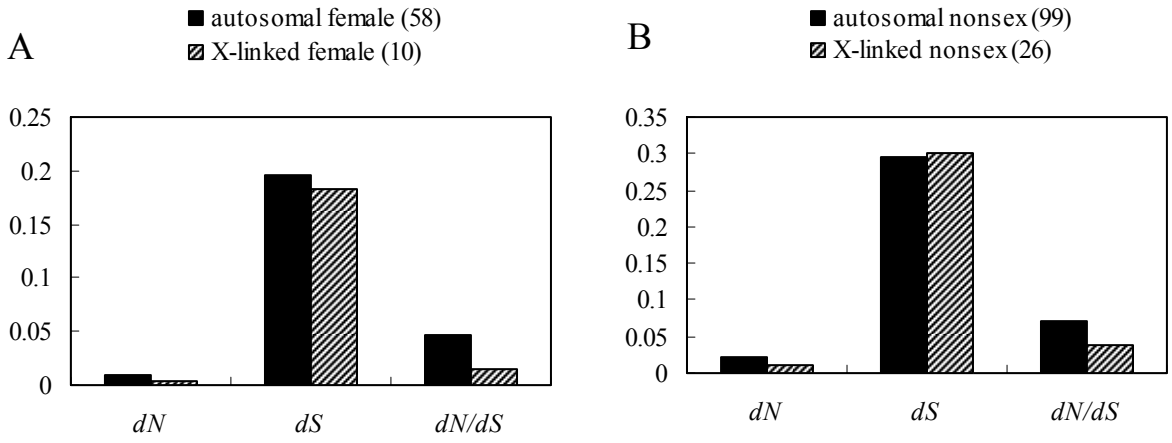

Figure 4. Nonsynonymous substitution rates (d_N), synonymous substitution rates (d_S) and evolutionary rates (d_N/d_S) estimated by divergence between *D. melanogaster* and *D. yakuba* (Domazet-Loso and Tautz 2003) for comparison of (A) autosomal female-biased genes and X-linked female-biased genes. (B) autosomal nonsex-biased genes and X-linked nonsex-biased genes, respectively. (2-fold cutoff used)

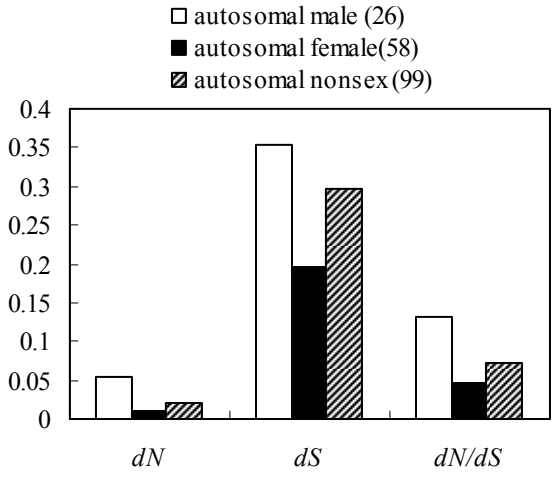

Figure 5. Nonsynonymous substitution rates (d_N), synonymous substitution rates (d_S) and evolutionary rates (d_N/d_S) estimated by divergence between *D. melanogaster* and *D. yakuba* (Domazet-Loso and Tautz 2003) for autosomal sex-biased genes (2-fold cutoff used).

the limited sample size. For 30 genes (eight male-, four female-, and 18 nonsex-biased) for which sequences from three or more species were available, I applied a "free-ratio" model (Yang 1998) that allowed d_N/d_S to vary over all branches of the phylogenetic tree. This model did not provide a significantly better fit to the data for any of the female-biased genes, but did provide a significantly better fit for four (22%) of the nonsex-biased genes and five (63%) of the male-biased genes. This indicates significant evolutionary rate heterogeneity among lineages, particularly for male-biased genes. If d_N/d_S is calculated as an average over all branches (scaled by their length), then the average d_N/d_S values for male-, female-, and nonsex-biased genes are 0.14, 0.02, and 0.10, respectively. The difference between male- and female-

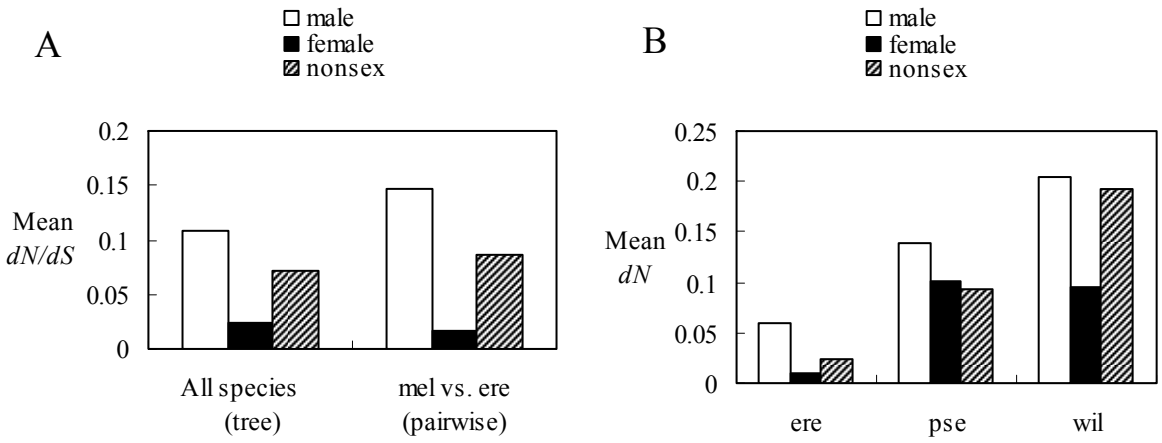

Figure 6. Evolutionary rates of sex-biased genes in the *apterous, even-skipped, fushi tarazu, twist,* and *Rhodopsin 1, 2, 3,* and *4* genomic regions (Bergman et al. 2002). (A) Mean d_N/d_S assuming a constant evolutionary rate over all branches of the phylogenetic tree for all available sequences (*D. melanogaster, D. erecta, D. pseudoobscura, D. willistoni,* and *D. virilis*) and for pairwise comparisons of *D. melanogaster* (mel) and *D. erecta* (ere). (B) Mean d_N for pairwise comparisons of *D. melanogaster* versus *D. erecta* (ere), *D. pseudoobscura* (pse), and *D. willistoni* (wil).

biased genes is significant (Mann-Whitney test, $P = 0.03$), though other comparisons are not significant.

I also calculated d_N/d_S for all pairwise comparisons of species. However, these values were not informative for most comparisons due to saturation of d_S. For example, comparisons of *D. melanogaster* to *D. erecta*, *D. pseudoobscura*, and *D. willistoni* produced average d_S values of 0.31, 4.9, 16.0, respectively. The only pairwise comparison that did not show saturation at synonymous sites was between *D. melanogaster* and *D. erecta*; this comparison also indicated an increased evolutionary rate in male-biased genes relative to female-biased genes (Mann-Whitney test, $P = 0.02$; Figure 6A). Female-biased genes had a lower average d_N/d_S than nonsex-biased genes, though this difference was not significant. Due to the saturation of d_S, I considered only d_N for the other pairwise species comparisons. The average d_N values for comparisons of *D. melanogaster vs. D. erecta*, *D. pseudoobscura*, and *D. willistoni* are shown in figure 6B (*D. virilis* was not included because it shared only one female-biased gene in common with *D. melanogaster*). In all cases, male-biased genes had higher nonsynonymous substitution rates than female-biased and nonsex-biased genes, while female-biased genes tended to have nonsynonymous substitution rates lower than nonsex-biased genes. However, these differences were not significant due to the limited number of genes in each comparison.

1.3.3 Comparison of highly sex-biased genes between *D. melanogaster* and *D. pseudoobscura*

In order to investigate the evolutionary rates of genes showing the most extreme levels of sex-biased expression, I extracted the 50 genes with the highest and lowest male/female expression ratios from both the Parisi et al. (2003) and Ranz et al. (2003) data sets. As a control, the 50 genes showing male/female ratios closest to one were extracted from each data set. After removing redundancies, the final list contained 93 male-, 92 female-, and 99 nonsex-biased genes (Table 3 and S3). The coding sequences from these genes were used for a BLAST search of the recently completed *D. pseudoobscura* genome. The three sex-bias classes showed significant differences in the number of genes with BLAST matches over a wide range of e-value cut-offs (Figure 7). In all cases, male-biased genes showed the least conservation between the two species. For example, using a conservative e-value cut-off of 10^{-9}, 46% of the male-biased genes did not have a significant BLAST match. The corresponding numbers for female-biased and nonsex-biased genes were 20% and 4%, respectively. Female-biased genes were less conserved than nonsex-biased genes for all cut-off values (Figure 7), but even in the most extreme case (e-value of cut-off of 10^{-6}) the difference between female- and nonsex-biased genes was not significant ($X^2 = 1.51$, $P = 0.22$).

Due to the high divergence of sex-biased genes (particularly those with male bias) between *D. melanogaster* and *D. pseudoobscura*, I was unable to align open reading frames for many of the genes, and thus could not quantify divergence in terms of d_N or d_S. However, I could use the BLAST score (Altschul et al. 1997) to get an estimate of combined synonymous and nonsynonymous divergence in exon sequences between the two species. Male-biased genes had significantly lower scores than both female and nonsex-biased genes (table 3), indicating a greater divergence of male-biased genes relative to genes of the other two expression classes. Female-biased genes had slightly higher scores than nonsex-biased genes, though this difference was not significant (Table 3). Levels of codon bias in genes of the three classes were inversely related to divergence, with male-biased genes having significantly less codon bias than female- and nonsex-biased genes (Table 3). Female-biased and nonsex-biased genes were nearly identical in their levels of codon bias (Table 3).

The *X*/autosome distribution of the highly sex-biased genes was even more

Table 3. Sequence conservation and levels of codon bias for highly sex-biased genes compared between *D. melanogaster* and *D. pseudoobscura.*

		Male-biased	Female-biased	Nonsex-biased	P_{MF}[a]	P_{MN}[a]	P_{FN}[a]
All genes	Number[b]	93	92	99			
	Score[c]	131	302	287	<0.001	<0.001	0.429
	ENC[d]	52.1	46.5	47.2	<0.001	<0.001	0.944
	*F*op[e]	0.47	0.57	0.55	<0.001	<0.001	0.936
Autosomal genes	Number[b]	92	68	76			
	Score[c]	132	288	268	<0.001	<0.001	0.496
	ENC[d]	52.1	47.04	47.53	<0.001	<0.001	0.2611
	*F*op[e]	0.47	0.56	0.55	<0.001	<0.001	0.2483

[a] Two-tailed *P* from Mann-Whitney test of male-biased versus female-biased (P_{MF}), male-biased versus nonsex-biased (P_{MN}), and female-biased versus nonsex-biased (P_{FN}).
[b] Combination of top 50 genes in each expression class from Parisi et al. (2003) and Ranz et al. (2003) after removing overlapping genes.
[c] Mean Blast score (Altschul et al. 1997).
[d] Effective number of codons (Wright 1990).
[e] Frequency of optimal codons (Ikemura 1981).

extreme than for the *D. yakuba* cDNA comparisons (see above). One percent (1/93) of the highly male-biased genes were on the *X* chromosome, while 26% (24/92) of the highly female-biased genes were on the *X*. However, these differences in chromosomal distribution cannot explain the observed differences in evolutionary rates. If only autosomal genes are considered, male-biased genes still show significantly fewer BLAST matches between species (χ^2 = 6.1, P = 0.01) and have significantly lower BLAST scores (Mann-Whitney test, $P < 0.001$) than female genes (Figure 7B).

1.3.4 Polymorphism and divergence in sex-biased genes

The most plausible explanation for the observed evolutionary rate differences among male, female, and nonsex-biased genes is variation in the strength and/or type of natural selection acting on genes of the three expression classes. One possibility is that male-biased genes are under relaxed selective constraints relative to genes of the other two classes, and thus accumulate a larger fraction of neutral amino acid

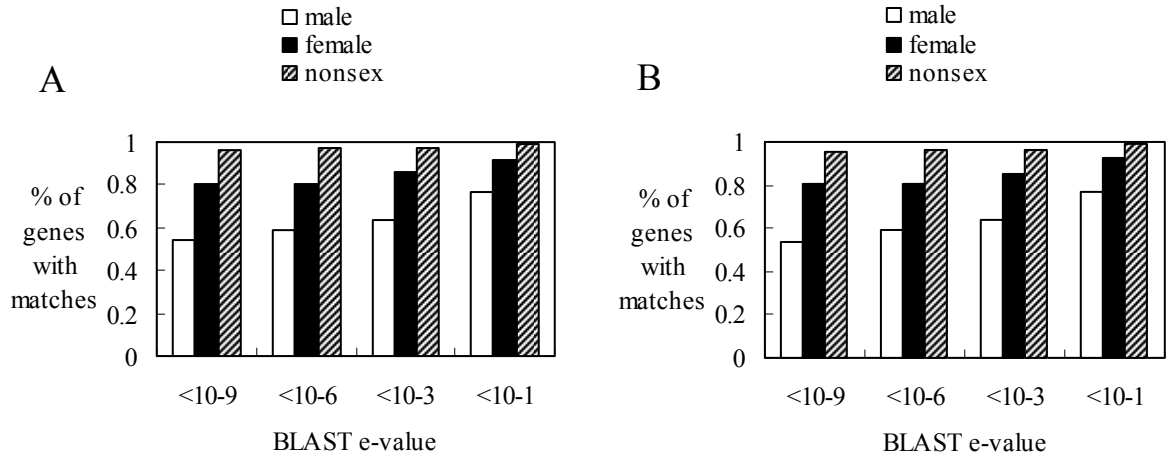

Figure 7. Conservation of sex-biased genes between *D. melanogaster* and *D. pseudoobscura*. Coding sequences of *D. melanogaster* genes showing the strongest male-biased or female-bias in expression were used for a Blast search of the *D. pseudoobscura* genome. The 50 genes showing male/female expression ratios closest to one in the same data sets were used as nonsex-biased controls. (A) all genes. The distribution of genes among expression classes differs significantly from the random expectation (χ^2 test; $P < 0.05$) for all e-values shown except 10^{-1} ($\chi^2 = 2.59$; $P = 0.27$). (B) only autosomal genes considered. The distribution of genes among expression classes differs significantly from the random expectation (χ^2 test; $P < 0.05$) for e-values 10^{-9} and 10^{-6}. 10^{-3} with $\chi^2 = 5.12$ ($P = 0.08$) and 10^{-1} with $\chi^2 = 2.29$ ($P = 0.32$).

Table 4. Comparison of divergence (*D. melanogaster* versus *D. simulans*) and polymorphism (*D. melanogaster*) in sex-biased genes.

	Male-biased	Female-biased	Nonsex-biased	P_{MF}[a]	P_{MN}[a]	P_{FN}[a]
Number	13	12	30			
d_N/d_S	0.208	0.105	0.111	0.134	0.049	0.772
d_N	0.025	0.010	0.012	0.060	0.018	0.795
d_S	0.115	0.091	0.104	0.047	0.055	0.810
π_N/π_S	0.137	0.183	0.261	0.976	0.197	0.542
Sig. MK[b]	6 (46)	3 (25)	4 (13)	0.185	0.025	0.220
Pos.[c]	4 (31)	1 (8)	1 (3)	0.161	0.022	0.418

[a] *P*-value for comparisons of male versus female (P_{MF}), male versus nonsex-biased (P_{MN}), and female versus nonsex-biased (P_{FN}) by two-tailed Mann-Whitney test (top 4 rows) or Fisher's exact test (bottom 2 rows).

[b] Number (percentage) of genes with a significant MK (McDonald and Kreitman 1991) test result.

[c] Number (percentage) of genes with a significant MK test result consistent with positive selection (i.e., a relative excess of nonsynonymous divergence).

replacements between species. Alternatively, male genes could be subject to increased positive selection due to male-male or male-female interactions/conflicts and thus accumulate more adaptive amino acid substitutions between species. To distinguish between these two possibilities, I examined DNA sequence polymorphism in 55 *D. melanogaster* protein-encoding genes that could be classified as male-, female-, or nonsex-biased in their expression based on the microarray data (Table 4 and S4). The divergence of these genes between *D. melanogaster* and *D. simulans* showed the same pattern observed for the other comparative genomic data sets, with male genes showing the greatest divergence and female genes showing the least (Table 4). If the increased d_N/d_S ratio observed for male-biased genes were due to relaxed selective constraints, then one would expect male genes to show a corresponding increase in their ratio of nonsynoymous/synonymous polymorphism (d_N/d_S) relative to female and nonsex-biased genes. However, the opposite pattern was observed: Male genes had lower average d_N/d_S than both female and nonsex-biased genes (Table 4). Thus, the polymorphism data do not support a general reduction of selective constraint on male-biased genes.

The type of selection affecting a particular gene can be inferred by the MK test (McDonald and Kreitman 1991), which compares ratios of polymorphism to divergence for synonymous and nonsynonymous sites. An excess of nonsynonymous divergence relative to nonsynonymous polymorphism is indicative of positive selection, while the opposite is indicative of balancing selection. Six of 13 (46%) male-biased genes showed a significant departure from neutrality by the MK test, including four (31%) with departures consistent with positive selection (Table 4). This is a higher fraction than observed for the female- or nonsex-biased genes (Table 4) and suggests that the increased rate of amino acid replacement observed for male-biased genes may be due to increased positive selection on these genes.

1.4 Discussion

My analysis of the polymorphism data available in the literature suggests that positive selection is responsible for the accelerated rate of evolution observed for male-biased genes. However, these results should be interpreted cautiously for several reasons. First, they are based on published surveys of DNA polymorphism that used different sample sizes and population sampling schemes, including African and non-African samples. Thus the results may be affected by demographic factors, such as bottlenecks or population subdivision. Second, polymorphism is known to be much more sensitive to chromosomal environment (e.g. local recombination rate) than is divergence (Begun and Aquadro 1992), and due to the limited available data I was unable to partition genes based on chromosomal location. By considering only the ratio of nonsynonymous to synonymous polymorphism (and not the separate values), I could partially control for the above two factors. However, it is possible that these factors also influence the d_N/d_S ratio. Finally, there is likely an ascertainment bias is the polymorphism data present in the literature. Some of the genes may have been surveyed with an *a priori* expectation of positive (or balancing) selection based on functional or divergence data. In addition, there may be a publication bias towards genes that depart from neutrality rather than those that fit the neutral model. These limitations can be addressed in future studies that use common population samples and select genes based only on expression class without *a priori* expectations of selection.

In summary, I used results from two male *vs*. female competitive cDNA microarray hybridization experiments and four comparative genomic data sets to investigate evolutionary rates of genes with sex-biased expression in *Drosophila*. The results consistently indicated an accelerated rate of evolution in male-biased genes relative to female-biased and nonsex-biased genes. Furthermore, the available polymorphism data suggested that male-biased genes are more often targets of positive selection than genes of the other two expression classes. Taken together, these observations suggest that the rapid evolution of male-biased genes is driven more by male-male competition than by antagonistic co-evolution of male- and female-biased genes. However, since my classification of male and female genes was based solely

on relative expression levels, it is possible that the female counterparts of sexually antagonistic gene interactions were systematically underrepresented. For example, male-biased genes that influence female reproduction and behavior (such as accessory protein genes; Wolfner 1997) may be expressed at high levels in male reproductive tissues, while their female counterparts may show less sex-specificity or may be expressed in non-reproductive tissues. Additional functional studies are needed to determine if such expression asymmetries are common among genes with sexually antagonistic interactions.

Chapter 2

2. Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression

2.1 Introduction

Sexual dimorphism is common among higher eukaryotes and is thought to result from the differential action of natural (or sexual) selection on individuals of the two sexes. Darwin (1871) proposed that sexual selection, either through direct male-male competition or female mate choice, was responsible for the extravagant secondary sexual characteristics present in the males of many species. The rapid evolution of male reproductive traits also may play a role in the frequent occurrence of hybrid male sterility (Wu and Davis 1993). Modern molecular evolutionary studies suggest that sexual selection may affect a broad spectrum of sex-related genes (Civetta and Singh 1999; Singh and Kulathinal 2000; Swanson and Vacquier 2002). A recent study that used microarray data to identify *Drosophila* genes with sex-biased expression and compared their rates of evolution between species found that genes with male-biased expression had significantly higher rates of nonsynonymous substitution (d_N) than genes with female- or nonsex-biased expression (Zhang et al. 2004). The accelerated evolutionary rate of male-biased genes could have two different explanations. One possibility is that they are subject to less selective constraint than female- or nonsex-biased genes, allowing them to accumulate more neutral (or slightly deleterious) amino acid changes. Alternatively, male-biased genes could be frequent targets of positive (or sexual) selection, and thus accumulate more adaptive amino acid changes. An analysis of the available polymorphism data from *D. melanogaster* supported the latter explanation: male-biased genes did not show an elevated level of nonsynonymous polymorphism as would be expected under the relaxed selective constraint hypothesis, but instead showed evidence of being subject to increased positive selection. However, there were several limitations to this analysis that made

this conclusion questionable. For example, the polymorphism comparison used a small number of sex-biased genes that were collected from a survey of the literature and, thus, included only a small fraction of the sex-biased genes in the genome. Furthermore, these genes likely were not a random sample because some were investigated with an a priori expectation of positive selection.

A further way to distinguish the evolutionary forces responsible for the increased d_N in genes with male-biased expression is to examine the relationship between d_N and local recombination rate. 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 (Hill and Robertson 1966; Marais and Charlesworth 2003). In contrast, if most amino acid replacements are neutral (or slightly deleterious), then there should be no correlation (or a negative correlation) between d_N and recombination rate. This is because purifying selection is less effective in regions of lower recombination for the same reason given above.

Betancourt and Presgraves (2002) investigated the relationship between d_N and recombination rate for 255 orthologous genes of *D. melanogaster* and *D. simulans*. They found that genes in regions of high recombination had significantly higher d_N than those in regions of low recombination. Their study focused primarily on a subset of male-expressed genes, namely those encoding accessory gland proteins (*Acps*), and suggested that most of the amino acid replacements in these genes were adaptive. When the *Acp* genes were removed from the analysis, there was no longer a significant difference in d_N between genes in regions of high and low recombination. Recently, Marais et al. (2004) investigated the relationship between d_N and recombination rate for 630 genes compared between *D. melanogaster* and *D. yakuba*. For this data set, which was not enriched for male-expressed genes, a slightly negative correlation between d_N and recombination rate was observed. The results of the two above studies suggest that the evolution of most genes is governed primarily by purifying selection, but that positive selection is responsible for the rapid evolution seen in male *Acp* genes. To investigate whether such adaptive evolution occurs in male-biased genes in general, I classified the genes used in Marais et al. (2004) as male-, female-, or nonsex-biased in their expression and examined the relationship between d_N and recombination rate separately for each group.

2.2 Materials and methods

2.2.1 Sequence data.

The data set analyzed in this study comes from Marais et al. (2004). Their study used 774 orthologous sequence pairs from *D. melanogaster* and *D. yakub*a. The latter were derived from two cDNA libraries (adult and embryo) of *D. yakuba* (Domazet-Loso and Tautz 2003), while the former were from the complete genome sequence of *D. melanogaster* (release 2). After correcting for redundant genes between the two cDNA libraries, they arrived at a final data set consisting of 630 orthologous sequence pairs from *D. melanogaster* and *D. yakuba*. This data set presumably is overrepresented for highly expressed genes, which are more abundant in cDNA libraries. The d_N and d_S values were obtained using the codeml program of the PAML package (Yang 1997) with the default parameters. Eight measures of the rate of crossing-over in *D. melanogaster* were taken from various published studies, compiled by Marais et al. (2003). In all cases, recombination rate is estimated by using Marey maps: the genetic positions (in centiMorgans, cM) and physical positions (in megabases, Mb) of markers that have been localized on both kinds of maps are plotted. The recombination rate (cM/Mb) at a given position of the chromosome is derived from the slope of the curve at that position. Two approaches are implemented to estimate the slope: the polynomial method (CK00, HK-p, KH93, MMD01) and the sliding window method (CC99, HK-w, RTE, ACE). There are large differences between these estimates, although they all co-vary, as Hey and Kliman (2002) noticed. By now it is not yet clear which of these approaches is the most accurate. To ensure that my results are not simply an artifact of the method used to estimate recombination rate, I perform my analyses using all eight of the recombination rate estimates. Of the 630 genes in the Marais et al. (2004) data set, 517 had information for both local recombination rate and evolutionary rate. These 517 genes were used for my analyses.

2.2.2 Identification of genes with sex-biased expression

The method used here to identify genes with sex-biased expression is the same as in Chapter 1. Briefly, two independent data sets that compared male and female gene

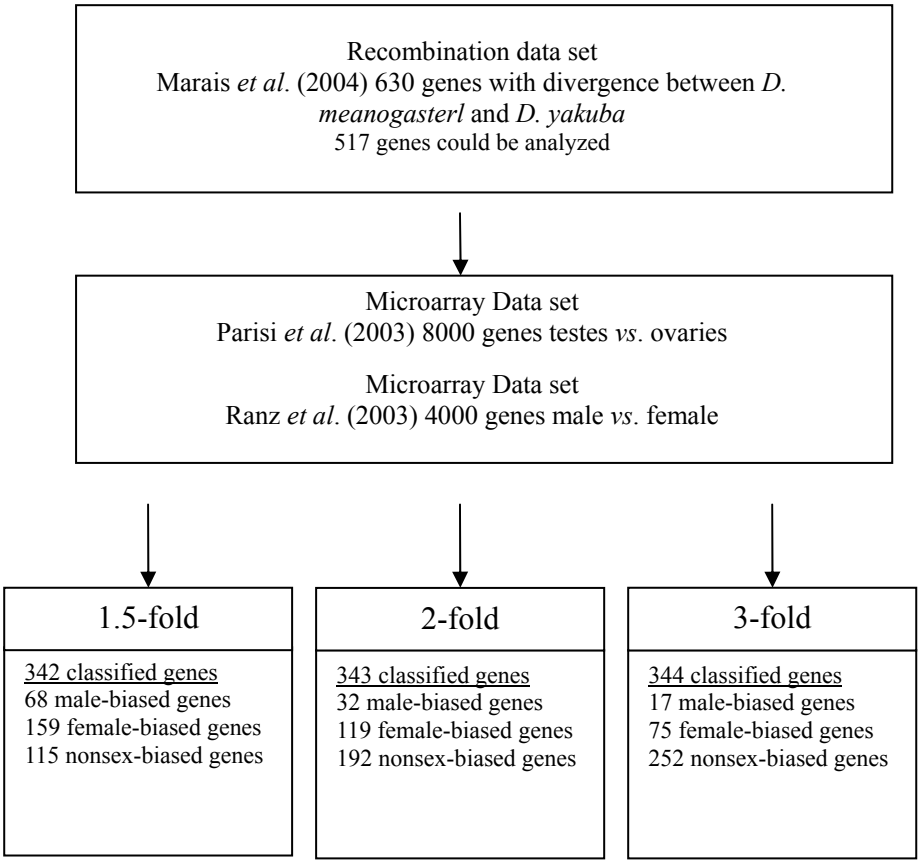

Figure 8. The construction of the data set analyzed in this study. Sex-biased genes were defined as those with greater than 1.5-fold, twofold and threefold expression difference between the sexes in either data set. Nonsex-biased genes were those with less than the correspondig fold difference in both data sets. The number of classified genes for each cutoff is given at the bottom.

expression in *D. melanogaster* through competitive microarray hybridizations were used to classify genes as male-, female-, or nonsex-biased in their expression (Parisi et al. 2003; Ranz et al. 2003). To be conservative, I used three different fold-change cutoffs of varying stringency (1.5-fold, twofold, and threefold) to classify genes with sex-biased expression. Of the 517 orthologous gene pairs, 342, 343 and 344 genes could be classified as male-, female-, or nonsex-biased using the 1.5-fold, twofold, and threefold cut-offs, respectively (Figure 8).

2.2.3 Analysis of the correlation between recombination rate and evolutionary rate

A correlation coefficient test (http://www.fon.hum.uva.nl/Service/Statistics/Correlation_coefficient.html) was used

to determine the degree of correlation between recombination rate and evolutionary rate under the different fold-change cutoffs used to classify sex-biased genes. The correlation coefficient is also known as the product-moment coefficient of correlation or Pearson's correlation. For a set of variable pairs, the correlation coefficient gives the strength of the association. It is assumed that the values of both members of the pairs are Normal (bivariate) distributed. The value of the correlation coefficient is between -1 and 1. The correlation coefficient R of the pairs (x , y) is calculated as:

$$R = \frac{\sum_i (x_i y_i) - \frac{\sum_i x_i \sum_i y_i}{N}}{\sqrt{(\sum_i x_i^2 - \frac{(\sum_i x_i)^2}{N})(\sum_i y_i^2 - \frac{(\sum_i y_i)^2}{N})}}$$

The regression line $y = a * x + b$ is calculated as:

$$a = \frac{\sum_i (x_i y_i) - \frac{\sum_i x_i \sum_i y_i}{N}}{\sum_i x_i^2 - \frac{(\sum_i x_i)^2}{N}} \qquad b = \frac{\sum_i y_i}{N} - \frac{a \sum_i x_i}{N}$$

The null hypothesis is that the values of the members of the pairs are uncorrelated, i.e., there are no linear dependencies and the correlation coefficient is close to zero. A positive value of the correlation coefficient indicates positive correlation, with higher values indicating stronger correlation. A negative value of the correlation indicates negative correlation, with more negative values indicating stronger correlation.

2.3 Results and discussion

My final data set consisted of 343 genes for which both the ratio of male to female expression (Parisi et al. 2003; Ranz et al. 2003) and an estimate of the local recombination rate were available (Marais et al. 2004). Because comparable estimates of gene expression and local recombination rates were not available for *D. yakuba*, all of my estimates are for *D. melanogaster*. The average d_N of the male-biased genes was about four-fold greater than that of the female-biased genes and about twofold greater than that of the non-sex-biased genes (Table 5). Comparisons of d_N for all pairwise combinations of the three groups were significant (Mann-Whitney test, $P \leq 0.01$), and these results held whether a 1.5-fold, 2-fold, or 3-fold expression difference between the sexes was used to define genes as sex-biased. Similarly, the synonymous substitution rate (d_S) was significantly higher for male-biased genes than for female- and non-sex-biased genes over all cutoffs ($P \leq 0.01$). These results cannot be explained solely by an increased mutation rate in male-biased genes, because d_N/d_S for male-biased genes was also significantly higher than that of both female- and non-sex-biased genes ($P \leq 0.05$; with the exception of male- vs. nonsex-biased at 1.5-fold where $P = 0.18$). The observed differences in evolutionary rate are consistent with those reported previously for a partially overlapping set of genes compared between *D. melanogaster* and *D. yakuba* (Zhang et al. 2004). In total, 161 genes are common to the two studies.

To examine the type of natural selection acting on genes of the three sex-biased expression classes, I looked at the relationship between d_N and local recombination rate within each class of genes. For this, I used eight different estimators of local recombination rate (Marais et al. 2004). All eight estimators were positively correlated with d_N for male-biased genes, and in most cases this correlation was significant (Table 6, Figure 9, and Figure 11). In contrast, all eight estimators were negatively correlated with d_N for female- and non-sex-biased genes, and this correlation was significant for many of the estimators (Table 6, Figure 9, and Figure 11). Furthermore, the correlation coefficients of the male- and female-biased genes were significantly different ($P < 0.05$) for seven of the eight estimators (Table 6). Differences between the correlation coefficients of female- and nonsex-biased genes were not significant

Table 5. Evolutionary rates and levels of codon bias for genes with sex-biased expression compared between *D. melanogaster* and *D. yakuba.*

		Male-biased	Female-biased	Nonsex-biased	P_{MF}[b]	P_{MN}[b]	P_{FN}[b]
1.5-fold[a]	Number	68	159	115			
	d_N	0.037	0.012	0.026	<.0001	0.0122	0.0003
	d_S	0.381	0.248	0.284	<.0001	<.0001	0.0007
	d_N/d_S	0.094	0.048	0.071	0.0006	0.1814	0.0026
	F_{op}[c]	0.605	0.679	0.635	<.0001	0.0294	0.0006
2-fold[a]	Number	32	119	192			
	d_N	0.043	0.012	0.021	<.0001	0.0012	<.0001
	d_S	0.381	0.233	0.305	<.0001	0.0002	<.0001
	d_N/d_S	0.114	0.046	0.069	0.0001	0.017	0.001
	F_{op}[c]	0.601	0.697	0.629	0.0002	0.1056	<.0001
3-fold[a]	Number	17	75	252			
	d_N	0.045	0.010	0.021	<.0001	0.0087	<.0001
	d_S	0.393	0.208	0.303	<.0001	0.0012	<.0001
	d_N/d_S	0.116	0.042	0.069	0.0003	0.0526	<.0001
	F_{op}[c]	0.574	0.734	0.630	0.0003	0.0485	<.0001

NOTE: Divergence data are from Marais et al. (2004). Expression data are from Parisi et al. (2003) and Ranz et al. (2003). Values represent the mean for each expression class.

[a] Fold expression difference used to classify genes as sex-bias.

[b] Two-tailed *P* from Mann-Whitney test of male *vs*. female (P_{MF}), male *vs*. non-sex-biased (P_{MN}), and female *vs*. non-sex-biased (P_{FN}).

[c] Frequency of optimal codons (Ikemura 1981).

for any recombination rate estimator.

I also observed a positive correlation between d_S and recombination rate for the male-biased genes (Table 6). A possible explanation for this is that there is an elevated rate of mutation in male-biased genes in regions of high recombination. However, mutational effects alone cannot explain my results, because d_N/d_S is also positively correlated with recombination rate (Table 6, Figure 9, and Figure 11). My results also cannot be explained by differences between X-linked and autosomal genes. Although male-biased genes are significantly underrepresented on the X chromosome, all of the above results are unchanged when only autosomal genes are considered (Figure 10).

Table 6. Correlation between d_N, d_S, d_N/d_S and recombination rate for genes with male-, female-, and nonsex-biased expression.

		Bias	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
1.5-fold	d_N	male	0.24*	0.26*	0.37**	0.32**	0.21	0.21	0.25*	0.12
		female	-0.14	-0.15	-0.18*	-0.18*	-0.12	-0.17*	-0.10	-0.20*
		nonsex	-0.04	-0.22*	-0.22*	-0.24**	-0.17	-0.23*	-0.14	-0.28**
	d_S	male	0.17	0.18	0.27*	0.26*	0.10	0.08	-0.05	0.24
		female	-0.13	-0.11	-0.11	-0.08	-0.12	-0.05	-0.08	-0.11
		nonsex	0.1	-0.04	0.04	-0.01	-0.06	-0.04	0.05	-0.20*
	d_N/d_S	male	0.19	0.21	0.29*	0.28*	0.14	0.18	0.25*	0.07
		female	-0.15	-0.14	-0.19*	-0.16*	-0.11	-0.17*	-0.13	-0.19*
		nonsex	-0.07	-0.19*	-0.19*	-0.21*	-0.10	-0.14	-0.09	-0.14
2-fold	d_N	male	0.28	0.24	0.24	0.35	0.28	0.29	0.37*	0.19
		female	-0.14	-0.13	-0.16	-0.14	-0.12	-0.14	-0.09	-0.21*
		nonsex	0.02	-0.01	0.03	-0.07	-0.03	-0.11	-0.06	-0.12
	d_S	male	0.33	0.22	0.29	0.29	0.20	0.19	0.07	0.26
		female	-0.23*	-0.12	-0.14	-0.10	-0.12	-0.09	-0.09	-0.15
		nonsex	0.12	0.07	0.11	0.09	-0.00	0.02	0.00	-0.00
	d_N/d_S	male	0.22	0.24	0.24	0.31	0.22	0.26	0.39*	0.21
		female	-0.13	-0.10	-0.16	-0.11	-0.11	-0.14	-0.10	-0.20*
		nonsex	-0.05	-0.08	-0.06	-0.12	-0.05	-0.11	-0.07	-0.18*
3-fold	d_N	male	0.48	0.26	0.23	0.41	0.47	0.30	0.37	0.17
		female	-0.11	-0.08	-0.06	-0.07	-0.09	-0.13	-0.09	-0.18
		nonsex	-0.01	-0.04	-0.01	-0.07	-0.06	-0.09	-0.03	-0.10
	d_S	male	0.02	-0.02	-0.03	0.04	0.02	0.11	-0.00	0.20
		female	-0.24*	-0.19	-0.15	-0.14	-0.18	-0.19	-0.23*	-0.16
		nonsex	0.09	0.07	0.08	0.08	0.00	0.03	0.01	0.01
	d_N/d_S	male	0.51*	0.34	0.30	0.45	0.49*	0.31	0.42	0.25
		female	-0.09	-0.03	-0.07	-0.03	-0.07	-0.10	-0.08	-0.17
		nonsex	-0.08	-0.11	-0.09	-0.13*	-0.09	-0.10	-0.05	-0.05

NOTE: The eight estimators of local recombination rate correspond to those from Marais et al. (2004). 1.5-, 2- and 3-fold cutoffs were used to classify genes as sex-biased. Gene number is the same as in Table 5. Numbers indicate Pearson's correlation coefficients.
* $p < 0.05$
** $p < 0.01$

My analyses indicate that male-biased genes are subject to different selective forces than female- and non-sex-biased genes. The positive correlation between d_N and recombination rate seen for male-biased genes suggests that they are often targets

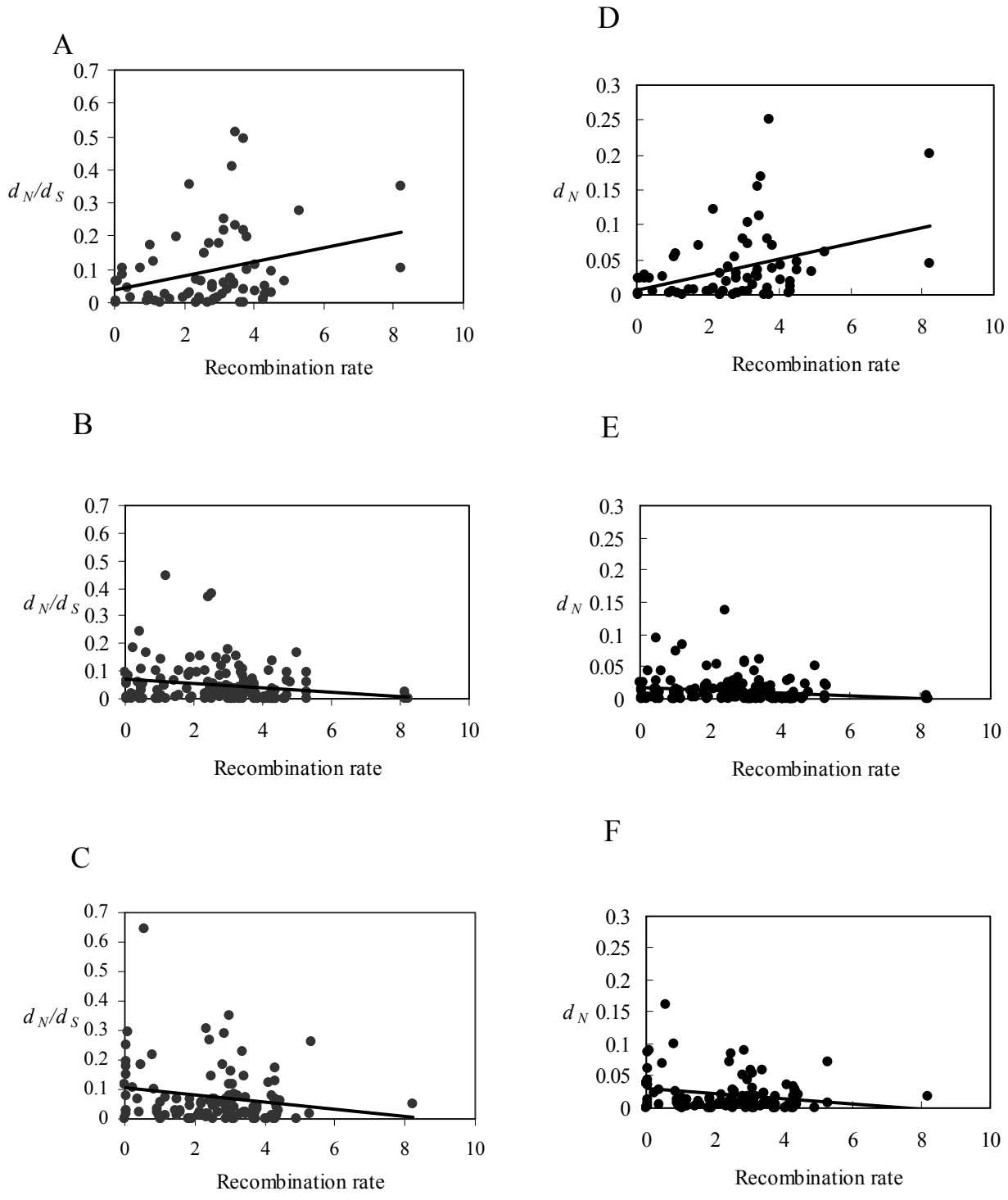

Figure 9. Correlation between evolutionary rate (d_N/d_S) and local recombination rate (measured using the HKw method; Marais et al. 2004) for (A) male-biased genes (R = 0.30, P = 0.01), (B) female-biased genes (R = -0.19, P = 0.01), and (C) nonsex-biased genes (R = -0.19, P = 0.04), and correlation between nonsynonymous substitution rate (d_N) and local recombination rate for (D) male-biased genes (R = 0.37, P = 0.01), (E) female-biased genes (R = -0.18, P = 0.02), and (F) nonsex-biased genes (R = -0.22, P = 0.02). (All graphs shown use the 1.5-fold cutoff to define sex-biased genes.)

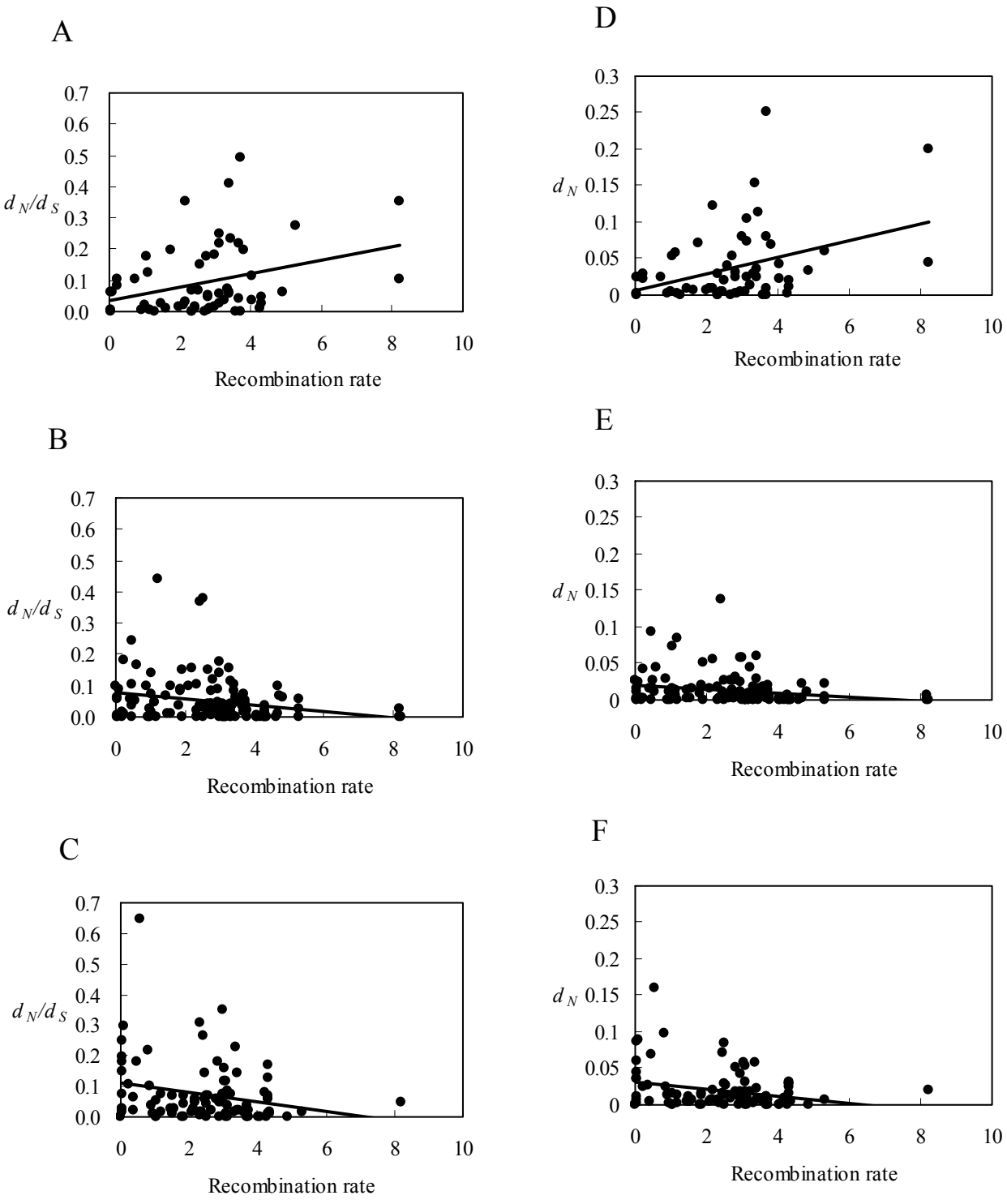

Figure 10. Correlation between evolutionary rate (d_N/d_S) and local recombination rate (measured using the HKw method; Marais et al. 2004) for (A) autosomal male-biased genes (R = 0.27, P = 0.04), (B) autosomal female-biased genes (R = -0.27, P = 0.002), and (C) autosomal nonsex-biased genes (R = -0.21, P = 0.04), and correlation between nonsynonymous substitution rate (d_N) and local recombination rate for (D) autosomal male-biased genes (R = 0.38, P = 0.002), (E) autosomal female-biased genes (R = -0.22, P = 0.01), and (F) autosomal nonsex-biased genes (R = -0.28, P = 0.006). (All graphs shown use the 1.5-fold cutoff to define sex-biased genes.)

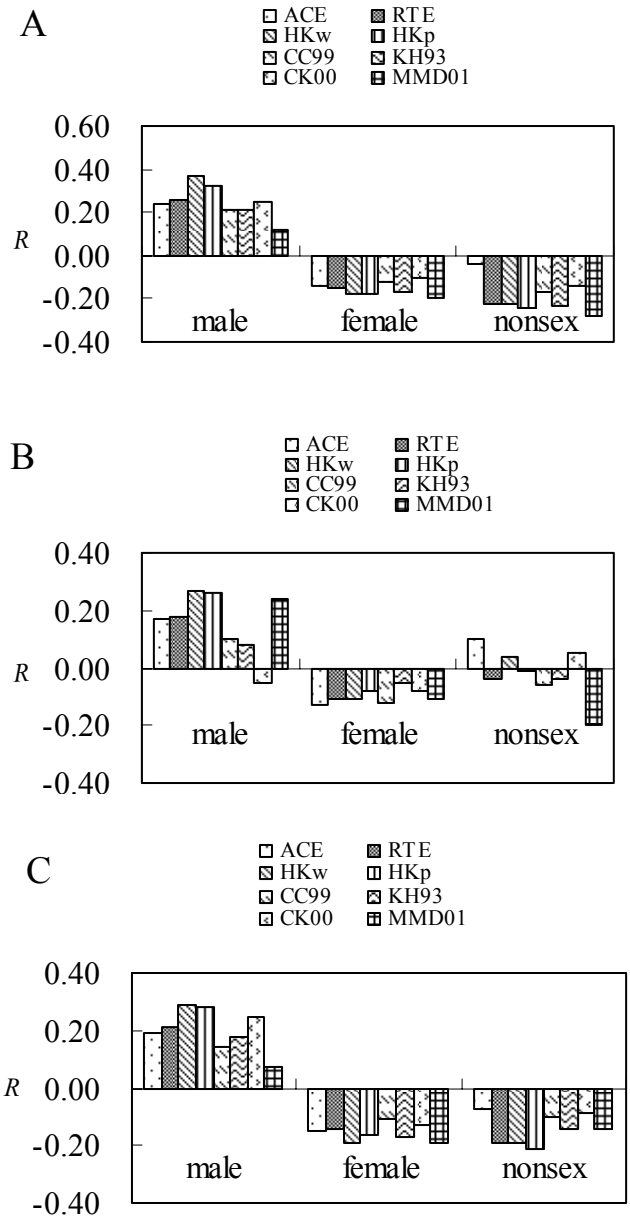

Figure 11. Correlation coefficient (R) between substitution rate and recombination rate for genes with male-, female-, and nonsex-biased expression. The eight estimators of local recombination rate correspond to those from Marais et al. (2004). (A) Correlation between recombination rate and d_N, (B) Correlation between recombination rate and d_S, (C) Correlation between recombination rate d_N/d_S. Genes classified using a 1.5-fold cutoff.

of positive selection, which is expected to be more effective in genomic regions with higher recombination rates due to a reduction of Hill-Robertson interference. In contrast, the negative correlation between d_N and recombination rate seen for female- and non-sex-biased genes suggests that these genes are predominantly subject to purifying selection, which is expected to be less effective in regions of lower recombination allowing fixation of more slightly deleterious mutations. The reduced efficacy of purifying selection in regions of reduced recombination is also expected to affect male-biased genes and would counteract the positive correlation between d_N

and recombination rate. Thus a significantly positive correlation, as is seen in my data, is a conservative criterion for the inference of positive selection.

Marais et al. (2004) observed a slightly negative correlation between d_N and recombination rate for 630 genes compared between *D. melanogaster* and *D. yakuba*. Their analysis, however, did not consider male-, female-, and non-sex-biased genes separately. Since the vast majority of genes are female- or nonsex-biased (table 1), the negative correlation in these genes would obscure any positive correlation present in the male genes.

My results are similar to those seen for male-specific Acp genes. However, the genes analyzed here came from a random EST survey and were classified only by their degree of sex-biased expression. They are not enriched for genes of a particular functional class. In fact, none of the male-biased genes used in my analysis match annotated Acp genes or putative Acp genes identified in an accessory gland-specific EST screen (Swanson et al. 2001). Thus it appears that rapid evolution due to positive selection is a general feature of male-biased genes, and is not limited to a relatively small set of Acp genes.

Chapter 3

3. Widespread adaptive evolution of *Drosophila* genes with sex-biased expression

3.1 Introduction

Males and females of animal species often differ in many morphological and behavioral traits. This sexual dimorphism has long fascinated biologists and served as the inspiration for Darwin's theory of sexual selection (Darwin 1871). Recent microarray studies have revealed that sexual dimorphism is also common at the level of gene expression (Parisi et al. 2003; Ranz et al. 2003; Gibson et al. 2004). For example, about 30% of all genes in *Drosophila melanogaster* show a twofold or greater difference in expression between the sexes (Parisi et al. 2004). Comparative genomic studies have shown that such sex-biased genes, particularly those with male-biased expression, are among the most rapidly evolving genes between species (Swanson et al. 2001; Zhang et al. 2004; Richards et al. 2005; Khaitovich et al. 2005). This raises the possibility that adaptive processes, such as sexual selection, may drive the evolution of a large number of genes with sexually dimorphic expression (Civetta and Singh 1999; Singh and Kulathinal 2000). An alternate possibility, however, is that sex-biased genes evolve under relaxed selective constraint, which allows them to accumulate more neutral (or nearly-neutral) changes between species. For instance, the product of an autosomal gene with sex-specific expression only will be visible to selection over half of its evolutionary history when it is in the appropriate sex. The rest of the time, it will be in the sex where it is not expressed and will be invisible to selection. Thus, it may experience only half as much purifying selection as a gene expressed equally in the two sexes (Barker et al. 2005).

In some well-studied cases, the rapid evolution of male-biased genes has been attributed to positive selection (Swanson and Vacquier 2002). In particular, the male reproductive genes of *Drosophila*, including those encoding accessory gland proteins

(*Acps*), appear to be a rich source of adaptively evolving genes (Tsaur and Wu 1997; Tsaur et al. 1998; Aguade 1998; Ting et al. 1998; Nurminsky et al. 1998; Aguade 1999; Begun et al. 2000; Betran and Long 2003). However, the evolutionary forces affecting the vast majority of male-biased genes are unknown. Although they have been less studied than male-biased genes, there is also evidence for positive selection driving the rapid evolution of particular female-biased genes (Swanson and Vacquier 2002). In these cases, either cooperative or antagonistic co-evolution between male and female reproductive proteins is thought to play an important role (Civetta and Singh 2005). For example, a survey of expressed sequence tags (ESTs) from the female reproductive tract of *D. simulans* uncovered a number adaptively evolving genes that may be the female counterparts of rapidly evolving male reproductive genes (Swanson et al. 2004).

A powerful method to distinguish the selective forces influencing a gene's evolution is to use combined polymorphism and divergence data (Parsch et al. 2005). Genes that have evolved adaptively are expected to show relatively little polymorphism within species, but high divergence between species. Genes under relaxed selective constraint, in contrast, should show higher levels of polymorphism within species that are proportional to their divergence between species. I have used this approach to determine the selective forces influencing the evolution of sex-biased genes. I surveyed DNA sequence polymorphism in 91 *D. melanogaster* genes with male-, female-, or nonsex-biased expression and also determined their divergence from the sister species *D. simulans*. Using statistical tests that compare ratios of polymorphism and divergence at synonymous and nonsynonymous sites, I inferred the type and strength of selection affecting the proteins encoded by genes of the three expression classes. I find that adaptive evolution is common among sex-biased genes (both male and female), but rare among nonsex-biased genes. This suggests that sexual selection and intersexual co-evolution play major roles in the genetic differentiation of species.

3.2 Materials and methods

3.2.1 Gene selection

Genes with sex-biased expression were selected on the basis of their male/female (or testes/ovaries) expression ratios, as determined by microarray experiments that used *D. melanogaster* (Parisi et al. 2003; Ranz et al. 2003; Gibson et al. 2004). For male-biased genes, I required that the ratio be greater than 2.0 (mean = 15.2), while for the female-biased genes I required a ratio less than 0.5 (mean = 0.23). In other words, I required at least a twofold expression difference between the sexes for a gene to be classified as sex-biased. Nonsex-biased genes were required to have a male/female expression ratio between 0.75 and 1.25 (mean = 1.01). In general, the male-biased genes showed more extreme expression differences between the sexes than the female-biased genes, reflecting the pattern that is seen genome-wide (Gibson et al. 2004; Parisi et al. 2004). Because the above three experiments used different microarray platforms, not all genes were represented in each experiment. However, for 44 (48%) of the genes, the sex-bias classification could be confirmed by all three experiments. An additional 43 (47%) genes were confirmed by two of the three experiments. The remaining genes (four male-biased genes) were confirmed by additional microarray experiments (Andrews et al. 2000; Stolc et al. 2004). Because only one of the above experiments also compared male and female expression in *D. simulans* (Ranz et al. 2003), I could not confirm the bias of all genes in this species. However, of the 61 genes with data from both species, 60 (98%) showed the same sex-bias classification. This included 22 male-biased genes, 25 female-biased genes, and 13 nonsex-biased genes. The one conflicting gene (CG4570) was female-biased in *D. melanogaster*, but nonsex-biased in *D. simulans*. This gene showed no evidence for selection (Table S6) and removing it from my analysis does not affect my results or conclusions. In addition to the expression criteria, genes were also selected to fall within a relatively narrow size distribution and to have similar intron/exon structures. This was done to remove the influence of coding sequence or intron length on the ratio of nonsynonymous/synonymous polymorphism or divergence (Comeron and Kreitman 2002; Comeron and Guthrie 2005). The mean lengths (standard deviations) for male-, female-, and nonsex-biased genes were 1006 (325), 1098 (372), and 821

Table 7. Overview of genes studied in the project.

Gene Group	M/F[a]	Overall expr. male[b]	Overall expr. female[c]	Overall expr. both[d]	Length[e]	Intron Number[f]	Intron Sites[g]	Rec. (Hkw)[h]	Rec. (Comeron)[i]
Male-biased (*N*=33)	15.22	12 (23.03)	1.09 (1.44)	6.54 (12)	1006 (325)	1.33 (1.02)	92 (80)	2.55 (1.62)	2.75 (1.62)
Female-biased (*N*=28)	0.23	1.67 (4.83)	7.21 (8.15)	4.44 (4.95)	1098 (372)	1.25 (1.32)	148 (234)	2.13 (1.19)	2.13 (1.42)
Nonsex-biased (*N*=30)	1.01	4.08 (5.29)	3.82 (4.68)	3.95 (4.98)	821 (167)	1.40 (1.13)	147 (127)	2.51 (1.52)	2.20 (1.51)

Note: the numbers in the table represent the mean (standard deviation) for each group.
[a] The average male/female expression ratio.
[b] Expression level relative to that of all other genes in the male.
[c] Expression level relative to all that of other genes in the female.
[d] Expression level relative to that of all other genes in both sexes.
[e] Gene length (including introns).
[f] Number of introns per gene.
[g] Number of intron sites per gene.
[h] Recombination rate (cM/Mb) from Hey and Kliman (2002) using sliding window method.
[i] Recombination rate (cM/Mb) from Comeron et al. (1999).

(167) bp, respectively (Table 7). Because male-biased genes are known to be underrepresented on the X chromosome (Parisi *et al*. 2003; Ranz et al. 2003), I limited my analysis to autosomal genes. It is important to note that functional information or measures of interspecific divergence were not considered in gene selection. Thus, aside from the selection criteria outlined above, my sample represents a random collection of sex-biased (and nonsex-biased) genes that is expected to be representative of the genome as a whole.

3..2.2 PCR and DNA sequencing

Oligonucleotide primers flanking the coding sequence of each gene were designed on the basis of the complete *D. melanogaster* genome sequence (release 4.0; http://www.flybase.org) and used for PCR with genomic DNA from 12 highly inbred *D. melanogaster* lines derived from Lake Kariba, Zimbabwe (Glinka et al. 2003), and one highly inbred *D. simulans* line derived from Chapel Hill, NC (Meiklejohn et al. 2004). A complete list of the PCR primers, as well as the cycling conditions used for each gene, is provided in table S8. PCR products were purified with ExoSAP-IT

Table 8. Summary of genes sequenced in my part of the project.

Gene	CGnum[a]	M/F Avg[b]	All sites[c]	Syn[d] sites	Non[e] sites	*Zim* lines[f]
Male-biased						
janB	CG7931	9.20	607	97	322	12
ocn	CG7929	6.64	515	71	265	12
T7	CG17956	5.03	192	44	124	12
T9	CG17376	12.50	594	51	150	11
T10	CG1980	6.95	853	132	603	11
T26	CG6332	23.88	1267	256	802	7
T27	CG3483	10.77	1185	276	897	11
T28	CG10307	7.94	1145	243	780	12
T29	CG10750	12.73	1176	222	771	9
T30	CG3085	19.81	1417	298	998	11
T40	CG5045	19.81	876	187	572	11
T41	CG5276	5.67	1314	301	956	12
T54	CG6036	13.66	1194	251	862	10
T55	CG13527	15.87	1046	210	660	9
Averages		12.18	955.79	188.50	625.86	10.71
Female-biased						
O25	CG1239	0.34	1134	204	696	12
O26	CG6554	0.24	1395	250	878	11
O27	CG7840	0.18	978	230	748	11
O28	CG9135	0.16	1586	345	1115	10
O29	CG5363	0.21	1357	203	688	11
O35	CG10206	0.18	1714	354	1176	8
O37	CG4299	0.24	1377	167	640	10
O38	CG4236	0.32	1475	287	1003	12
O49	CG17361	0.32	546	117	420	11
O51	CG32409	0.26	781	149	514	11
O52	CG6459	0.18	969	179	610	11
O53	CG3975	0.25	1299	292	1001	11

(Continues...)

Table 8. *(Continues)*

Gene	CGnum[a]	M/F Avg[b]	All sites[c]	Syn[d] sites	Non[e] sites	*Zim* lines[f]
Averages		0.24	1217.58	231.42	790.75	10.75
Nonsex-biased						
U11	CG13419	0.98	653	125	394	12
U12	CG7484	0.95	935	122	412	12
U13	CG9283	0.98	627	122	490	12
U14	CG13934	0.92	779	106	323	10
U15	CG17404	0.93	944	226	659	11
U16	CG10623	0.95	1139	241	752	12
U20	CG9822	1.02	920	181	608	12
U21	CG13845	1.02	894	151	548	11
U22	CG3683	1.04	822	121	405	12
U23	CG3476	0.98	1094	235	662	11
Averages		0.98	880.70	163.00	525.30	11.50

[a] CG number (*D. melanogaster* genome release 4.0).
[b] The average male/female expression ratio.
[c] Total number of sites sequenced, including introns.
[d] Number of synonymous sites.
[e] Number of nonsynonymous sites.
[f] Sample size (number of *D. melanogaster* alleles sequenced).

(USB, Cleveland, OH). Sequencing of PCR products (both strands) was carried out using BigDye chemistry and a 3730 automated sequencer (Applied Biosystems, Foster City, CA). The PCR primers were also used as sequencing primers. When necessary to get complete sequence coverage of the entire coding region, additional internal sequencing primers were used. A complete list of PCR and sequencing primers is available in supplementary Table S6. All new sequences have been deposited in the EMBL/GenBank databases under accession numbers AM293861-AM294919. I personally did PCR and sequencing of 14 male-biased, 12 female-biased, and 10 nonsex-biased genes (Table 8). For some genes, I was unable to get successful PCR or DNA sequence from all 12 *D. melanogaster* strains (see Table S5). The average number of strains sequenced per gene was 11. For 25 genes, I was unable to obtain a PCR product from *D. simulans* using my primers designed to *D. melanogaster*. In these cases, I used the sequence from the *D. simulans* genome project (Washington University School of Medicine Genome Sequencing Center)

downloaded from the UCSC Genome Browser (http://genome-test.cse.ucsc.edu/).

3.2.3 Analysis

Sequences were edited using either Sequencher (Gene Codes, Ann Arbor, MI) or DNAstar (Madison, WI) software with manual adjustments to the alignments. Polymorphism and divergence statistics were calculated using DnaSP 4 (Rozas et al. 2003). For MK table data, I used the number of segregating mutations (instead of the number of segregating sites), because some genes had sites with three segregating variants. In these cases, the frequency of each mutation was considered separately for calculation of Tajima's *D* and the identification of singleton polymorphisms. For divergence, I considered only sites with fixed differences between all *D. melanogaster* lines and *D. simulans*. The fraction of positively selected amino acid substitutions, α, its 95% confidence intervals, and a likelihood ratio test for positive selection were calculated using the program DoFE (kindly provided by A. Eyre-Walker). The selection parameter, γ, its 95% confidence intervals, and the proportion of the distribution falling below zero were calculated using the MKPRF web server (http://cbsuapps.tc.cornell.edu/mkprf.aspx). Multi-locus HKA and Tajima's *D* tests were performed using the program HKA, which was kindly provided by J. Hey (http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm).

My polymorphism survey revealed a few potential annotation errors in genome release 4.0. One female-biased gene (CG17361) had a frameshift-causing insertion (relative to the annotated ORF) in some *D. melanogaster* lines (2 bp) and in *D. simulans* (1 bp). This occurred 42 bp downstream of the start codon. The ORF was otherwise intact with both d_N/d_S and $\pi_N/\pi_S < 1$, suggesting that it is maintained by purifying selection. Another in-frame ATG codon is present 90 bp downstream of the annotated start codon and I used this as the starting point of my alignment. Two nonsex-biased genes (CG17404 and CG18553) had frameshift-causing deletions (1 and 2 bp, respectively) segregating in *D. melanogaster*. Both genes had otherwise intact ORFs with d_N/d_S and $\pi_N/\pi_S < 1$, suggesting functional constraint on the coding sequence. It is possible that these deletions fall within unannotated introns. For my analyses, I ignored these sites with deletions. Elimination of the three above genes from my analyses has negligible effect on my results and does not alter the conclusions of this paper.

3.3 Results

3.3.1 MK test for sex-biased genes

To investigate the type and strength of selection influencing the evolution of sexbiased genes, I surveyed DNA sequence polymorphism in 91 protein-encoding genes in a sample of 12 highly inbred *D. melanogaster* isofemale lines from Zimbabwe, Africa (Table 9 and Table S5). The genes were selected on the basis of previously published microarray results (Parisi et al. 2003; Ranz et al. 2003; Gibson et al. 2004), which allowed them to be separated into three expression classes: male-biased, female-biased, and nonsex-biased. For the sex-biased genes, I required at least a twofold difference in expression between the sexes, while for the nonsex-biased genes I required the difference to be less than 1.25-fold. In all cases, the expression difference was confirmed by at least two independent microarray experiments. The Zimbabwe population of *D. melanogaster* was chosen because it is an ancestral, near-equilibrium population that is expected to be largely free from confounding demographic factors, such as population expansion or subdivision (Glinka et al. 2003; Ometto et al. 2005). For each gene, I also determined interspecific divergence using a single sequence from *D. simulans*.

The combination of within-species polymorphism and between-species divergence data allows the application of powerful statistical methods to detect

Table 9. Summary of polymorphism and divergence statistics.

Bias	Number of genes	Significant MK tests[a]	Positive selection[b]	Ds[c]	Ps[d]	Dn[e]	Pn[f]	P-value[g]
Male	33	7	7	744	447	370	112	<0.0001
Female	28	6	3	631	233	299	90	0.15
Nonsex	30	1	0	436	267	121	83	0.51

[a] McDonald-Kreitman (MK) tests were performed for each gene and considered significant if $P < 0.05$.
[b] Genes with significant MK tests showing a relative excess of nonsynonymous fixed differences.
[c] The total number of synonymous fixed differences.
[d] The total number of synonymous polymorphisms.
[e] The total number of nonsynonymous fixed differences.
[f] The total number of nonsynonymous polymorphisms.
[g] The P-value of the summed data as determined by a $\chi 2$-test.

Table 10. Genes with significant McDonald-Kreitman tests.

Gene	Bias	Ds	Ps	Dn	Pn	P-value[a]	Positive Selection?	TD_{syn}[b]	TD_{non}[c]
CG3085	male	25	41	5	1	0.028	yes	-0.45	-0.07
CG5565	male	11	15	13	5	0.047	yes	-0.07	-1.30
CG6255	male	28	25	10	1	0.011	yes	0.57	1.83
CG8564	male	30	23	27	7	0.026	yes	-0.45	-0.69
CG10750	male	21	20	10	0	0.004	yes	-0.33	–
CG11475	male	31	39	22	4	0	yes	0.06	-1.04
CG18418	male	27	33	13	5	0.04	yes	-0.01	0.25
CG3509	female	26	13	35	6	0.048	yes	0.36	-0.24
CG3975	female	22	30	31	17	0.029	yes	-0.68	-1.11
CG4973	female	41	6	9	12	0	no	-0.66	-1.30
CG6874	female	17	4	22	0	0.048	yes	0.11	–
CG9273	female	18	9	2	6	0.035	no	-0.19	-1.92
CG12276	female	41	3	8	5	0.008	no	0.02	-0.91
CG3476	nonsex	17	16	0	6	0.027	no	-0.99	-1.56

NOTE: Symbols are the same as in Table 9.
[a]P-value was determined by *G*-test when applicable, otherwise by Fisher's exact test.
[b]Tajima's *D* for synonymous sites.
[c]Tajima's *D* for nonsynonymous sites.

departures from neutral evolution. For example, The HKA test (Hudson et al. 1987) compares the ratio of polymorphism to divergence at two (or more) loci. Under neutrality, these ratios are expected to be equal. A departure from the neutral expectation could be caused by selective or demographic factors. For the 91 genes in our survey, a multi-locus HKA test was highly significant ($\chi^2 = 181.1$, $P < 0.001$). In contrast, Ometto et al. (2005) detected no significant departure from neutrality for 232 non-coding loci (introns and intergenic regions) sequenced in the same Zimbabwe

population sample. This suggests that the departure observed for our genes is caused by selection and not the demographic history of the population. test (Mcdonald and Kreitman 1991), which compares the ratio of polymorphism and divergence at synonymous sites to that at nonsynonymous sites. Under a neutral model of molecular evolution, the two ratios are expected to be equal. A relative excess of nonsynonymous divergence is indicative of positive selection favoring amino acid replacements between species (Table 10). A relative excess of nonsynonymous polymorphism could be caused either by balancing selection, which maintains amino acid polymorphism within a species, or by weak purifying selection, which allows slightly deleterious nonsynonymous mutations to segregate as low-frequency polymorphisms, but not become fixed between species.

3.3.2 Tajima' *D* test for sex-biased genes with significant MK result

Application of individual MK tests to the genes in my survey revealed interesting selective differences among genes of the three expression classes. Strikingly, about 20% of the genes in both the male- and female-biased classes gave a significant MK test result (Table 10). All of the significant male-biased genes departed from neutrality in the direction of positive selection, while only half of the significant female-biased genes were indicative of positive selection (Table 10). The other half departed from neutrality in a pattern consistent with either balancing or weak purifying selection. The former should increase the frequency of polymorphic amino acids within a population and, thus, increase Tajima's *D* statistic (Tajima 1989) at nonsynonymous sites. However, there was no evidence for this within the

Table 11. Average values of Tajima's *D*.

Bias	Synonymous	Nonsynonymous
Male	-0.27 (0.10)	-0.76 (<0.001)
Female	-0.03 (0.59)	-0.51 (<0.001)
Nonsex	-0.11 (0.40)	-0.62 (<0.001)

NOTE: *P*-values (in parantheses) were determined as the proportion of 1000 simulations giving a *D* value equal to or lower than the observed.

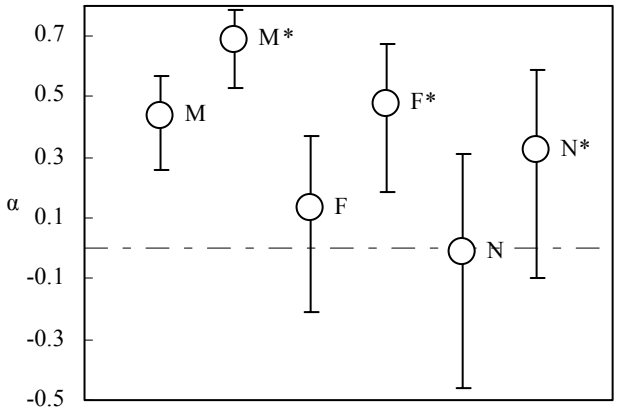

Figure 12. The fraction of positively selected amino acid replacements between species, α, for genes with male-biased (M), female-biased (F), and nonsex-biased (N) expression. The corresponding estimates for each group with all low frequency (singleton) polymorphism excluded are indicated by asterisks. Error bars represent 95% confidence intervals.

female-biased genes in general (Table 11), or within the individual genes showing significant MK tests in this direction (Table 10). For the female-biased genes with a significant excess of nonsynonymous polymorphism, the average Tajima's *D* at nonsynonymous sites was –1.38, which is far lower than the average for all other female-biased genes of -0.39. This suggests that the observed departures from the neutral expectation are due to weak purifying selection against nonsynonymous mutations. Only one of the nonsex-biased genes showed a significant departure from neutrality by the MK test (Tables 9 and 10), and this gene was also consistent with weak purifying selection. Thus, both groups of sex-biased genes showed evidence for increased positive selection relative to nonsex-biased genes. For the genes showing significant evidence for positive selection, the average Tajima's *D* at nonsynonymous sites was –0.30, which is well above the average for male- and female-biased genes (see Table 11), but still lower than the average *D* at synonymous sites in these same genes (-0.09). Thus, it may be that some amino acid positions in these genes have been subject to weak purifying selection, while others have been subject to positive selection.

An MK test using the summed polymorphism and divergence values within each class of genes indicated a significant departure from neutrality in the direction of positive selection for the male-biased genes (Table 10). Female-biased genes also showed an excess of nonsynonymous divergence consistent with positive selection, although this was not significant. Nonsexbiased genes did not differ from the neutral expectation and showed a slight, though insignificant, excess of within-species

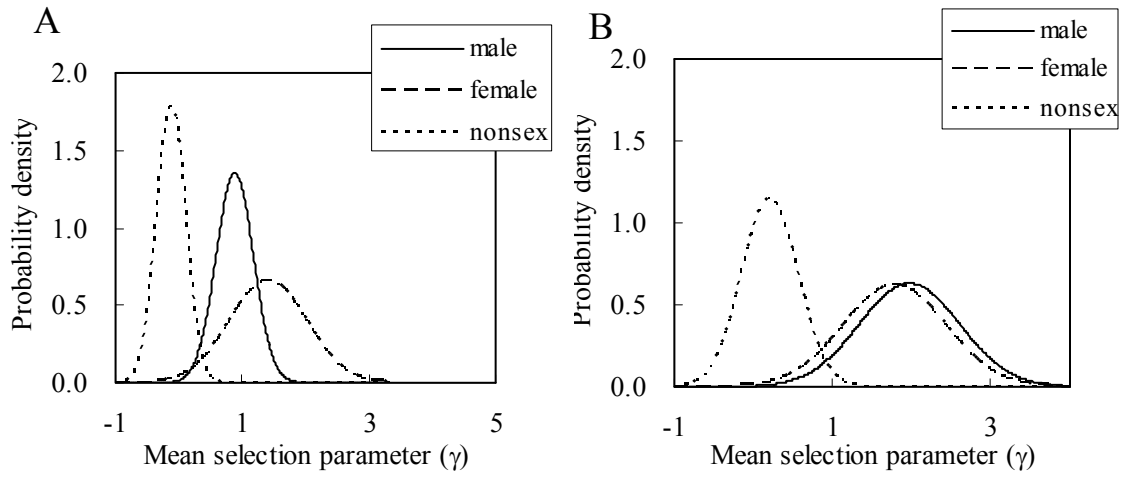

Figure 13. Bayesian posterior distribution of the mean selection parameter (γ) for male-biased (solid), female-biased (dash), and nonsex-biased (dot) genes. (A) Distribution of mean using all polymorphic sites. (B) Distribution of mean after excluding all low frequency (singleton) polymorphisms.

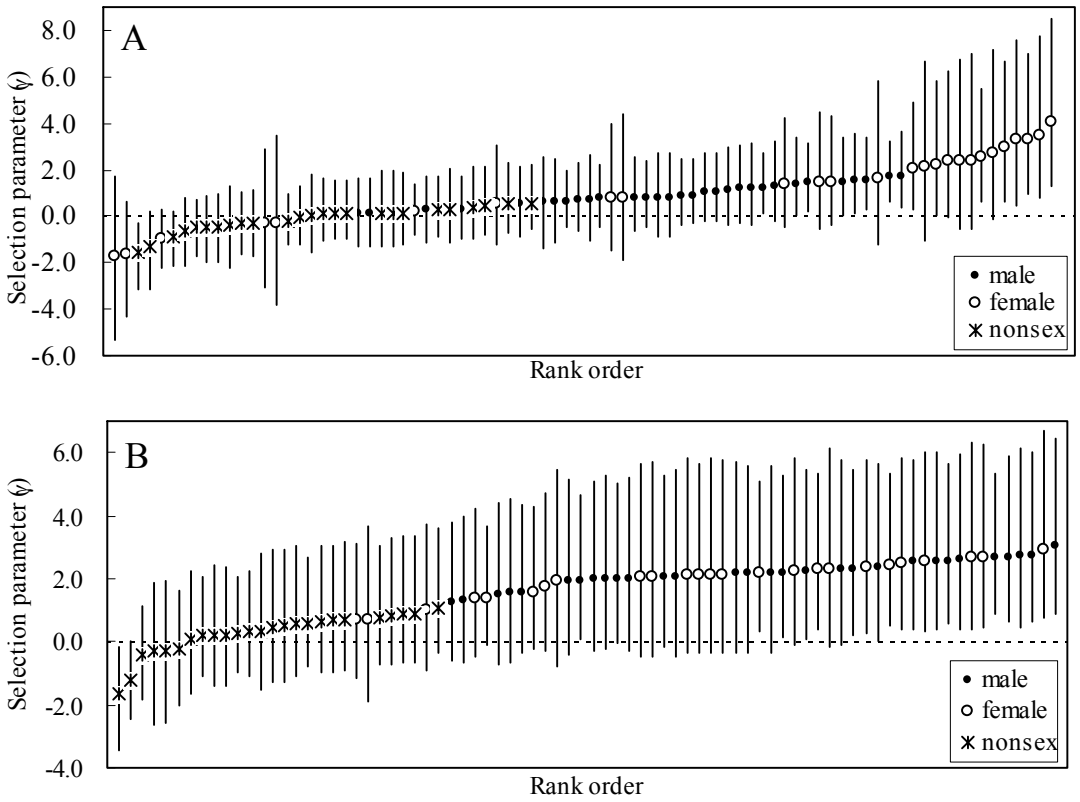

Figure 14. Estimated selection parameter (γ) for each gene. Male-biased genes are indicated by solid circles, female-biased genes by empty circles, and nonsex-biased genes by stars. Error bars represent 95% confidence intervals. (A) Estimated γ for each gene using all polymorphic sites. (B) Estimated γ for each gene after excluding all low frequency (singleton) polymorphisms.

nonsynonymous polymorphism.

3.3.3 Estimate α by multilocus MK test with maximum likelihood method

The MK test framework can be expanded to multilocus polymorphism and divergence data to estimate the average type and strength of selection affecting groups of genes. I used a maximum likelihood method (Bierne and Eyre-Walker 2004) to estimate α, the fraction of amino acid replacements between species that can be attributed to positive selection, within each class of genes (Figure 12). For the male-biased genes, I estimate that 44% of all amino acid replacements were driven by positive selection, while for female-biased genes the estimate is 13%. This fraction is significantly greater than zero for the male-biased genes (likelihood ratio test, $P < 0.001$), but not for the female-biased genes. Nonsex-biased genes, in contrast, showed no evidence for positive selection and, if anything, showed evidence for weak purifying selection ($\alpha < 0$). If weak purifying selection is common in all classes of genes, then the above values of α will be underestimates. Indeed, the observation that nonsynonymous polymorphisms segregate at lower frequency than synonymous polymorphisms, as indicated by Tajima's *D* statistic (Table 11), suggests that weak purifying selection affects all three classes of genes. To reduce the effect of weak purifying selection I repeated the above analysis after removing all low frequency (singleton) polymorphisms at both synonymous and nonsynonymous sites (Figure 12). This led to estimates of 69%, 47%, and 33% for male- and female-, and nonsex-biased genes, respectively. These fractions are significantly greater than zero for the male- and female-biased genes (likelihood ratio test, $P < 0.001$ and $P < 0.01$, respectively), but not for the nonsex-biased genes.

3.3.4 Estimate γ by multilocus MK test with MKPRF method

I also estimated the average strength of selection for amino acid replacements within each group of genes using a Bayesian analysis method (Bustamante et al. 2002). With this approach, the MK table data is used to estimate a selection parameter, $\gamma = 2N_e s$, where N_e is the effective population size and s is the selection coefficient. The estimated selection parameters were greater than zero for both male- and female-biased genes, with mean values of 0.9 and 1.4, respectively (Figures 13A and 14A). For both male- and femalebiased genes, the proportion of the distribution of

mean γ falling below zero was less than 1% ($P_{(\gamma<0)} < 0.001$ and $P_{(\gamma<0)} < 0.01$, respectively). This indicates positive selection favoring amino acid replacements, with the strongest selection occurring in female-biased genes. However, the variance in the mean γ was quite large for female-biased genes and its distribution showed considerable overlap with that of the male-biased genes (Figure 14A). Nonsex-biased genes had a mean γ that was slightly (but not significantly) less than zero (γ= -0.1), again suggesting that there is weak purifying selection against nonsynonymous mutations. As above, I repeated my analysis after excluding all low-frequency polymorphisms (Figures 13B and 14B). This resulted in more similar estimates of the mean γ for male- and female-biased genes (2.0 and 1.8, respectively), and in both cases the proportion of the distribution of mean γ falling below zero was less than 0.01%. Nonsex-biased genes had a positive value of γ (0.2), though this was not significantly greater than zero. The removal of singleton polymorphisms had a particularly strong effect on the female-biased genes, where the ratio of nonsynonymous to synonymous singletons (56/94 = 0.60) was greater than that for male-biased genes (75/195 = 0.38; $\chi^2 = 4.1$, $P = 0.04$) and nonsex-biased genes (44/103 = 0.43; $\chi^2 = 1.8$, $P = 0.18$).

3.4 Discussion

My analyses of polymorphism and divergence indicate that adaptive evolution occurs more frequently in sex-biased genes (both male and female) than in nonsex-biased genes. Male-biased genes, in particular, appear to be consistent targets of positive selection. Female biasedgenes show more variance in the type of selection they experience, with positive selection affecting some genes and purifying selection affecting others. Nonsex-biased genes appear to evolve primarily under purifying selection and have undergone relatively little adaptive evolution since the split of *D. melanogaster* and *D. simulans*. These results argue against the hypothesis that the rapid evolution of sex-biased genes is the result of relaxed selective constraint (see Introduction). This hypothesis predicts that the ratio of nonsynonymous to synonymous polymorphism within species should equal the ratio of nonsynonymous to synonymous divergence between species. However, I find a general excess of nonsynonymous divergence in the sex-biased genes that is reflected in their positive values of the selection parameters α and γ (Figures 12 and 13) and indicates that positive selection has driven their evolution at the protein level. The finding that male-biased genes show high rates of adaptive evolution is consistent with previous reports that looked at interspecific divergence and the relationship between protein divergence and local recombination rate (Zhang et al. 2004; Zhang and Parsch 2005). However, those studies did not find evidence for adaptive evolution in female-biased genes. A possible explanation for this is that the previous studies used a set of genes cloned from an EST survey (Domazet-Loso and Tautz 2003) that was enriched for highly expressed genes. The female-biased genes, in particular, showed exceptionally high levels of both absolute expression and synonymous codon usage bias (Hambuch and Parsch 2005). This suggests that the EST collection was comprised of an unusually constrained set of female-biased genes subject to strong purifying selection. A further difference between the present and the previous studies is that the latter did not include extensive within-species polymorphism data. Thus, the previous studies had less power to detect adaptive evolution and could not account for differences in selective constraint among genes.

Although the selection parameters α and γ are defined differently (the former as

the fraction of positively selected amino acid substitutions and the latter as their average scaled selection coefficient), both are calculated from the same MK-table data. Thus, one would expect the two measures to be highly correlated. However, we observe a marked difference between the two with respect to the female-biased genes, where the relative level of positive selection is greater when measured by γ. (compare Figure 12 and 13). The reason for this appears to be in the way the two methods are implemented. Both assume that synonymous sites evolve neutrally and use the ratio of divergence to polymorphism at these sites to determine a neutral standard. In the method of Bierne and Eyre-Walker (2004), α is calculated separately for each group of genes (male-, female-, and nonsex-biased) using only the synonymous sites from that particular group, while in the method of Bustamante et al. (2002), γ is calculated for each group of genes using the combined synonymous sites of all genes as the neutral standard. In our data, the female-biased genes have a higher ratio of divergence to polymorphism at synonymous sites (631/233 = 2.71) than both the male-biased (744/447 = 1.66) and the nonsex-biased genes (436/267 = 1.63). This can explain the observed discordance in selection parameter between the two methods. If we re-calculate γ using the synonymous sites of each group of genes separately, we obtain estimates of 1.23, 0.62, and –0.02 for male-, female, and nonsex-biased genes, respectively, which agrees well with the estimates of α. Using this approach, γ for female-biased genes is no longer significantly greater than zero ($P = 0.065$). When all singleton polymorphisms are removed, the γ estimates increase to 2.39, 0.89, and 0.36, for male-, female, and nonsex-biased genes, respectively, and γ for the female biased genes is significantly greater than zero ($P = 0.002$).

It is not clear why the ratio of polymorphism to divergence at synonymous sites is elevated in the female-biased genes relative to the other two groups. One possibility is that the three groups of genes experience differential selection for synonymous codon usage. On a genome-wide scale, significant differences in codon bias have been observed among groups of sex-biased genes (Hambuch and Parsch 2005). However, it was male-biased genes that differed significantly from female- and nonsex-biased genes, while the latter two groups showed equal levels of codon bias. This pattern does not correspond to the pattern seen for polymorphism and divergence at synonymous sites. Furthermore, for the genes included in the present study the frequency of optimal codon usage (F_{op}; Ikemura 1981) is 0.51, 0.53, and 0.56, for the

male-, female-, and nonsex-biased genes, respectively, which also does not correspond to the pattern seen for polymorphism and divergence at synonymous sites.

Why does adaptive evolution occur so frequently in sex-biased genes? I first consider the male-biased genes. In a highly polygamous species, such as *D. melanogaster*, in which there is no paternal investment in offspring and females are able to store the sperm from a single mating to fertilize a lifetime's worth of eggs, sexual selection among males is expected to be very strong. This is evident in the intense sperm competition that occurs among males, which is influenced by accessory gland proteins and other male-expressed genes (Clark et al. 1995, 1999). Indeed, some of these proteins are known to affect a male's reproductive output and show clear signs of adaptive evolution (Herndon and Wolfner 1995; Tsaur and Wu 1997; Tsaur et al. 1998; Aguade 1998, 1999; Begun et al. 2000; Chapman et al. 2000). Acps, however, represent only a small fraction (less than 10%) of genes with male-biased expression (Swanson et al. 2001), and none of the genes in the current study are known Acps. This suggests that many other male-biased genes may be either directly or indirectly involved in determining reproductive success and, thus, subject to sexual selection (Zhang and Parsch 2005). Indeed, laboratory evolution experiments have shown that, when subject to strong male-male competition (or released from it), *Drosophila* males show heritable changes in many aspects of their reproductive biology and behavior (Rice 1996; Holland and Rice 1999), which are presumably controlled by a wide variety of genes. What drives the adaptive evolution of female-biased genes? Because there is less variation in reproductive success among *Drosophila* females than males, sexual selection is expected to be much weaker in females. However, sexual selection on male traits may lead to rapid co-evolution of female reproductive traits or *vice versa*. In some cases, the co-evolution may be considered cooperative, with males and females sharing the same evolutionary interests. For example, male sperm length and female seminal receptacle length appear to evolve co-adaptively in *Drosophila* species (Pitnick et al. 1999; Miller and Pitnick 2002). In many other cases, however, conflict between male and female reproductive interests may drive co-evolution. For example, the strong selection on males to maximize paternity can lead to the fixation of traits that are harmful to females, which, in turn, leads to selection for females that can counteract their effect. Indeed, components of male seminal fluid, including Acps, are known to have deleterious effects on mated females (Chapman et al. 1995; Wigby and Chapman

2005). Furthermore, sexually antagonistic (or "arms race") co-evolution has been demonstrated in laboratory populations of *D. melanogaster*, where sexually selected males are known to shorten the lifespan of their naive female mates (Rice 1996). Females that have co-evolved with males, however, are able to avoid these damaging consequences, indicating that they adapt in response to the males in their environment. Although the genes underlying these coadapted female traits are unknown, several female-expressed genes showing the molecular hallmarks of sexually antagonistic co-evolution, including a significant excess of nonsynonymous divergence between species, have been recently identified (Swanson et al. 2004).

In summary, I propose that the increased signal of positive selection seen for genes with sex-biased expression results from the combined action of sexual selection and intersexual co-evolution. The former should affect primarily males, while the latter will affect both males and females. This provides a biological explanation for why the signal of selection is stronger and more consistent for male-biased genes, but weaker and more variable for the female-biased genes. The finding that sex-biased genes are more frequent targets of positive selection than nonsex-biased genes suggests that sex-biased genes play a predominant role inthe genetic and reproductive differentiation of species.

Chapter 4

4. Preliminary study of sex-biased gene expression using microarrays

4.1 Introduction

The first three chapters of this dissertation use comparative genomic and population genetic approaches to study the evolution of genes with sex-biased expression. The sex-biased genes were identified from previously published microarray experiments (Ranz et al. 2003; Parisi et al. 2003; Gibson et al. 2004), most of which used only one or two laboratory strains of a single species, *D. melanogaster*. The assumption made in the first three chapters is that the expression pattern observed in these studies is conserved across all strains of *D. melanogaster* and in other *Drosophila* species. The experimental evidence to support this assumption, however, is limited.

The experiments of Ranz et al. (2003) compared male *versus* female gene expression in both *D. melanogaster* (strain *Canton-S*) and *D. simulans* (strain *s1*). Here they found that genes with sex-biased expression, especially male-biased expression, showed a relatively high level of gene expression divergence between species. However, many of the expression differences were not in the sex-bias classification of a gene (i.e. male-, female-, or nonsex-biased), but instead in the expression level of the gene between *D. melanogaster* and *D. simulans* flies of the same sex. For example, a gene showing a 4-fold male bias in *D. melanogaster* and a 6-fold male-bias in *D. simulans* may show a significant expression difference between males of the two species, yet still be classified as male-biased in both species. Of the 5000 genes surveyed by Ranz et al. (2003), 951 (19%) showed a difference in their sex-bias classification, with the vast majority being cases where a gene was sex-biased in one species and nonsex-biased in the other [930 genes (18.6%)]. Only 20 genes (4%) showed a change from male-biased to female-biased or *vice versa*.

The microarray experiments of Meiklejohn et al. (2003) used eight strains of *D. melanogaster*, including four cosmopolitan and four African strains. These authors found that male-biased genes show a relatively high level of gene expression polymorphism. However, these experiments compared only expression in adult males among the eight strains and relied on the data from Ranz et al. (2003) for sex-bias classification. Thus, they could not determine whether or not the genes had changed in their sex-bias classification among strains of *D. melanogaster*.

Because of the limitations of the above experiments, I performed new microarray experiments to compare levels of male *versus* female gene expression in four strains of *D. melanogaster* (two from a European population and two from Zimbabwe, Africa) and in one strain of *D. simulans*. The experiments were designed so that the sex-bias classification of the genes could be compared between strains or species. Furthermore, I used two different microarray platforms. One platform was custom made and contained probes specific to the 91 genes surveyed in chapter 3. The other platform was commercially available and contained probes to almost all genes in the *D. melanogaster* genome (approximately 14,000 genes). In general, I find that sex-bias classification is well conserved among strains and between species. However, technical problems with the microarray experiments, especially problems with weak signal intensity, make it difficult to compare genes with low expression levels and/or weak hybridization to the array probes.

4.2 Materials and Methods

4.2.1 Fly strains and cDNA preparations

Four strains of *D. melanogaster* (*ZB82* and *ZB398* from Zimbabwe, Africa, and *EU01* and *EU20* from Leiden, The Netherlands) were raised on standard medium at 22 degrees centigrade. After eclosion, adult males and females were isolated and aged 4-6 days. Total RNA was extracted using TRIzol reagent (Invitrogen), following the manufacturer's protocol. Total RNA was quantified by spectrophotometer, using A260 to determine the RNA concentration and the A260/A280 ratio to ensure quality of the RNA. For reverse transcription, the amount of total RNA used ranged from 22-60 μg per reaction per sex, with smaller amounts used for the custom arrays and larger amounts for the whole-genome arrays (see next section). cDNA synthesis was performed using SuperScript reverse transcriptase (Invitrogen). After cDNA synthesis and purification, male- and female-derived cDNA were labeled with Alexa Fluor dyes (Invitrogen, 555 and 647 reactive dye pack) overnight. Labeled cDNA was purified following the protocol of the SuperScript Plus Indirect cDNA Labeling System (Invitrogen). Microarray hybridization and post-hybridization were performed using the Pronto! Microarray Hybridization Kit (Corning) with slight modification. After post-hybridization washing, the final array was scanned immediately with a 2-laser microarray scanner (Genetix aQuire).

4.2.2 Microarrays

Two types of microarray were used in this study. The first was custom made and is here referred to as the MFU array platform. The second was purchased from the Drosophila Genome Resource Center (DGRC) and is referred to as the DGRC-1 platform. The MFU arrays were constructed in-house and printed with an OmniGrid Accent Microarrayer (GeneMachines) on UltraGAPS-coated glass slides (Corning). Each array contained probes to the 91 genes surveyed in chapter 3 (33 male-biased, 28 female-biased, and 30 nonsex-biased). The probes were amplified by PCR using the primers given in Table S8 and confirmed by gel electrophoresis. Each of the 91 probes was spotted in eight replicates on each microarray. In addition, eight replicate spots of five control probes were spotted on each array: 1) male fly whole genomic DNA, 2)

female fly whole genomic DNA, 3) a PCR amplicon of the alcohol dehydrogenase gene (*Adh*) of *D. melanogaster*, 4) a PCR amplicon of the ribosomal protein gene (*RpL23*) of *D. melanogaster*, and 5) a PCR-amplified fragment of mouse DNA.

The second type of microarray used in this study is the DGRC-1 whole fly genome array (version A1.4.13.33). The DGRC-1 arrays were originally designed using annotation v.1 by B. Oliver (NIH) and Incyte Genomics, who donated the primers to the DGRC. Re-analysis showed that they contain amplicons corresponding to around 88% of the genes in the *D. melanogaster* annotation v. 4.2. For the DGRC-1 arrays, the DNAs were spotted on GAPS II-coated slides (Corning). Of the 91 genes used for the MFU arrays (described above), 85 were also included on the DGRC-1 arrays (33 male-biased, 26 female-biased and 26 nonsex-biased genes). Two female-biased genes (CG5272 and CG32409) and four nonsex-biased genes (CG5919, CG6913, CG13419 and CG11785) were not included on the DGRC-1 arrays.

For both types of arrays, I carried out two biological replicates (from different RNA extractions) and two "dye-swaps", in which the dyes used to label the male and female cDNA were reversed. This compensates for the possibility that different samples (or particular genes) had different incorporation efficiency for one of the two dyes.

4.2.3 Statistical analysis

For the MFU arrays, I performed the analysis using a Microsoft Excel template designed by John Parsch. This provided a quick and easy way to analyze the results from microarray experiments using the MFU arrays. However, it was inflexible and required the data to have exactly the same layout for each experiment. Background correction was performed by subtracting the mean background intensity from the mean foreground intensity for each channel at each spot. The mean value of the eight replicate spots per gene was then used as the raw intensity value. The intensities were not logarithm transformed before analysis.

To normalize the intensities of the two channels (Cy3 = green and Cy5 = red), an adjustor was calculated from the ratio of the two channels of the control spots (male genomic DNA and female genomic DNA), i.e. the raw intensity of Cy5 was multiplied by the adjustor to make the average signal intensity of the two channels equal. For each slide, genes were assigned to sex-bias categories using either a fold-change cutoff or the results of a paired t-test, or both. For example, a gene with a

male/female ratio greater than 2 and *P*-value less than 0.05 from the paired t-test was classified as male-biased. A nested ANOVA test, which was similar to the paired t-test, was also performed to classify genes into different sex-bias categories. Assuming the sex bias classifications from previous studies (Ranz et al 2003; Parisi et al. 2003; Gibson et al. 2004) were "correct", I could estimate the frequency of type I and type II errors in my results.

A nested ANOVA test was also performed over all four replicates (including dye swaps and biological replicates) to classify genes as male-, female-, and nonsex-biased in their expression. The numbers of false negative errors (type II) and false positive errors (type I) were estimated in the same way as described above. Overall, every gene could be classified into a sex bias group by this analysis, although some classifications were not consistent across the different analysis methods. Gene expression in adult males and females of two highly inbred *D. melanogaster* lines from Africa (*ZB82* and *ZB398*, derived from Lake Kariba, Zimbabwe) and two highly inbred *D. melanogaster* lines from Europe (*EU01* and *EU20,* derived from Leiden, Netherlands) was analyzed in this manner. To further investigate inconsistencies with previous studies (Ranz et al 2003; Parisi et al. 2003; Gibson et al. 2004), the 91 genes were classified into two groups: genes with high quality signal, whose intensity value in at least one of the channels was one standard deviation above the local background in any replicate, and genes with low quality signal, whose intensity value in neither channel was one standard deviation above the local background in any replicate. To quantify the quality of signal of each gene, a signal score was defined as follows: a value of 1 was given to a gene if the intensity value in either channel was one standard deviation above the local background at a spot, otherwise a value of 0 was given. Because there were eight replicate spots for each gene, the maximum signal score for each slide was eight, meaning all eight replicate spots from that slide had a high quality signal. A signal score of zero means that all eight replicate spots were of low quality. In the end, the signal score for each gene was calculated as the mean signal score for that gene over the four replicate slides.

For the DGRC-1 arrays, the spot signal intensities were normalized using the CARMAweb server (https://carmaweb.genome.tugraz.at/carma/), which supports an intuitive graphical interface for the normalization and analysis of microarray data derived from current microarray platforms. In general, the microarray analysis can be split up into three main steps: 1) data upload, 2) preprocessing, normalization,

replicate handling, and 3) detection of differentially expressed genes. A schematic of how this is achieved in CARMAweb is shown in Figure 15. The preprocessing of two color microarrays consists of three steps: background correction, within array normalization and between array normalization. For my case, the parameters "subtract", "print tip loess", and "quantile method" were set for the above three steps, respectively. To exclude low quality spots, my analysis used only those spots whose intensity value in at least one of the channels was one standard deviation above the local background in any replicate slide. After normalization, the ratio of the fluorescence intensity of the two channels for each of the spots from each of the replicates was used as input for the Bayesian analysis software, BAGEL (Townsend and Hartl 2002), to detect differentially expressed genes between adult males and females. Two highly inbred *D. melanogaster* lines from Africa (*ZB82* and *ZB398*, derived from Lake Kariba, Zimbabwe), two highly inbred *D. melanogaster* lines from Europe (*EU01* and *EU20,* derived from Leiden, The Netherlands), and one lab strain of *D. simulans* (*s1*, derived from Chapel Hill, NC; Meiklejohn et al. 2004) were used for this analysis.

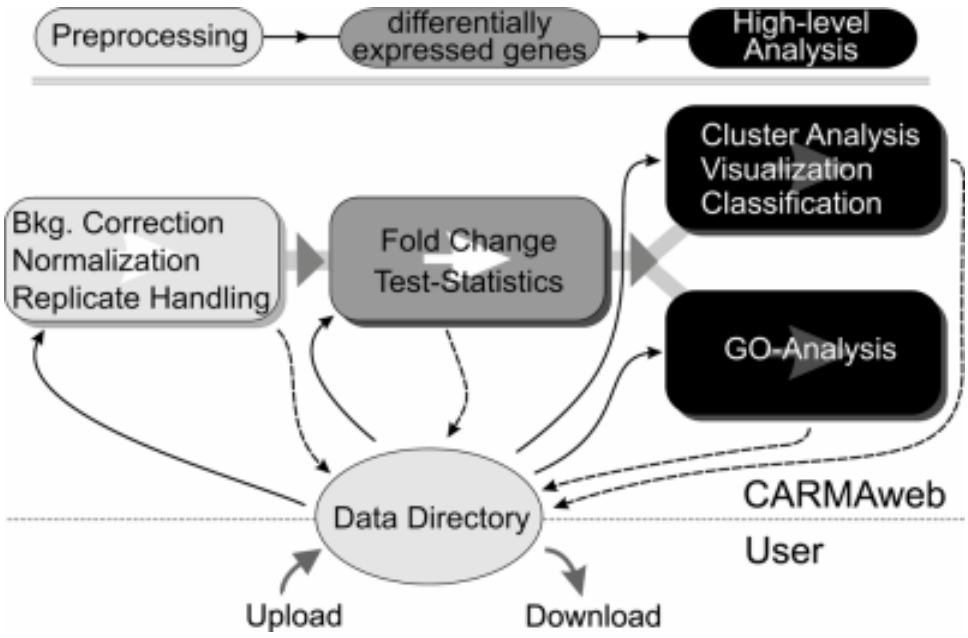

Figure 15. The picture above shows how this analysis workflow is modeled in CARMAweb and how the different components can be linked together. The central part is the Data directory. All files listed there can be used as input for a new analysis step. After each analysis step the results can be returned to this Data directory, or further analyses can be performed directly on the result files from a previous step. (From https://carmaweb.genome.tugraz.at/carma/)

4.3 Results

4.3.1. Identification of of sex-biased genes using custom arrays

In general, I use a fold-change cutoff of 2 and a *P*-value cutoff of 0.05 (for both paired t-test and ANOVA analyses) to classify genes as male-, female-, or nonsex-biased. These two cutoffs can, however, be changed depending on the desired stringency of analysis. Overall, the percentage of male-biased and female-biased genes that could be "correctly" classified ranged from 36% to 79% for the four *D. melanogaster* strains (Table 12). This assumes that the results from previously published experiments are "correct" (Ranz et al 2003; Parisi et al. 2003; Gibson et al. 2004). Most undetected sex-biased genes appear to be missed due to low signal intensity, because genes with little or no signal will by default be classified as nonsex-biased. For some slides, the adjustor for normalization was quite different from one, which suggests that either the use of genomic DNA spots for normalization is not reliable or that there are very large differences in the signal strength of the two dyes. Thus, difficulty in the normalization of the two dyes is a disadvantage of the custom microarrays. When compared with the paired t-test, the ANOVA test shows about the same power for detecting sex-biased genes with the same cutoff parameters (Table 12). If the stringency of analysis is decreased (by lowering the fold-change cutoff or raising the *P*-value cutoff), more sex-biased genes can be detected, although at the same time the frequency of false positive errors (type I; the classification of nonsex-biased genes as sex-biased) increases. To investigate the reason for the inconsistencies in the classification of sex-biased genes between my experiments and previous studies (Ranz et al 2003; Parisi et al. 2003; Gibson et al. 2004), I examined the signal intensity of the 91 genes included on my custom arrays. For this, I separated the genes into two groups: those with sex-bias classification consistent with the previous studies and those with sex-bias classification inconsistent with previous studies. The inconsistent group showed significantly lower signal intensity than the consistent group (Mann-Whitney test; $P<0.05$; Figure 16). This suggests that most misclassified genes are those with either low overall expression or with poor hybridization to the array probes. One difference between my custom arrays and those of the other studies is that I used probes of PCR-amplified genomic DNA that spanned

Table 12. Identification of sex-biased genes using *D. melanogaster* custom arrays.

		paired t-test			ANOVA		
		Male	Female	Nonsex	Male	Female	Nonsex
ZB82	Total[a]	18	19	59	18	17	56
	"correct"[b]	17	18	30	18	17	30
	(percent)	(52)	(64)	(100)	(55)	(61)	(100)
	false neg[c]	16	10	0	15	11	0
	false pos[d]	1	1	29	0	0	26
ZB398	Total[a]	13	20	63	14	14	63
	"correct"[b]	13	18	29	14	14	30
	(percent)	(39)	(64)	(97)	(42)	(50)	(100)
	false neg[c]	20	10	1	19	14	0
	false pos[d]	0	2	34	0	0	33
EU01	Total[a]	27	19	50	26	19	46
	"correct"[b]	26	18	30	26	19	30
	(percent)	(79)	(64)	(100)	(79)	(68)	(100)
	false neg[c]	7	10	0	7	9	0
	false pos[d]	1	1	20	0	0	16
EU20	Total[a]	12	19	65	18	16	57
	"correct"[b]	12	18	30	17	16	29
	(percent)	(36)	(64)	(100)	(52)	(57)	(97)
	false neg[c]	21	10	0	16	12	1
	false pos[d]	0	1	35	1	0	28

[a]the number of genes assigned to each category using a fold-change cutoff of 2 and a *P*-value cutoff of 0.05.
[b]the number genes with classification consistent with previous studies (Parisi et al. 2003; Ranz et al. 2003; Gibson et al. 2004).
[c]type II error.
[d]type I error.

the entire coding region of the target genes, while the others used cDNAs (Ranz et al. 2003), PCR-amplified single exons (Parisi et al. 2003), or short oligonucleotides complimentary to single exons (Gibson et al. 2004). My probes, therefore, often include sequences complementary to multiple exons, as well as introns and UTRs. It is possible that these extra sequences alter the strength or specificity of hybridization.

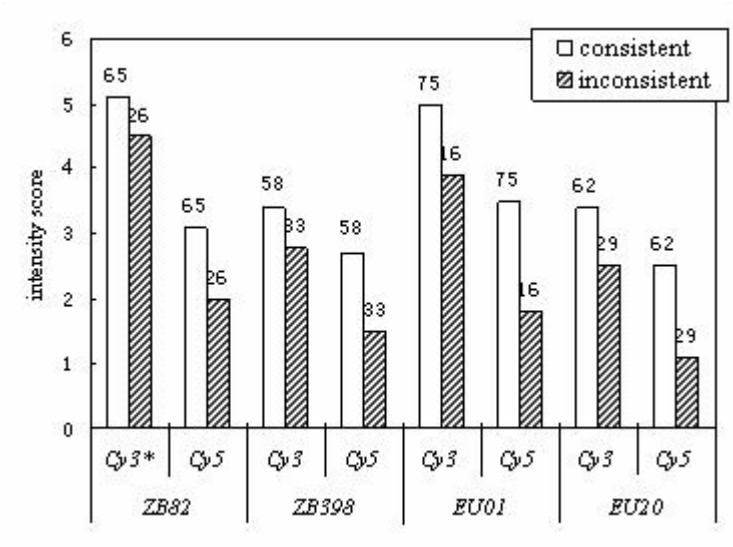

Figure 16. Comparison of signal intensity between the consistent group and the inconsistent group. The *P*-value of Mann-Whitney tests for all comparisons is significant ($P < 0.05$) except the comparison of Cy3 of *ZB82* ($P = 0.069$). The number above each bar represents the gene number in the corresponding category.

For example, if a gene is alternatively spliced in males and females, the sex-specific forms may not be individually recognized in my experiments. In any case, it should be noted that my approach to identify sex-biased genes is conservative. That is, the number of false negatives is much greater than the number of false positives (Table 12).

4.3.2. Identification of sex-biased genes using DGRC-1 arrays

An analysis of sex-biased gene expression in *D. simulans* and in two lines of *D. melanogaster* (*EU01* and *ZB398*) was carried out using DGRC-1 whole-genome arrays and the BAGEL statistical software. Prior to the statistical analysis, normalization of the two dye channels was performed using the CARMAweb server. Background correction was also performed in CARMAweb by subtracting the mean background intensity from the mean foreground intensity in each channel for each spot (Figure 17). The within array normalization corrects for dye signal bias and other systematic errors within each slide. In addition, the print tip loess correction was used for normalization, which performs a normalization for each print tip group separately to account for any biases caused by differences in the printing pins. The between array normalization adjusts the expression values for each spot across the separate replicate arrays. For this, the quantile method was used (Figure 18). In the end, the signal of the two channels was completely balanced (Figure 18). The average regulation (M) and average expression (A) values between the two groups (male and female) was then calculated. The average M value for each gene is calculated by subtracting the $\log_2$

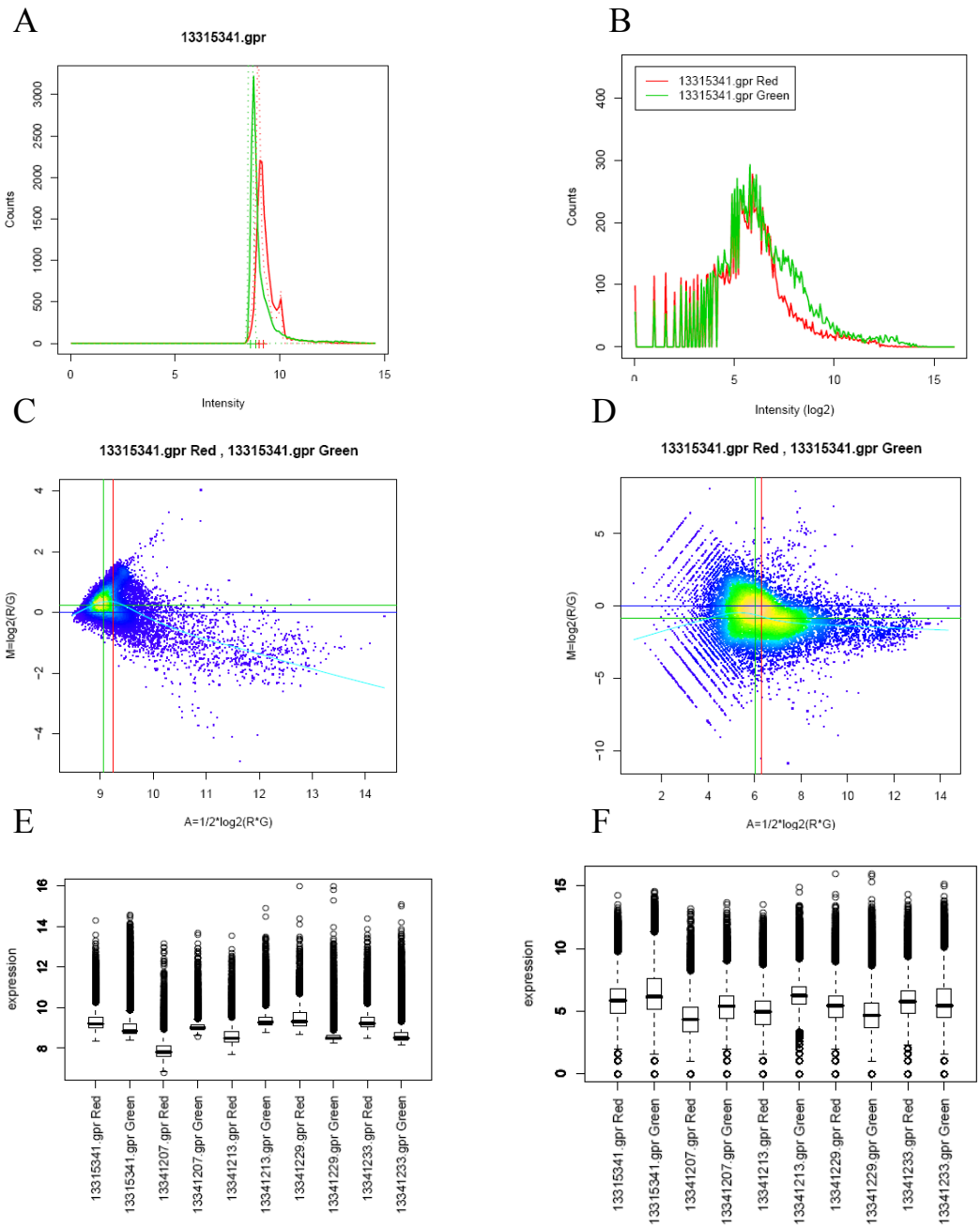

Figure 17. Comparison of DGRC-1 arrays before background correction (A C E) and after background correction (B D F). (A) (B) Histogram of the array 1 (13315341.gpr). The green line corresponds to the green signal channel and the red line to the red channel. Dotted lines represent the background intensities. (C) (D) MA plot of array 1 (13315341.gpr). (E) (F) Boxplots of the signal intensities of each signal channel of the microarrays (5 arrays). All hybridizations are with *D. simulans* (strain *s1*) cDNA.

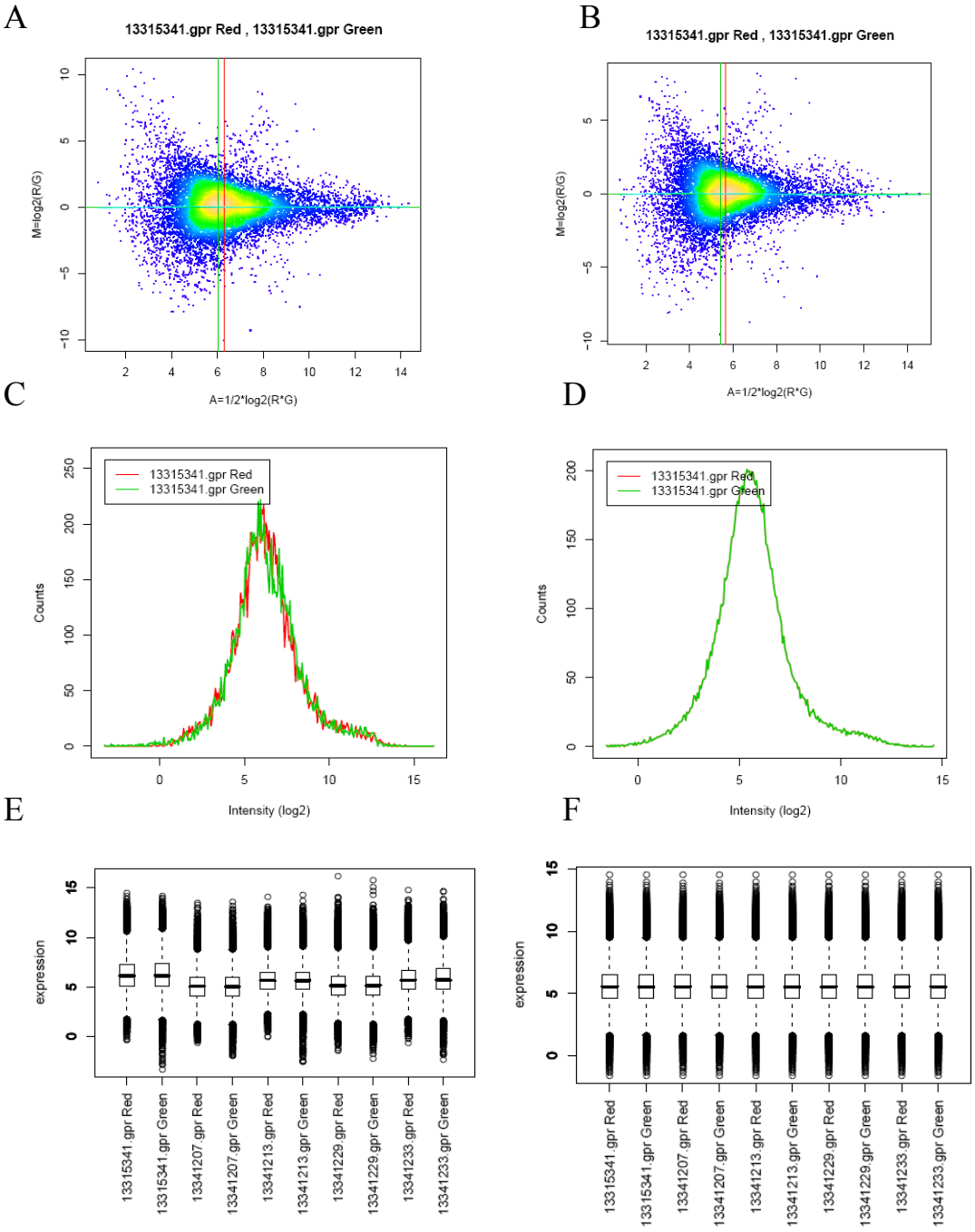

Figure 18. Comparison of DGRC-1 arrays of using within array normalization (A C E) and between array normalization (B D F). (A) (B) MA plot of array 1 (13315341.gpr). (C) (D) Histogram of the array 1 (13315341.gpr). (E) (F) Boxplots of the signal intensities of each signal channel of the microarrays (5 arrays). All hybridizations are with *D. simulans* (strain *s1*) cDNA.

Table 13. Identification of sex-biased genes of *D. simulans* and *D. melanogaster* (strains *ZB398* and *EU01*) using DGRC-1 arrays.

D. sim	Valid genes[a]	8095		
	BAGEL[b]	7517		
	P[c]	0.05	0.025	0.01
	male-biased	1396 (19)	1047 (14)	751 (10)
	female-biased	1291 (17)	973 (13)	654 (9)
	nonsex-biased	4830 (64)	5497 (73)	6112 (81)
ZB398	Valid genes[a]	9049		
	BAGEL[b]	8608		
	P[c]	0.05	0.025	0.01
	male-biased	1365 (16)	940 (11)	582 (7)
	female-biased	1532 (18)	1116 (13)	711 (8)
	nonsex-biased	5711 (66)	6552 (76)	7315 (85)
EU01	Valid genes[a]	7307		
	BAGEL[b]	7055		
	P[c]	0.05	0.025	0.01
	male-biased	1561 (22)	1238 (18)	891 (13)
	female-biased	1694 (24)	1279 (18)	836 (12)
	nonsex-biased	3800 (54)	4538 (64)	5328 (75)

NOTE: the number i parentheses represents the percentage of valid genes in each category.
[a]genes normalized by CARMAweb with valid MA values that could be used as input for BAGEL.
[b]number of genes in the output file after BAGEL analysis.
[c]the P-value cutoff (critical value) used to classify genes as sex-biased.

average expression value of the gene in females from the $\log_2$ average expression value of the gene in males (so a mean M of 1 indicates a twofold increase in the expression level of the gene in males relative to females). The average of the expression values is calculated using the median function. The average MA plot is drawn using M and A values that are calculated from the average expression values of each gene in each sample group. The normalized data serve as the starting point for statistical analysis by BAGEL to detect genes with differential expression between males and females. Using the default parameters for BAGEL, a total of 8095, 9049 and 7307 genes could be analyzed for *D. simulans*, *ZB398* and *EU01*, respectively (Table 13). The BAGEL analysis provides *P*-values, which give information about

how likely a gene is to be differentially expressed between males and females. In a so-called "volcano plot", the *P*-value of each gene is plotted against the regulation value (average M value) (Figure 19). Thus, the volcano plot allows one to see both the statistical significance and the male/female expression ratio for all genes analyzed. The most interesting genes are those with low *P*-values and high (or low) average M values. Overall, for both strains and species I found that about 40% of the genes showed sex-biased expression (using a cutoff of $P < 0.05$), with approximately equal numbers of male- and female-biased genes (Table 13). The genes of both *D. simulans* and *D. melanogaster* showed a similar distribution in their male/female expression ratios (Figure 19). To examine the consistency of sex-biased expression among strains and between experiments, I performed two comparisons. First, I compared my *D. simulans* data to the *D. simulans* data of Ranz et al. (2003), which I downloaded from the Sex Bias Database, Sebida (Gnad and Parsch 2006; http://www.sebida.de). Second, I compared my DGRC-1 array results for two *D. melanogaster* strains, one African (*ZB398*) and one European (*EU01*). Both of these comparisons showed a strong correlation in the male/female ratio of the individual genes (Figure 20). This suggests that the DGRC-1 arrays are a reliable tool for measuring differential gene expression and that ratios of male to female gene expression are typically conserved among *D. melanogaster* strains.

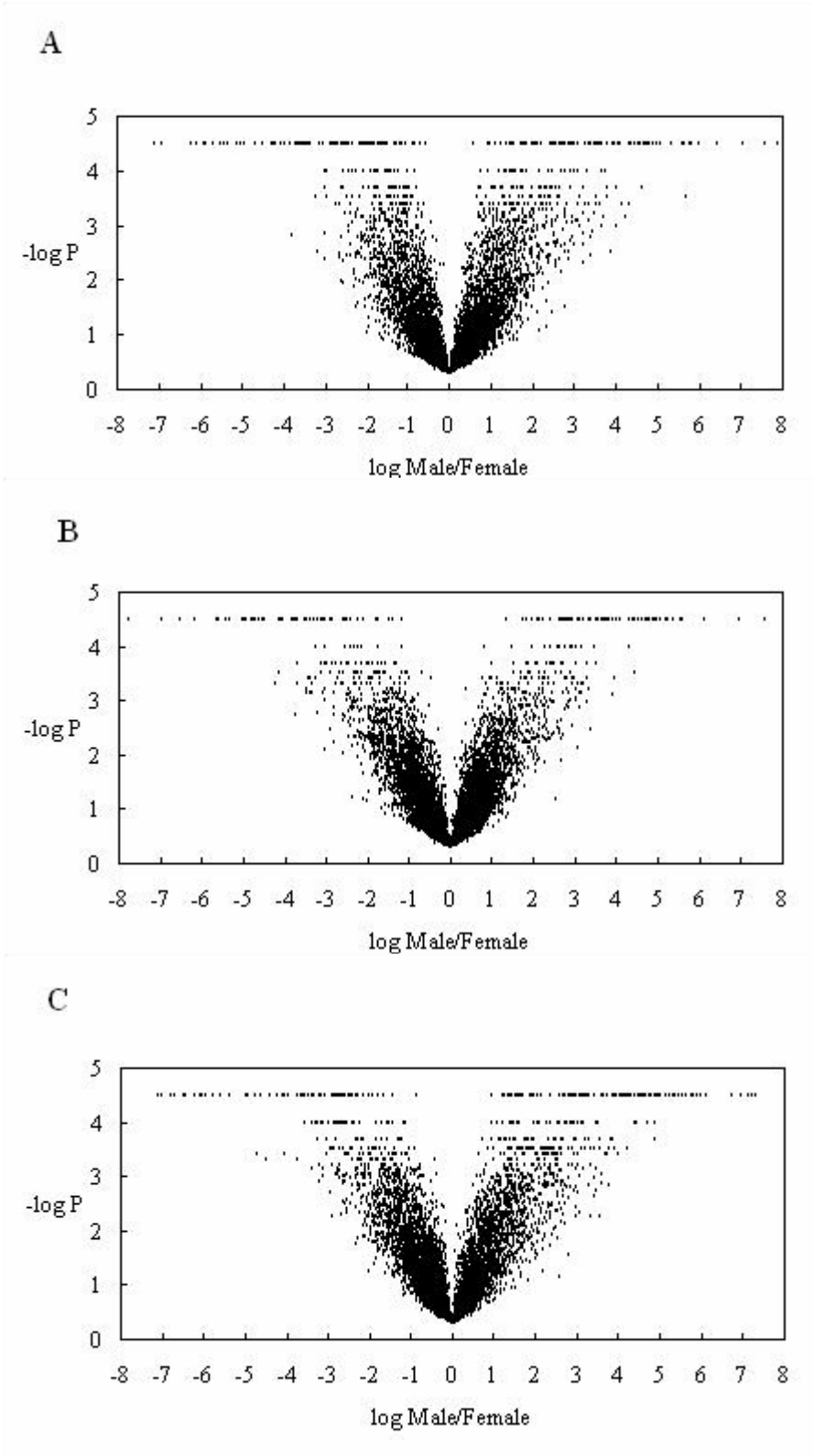

Figure 19. Volcano plot of the male/female expression ratio *versus* the *P*-value of each gene from the DGRC-1 arrays. The *P*-value was determined using the BAGEL software (x axis is the $\log_2$ scale of the ratio of male/female, y axis is the $-\log_{10}$ scale of *P*-value). (A) Strain *s1* of *D. simulans*. (B) Strain *ZB398* of *D. melanogaster*. (C) Strain *EU01* of *D. melanogaster*.

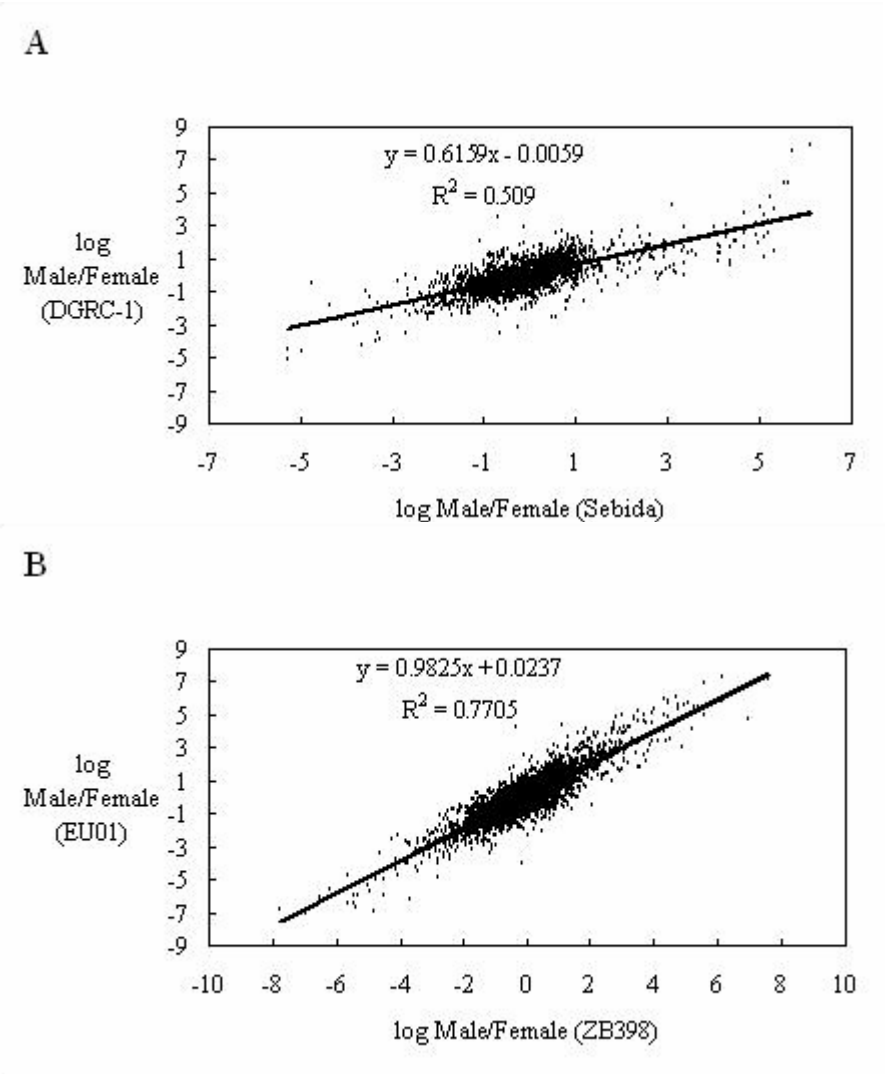

Figure 20. Correlation of the male/female gene expression ratio between different experiments (Both x and y axis are the $\log_2$ scale of the ratio of male/female). (A) Comparison of *D. simulans* data from Sebida (Gnad and Parsch 2006; derived from Ranz et al. 2003) and my DGRC-1 arrays. (B) Comparison of two *D. melanogaster* strains

4.4 Discussion

The central dogma of molecular biology, which states that information flows only from nucleic acid to nucleic acid or protein, constrains the association between genotype (DNA) and phenotype (gene expression). Differences in genes and the interactions among sets of genes give rise to differences in phenotype. Natural selection can act directly on these phenotypic differences and, thereby, indirectly on genes and gene expression. Males and females share nearly the same genome, typically differing only by the nearly-inert Y chromosome. However, there are often very large phenotypic differences between the two sexes. It is thus clear that differences in gene regulation must play an important role in sexual dimorphism and may also be subject to sexual selection. Recently, it has become possible to study differences in gene expression on a large scale. This has been made possible by microarray technologies, which allow one to examine global differences in gene expression between two samples (e.g. male and female) in a single experiment. To date, such comparisons have been performed only for a small number of species, including *D. melanogaster* and *D. simulans* (Ranz et al. 2003; Parisi et al. 2003; Gibson et al. 2004). These studies, however, have only examined expression in a few laboratory strains. Furthermore, each study has used a different microarray platform. In this chapter, I have used custom made microarrays, as well as DGRC-1 whole-genome microarrays, to analyze male *versus* female gene expression in four strains of *D. melanogaster*, two derived from an African (Zimbabwe) population and two from a European (The Netherlands) population. These strains had not been studied previously. Furthermore, I used the whole-genome arrays to analyze gene expression in one strain of *D. simulans*. This strain had been analyzed previously (Ranz et al. 2003), but only using cDNA arrays corresponding to about 5,000 genes. In contrast, my arrays covered about 14,000 genes.

The custom arrays had the advantage of providing high replication for a small number of target genes. In my case, they were designed for the 91 genes analyzed in Chapter 3. All of these genes had been previously surveyed on microarrays and could be classified *a priori* as male-, female-, or nonsex-biased. This expression pattern could be confirmed for many of the genes in the four strains I examined, even with the

use of conservative fold-change and *P*-value cutoffs. However, there was also a relatively high number of false negatives (genes that were expected to be sex-biased, but showed no bias in my experiments). The main reason for this appears to be genes with weak signal intensity. This could be caused by low absolute expression levels or by poor hybridization between the cDNA and the PCR-amplicons on my arrays. An additional problem with the custom arrays was that it was difficult to normalize the signal between the two dye channels. This is because one cannot assume that there is an equal total signal intensity over all spots in males and females as one can with whole genome arrays. To get around this problem, I used genomic DNA spots as standards that should show equal hybridization to male and female cDNA. However, these spots showed high variation in signal intensity in both channels, which made normalization difficult and unreliable.

The DGRC-1 whole genome arrays, on the other hand, could be normalized easily and accurately using standard procedures. These arrays were very efficient for examining global patterns of gene expression, although they did not provide as much replication for each individual gene. The DGRC-1 arrays also were more expensive than the custom arrays and required more cDNA and fluorescent dyes than the custom arrays. This leads to much higher costs per gene replicate.

In summary, both types of arrays had advantages and disadvantages. Both could be used to confirm or identify sex-biased genes in *Drosophila*. My results indicate that many genes (around 40%) show differential expression between adult males and females in both *D. melanogaster* and *D. simulans*. In general, I find that the overall pattern of sex-biased gene expression is well conserved among strains and between species. However, technical problems with the microarray experiments, especially problems with weak signal intensity, make it difficult to compare genes with low expression levels and/or weak hybridization to the array probes.

5. Summary

Sexual dimorphism, i.e., phenotypic differences between male and female members of the same species, is plentiful in nature. It forms a major component of the total biological diversity among sexual organisms. On the surface, some sexually dimorphic traits appear maladaptive and presumably could not have evolved by natural selection. To account for this, Charles Darwin proposed his theory of sexual selection. Sexual selection results from differential mating success among individuals within a population. It consists of the direct competition between members of the same sex (e.g. male-male competition and sperm competition) and/or the attraction of one sex to the other (e.g. female choice). It has long been known that in multicellular eukaryotes, phenotypes associated with sex and reproduction accumulate interspecific differences more rapidly than other phenotypes (Darwin 1871). Such rapid evolution is often evident in male secondary sexual characteristics (e.g. feather coloration or song in male birds), and also in primary sexual characteristics. Often species of a genus can only be distinguished by their different primary sexual characteristics. For example primary sexual characteristics, such as sperm size or genital morphology, can differ greatly among closely related insect species and are often used for taxonomic classification (Jamieson 1987; Eberhard 1985; Hosken and Stockley 2004). It is assumed that sexual selection and/or sexual antagonism drive the evolution of sexual traits. The rapid evolution of sexual characters has recently been extended to the molecular level. Several studies have found increased rates of evolution in reproductive genes, suggesting that sexual selection plays an important role in the evolution of many sex-related genes (Civetta and Singh 1999; Singh and Kulathinal 2000; Swanson and Vacquier 2002). The major question facing us now is whether this faster evolution is caused by positive selection, or if it is due to the relaxation of selective constraint on sex-related genes. Here positive selection is defined as selection that drives changes in the sequences of genes (or in noncoding sequences) contrary to purifying selection, which acts to preserve sequences that are already present in the population. The cause for positive selection can be adaptation to changing environmental conditions (“natural” or “Darwinian” selection) or to

requirements for successful reproduction ("sexual selection"). In some cases, the rapid evolution of male-expressed genes has been shown to be caused by positive selection. Some examples of this include *Acp26Aa* (Tsaur and Wu 1997; Aguadé 1998; Tsaur, Ting and Wu 1998), *OdsH* (Ting et al. 1998), *Sdic* (Nurminsky et al. 1998; Nurminsky et al. 2001), *Dntf-2r* (Betrán and Long 2003) and *jan-ocn* (Parsch et al. 2001a; Parsch et al. 2001b). It is predicted that incompatibility involving sex-related genes will cause population differentiation, reproductive isolation, and eventually speciation (Orr and Presgraves 2000). Furthermore, it is known that these incompatibilities often affect one sex more than the other. In the case of *Drosophila*, it is the males that are most affected by incompatibilities leading to hybrid sterility or lethality, a result consistent with Haldane's rule. Therefore, the possibility arises that different classes of sex-biased genes and nonsex-biased genes may exhibit different patterns of evolution. My hypothesis is that these different classes of genes are subject to different evolutionary forces, which shape their molecular evolution. By investigating the type and strength of selection acting on sex-biased genes and nonsex-biased genes, I hope to shed light on the mechanisms responsible for genetic differentiation between species.

In my thesis, I studied the evolutionary patterns of male-biased, female-biased, and nonsex-biased genes in the genus *Drosophila* in detail. If sexual selection is common in nature and affects a wide variety of sex-related genes, it should leave a detectable footprint at the molecular genetic level. By analyzing the DNA sequence divergence of sex-biased genes among *Drosophila* species, as well as combined polymorphism and divergence data for a large sample of sex-biased genes in *D. melanogaster* and *D. simulans*, I was able to show that different classes of sex-biased genes evolve under different types and strengths of selection. These differences are most likely caused by sexual selection and sexual antagonism. The major results presented in this dissertation are:

1. Male-biased genes show greater divergence between *Drosophila* species than both female-biased and nonsex-biased genes (Chapter 1)

Genomic comparisons between species provide a powerful tool for inferring the historical effects of selection on various classes of genes and determining their rates of evolution. By combining microarray expression data with interspecific genome comparisons, I found different evolutionary rates among male-biased, female-biased,

and nonsex-biased genes on different timescales. Comparisons between species of the *D. melanogaster* subgroup revealed that genes with male-biased expression had significantly faster rates of evolution (as measured by d_N/d_S) than genes with female-biased or nonsex-biased expression, which was mainly due to a higher d_N in the male-biased genes. The same pattern was observed for comparisons among more distantly related species. In comparisons between *D. melanogaster* and *D. pseudoobscura*, male-biased genes were significantly more divergent (as measured by the fraction of genes with a significant BLAST hit) than female-biased genes and nonsex-biased genes. Male-biased genes did not show an increased ratio of nonsynonymous/synonymous polymorphism within *D. melanogaster*, as would be expected if they were under less selective constraint. Comparisons of polymorphism/divergence ratios among the three groups of genes suggested that the rapid evolution of male-biased genes was due to positive selection.

2. Male-biased genes show a positive correlation between evolutionary rate and local recombination rate, while female-biased and nonsex-biased genes do not (Chapter 2).

Recombination rates typically vary across eukaryotic genomes, and the local rate of recombination may influence a gene's rate of evolution. It is assumed that recombination enhances the efficiency of adaptation and selection. Thus, if positive selection occurs frequently, one expects to observe a positive correlation between evolutionary rate and recombination rate. This is the pattern I observed for male-biased genes. Female-biased and nonsex-biased genes, in contrast, showed a negative correlation between evolutionary rate and local recombination rate. These patterns cannot be explained by differences in mutation rates. Instead, they suggest frequent adaptive evolution in male-biased genes, which is limited by Hill-Robertson interference in regions of low recombination.

3. Sex-biased genes (both male and female) are more often targets of adaptive evolution than nonsex-biased genes (Chapter 3).

To further investigate the selective forces affecting sex-biased genes, I surveyed DNA sequence polymorphism in 91 protein-encoding genes in a sample of 12 *D. melanogaster* strains from Zimbabwe, Africa. In addition, I used a single *D. simulans* strain for interspecific comparisons. The combination of polymorphism and

divergence data allowed the application of single-locus McDonald-Kretiman (MK) tests, as well as two recently developed multi-locus versions of the MK test. My analysis of 33 male-biased genes, 28 female-biased genes and 30 nonsex-biased genes indicated that adaptive evolution occurs more frequently in sex-biased genes (both male and female) than in nonsex-biased genes. Male-biased genes, in particular, appear to be consistent targets of positive selection. Female-biased genes showed more variance in the type of selection they experience, with positive selection affecting some genes and purifying selection affecting others. Nonsex-biased genes appear to evolve primarily under purifying selection and have undergone relatively little adaptive evolution since the split of *D. melanogaster* and *D. simulans*.

4. The sex-bias classification of genes is typically well conserved between *D. melanogaster* strains and in *D. simulans* (Chapter 4).

I performed a preliminary study of sex-biased gene expression using several strains of *D. melanogaster* and one strain of *D. simulans*. I used both custom arrays and commercial whole-genome arrays to compare expression levels between adult males and females. The custom arrays could be used to efficiently test the expression of the 91 genes analyzed in Chapter 3, while the commercial arrays provided a large-scale view of gene expression for the whole genome. The expression results were in good agreement across *D. melanogaster* strains, indicating that the sex-biased expression pattern of genes is typically conserved throughout the species. In addition, my results from *D. simulans* agreed well with previously published results that used a different microarray platform, suggesting that microarray hybridizations are a reliable method to identify sex-biased genes.

Literature Cited

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

Aguade, M. 1999. Positive selection drives the evolution of the *Acp29AB* accessory gland protein in Drosophila. Genetics **152**:543-551.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25**:3389-3402.

Andrews, J., G. G. Bouffard, C. Cheadle, J. Lu, K. G. Becker, and B. Oliver. 2000. Gene discovery using computational and microarray analysis of transcription in the *Drosophila melanogaster* testis. Genome Res **10**:2030-2043.

Arguello, J. R., Y. Chen, S. Yang, W. Wang, and M. Long. 2006. Origination of an X-linked testes chimeric gene by illegitimate recombination in *Drosophila*. PLoS Genet **2**:e77.

Barker, M. S., J. P. Demuth, and M. J. Wade. 2005. Maternal expression relaxes constraint on innovation of the anterior determinant, bicoid. PLoS Genet **1**:e57.

Barraclough, T. G., P. H. Harvey, and S. Nee. 1995. Sexual selection and taxonomic diversity in passerine birds. Proc R Soc London B Biol Sci **259**:211-215.

Begun, D. J., and C. F. Aquadro. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature **356**:519-520.

Begun, D. J., P. Whitley, B. L. Todd, H. M. Waldrip-Dail, and A. G. Clark. 2000. Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics **156**:1879-1888.

Bergman, C. M., B. D. Pfeiffer, D. E. Rincon-Limas, R. A. Hoskins, A. Gnirke, C. J. Mungall, A. M. Wang, B. Kronmiller, J. Pacleb, S. Park, M. Stapleton, K. Wan, R. A. George, P. J. de Jong, J. Botas, G. M. Rubin, and S. E. Celniker. 2002. Assessing the impact of comparative genomic sequence data on the functional annotation of the *Drosophila* genome. Genome Biol **3**:RESEARCH0086.

Betancourt, A. J., and D. C. Presgraves. 2002. Linkage limits the power of natural selection in *Drosophila*. Proc Natl Acad Sci U S A **99**:13616-13620.

Betancourt, A. J., D. C. Presgraves, and W. J. Swanson. 2002. A test for faster X evolution in *Drosophil*a. Mol Biol Evol **19**:1816-1819.

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

Betran, E., K. Thornton, and M. Long. 2002. Retroposed new genes out of the X in *Drosophila*. Genome Res **12**:1854-1859.

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

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

Celniker, S. E., D. A. Wheeler, B. Kronmiller, J. W. Carlson, A. Halpern, S. Patel, M. Adams, M. Champe, S. P. Dugan, E. Frise, A. Hodgson, R. A. George, R. A. Hoskins, T. Laverty, D. M. Muzny, C. R. Nelson, J. M. Pacleb, S. Park, B. D. Pfeiffer, S. Richards, E. J. Sodergren, R. Svirskas, P. E. Tabor, K. Wan, M. Stapleton, G. G. Sutton, C. Venter, G. Weinstock, S. E. Scherer, E. W. Myers, R. A. Gibbs, and G. M. Rubin.

2002. Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. Genome Biol **3**:RESEARCH0079.

Chapman, T., L. F. Liddle, J. M. Kalb, M. F. Wolfner, and L. Partridge. 1995. Cost of mating in *Drosophila melanogaster* females is mediated by male accessory gland products. Nature **373**:241-244.

Chapman, T., D. M. Neubaum, M. F. Wolfner, and L. Partridge. 2000. The role of male accessory gland protein *Acp36DE* in sperm competition in *Drosophila melanogaster*. Proc Biol Sci **267**:1097-1105.

Civetta, A., S. A. Rajakumar, B. Brouwers, and J. P. Bacik. 2006. Rapid evolution and gene-specific patterns of selection for three genes of spermatogenesis in *Drosophila*. Mol Biol Evol **23**:655-662.

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

Civetta, A., and R. S. Singh. 1998. Sex-related genes, directional sexual selection, and speciation. Mol Biol Evol **15**:901-909.

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

Clark, A. G., M. Aguade, T. Prout, L. G. Harshman, and C. H. Langley. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. Genetics **139**:189-201.

Clark, A. G., D. J. Begun, and T. Prout. 1999. Female x male interactions in *Drosophila* sperm competition. Science **283**:217-220.

Comeron, J. M., and T. B. Guthrie. 2005. Intragenic Hill-Robertson interference influences selection intensity on synonymous mutations in *Drosophila*. Mol Biol Evol

22:2519-2530.

Comeron, J. M., and M. Kreitman. 2002. Population, evolutionary and genomic consequences of interference selection. Genetics **161**:389-410.

Comeron, J. M., M. Kreitman, and M. Aguade. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. Genetics **151**:239-249.

Coulthart, M. B., and R. S. Singh. 1988. High level of divergence of male-reproductive-tract proteins, between *Drosophila melanogaster* and its sibling species, *D. simulans*. Mol Biol Evol **5**:182-191.

Darwin, C. 1859. The origin of species. John Murray, London, UK.

Darwin, C. 1871. The descent of man, and selection in relation to sex. John Murray, London, UK.

Domazet-Loso, T., and D. Tautz. 2003. An evolutionary analysis of orphan genes in *Drosophila*. Genome Res **13**:2213-2219.

Eberhard, W. G. 1985. Sexual selection and animal genitalia. Harvard University Press, Cambridge, Mass, US.

Galindo, B. E., V. D. Vacquier, and W. J. Swanson. 2003. Positive selection in the egg receptor for abalone sperm lysin. Proc Natl Acad Sci U S A **100**:4639-4643.

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

Glinka, S., L. Ometto, S. Mousset, W. Stephan, and D. De Lorenzo. 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. Genetics **165**:1269-1278.

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

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

Hambuch, T. M., and J. Parsch. 2005. Patterns of synonymous codon usage in *Drosophila melanogaster* genes with sex-biased expression. Genetics **170**:1691-1700.

Herndon, L. A., and M. F. Wolfner. 1995. A *Drosophila* seminal fluid protein, *Acp26Aa*, stimulates egg laying in females for 1 day after mating. Proc Natl Acad Sci U S A **92**:10114-10118.

Hey, J., and R. M. Kliman. 1993. Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol Biol Evol **10**:804-822.

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

Hill, W. G., and A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genet Res **8**:269-294.

Hodges, S. A., and M. L. Arnold. 1995. Spurring plant diversification: are floral nectar spurs a key innovation? Proc R Soc London B Biol Sci **262**:343-348.

Holland, B., and W. R. Rice. 1999. Experimental removal of sexual selection reverses intersexual antagonistic coevolution and removes a reproductive load. Proc Natl Acad Sci U S A **96**:5083-5088.

Hosken, D. J., and P. Stockley. 2004. Sexual selection and genital evolution. Trends Ecol Evol **19**:87-93.

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

Ikemura, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J Mol Biol **146**:1-21.

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

Jansa, S. A., B. L. Lundrigan, and P. K. Tucker. 2003. Tests for positive selection on immune and reproductive genes in closely related species of the murine *genus mus*. J Mol Evol **56**:294-307.

Kern, A. D., C. D. Jones, and D. J. Begun. 2004. Molecular population genetics of male accessory gland proteins in the *Drosophila* simulans complex. Genetics **167**:725-735.

Khaitovich, P., S. Paabo, and G. Weiss. 2005. Toward a neutral evolutionary model of gene expression. Genetics **170**:929-939.

Kim, Y. 2004. Effect of strong directional selection on weakly selected mutations at linked sites: implication for synonymous codon usage. Mol Biol Evol **21**:286-294.

Levine, M. T., C. D. Jones, A. D. Kern, H. A. Lindfors, and D. J. Begun. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. Proc Natl Acad Sci U S A **103**:9935-9939.

Mack, P. D., A. Kapelnikov, Y. Heifetz, and M. Bender. 2006. Mating-responsive genes in reproductive tissues of female *Drosophila melanogaster*. Proc Natl Acad Sci U S A **103**:10358-10363.

Marais, G., and B. Charlesworth. 2003. Genome evolution: recombination speeds up

adaptive evolution. Curr Biol **13**:R68-70.

Marais, G., T. Domazet-Loso, D. Tautz, and B. Charlesworth. 2004. Correlated evolution of synonymous and nonsynonymous sites in *Drosophila*. J Mol Evol **59**:771-779.

Mayr, E. 1963. Animal species and evolution. The Belknap Press, Cambridge, MA.

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

Meiklejohn, C. D., Y. Kim, D. L. Hartl, and J. Parsch. 2004. Identification of a locus under complex positive selection in *Drosophila simulans* by haplotype mapping and composite-likelihood estimation. Genetics **168**:265-279.

Meiklejohn, C. D., J. Parsch, J. M. Ranz, and D. L. Hartl. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. Proc Natl Acad Sci U S A **100**:9894-9899.

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

Miller, G. T., and S. Pitnick. 2002. Sperm-female coevolution in *Drosophila*. Science **298**:1230-1233.

Moller, A. P., and J. J. Cuervo. 1998. Speciation and feather ornamentation in birds. Evolution **52**:859-869.

Musters, H., M. A. Huntley, and R. S. Singh. 2006. A genomic comparison of faster-sex, faster-X, and faster-male evolution between *Drosophila melanogaster* and *Drosophila pseudoobscura*. J Mol Evol **62**:693-700.

Nurminsky, D., D. D. Aguiar, C. D. Bustamante, and D. L. Hartl. 2001. Chromosomal

effects of rapid gene evolution in *Drosophila melanogaster*. Science **291**:128-130.

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

Ometto, L., S. Glinka, D. De Lorenzo, and W. Stephan. 2005. Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. Mol Biol Evol **22**:2119-2130.

Orr, H. A., and A. J. Betancourt. 2001. Haldane's sieve and adaptation from the standing genetic variation. Genetics **157**:875-884.

Orr, H. A., and D. C. Presgraves. 2000. Speciation by postzygotic isolation: forces, genes and molecules. Bioessays **22**:1085-1094.

Panhuis, T. M., R. Butlin, M. Zuk, and T. Tregenza. 2001. Sexual selection and speciation. Trends Ecol Evol **16**:364-371.

Panhuis, T. M., and W. J. Swanson. 2006. Molecular evolution and population genetic analysis of candidate female reproductive genes in *Drosophila*. Genetics. **173**:2039-2047

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

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

Parsch, J., C. D. Meiklejohn, and D. L. Hartl. 2001. Patterns of DNA sequence variation suggest the recent action of positive selection in the *janus-ocnus* region of *Drosophila simulans*. Genetics **159**:647-657.

Parsch, J., C. D. Meiklejohn, and D. L. Hartl. 2005. Selective sweep. Landes Bioscience, Georgetown, TX.

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

Pitnick, S., T. Markow, and G. S. Spicer. 1999. Evolution of multiple kinds of sperm storage organs in *Drosophila*. Evolution **53**:1804-1822.

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

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

Rice, W. R. 1996. Sexually antagonistic male adaptation triggered by experimental arrest of female evolution. Nature **381**:232-234.

Rice, W. R., and A. K. Chippindale. 2002. The evolution of hybrid infertility: perpetual coevolution between gender-specific and sexually antagonistic genes. Genetica **116**:179-188.

Richards, S., Y. Liu, B. R. Bettencourt, P. Hradecky, S. Letovsky, R. Nielsen, K. Thornton, M. J. Hubisz, R. Chen, R. P. Meisel, O. Couronne, S. Hua, M. A. Smith, P. Zhang, J. Liu, H. J. Bussemaker, M. F. van Batenburg, S. L. Howells, S. E. Scherer, E. Sodergren, B. B. Matthews, M. A. Crosby, A. J. Schroeder, D. Ortiz-Barrientos, C. M. Rives, M. L. Metzker, D. M. Muzny, G. Scott, D. Steffen, D. A. Wheeler, K. C. Worley, P. Havlak, K. J. Durbin, A. Egan, R. Gill, J. Hume, M. B. Morgan, G. Miner, C. Hamilton, Y. Huang, L. Waldron, D. Verduzco, K. P. Clerc-Blankenburg, I. Dubchak, M. A. Noor, W. Anderson, K. P. White, A. G. Clark, S. W. Schaeffer, W. Gelbart, G. M. Weinstock, and R. A. Gibbs. 2005. Comparative genome sequencing of Drosophila

pseudoobscura: chromosomal, gene, and cis-element evolution. Genome Res 15:1-18.

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

Rubin, G. M., L. Hong, P. Brokstein, M. Evans-Holm, E. Frise, M. Stapleton, and D. A. Harvey. 2000. A *Drosophila* complementary DNA resource. Science **287**:2222-2224.

Singh, R. S., and R. J. Kulathinal. 2000. Sex gene pool evolution and speciation: a new paradigm. Genes Genet Syst **75**:119-130.

Singh, R. S., and R. J. Kulathinal. 2005. Male sex drive and the masculinization of the genome. Bioessays **27**:518-525.

Stolc, V., Z. Gauhar, C. Mason, G. Halasz, M. F. van Batenburg, S. A. Rifkin, S. Hua, T. Herreman, W. Tongprasit, P. E. Barbano, H. J. Bussemaker, and K. P. White. 2004. A gene expression map for the euchromatic genome of *Drosophila melanogaster*. Science **306**:655-660.

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner, and C. F. Aquadro. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci U S A **98**:7375-7379.

Swanson, W. J., and V. D. Vacquier. 2002. The rapid evolution of reproductive proteins. Nat Rev Genet **3**:137-144.

Swanson, W. J., A. Wong, M. F. Wolfner, and C. F. Aquadro. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. Genetics **168**:1457-1465.

Swanson, W. J., Z. Yang, M. F. Wolfner, and C. F. Aquadro. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci U S A **98**:2509-2514.

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

Tao, Y., S. Chen, D. L. Hartl, and C. C. Laurie. 2003. 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.

Ting, C. T., S. C. Tsaur, M. L. Wu, and C. I. Wu. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282**:1501-1504.

Townsend, J. P., and D. L. Hartl. 2002. Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments. Genome Biol **3**:RESEARCH0071.

True, J. R., B. S. Weir, and C. C. Laurie. 1996. A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila mauritiana* chromosomes into *Drosophila simulans*. Genetics **142**:819-837.

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

Tsaur, S. C., and C. I. Wu. 1997. Positive selection and the molecular evolution of a gene of male reproduction, *Acp26Aa* of *Drosophila*. Mol Biol Evol **14**:544-549.

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

Wigby, S., and T. Chapman. 2005. Sex peptide causes mating costs in female *Drosophila melanogaster*. Curr Biol **15**:316-321.

Wolfner, M. F. 1997. Tokens of love: functions and regulation of *Drosophila* male

accessory gland products. Insect Biochem Mol Biol **27**:179-192.

Wright, F. 1990. The 'effective number of codons' used in a gene. Gene **87**:23-29.

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

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci **13**:555-556.

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol **15**:568-573.

Zhang, Z., T. M. Hambuch, and J. Parsch. 2004. Molecular evolution of sex-biased genes in *Drosophila*. Mol Biol Evol **21**:2130-2139.

Zhang, Z., and J. Parsch. 2005. Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression. Mol Biol Evol **22**:1945-1947.

Table S1. The divergence data between *D. melnogaster* and *D. yakuba.* (2-fold cutoff used)

Gene	Loc[a]	d_N	d_S	d_N/d_S	F_{op}	ENC
Male-biased						
AttA	2R	0.0310	0.4254	0.0729	0.459	48.86
CG10252	3R	0.0019	0.4091	0.0047	0.434	58.30
CG10589	3L	0.0241	0.3783	0.0638	0.485	55.23
PebIII	2R	0.0248	0.4436	0.0559	0.805	29.62
to	3R	0.0220	0.6117	0.0360	0.490	53.22
CG11876	3R	0.0021	0.1847	0.0116	0.700	37.95
CG12699	2R	0.2108	0.5849	0.3604	0.387	47.09
chp	3R	0.0002	0.2454	0.0010	0.593	43.16
regucalcin	X	0.0309	0.3765	0.0820	0.723	34.97
CG18284	2L	0.0609	0.2202	0.2766	0.465	57.76
CG2254	X	0.0461	0.4839	0.0953	0.651	40.48
BcDNA:GH06048	2R	0.2684	0.6137	0.4373	0.216	54.58
CG17022 (CG31883)	2L	0.2510	0.5107	0.4914	0.166	47.62
BG:DS06874.1	2L	0.0489	0.3545	0.1381	0.540	43.10
Pdh	3L	0.0082	0.3305	0.0249	0.522	54.13
CG5547	2L	0.0036	0.1407	0.0253	0.530	47.35
Arr1	2L	0.0090	0.1167	0.0775	0.677	41.14
sp2	3R	0.1345	0.3048	0.4413	0.658	42.30
CG6467	3L	0.1035	0.4141	0.2499	0.735	31.51
fau, anoxia	3R	0.0032	0.0616	0.0512	0.537	46.39
CG6910	3L	0.0056	0.3569	0.0156	0.596	44.38
smp-30	3R	0.0409	0.2764	0.1480	0.590	44.31
CG7584	3R	0.0238	0.5235	0.0454	0.759	38.13
ocn	3R	0.0295	0.4348	0.0679	0.454	53.46
CG8043	3R	0.0249	0.2457	0.1012	0.519	51.46
CG8701	2R	0.0581	0.4685	0.1241	0.564	51.71
CG9259	2L	0.0281	0.2684	0.1048	0.453	56.79
CG9920	3R	0.0045	0.2752	0.0165	0.663	35.91
CG17494	na	0.0934	0.3269	0.2856	0.313	54.25
CG6921	3R	0.0122	0.4661	0.0263	0.243	46.61
CG7768	3L	0.0032	0.4760	0.0067	0.250	59.47
CG8417	3R	0.0028	0.4495	0.0063	0.218	47.76
yip7	na	0.0727	0.3367	0.2159	0.633	40.81
Female-biased						
CG10423	3R	0.0001	0.0922	0.0010	0.707	37.74
Uev1A, CG10640	3L	0.0001	0.0912	0.0010	0.478	51.62
RpL30, CG10652	2L	0.0037	0.2332	0.0157	0.750	35.15
RpS6	X	0.0025	0.1688	0.0151	0.779	32.06

(Continues...)

Table S1. *(Cont.)*

Gene	Loc[a]	d_N	d_S	d_N/d_S	F_{op}	ENC
CG11015	2L	0.0075	0.1196	0.0625	0.676	37.69
RpS12	3L	0.0001	0.0983	0.0010	0.743	34.26
CG11342	3L	0.0568	0.4070	0.1395	0.445	48.12
CG12253, *BEST:LD08487*	2L	0.0222	0.3733	0.0596	0.361	59.24
RpS13	2L	0.0055	0.1348	0.0408	0.752	32.48
CG1475	3R	0.0002	0.1675	0.0010	0.786	34.78
sta	X	0.0002	0.1954	0.0010	0.783	30.76
RpS14a	X	0.0001	0.1499	0.0010	0.701	33.50
RpL27a	2L	0.0001	0.1474	0.0010	0.759	31.37
RpS20	3R	0.0032	0.2128	0.0151	0.733	37.23
CG15697	3R	0.0064	0.0551	0.1165	0.769	33.45
CG17280	2R	0.0146	0.2480	0.0588	0.846	26.22
RpL15	3h	0.0003	0.3088	0.0010	0.367	48.37
yip6	2h	0.0022	0.3254	0.0067	0.495	48.30
CG17508	2R	0.0714	0.2306	0.3095	0.263	57.03
cta	2h	0.0123	0.2367	0.0519	0.304	56.83
CG18543	3L	0.0546	0.3520	0.1551	0.569	42.81
CG2099	3R	0.0249	0.2908	0.0858	0.784	34.89
CG2108	3R	0.0127	0.2053	0.0617	0.636	41.54
Gip	X	0.0074	0.4336	0.0171	0.564	48.23
RpL19	2R	0.0021	0.2689	0.0079	0.754	39.88
CG2852	2R	0.0086	0.2375	0.0361	0.794	28.65
RpL40	2L	0.0001	0.1469	0.0010	0.786	35.30
oho23B	2L	0.0002	0.1527	0.0010	0.827	26.23
CG2998	X	0.0001	0.0982	0.0010	0.694	29.36
eIF-5A	2R	0.0266	0.0702	0.3788	0.414	54.69
CG3203	X	0.0001	0.0765	0.0010	0.794	34.36
RpS9	3L	0.0001	0.0948	0.0010	0.797	34.50
bic	2R	0.0078	0.2179	0.0359	0.719	34.32
CG3751	2R	0.0002	0.1638	0.0010	0.802	31.19
RpL46	2R	0.0000	0.0209	0.0010	0.646	33.86
CG18001	2h	0.0004	0.3571	0.0010	0.348	59.07
CG4046	2R	0.0000	0.0436	0.0010	0.735	35.31
tsr	2R	0.0002	0.1786	0.0010	0.538	50.40
RpS19	X	0.0001	0.1126	0.0010	0.712	35.67
FKBP59	2L	0.0076	0.3016	0.0252	0.571	45.32
RpL13	2L	0.0002	0.1547	0.0010	0.793	32.98
CG4759	3R	0.0001	0.1168	0.0010	0.800	32.05
CG4800	3R	0.0105	0.4502	0.0233	0.822	27.66
RpL7	2L	0.0021	0.2026	0.0105	0.805	33.36
eEF1delta	2L	0.0417	0.3151	0.1324	0.516	54.26

(Continues...)

Table S1. *(Cont.)*

Gene	Loc[a]	d_N	d_S	d_N/d_S	F_{op}	ENC
me31B	2L	0.0033	0.0342	0.0961	0.408	59.69
RpS27A	2L	0.0001	0.0639	0.0010	0.662	44.06
Nap1	2R	0.0285	0.2734	0.1042	0.690	39.75
SmB	2L	0.0047	0.1838	0.0257	0.514	41.90
l(2)04154	2R	0.0623	0.1534	0.4060	0.442	57.38
RpL14	3L	0.0024	0.0777	0.0314	0.747	38.22
Ef1beta	2R	0.0023	0.2396	0.0098	0.704	38.91
porin	2L	0.0000	0.0398	0.0010	0.690	35.07
RpS25	3R	0.0001	0.0591	0.0010	0.783	35.47
RpS3	3R	0.0059	0.1449	0.0410	0.724	35.41
CG7033	X	0.0211	0.2290	0.0923	0.711	36.39
CG7283	3L	0.0019	0.1616	0.0116	0.737	35.17
CG7380	2L	0.0138	0.5779	0.0238	0.583	47.72
CG7787	2L	0.0130	0.3072	0.0424	0.595	45.66
CG7808	3R	0.0059	0.1847	0.0320	0.743	32.37
RpL32	3R	0.0001	0.1065	0.0010	0.798	30.41
CG8415	2R	0.0001	0.0655	0.0010	0.714	40.81
CG8495	3R	0.0002	0.1614	0.0010	0.704	41.94
Nacalpha	2R	0.0065	0.2382	0.0271	0.670	37.81
CG9091	X	0.0001	0.1476	0.0010	0.822	30.13
CG9354	3R	0.0024	0.1448	0.0167	0.765	32.43
mago	2R	0.0003	0.2770	0.0010	0.697	35.84
Cyp1	X	0.0026	0.2192	0.0118	0.755	35.81
Arc-p34	2L	0.0023	0.2130	0.0109	0.481	53.81
CG10424	3L	0.0291	0.2889	0.1007	0.507	52.44
CG13603	3R	0.0129	0.2512	0.0514	0.577	46.28
CG17672	3L	0.0100	0.2013	0.0495	0.743	34.26
CG4338	3R	0.0108	0.3292	0.0329	0.487	49.43
CSN5	3R	0.0004	0.3649	0.0010	0.591	46.05
ran	X	0.0004	0.3514	0.0010	0.676	42.07
RpL23a	3L	0.0026	0.1385	0.0189	0.752	32.82
Scsalpha	3L	0.0023	0.1436	0.0161	0.688	36.47
thioredoxin	na	0.0254	0.2266	0.1122	0.733	33.05
Nonsex-biased						
CG10219	3R	0.0182	0.2417	0.0754	0.638	41.79
Rpn9	3R	0.0043	0.2978	0.0143	0.604	43.68
Hs2st	2L	0.0050	0.2340	0.0213	0.552	50.49
CG10320	2R	0.0082	0.0956	0.0854	0.682	34.86
CG10472	3L	0.0238	0.4144	0.0574	0.695	36.60
CG10527	2R	0.0027	0.3610	0.0075	0.682	36.27
Os9,CG10658	2L	0.0690	0.3794	0.1817	0.495	53.68

(Continues...)

Table S1. *(Cont.)*

Gene	Loc[a]	d_N	d_S	d_N/d_S	F_{op}	ENC
CG10837	U	0.0686	0.4470	0.1534	0.223	48.75
ProsMA5	2R	0.0048	0.2970	0.0160	0.586	43.59
CG10992	X	0.0209	0.3374	0.0620	0.580	42.04
CG11024	2L	0.0162	0.2902	0.0559	0.681	38.33
CG11151	X	0.0035	0.1638	0.0211	0.809	30.67
CG1124	3R	0.0174	0.3195	0.0543	0.523	51.71
CG11455	2L	0.0003	0.2837	0.0010	0.634	40.52
LysS	3L	0.0061	0.5932	0.0104	0.718	31.97
CG11752	X	0.0272	0.2751	0.0989	0.551	33.18
CG11892	3R	0.0427	0.3810	0.1122	0.482	53.47
CG12203	X	0.0103	0.3624	0.0283	0.719	34.64
CG12292	2L	0.0192	0.2314	0.0828	0.552	45.55
Prosbeta5	2R	0.0233	0.2743	0.0849	0.620	45.90
CG12848	2R	0.0086	0.5103	0.0168	0.716	35.99
CG13095	2L	0.0231	0.5249	0.0440	0.683	34.56
EG:34F3.5	X	0.0065	0.6044	0.0107	0.699	38.85
CG13618	3R	0.0089	0.1147	0.0778	0.556	49.17
CG14619	X	0.0040	0.2493	0.0160	0.446	52.02
CG15027	X	0.0171	0.3710	0.0460	0.587	46.01
CG1532	X	0.0072	0.3658	0.0197	0.468	51.57
Scp1	3h	0.0047	0.1218	0.0388	0.286	55.26
guf	2R	0.0111	0.1004	0.1102	0.446	36.68
CG17202	3R	0.0737	0.5258	0.1402	0.687	41.06
Pglym	3R	0.0072	0.2888	0.0248	0.811	31.14
CG17327	3R	0.0006	0.5606	0.0010	0.615	48.42
CG17333	X	0.0377	0.3886	0.0971	0.538	47.43
Qm	3L	0.0003	0.2801	0.0010	0.374	47.32
CG17571	2L	0.0322	0.2961	0.1086	0.696	38.51
CG18081	3L	0.0053	0.4120	0.0128	0.676	46.09
Rpn11	2L	0.0001	0.1338	0.0010	0.661	42.74
CG18180	3L	0.0308	0.4039	0.0762	0.702	36.33
CG1883	3R	0.0021	0.0947	0.0227	0.808	30.11
Mlc2	3R	0.0000	0.0287	0.0010	0.657	37.91
Fer1HCH	3R	0.0121	0.1936	0.0624	0.791	30.84
CG2471	X	0.0110	0.4113	0.0267	0.420	54.36
Transaldolase, CG2827	2R	0.0002	0.2477	0.0010	0.740	35.27
Alas	2R	0.0336	0.5462	0.0615	0.607	44.00
CG3040	X	0.0171	0.2832	0.0603	0.621	40.42
l(2)k09913	2R	0.0155	0.3109	0.0498	0.657	37.92
CG14558	3R	0.0087	0.3646	0.0238	0.573	42.10
CG9645	3R	0.0717	0.2695	0.2659	0.409	54.97

(Continues...)

Table S1. *(Cont.)*

Gene	Loc[a]	d_N	d_S	d_N/d_S	F_{op}	ENC
Vha16	2R	0.0000	0.0477	0.0010	0.533	41.87
CG9332	2L	0.0433	0.2357	0.1836	0.474	49.10
CG5453	2L	0.0677	0.2313	0.2926	0.511	51.14
CG3214	2L	0.0140	0.3646	0.0383	0.712	37.42
CG1534	X	0.0001	0.0899	0.0010	0.578	44.40
CG7217	3R	0.0053	0.4650	0.0115	0.559	42.85
rab1	3R	0.0002	0.1601	0.0010	0.483	61.00
CG3321	3R	0.0082	0.2850	0.0288	0.734	34.97
CG3344	3L	0.0249	0.2449	0.1018	0.588	42.15
CG3446	X	0.0027	0.2527	0.0108	0.707	39.88
Cyp4d2	X	0.0225	0.4900	0.0459	0.514	51.80
sqh	X	0.0026	0.2378	0.0110	0.738	36.72
CG3609	2L	0.0122	0.1882	0.0647	0.728	34.19
CG3683	2R	0.0166	0.2409	0.0690	0.635	45.08
CG3760 *anon2C9*	2R	0.0155	0.1572	0.0986	0.638	48.62
Spat	X	0.0045	0.5743	0.0078	0.614	41.64
CG4108	3L	0.0004	0.4378	0.0010	0.586	46.70
mbf1	3L	0.0030	0.2238	0.0132	0.549	46.00
CG4370	3R	0.0035	0.4459	0.0078	0.522	50.68
CG4413	3R	0.0290	0.4237	0.0685	0.512	49.83
crl	X	0.0083	0.1523	0.0543	0.516	49.18
CG4666	X	0.0171	0.4531	0.0378	0.676	36.25
CG4692	2R	0.0002	0.1892	0.0010	0.772	31.43
CG4716	2R	0.1536	0.3765	0.4081	0.498	56.67
CG5134	2R	0.0031	0.3535	0.0088	0.606	44.70
CG5317	2L	0.0158	0.1635	0.0964	0.632	40.70
Rpn7	3R	0.0049	0.4747	0.0104	0.684	37.32
CG5390	2L	0.2004	0.5724	0.3501	0.524	52.34
CG5445	X	0.0268	0.2414	0.1112	0.520	43.48
CG5548	X	0.0064	0.1890	0.0337	0.829	25.76
CG5703	X	0.0003	0.2741	0.0010	0.672	38.34
fbl	3L	0.0051	0.2527	0.0201	0.553	49.65
CG5770	2R	0.0588	0.3640	0.1614	0.593	49.55
CG5844	3R	0.0022	0.3686	0.0060	0.590	41.43
CG5885, *BEST:CK01296*	2L	0.0003	0.2795	0.0010	0.738	33.99
CG5902	3R	0.0056	0.2836	0.0196	0.637	46.51
CG6084	3L	0.0048	0.1678	0.0289	0.550	51.92
l(2)06225 , CG6105	2L	0.0002	0.2014	0.0010	0.656	42.66
CG6115	2L	0.0049	0.2268	0.0217	0.687	43.41
Vha13	3R	0.0033	0.2048	0.0163	0.772	37.46
CG6364	3R	0.0003	0.3112	0.0010	0.561	50.51

(Continues...)

Table S1. *(Cont.)*

Gene	Loc[a]	d_N	d_S	d_N/d_S	F_{op}	ENC
BM-40/SPARC	3R	0.0188	0.5566	0.0337	0.760	34.25
CG6398	X	0.0002	0.1691	0.0010	0.651	40.46
CG6543	2R	0.0084	0.2788	0.0302	0.750	34.67
CG6666	3R	0.0029	0.2708	0.0106	0.607	39.17
CG6746	2L	0.0041	0.3007	0.0136	0.664	39.97
ctp	X	0.0001	0.1137	0.0010	0.541	39.81
CG7118	3L	0.0087	0.2932	0.0297	0.741	32.92
CG7170	3L	0.0205	0.4136	0.0497	0.690	36.64
CG7224	2L	0.0183	0.4779	0.0382	0.559	44.72
RFeSP, CG7361	2L	0.0081	0.2139	0.0376	0.556	46.51
fln	3L	0.0026	0.1637	0.0158	0.407	54.82
CG7542	3L	0.0531	0.3035	0.1750	0.542	49.81
ATPsyn-gamma	3R	0.0022	0.1175	0.0190	0.756	32.12
CG7770	3L	0.0439	0.2962	0.1482	0.723	37.26
CG7778	2L	0.0534	0.3010	0.1772	0.510	46.90
BG:DS00941.14	2L	0.0192	0.1994	0.0962	0.711	36.01
Cys	3R	0.0838	0.5773	0.1451	0.525	48.07
Mlf	2R	0.0099	0.2406	0.0413	0.497	50.55
CG8309	2R	0.0001	0.1482	0.0010	0.635	44.55
Cam	2R	0.0001	0.0790	0.0010	0.626	43.79
CG8586	2R	0.0988	0.4586	0.2155	0.550	48.20
CG8588	3L	0.1116	0.3057	0.3650	0.425	57.60
Ser4	2L	0.0282	0.1659	0.1698	0.773	29.03
CG8869	2L	0.0279	0.4038	0.0692	0.786	29.07
Trip1	2L	0.0025	0.1642	0.0152	0.615	42.46
wal	2R	0.0040	0.2270	0.0175	0.599	40.23
CG9066	X	0.0275	0.2820	0.0976	0.461	54.64
Kisir	2L	0.0003	0.3065	0.0010	0.660	34.80
mtacp1	3L	0.0221	0.0439	0.5032	0.586	45.68
eIF5, CG9177	X	0.0002	0.2180	0.0010	0.500	48.20
CG9288	3R	0.0703	0.3548	0.1982	0.519	52.03
CG9306	2L	0.0126	0.2870	0.0439	0.696	43.29
CG9336	2L	0.0230	0.2787	0.0824	0.692	41.16
CG9344	2R	0.0003	0.2847	0.0010	0.727	33.23
CG9471	3R	0.0130	0.3651	0.0356	0.641	37.69
Ag5r	X	0.0361	0.2987	0.1207	0.557	47.98
CG13356	na	0.0381	0.3197	0.1192	na	na

[a] Loc = chromosome arm of locus

Table S2. Evolutionary rates of sex-biased genes in the *apterous, even-skipped, fushi tarazu, twist,* and *Rhodopsin 1, 2, 3,* and *4* genomic regions (Bergman et al. 2002).

Gene	d_N/d_S (mode 0)	pairwise (mode -2)	d_N (mode -2)	d_S (mode -2)	d_N/d_S (mode -2)
Male-biased					
CG10887	0.4168	2 (ere) ... 1 (mel)	0.2644	0.5294	0.4994
		3 (wil) ... 1 (mel)	0.5710	6.7859	0.0841
		3 (wil) ... 2 (ere)	0.6222	9.7285	0.0640
CG13030	0.1143	2 (ere) ... 1 (mel)	0.0660	0.3661	0.1802
		3 (pse) ... 1 (mel)	0.1589	12.3600	0.0129
		3 (pse) ... 2 (ere)	0.1290	6.2290	0.0207
		4 (wil) ... 1 (mel)	0.1587	53.2832	0.0030
		4 (wil) ... 2 (ere)	0.1192	24.1942	0.0049
		4 (wil) ... 3 (pse)	0.1365	56.6106	0.0024
CG14297	0.1222	2 (ere) ... 1 (mel)	0.0445	0.3421	0.1300
		3 (pse) ... 1 (mel)	0.2482	3.0802	0.0806
		3 (pse) ... 2 (ere)	0.2364	3.0430	0.0777
		4 (wil) ... 1 (mel)	0.2267	5.0364	0.0450
		4 (wil) ... 2 (ere)	0.2405	4.6146	0.0521
		4 (wil) ... 3 (pse)	0.2975	3.7390	0.0796
		5 (vir) ... 1 (mel)	0.3050	43.6872	0.0070
		5 (vir) ... 2 (ere)	0.3133	30.7271	0.0102
		5 (vir) ... 3 (pse)	0.2961	13.5080	0.0219
		5 (vir) ... 4 (wil)	0.3267	3.7462	0.0872
CG3085	0.0072	2 (ere) ... 1 (mel)	0.0037	0.5077	0.0072
CG4390	0.0356	2(pse) ... 1 (mel)	0.0857	3.4262	0.0250
		3 (wil) ... 1 (mel)	0.0786	55.3263	0.0014
		3 (wil) ... 2 (pse)	0.0824	2.3009	0.0358
CG4538	0.0822	2 (ere) ... 1 (mel)	0.0172	0.2471	0.0696
		3 (pse) ... 1 (mel)	0.1093	1.5461	0.0707
		3 (pse) ... 2 (ere)	0.1115	1.7953	0.0621
		4 (wil) ... 1 (mel)	0.1249	8.2559	0.0151
		4 (wil) ... 2 (ere)	0.1165	48.7635	0.0024
		4 (wil) ... 3 (pse)	0.1376	3.6958	0.0372
		5 (vir) ... 1 (mel)	0.1239	2.3396	0.0529
		5 (vir) ... 2 (ere)	0.1272	2.2982	0.0554
		5 (vir) ... 3 (pse)	0.1238	1.6180	0.0765
		5 (vir) ... 4 (wil)	0.1312	2.4726	0.0531
CG4562	0.0247	2 (pse) ... 1 (mel)	0.0943	1.6961	0.0556
		3 (wil) ... 1 (mel)	0.1172	3.2957	0.0355
		3 (wil) ... 2 (pse)	0.0304	1.7437	0.0175

(Continues...)

Table S2. *(Cont.)*

Gene	d_N/d_S (mode 0)	pairwise (mode -2)	d_N (mode -2)	d_S (mode -2)	d_N/d_S (mode -2)
Indy-2	0.0716	2 (pse) ... 1 (mel)	0.1580	2.2220	0.0711
twi	0.0609	2 (ere) ... 1 (mel)	0.0152	0.2468	0.0618
		3 (pse) ... 1 (mel)	0.1154	2.3721	0.0487
		3 (pse) ... 2 (ere)	0.1190	1.4455	0.0823
		4 (vir) ... 1 (mel)	0.1979	6.3920	0.0310
		4 (vir) ... 2 (ere)	0.1934	6.0435	0.0320
		4 (vir) ... 3 (pse)	0.1722	68.0027	0.0025
CG4468	0.1498	2 (ere) ... 1 (mel)	0.0108	0.1401	0.0768
		3 (wil) ... 1 (mel)	0.1540	1.2457	0.1236
		3 (wil) ... 2 (ere)	0.1743	1.0587	0.1646
		4 (vir) ... 1 (mel)	0.0797	1.0462	0.0761
		4 (vir) ... 2 (ere)	0.0640	1.1747	0.0545
		4 (vir) ... 3 (wil)	0.2075	0.9167	0.2263
Female-biased					
Arc42	0.0142	2 (ere) ... 1 (mel)	0.0004	0.3953	0.0010
		3 (pse) ... 1 (mel)	0.0327	2.1915	0.0149
		3 (pse) ... 2 (ere)	0.0323	2.4208	0.0133
		4 (wil) ... 1 (mel)	0.0383	8.1127	0.0047
		4 (wil) ... 2 (ere)	0.0384	6.5773	0.0058
		4 (wil) ... 3 (pse)	0.0379	2.7510	0.0138
CG4973	0.0346	2 (pse) ... 1 (mel)	0.0888	2.9399	0.0302
		3 (wil) ... 1 (mel)	0.1183	4.3380	0.0273
		3 (wil) ... 2 (pse)	0.0939	3.3332	0.0282
Fib	0.0235	2 (ere) ... 1 (mel)	0.0044	0.1860	0.0234
sina	0.0102	2 (ere) ... 1 (mel)	0.0002	0.1581	0.0010
		3 (wil) ... 1 (mel)	0.0238	23.8428	0.0010
		3 (wil) ... 2 (ere)	0.0242	24.1674	0.0010
Surf6	0.0383	2 (ere) ... 1 (mel)	0.0201	0.5440	0.0370
		3 (pse) ... 1 (mel)	0.1806	6.9374	0.0260
		3 (pse) ... 2 (ere)	0.1744	4.9649	0.0351
		4 (wil) ... 1 (mel)	0.2023	13.4674	0.0150
		4 (wil) ... 2 (ere)	0.1967	13.8329	0.0142
		4 (wil) ... 3 (pse)	0.2296	8.9342	0.0257
		5 (vir) ... 1 (mel)	0.1995	63.9241	0.0031
		5 (vir) ... 2 (ere)	0.1980	66.9457	0.0030
		5 (vir) ... 3 (pse)	0.2509	6.1731	0.0407
		5 (vir) ... 4 (wil)	0.2713	7.8624	0.0345
Nonsex-biased					
CG10505	0.0050	2 (wil) ... 1 (mel)	0.1273	28.5055	0.0045
CG10754	0.0008	2 (vir) ... 1 (mel)	0.0238	2.2028	0.0108

(Continues...)

Table S2. *(Cont.)*

Gene	d_N/d_S (mode 0)	pairwise (mode -2)	d_N (mode -2)	d_S (mode -2)	d_N/d_S (mode -2)
CG10967	0.0343	2 (vir) ... 1 (mel)	0.0408	1.1958	0.0341
CG11008	0.1724	2 (vir) ... 1 (mel)	0.7706	4.5045	0.1711
CG12848	0.0059	2 (vir) ... 1 (mel)	0.0423	7.1299	0.0059
CG14291	0.0670	2 (ere) ... 1 (mel)	0.0338	0.5838	0.0578
		3 (pse) ... 1 (mel)	0.0984	1.3293	0.0740
		3 (pse) ... 2 (ere)	0.0926	1.7069	0.0542
		4 (wil) ... 1 (mel)	0.1311	2.4634	0.0532
		4 (wil) ... 2 (ere)	0.1408	4.5353	0.0310
		4 (wil) ... 3 (pse)	0.0857	1.8937	0.0452
		5 (vir) ... 1 (mel)	0.1526	1.6880	0.0904
		5 (vir) ... 2 (ere)	0.1530	2.5580	0.0598
		5 (vir) ... 3 (pse)	0.0942	1.6021	0.0588
		5 (vir) ... 4 (wil)	0.1180	1.7068	0.0692
CG17667	0.1773	2 (vir) ... 1 (mel)	0.2446	1.3730	0.1781
CG2222	0.0142	2 (wil) ... 1 (mel)	0.1255	9.4355	0.0133
CG4465	0.1335	2 (ere) ... 1 (mel)	0.0616	0.6457	0.0954
		3 (wil) ... 1 (mel)	0.6306	51.8075	0.0122
		3 (wil) ... 2 (ere)	0.6272	51.3020	0.0122
CG4572	0.0285	2 (pse) ... 1 (mel)	0.0632	2.1957	0.0288
CG4686	0.0025	2 (pse) ... 1 (mel)	0.2102	2.1316	0.0986
		3 (wil) ... 1 (mel)	0.1233	48.0940	0.0026
		3 (wil) ... 2 (pse)	0.1980	47.5965	0.0042
CG8319	0.0031	2 (wil) ... 1 (mel)	0.4399	55.5112	0.0079
CG9715	0.1441	2 (ere) ... 1 (mel)	0.0561	0.3125	0.1797
		3 (wil) ... 1 (mel)	0.3356	4.8422	0.0693
		3 (wil) ... 2 (ere)	0.3224	5.0955	0.0633
CG9882	0.0482	2 (ere) ... 1 (mel)	0.0186	0.3866	0.0482
CG9890	0.0627	2 (ere) ... 1 (mel)	0.0293	0.4675	0.0627
CG9895	0.1161	2 (pse) ... 1 (mel)	0.1486	1.9786	0.0751
		3 (vir) ... 1 (mel)	0.1805	2.0294	0.0889
		3 (vir) ... 2 (pse)	0.1485	1.9626	0.0756
CG9951	0.0723	2 (ere) ... 1 (mel)	0.0239	0.2625	0.0909
		3 (pse) ... 1 (mel)	0.1526	2.4983	0.0611
		3 (pse) ... 2 (ere)	0.1553	2.4700	0.0629
		4 (wil) ... 1 (mel)	0.2351	54.0567	0.0043
		4 (wil) ... 2 (ere)	0.2357	54.8119	0.0043
		4 (wil) ... 3 (pse)	0.2079	3.7392	0.0556
Dox-A3	0.1022	2 (ere) ... 1 (mel)	0.0531	0.5192	0.1022
ftz	0.1605	2 (ere) ... 1 (mel)	0.0235	0.1978	0.1190
		3 (pse) ... 1 (mel)	0.2178	1.5275	0.1426

(Continues...)

Table S2. *(Cont.)*

Gene	d_N/d_S (mode 0)	pairwise (mode -2)	d_N (mode -2)	d_S (mode -2)	d_N/d_S (mode -2)
		3 (pse) ... 2 (ere)	0.2180	1.6811	0.1297
		4 (vir) ... 1 (mel)	0.2498	1.8886	0.1323
		4 (vir) ... 2 (ere)	0.2406	1.9731	0.1219
		4 (vir) ... 3 (pse)	0.2822	1.6796	0.1680
ninaE	0.0140	2 (ere) ... 1 (mel)	0.0002	0.1766	0.0010
		3 (pse) ... 1 (mel)	0.0102	0.8695	0.0118
		3 (pse) ... 2 (ere)	0.0102	0.9159	0.0111
		4 (wil) ... 1 (mel)	0.0075	1.1955	0.0062
		4 (wil) ... 2 (ere)	0.0075	1.3941	0.0054
		4 (wil) ... 3 (pse)	0.0151	1.4866	0.0102
		5 (vir) ... 1 (mel)	0.0107	0.9011	0.0119
		5 (vir) ... 2 (ere)	0.0107	0.9307	0.0115
		5 (vir) ... 3 (pse)	0.0106	1.2291	0.0086
		5 (vir) ... 4 (wil)	0.0139	1.0552	0.0132
Pk92B	0.0365	2 (ere) ... 1 (mel)	0.0926	0.6311	0.1466
		3 (pse) ... 1 (mel)	0.0629	1.2242	0.0513
		3 (pse) ... 2 (ere)	0.1076	2.1242	0.0506
		4 (wil) ... 1 (mel)	0.1469	2.0146	0.0729
		4 (wil) ... 2 (ere)	0.1799	8.3102	0.0217
		4 (wil) ... 3 (pse)	0.1738	2.0266	0.0857
		5 (vir) ... 1 (mel)	0.3132	2.0379	0.1537
		5 (vir) ... 2 (ere)	0.2966	1.9708	0.1505
		5 (vir) ... 3 (pse)	0.3660	2.8143	0.1304
		5 (vir) ... 4 (wil)	0.2636	1.7602	0.1497
RhoGAP92B	0.0467	2 (ere) ... 1 (mel)	0.0098	0.2473	0.0398
		3 (wil) ... 1 (mel)	0.0801	3.5550	0.0225
		3 (wil) ... 2 (ere)	0.0789	3.4214	0.0230
		4 (vir) ... 1 (mel)	0.0824	2.2043	0.0374
		4 (vir) ... 2 (ere)	0.0803	2.3629	0.0340
		4 (vir) ... 3 (wil)	0.0900	1.6595	0.0543
Scr	0.0318	2 (vir) ... 1 (mel)	0.0768	2.6792	0.0287
Adam	0.0624	2 (ere) ... 1 (mel)	0.0108	0.2613	0.0414
		3 (pse) ... 1 (mel)	0.1189	2.0641	0.0576
		3 (pse) ... 2 (ere)	0.1218	1.6587	0.0734
		4 (wil) ... 1 (mel)	0.1139	4.6528	0.0245
		4 (wil) ... 2 (ere)	0.1181	3.5641	0.0331
		4 (wil) ... 3 (pse)	0.1607	2.5049	0.0641
Antp	0.0001	2 (ere) ... 1 (mel)	0.0002	0.1866	0.0010
		3 (pse) ... 1 (mel)	0.0011	1.1246	0.0010
		3 (pse) ... 2 (ere)	0.0012	1.2357	0.0010

(Continues...)

Table S2. *(Cont.)*

Gene	d_N/d_S (mode 0)	pairwise (mode -2)	d_N (mode -2)	d_S (mode -2)	d_N/d_S (mode -2)
ap	0.0286	2 (ere) ... 1 (mel)	0.0053	0.0678	0.0786
		3 (pse) ... 1 (mel)	0.0142	0.4218	0.0337
		3 (pse) ... 2 (ere)	0.0142	0.3578	0.0396
		4 (wil) ... 1 (mel)	0.0195	0.4576	0.0427
		4 (wil) ... 2 (ere)	0.0195	0.4029	0.0483
		4 (wil) ... 3 (pse)	0.0112	0.2572	0.0434
		5 (vir) ... 1 (mel)	0.0223	0.3390	0.0657
		5 (vir) ... 2 (ere)	0.0220	0.3230	0.0681
		5 (vir) ... 3 (pse)	0.0199	0.2979	0.0668
		5 (vir) ... 4 (wil)	0.0206	0.2112	0.0975
CG11915	0.0281	2 (pse) ... 1 (mel)	0.1901	6.7040	0.0284
CG12130	0.1064	2 (ere) ... 1 (mel)	0.0353	0.2274	0.1551
		3 (pse) ... 1 (mel)	0.1230	2.0192	0.0609
		3 (pse) ... 2 (ere)	0.1359	1.7733	0.0766
CG17836	0.4077	2 (wil) ... 1 (mel)	0.4990	1.2190	0.4093
CG3880	0.0315	2 (pse) ... 1 (mel)	0.0777	63.1553	0.0012
		3 (vir) ... 1 (mel)	0.1365	2.6029	0.0524
		3 (vir) ... 2 (pse)	0.1129	3.7472	0.0301
CG9196	0.0341	2 (pse) ... 1 (mel)	0.0519	1.7455	0.0297
		3 (vir) ... 1 (mel)	0.0677	31.1692	0.0022
		3 (vir) ... 2 (pse)	0.0553	2.3463	0.0236
Lmpt	0.1115	2 (ere) ... 1 (mel)	0.0167	0.0510	0.3281
		3 (pse) ... 1 (mel)	0.0178	0.3462	0.0515
		3 (pse) ... 2 (ere)	0.0355	0.2739	0.1294
		4 (wil) ... 1 (mel)	0.0749	5.2242	0.0143
		4 (wil) ... 2 (ere)	0.0929	4.6650	0.0199
		4 (wil) ... 3 (pse)	0.1001	2.3491	0.0426
Pka-R2	0.0323	2 (ere) ... 1 (mel)	0.0016	0.0494	0.0319
Rh3	0.0394	2 (ere) ... 1 (mel)	0.0124	0.2352	0.0528
		3 (pse) ... 1 (mel)	0.0366	1.3520	0.0271
		3 (pse) ... 2 (ere)	0.0420	1.5967	0.0263
		4 (wil) ... 1 (mel)	0.0438	1.8085	0.0242
		4 (wil) ... 2 (ere)	0.0511	2.0721	0.0247
		4 (wil) ... 3 (pse)	0.0521	1.8855	0.0276
		5 (vir) ... 1 (mel)	0.0674	2.4471	0.0275
		5 (vir) ... 2 (ere)	0.0711	4.1356	0.0172
		5 (vir) ... 3 (pse)	0.0669	1.7249	0.0388
		5 (vir) ... 4 (wil)	0.0570	1.9351	0.0295
sina	0.0102	2 (ere) ... 1 (mel)	0.0002	0.1581	0.0010
		3 (wil) ... 1 (mel)	0.0246	2.1894	0.0112

(Continues...)

Table S2. *(Cont.)*

Gene	d_N/d_S (mode 0)	pairwise (mode -2)	d_N (mode -2)	d_S (mode -2)	d_N/d_S (mode -2)
		3 (wil) ... 2 (ere)	0.0242	24.1674	0.0010
vlc	0.1014	2 (wil) ... 1 (mel)	0.3055	3.0061	0.1016

Table S3. Conservation of sex-biased genes between *D. melanogaster* and *D. pseudoobscura*.

Gene	Loc[a]	score	e-value	F_{op}	ENC
Male-biased					
CG7363	2L	40	0.14	0.349	55.66
CG2955	2L	38	0.74	0.284	57.50
CG17210	3R	383	1.00E-105	0.629	44.27
CG5207	3R	377	1.00E-103	0.620	45.04
CG15635	2L	44	2.40E-02	0.158	41.85
CG32061	3L	40	7.00E-02	0.343	54.02
CG32063	3L	159	3.00E-37	0.447	52.63
CG32064	3L	119	2.00E-25	0.490	51.20
CG9254	2L	107	8.00E-22	0.385	61.00
CG13263	2L	109	4.00E-23	0.559	49.19
CG5790	2L	82	7.00E-14	0.412	56.34
CG31948	2L	42	2.50E-02	0.397	61.00
CG31684	2L	186	1.00E-45	0.432	56.76
CG3213	2L	412	1.00E-133	0.645	42.37
CG14098	3L	119	2.00E-25	0.472	56.38
CG31025	3R	90	4.00E-16	0.531	51.40
CG31029	3R	70	3.00E-10	0.534	53.26
CG14735	3R	48	1.00E-03	0.521	51.66
CG10252	3R	90	9.00E-17	0.434	58.30
CG31872	2L	50	4.00E-04	0.312	56.08
CG4836	3R	64	2.00E-08	0.537	49.33
CG32079	3L	40	0.14	0.303	57.75
CG1262	3L	36	0.58	0.389	60.26
CG7045	3R	36	0.64	0.558	54.91
CG8564	3L	190	7.00E-47	0.493	51.36
CG8526	3R	182	2.00E-44	0.548	50.28
CG18396	3R	424	1.00E-117	0.523	53.85
CG6663	3L	42	0.03	0.411	61.00
CG31431	3R	92	4.00E-17	0.448	54.65
CG31178	3R	36	1	0.360	55.84
CG9389	3L	62	5.00E-08	0.441	55.74
CG6279	3L	78	1.00E-12	0.376	58.39
CG1014	3L	38	0.13	0.578	36.25
CG3964	2L	761	0	0.562	44.02
CG2830	2L	170	9.00E-41	0.507	49.94
CG6917	3L	109	2.00E-22	0.391	54.65
CG11201	2L	391	1.00E-107	0.630	40.90
CG14305	3R	389	1.00E-107	0.474	56.46

(Continues...)

Table S3. *(Cont.)*

Gene	Loc[a]	score	e-value	F_{op}	ENC
CG7131	3R	315	2.00E-84	0.546	45.22
CG9314	2L	662	0.00E+00	0.566	48.50
CG6569	3R	40	1.40E-01	0.547	47.97
CG8565	X	62	7.00E-08	0.462	53.19
CG5565	2L	40	8.40E-02	0.432	60.64
CG6372	3L	400	1.00E-110	0.583	47.28
CG4434	3R	218	3.00E-55	0.449	56.49
CG8813	2L	194	3.00E-48	0.655	42.67
CG9389	3L	62	5.00E-08	0.499	56.37
CG18427	3R	109	6.00E-22	0.448	54.16
CG10177	3R	44	0.009	0.519	48.86
CG4161	2L	78	7.00E-13	0.376	58.88
CG8701	2R	113	7.00E-24	0.564	51.71
CG7157	2L	48	1.00E-03	0.300	51.48
CG4750	2R	436	1.00E-121	0.594	42.50
CG4983	2L	38	4.80E-01	0.352	60.83
CG4691	2L	40	8.60E-02	0.540	43.10
CG2668	2R	40	1.30E-01	0.216	54.58
CG12699	2R	36	7.40E-01	0.387	47.09
CG6304	2L	40	1.40E-01	0.377	55.49
CG14718	3R	96	3.00E-18	0.534	52.01
CG4439	2R	78	8.00E-13	0.540	48.63
CG5762	3R	40	6.60E-02	0.539	57.05
CG5538	3R	40	1.30E-01	0.354	57.60
CG5089	2R	40	1.40E-01	0.492	52.76
CG14740	3R	109	2.00E-22	0.503	52.95
CG3330	3R	70	1.00E-10	0.534	48.49
CG8136	3R	40	0.15	0.496	51.32
CG7742	2L	367	1.00E-100	0.500	51.17
CG17302	2L	98	3.00E-19	0.469	55.14
CG1380	2R	60	2.00E-07	0.458	50.88
CG7387	3L	44	1.00E-02	0.526	51.76
CG32388	3L	34	1.50E+00	0.453	51.98
CG18418	3L	40	1.00E-01	0.472	54.53
CG3565	2R	46	1.00E-03	0.541	47.33
CG18568	2R	72	3.00E-11	0.492	55.28
CG11475	2R	52	4.00E-05	0.513	53.94
CG3306	3L	38	3.80E-01	0.475	52.13
CG30438	2R	133	1.00E-29	0.347	57.67
CG4767	2L	119	3.00E-25	0.474	55.31
CG9218	2R	343	8.00E-93	0.575	47.14

(Continues...)

Table S3. *(Cont.)*

Gene	Loc[a]	score	e-value	F_{op}	ENC
CG32849	3R	137	9.00E-31	0.575	48.60
CG31231	3R	40	2.30E-01	0.535	49.31
CG4995	2L	42	3.40E-02	0.497	55.04
CG9010	2R	139	2.00E-31	0.526	54.51
CG5398	2R	40	7.80E-02	0.548	51.72
CG7557	3L	40	1.20E-01	0.415	56.96
CG30416	2R	38	7.10E-01	0.414	57.29
CG9803	2R	92	3.00E-17	0.699	38.59
CG3483	2R	157	8.00E-37	0.343	54.02
CG17401	2L	196	4.00E-49	0.166	47.62
CG17022	2L	38	4.30E-01	0.673	48.41
CG5443	3R	84	1.00E-14	0.436	58.54
CG4961	3R	105	3.00E-21	0.520	47.05
CG8278	2R	297	1.00E-78	0.571	44.40
Female-biased					
CG4193	X	38	0.14	0.650	40.88
CG7660	3R	115	6.00E-24	0.444	53.34
CG17489	2h	278	2.00E-73	0.495	48.30
CG2016	3R	216	6.00E-55	0.663	39.86
CG11674	X	42	2.90E-02	0.316	50.53
CG10901	3R	78	9.00E-13	0.486	50.47
CG4916	2L	159	2.00E-37	0.408	59.69
CG1372	X	283	3.00E-74	0.551	44.02
CG2979	X	624	1.00E-177	0.744	33.36
CG3510	2R	400	1.00E-110	0.581	45.38
CG6927	X	381	1.00E-104	0.521	47.51
CG13804	3L	630	1.00E-179	0.612	42.49
CG8893	X	482	1.00E-135	0.686	35.57
CG10071	2R	125	5.00E-28	0.743	38.11
CG14309	3R	216	2.00E-54	0.482	53.98
CG6293	3R	452	1.00E-125	0.486	51.60
CG6779	3R	894	0.00E+00	0.724	35.41
CG9809	3R	123	3.00E-26	0.376	59.37
CG17950	2R	157	2.00E-37	0.623	40.95
CG2033	X	460	1.00E-128	0.710	36.68
CG5371	2L	42	6.90E-02	0.519	51.00
CG7111	2L	922	0.00E+00	0.786	29.09
CG13849	3R	1007	0.00E+00	0.605	46.65
CG3506	2L	54	1.00E-05	0.397	55.76
CG4897	2L	406	1.00E-112	0.805	33.36
CG11271	3L	353	2.00E-96	0.743	34.26

(Continues...)

Table S3. *(Cont.)*

Gene	Loc[a]	score	e-value	F_{op}	ENC
CG7433	3L	38	0.64	0.520	52.03
CG14516	3R	414	1.00E-114	0.556	48.96
CG7840	2L	117	6.00E-25	0.528	50.15
CG11901	3R	842	0.00E+00	0.830	30.55
CG15442	2L	460	1.00E-128	0.759	31.37
CG2746	2R	541	1.00E-153	0.754	39.88
CG9091	X	165	8.00E-40	0.822	30.13
CG7808	3R	377	1.00E-103	0.750	31.11
CG3314	X	412	1.00E-114	0.798	32.18
CG14444	X	50	2.00E-04	0.474	52.78
CG1475	3R	472	1.00E-132	0.786	34.78
CG8231	X	1015	0.00E+00	0.723	36.23
CG14217	X	256	1.00E-66	0.588	44.57
CG7627	2L	440	1.00E-121	0.469	52.72
CG6141	2L	216	5.00E-55	0.770	37.18
CG17420	3h	222	8.00E-57	0.367	48.37
CG5717	3L	543	1.00E-153	0.741	34.46
CG4183	3L	244	2.00E-63	0.635	39.46
CG7033	X	533	1.00E-150	0.710	36.40
CG9057	X	464	1.00E-129	0.574	50.81
CG10944	X	511	1.00E-144	0.770	33.95
CG7939	3R	351	9.00E-96	0.798	30.41
CG16944	X	722	0.00E+00	0.293	52.73
CG4634	2R	385	1.00E-105	0.709	38.99
CG3509	3R	40	0.12	0.450	52.15
CG2207	2L	143	6.00E-33	0.500	47.81
CG5272	3L	44	5.00E-03	0.430	58.96
CG10895	2L	192	2.00E-47	0.526	50.78
CG10243	2R	54	1.00E-05	0.481	54.58
CG3679	X	52	6.00E-05	0.445	56.53
CG12109	X	317	5.00E-85	0.556	48.23
CG4978	3L	698	0.00E+00	0.543	46.83
CG18543	3L	42	0.018	0.569	42.81
CG5940	3L	174	4.00E-42	0.527	51.74
CG9135	2L	464	1.00E-129	0.557	47.65
CG7719	3R	392	1.00E-107	0.514	50.32
CG15737	X	121	2.00E-25	0.478	49.55
CG5263	3L	577	1.00E-163	0.488	50.44
CG9193	2R	607	1.00E-172	0.676	37.16
CG4039	X	852	0	0.598	40.27
CG10387	2L	343	1.00E-92	0.421	56.00

(Continues...)

Table S3. *(Cont.)*

Gene	Loc[a]	score	e-value	F_{op}	ENC
CG8975	2R	543	1.00E-153	0.699	37.58
CG7242	3R	46	0.001	0.421	61.00
CG2982	X	38	0.86	0.526	48.46
CG8961	2R	40	0.22	0.353	59.34
CG2050	3R	46	0.003	0.387	59.32
CG9924	3R	416	1.00E-115	0.423	56.95
CG3938	2L	420	1.00E-116	0.517	47.87
CG3238	2L	250	1.00E-64	0.529	51.53
CG9752	2R	101	4.00E-20	0.533	51.34
CG6122	2L	86	5.00E-15	0.379	57.57
CG11980	3R	38	0.41	0.474	57.48
CG8180	2R	228	4.00E-58	0.502	45.73
CG4991	X	182	2.00E-44	0.597	44.20
CG7670	3R	167	8.00E-40	0.444	53.90
CG2913	X	137	1.00E-30	0.511	52.89
CG11397	2L	345	6.00E-93	0.505	53.20
CG14814	X	82	1.00E-13	0.490	52.56
CG6226	3R	68	5.00E-10	0.618	42.52
CG5363	2L	46	0.002	0.423	56.26
CG1825	2R	133	1.00E-29	0.549	45.97
CG16838	3L	123	3.00E-26	0.437	56.69
CG4711	2L	168	9.00E-41	0.575	51.35
CG11744	3R	88	7.00E-16	0.488	56.07
CG31450	3R	147	3E-34	0.495	51.38
CG4771	3R	40	0.23	0.482	53.74
Nonsex-biased					
CG11245	X	212	6.00E-53	0.346	52.84
CG16986	3L	56	8.00E-07	0.586	52.71
CG32532	X	396	1.00E-109	0.562	44.00
CG6348	3R	240	6.00E-62	0.499	50.87
CG32809	X	385	1.00E-105	0.548	45.69
CG3200	3L	208	4.00E-52	0.635	42.82
CG3488	2L	226	1.00E-57	0.589	46.64
CG32790	X	111	5.00E-23	0.431	50.29
CG32778	X	299	2.00E-79	0.445	52.43
CG15585	3R	248	2.00E-64	0.608	44.20
CG7009	3R	109	1.00E-22	0.518	52.03
CG2256	X	373	1.00E-102	0.503	50.75
CG17776	X	161	9.00E-39	0.779	30.25
CG15629	2L	438	1.00E-121	0.531	48.70
CG14068	2L	163	5.00E-39	0.455	47.04

(Continues...)

Table S3. *(Cont.)*

Gene	Loc[a]	score	e-value	F_{op}	ENC
CG10874	2L	305	1.00E-81	0.597	42.09
CG16902	X	458	1.00E-127	0.532	44.65
CG12766	3L	109	1.00E-22	0.570	43.88
CG11098	2L	157	3.00E-36	0.444	55.31
CG5387	2L	115	3.00E-24	0.559	48.68
CG1308	3L	377	1.00E-102	0.425	55.16
CG5646	3R	301	2.00E-80	0.614	41.43
CG2930	X	587	1.00E-166	0.529	49.70
CG14938	2L	1084	0.00E+00	0.535	42.23
CG15887	3R	262	6.00E-69	0.654	39.16
CG3823	X	317	3.00E-85	0.528	49.42
CG16787	2R	141	3.00E-32	0.602	45.36
CG3182	2R	42	7.20E-02	0.467	54.02
CG7571	3L	505	1.00E-141	0.613	41.38
CG10539	3L	299	1.00E-79	0.502	51.79
CG12136	X	246	4.00E-63	0.512	48.46
CG13658	3R	80	2.00E-13	0.350	61.00
CG11444	X	127	4.00E-28	0.528	47.90
CG13314	3L	176	3.00E-43	0.500	46.68
CG1486	X	319	2.00E-85	0.530	46.74
CG10582	3L	141	6.00E-32	0.546	52.57
CG12800	3R	246	1.00E-63	0.516	48.79
CG13636	3R	204	2.00E-51	0.624	45.89
CG4097	3L	176	5.00E-43	0.626	41.80
CG10230	3R	337	4.00E-91	0.604	43.68
CG6020	3L	311	2.00E-83	0.676	38.47
CG13610	3R	381	1.00E-104	0.582	43.34
CG7900	3R	70	2.00E-10	0.383	56.23
CG5705	2L	220	6.00E-56	0.532	50.96
CG9326	2L	299	1.00E-79	0.604	46.38
CG1718	X	377	1.00E-102	0.479	52.43
CG8609	3L	153	8.00E-36	0.575	46.24
CG4203	3R	642	0.00E+00	0.548	47.31
CG13043	3L	167	3.00E-40	0.582	40.80
CG9356	3R	309	1.00E-82	0.498	50.35
CG1725	X	214	1.00E-53	0.412	55.58
CG7893	X	636	0.00E+00	0.652	38.95
CG1815	3R	133	5.00E-29	0.402	58.21
CG6483	3L	218	2.00E-55	0.687	36.59
CG10611	2L	434	1.00E-120	0.498	51.57
CG9165	3L	196	2.00E-48	0.496	49.81

(Continues...)

Table S3. *(Cont.)*

Gene	Loc[a]	score	e-value	F_{op}	ENC
CG14548	3R	281	1.00E-74	0.495	50.46
CG8605	3L	192	3.00E-47	0.645	39.55
CG11992	3R	240	2.00E-61	0.629	46.46
CG14548	3R	281	1.00E-74	0.522	50.28
CG2789	2L	143	6.00E-33	0.554	47.84
CG4791	2L	125	6.00E-27	0.629	46.46
CG8804	2R	147	6.00E-34	0.634	43.74
CG2789	2L	143	6.00E-33	0.647	43.16
CG8230	2R	404	1.00E-111	0.529	53.96
CG13853	3R	359	7.00E-98	0.628	42.41
CG4778	2L	242	2.00E-62	0.465	47.20
CG2790	2R	82	5.00E-14	0.507	53.03
CG4008	2L	515	1.00E-145	0.669	41.27
CG1408	3R	117	1.00E-24	0.606	39.32
CG4784	3L	103	1.00E-20	0.442	57.14
CG10509	2R	38	2.90E-01	0.409	55.67
CG30392	2R	204	3.00E-51	0.616	41.60
CG9155	3L	1179	0.00E+00	0.597	42.66
CG8206	X	186	4.00E-46	0.561	46.95
CG11907	2L	101	5.00E-20	0.597	43.33
CG6321	3L	44	0.009	0.791	30.84
CG9331	2L	220	5.00E-56	0.479	52.49
CG9224	X	96	7.00E-18	0.640	37.36
CG10079	2R	841	0.00E+00	0.514	49.43
CG9507	2L	127	1.00E-27	0.562	46.83
CG2216	3R	236	6.00E-61	0.494	50.15
CG7945	3L	184	2.00E-45	0.599	45.24
CG2061	X	680	0.00E+00	0.446	52.02
CG5462	3R	547	1.00E-153	0.531	49.49
CG6227	X	886	0.00E+00	0.736	39.12
CG9628	3L	143	5.00E-33	0.718	36.37
CG3093	X	141	1.00E-31	0.541	48.98
CG18742	2R	141	3.00E-32	0.365	54.40
CG14619	X	212	3.00E-53	0.537	46.71
CG9137	3L	307	1.00E-81	0.537	48.62
CG8808	2R	256	1.00E-66	0.603	43.98
CG12164	2R	311	2.00E-83	0.509	53.79
CG7107	X	628	1.00E-179	0.480	51.86
CG4106	2L	862	0.00E+00	0.507	53.03
CG17888	3L	377	2.00E-91	0.652	41.43
CG17705	3L	317	8.00E-85	0.518	47.39

(Continues...)

Table S3. *(Cont.)*

Gene	Loc[a]	score	e-value	F_{op}	ENC
CG9266	2L	68	1.00E-10	0.518	46.11
CG31125	3R	184	4.00E-15	0.517	51.94

[a] Loc = chromosome arm of locus

Table S4. Polymorphism and divergence of 55 loci in *D. melanogaster* and *D. simulans*. (2-fold cutoff used)

Gene	π_N	π_S	d_N	d_S	Ds	Ps	Dn	Pn	P-value	Reference
Male-biased										
Acp36DE	0.0028	0.0172	0.0517	0.1364	59	31	84	23	0.043	Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG. Genetics. 2000 156:1879-88.
Acp62F	0.0000	0.0187	0.0616	0.1323	8	6	13	0	0.016	Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG. Genetics. 2000 156:1879-88.
Acp76A	0.0000	0.0010	0.0224	0.1593	32	1	16	0	1	Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG. Genetics. 2000 156:1879-88.
CG17097	0.0022	0.0182	0.0174	0.1192	25	19	12	11	0.71	Kohn MH, Fang S, Wu CI. Mol Biol Evol. 2004 21:374-83.
CG8552	0.0223	0.0468	0.0461	0.1170	10	18	11	28	0.6	Kohn MH, Fang S, Wu CI. Mol Biol Evol. 2004 21:374-83.
Ddc	0.0012	0.0092	0.0033	0.0649	19	9	3	6	0.068	De Luca et al. Nat Genet. 2003 34:429-33.
Dntf-2r	0.0015	0.0125	0.0630	0.0935	7	5	18	2	0.038	Betran E, Long M. Genetics. 2003 164:977-88.
dpp	0.0002	0.0123	0.0022	0.0798	10	8	1	1	0.88	Mousset et al. Genetics. 2003 163:599-609.
est-6	0.0029	0.0181	0.0223	0.1319	43	30	21	15	0.95	Hasson E, Eanes WF. Genetics. 1996 144:1565-75.
gld	0.0006	0.0019	0.0046	0.1129	43	4	5	5	0.0036	Hamblin MT, Aquadro CF. Genetics. 1997 145:1053-62.
Hex-t1	0.0007	0.0347	0.0087	0.1275	28	36	8	2	0.028	Duvernell DD, Eanes WF. Genetics. 2000 156:1191-201.
Idgf-3	0.0016	0.0204	0.0116	0.1144	29	18	10	9	0.49	Zurovcova M, Ayala FJ. Genetics. 2002 162:177-88.
ksr	0.0004	0.0017	0.0044	0.1005	70	5	9	4	0.021	Riley RM, Jin W, Gibson G. Mol. Ecol. 2003 12:1315-23.
Female-biased										
anon1E9	0.0004	0.0011	0.0424	0.0896	33	3	57	4	0.74	Schmid KJ, Nigro L, Aquadro CF, Tautz D. Genetics. 1999 153:1717-29.
CG3038	0.0008	0.0017	0.0099	0.0879	17	1	6	1	0.49	Sheldahl LA, Weinreich DM, Rand DM. Genetics. 2003 165:1195-208.
cin	0.0004	0.0009	0.0228	0.1306	23	2	13	2	0.59	Sheldahl LA, Weinreich DM, Rand DM. Genetics. 2003 165:1195-208.
Dras85	0.0000	0.0006	0.0000	0.0381	5	1	0	0	na	Gasperini R, Gibson G. J Mol Evol. 1999 49:583-90.
Dsor1	0.0001	0.0067	0.0095	0.0862	23	8	8	1	0.32	Riley RM, Jin W, Gibson G. Mol. Ecol. 2003 12:1315-23.
hsp83	0.0006	0.0076	0.0017	0.0541	12	7	1	1	0.72	Hasson E, Eanes WF. Genetics. 1996 144:1565-75.
Nup96	0.0007	0.0193	0.0171	0.0957	48	43	36	5	0.00004	Presgraves DC, Balagopalan L, Abmayr SM, Orr HA. Nature. 2003 423:715-9.

(Continues...)

Table S4. *(Cont.)*

Gene	π_N	π_S	d_N	d_S	Ds	Ps	Dn	Pn	P-value	Reference
Pgm	0.0015	0.0055	0.0024	0.1160	46	19	0	16	0	Verrelli BC, Eanes WF. Genetics. 2000 156:1737-52.
ph (phl)	0.0003	0.0218	0.0002	0.0846	30	34	0	4	0.124	Riley RM, Jin W, Gibson G. Mol. Ecol. 2003 12:1315-23.
RhoGAP1A	0.0011	0.0000	0.0049	0.1346	15	0	1	2	0.019	Sheldahl LA, Weinreich DM, Rand DM. Genetics. 2003 165:1195-208.
Sod	0.0017	0.0081	0.0015	0.1087	11	5	0	2	0.14	Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ. Genetics. 1994 136:1329-40.
Yp2	0.0013	0.0146	0.0069	0.0628	12	7	5	2	0.69	Hey J, Kliman RM. Mol Biol Evol. 1993 10:804-22.
Nonsex-biased										
Adh	0.0006	0.0254	0.0038	0.0522	5	18	2	1	0.12	Kreitman M. Nature.1983 10:412-7; Begun et al. Mol Biol Evol. 1999 16:1816-9.
anon1A3	0.0018	0.0044	0.0404	0.0983	20	5	28	11	0.45	Schmid KJ, Nigro L, Aquadro CF, Tautz D. Genetics. 1999 153:1717-29.
cav	0.0016	0.0105	0.0517	0.1083	16	5	29	4	0.267	Schmid KJ, Nigro L, Aquadro CF, Tautz D. Genetics. 1999 153:1717-29.
bcd	0.0009	0.0002	0.0058	0.0937	32	1	6	6	0.003	Baines JF, Chen Y, Das A, Stephan W. Mol Biol Evol. 2002 19:989-98.
bt	0.0000	0.0008	0.0145	0.0866	22	1	11	0	1	Wang W, Thornton K, Berry A, Long M. Science. 2002 295:134-7.
CG5333	0.0014	0.0009	0.0222	0.1407	31	2	16	4	0.13	Kohn MH, Fang S, Wu CI. Mol Biol Evol. 2004 21:374-83.
csw	0.0005	0.0071	0.0057	0.0716	41	14	10	3	0.86	Riley RM, Jin W, Gibson G. Mol. Ecol. 2003 12:1315-23.
Delta	0.0018	0.0189	0.0025	0.0792	18	24	1	9	0.035	Genissel A, Pastinen T, Dowell A, Mackay TF, Long AD. Genetics. 2004 166:291-306.
drk	0.0000	0.0094	0.0000	0.0195	2	7	0	0	na	Riley RM, Jin W, Gibson G. Mol. Ecol. 2003 12:1315-23.
fbp2	0.0018	0.0385	0.0085	0.1300	14	13	4	5	0.7	Benassi V, Depaulis F, Meghlaoui GK, Veuille M. Mol Biol Evol. 1999 16:347-53.
G6pd (Zw)	0.0006	0.0184	0.0206	0.1017	32	37	22	3	0.00013	Eanes WF, et al. PNAS USA 1993 90:7475-9; Eanes WF, et al. Genetics. 1996 144:1027-41.
GH10711	0.0054	0.0272	0.0341	0.1437	8	6	7	5	0.95	Mousset et al. Genetics. 2003 163:599-609.
hairy	0.0006	0.0212	0.0017	0.0764	13	16	1	1	0.88	Robin C, Lyman RF, Long AD, Langley CH, Mackay TF. Genetics. 2002 162:155-64.
Idgf-1	0.0033	0.0499	0.0136	0.1573	34	43	11	9	0.39	Zurovcova M, Ayala FJ. Genetics. 2002 162:177-88.

(Continues...)

Table S4. *(Cont.)*

Gene	π_N	π_S	d_N	d_S	Ds	Ps	Dn	Pn	P-value	Reference
mlc1	0.0000	0.0000	0.0052	0.0780	4	0	1	0	na	Leicht BG, Muse SV, Hanczyc M, Clark AG. Genetics. 1995 139:299-308.
mth	0.0010	0.0085	0.0251	0.1165	38	10	28	9	0.7	Schmidt PS, Duvernell DD, Eanes WF. Proc Natl Acad Sci USA. 2000 97:10861-5.
mth2	0.0001	0.0071	0.0053	0.1297	44	8	6	1	0.94	Duvernell DD, Schmidt PS, Eanes WF. Mol Ecol. 2003 12:1277-85.
Notch	0.0001	0.0120	0.0013	0.2558	117	36	2	2	0.26	Bauer DuMont et al. Genetics. 2004 167:171-85.
per	0.0003	0.0201	0.0017	0.1255	42	19	2	1	0.94	Kliman RM, Hey J. Genetics. 1993 133:375-87.
Pgd	0.0007	0.0047	0.0049	0.1219	26	2	3	1	0.32	Begun DJ, Aquadro CF. Genetics. 1994 136:155-71.
Pgi	0.0005	0.0019	0.0011	0.0651	25	4	1	3	0.012	Unpublished data (J. H. McDonald, Personal communication)
Rel	0.0007	0.0052	0.0502	0.0948	50	7	92	4	0.066	Verrelli BC, Eanes WF. Genetics. 2000 156:1737-52.
runt	0.0011	0.0180	0.0043	0.0770	23	20	4	3	0.86	Labate JA, Biermann CH, Eanes WF. Mol Biol Evol. 1999 16:724-31.
Sos	0.0005	0.0103	0.0028	0.0404	7	8	2	3	0.79	Mousset et al. Genetics. 2003 163:599-609.
su(H)	0.0000	0.0133	0.0041	0.1228	6	3	1	0	1	Depaulis F, Brazier L, Veuille M. Genetics. 1999 152:1017-24.
tinman	0.0012	0.0053	0.0083	0.0639	10	4	4	4	0.32	Balakirev,E.S., Ayala,F.J. Genetics 2004 166:1845-56.
Tpi	0.0003	0.0292	0.0001	0.0685	8	28	0	2	1	Hasson E, Wang IN, Zeng LW, Kreitman M, Eanes WF. Mol Biol Evol. 1998 15:756-69.
vermillion	0.0002	0.0267	0.0012	0.1266	24	42	1	2	0.91	Begun DJ, Aquadro CF. Genetics. 1994 136:155-71.
Vha68-1	0.0002	0.0044	0.0001	0.1555	28	4	0	1	0.15	Depaulis F, Brazier L, Veuille M. Genetics. 1999 152:1017-24.
white	0.0005	0.0054	0.0044	0.1088	50	13	6	2	0.78	Sheldahl LA, Weinreich DM, Rand DM. Genetics. 2003 165:1195-208.

Table S5. Eight estimators of local recombination rates (Marais et al. 2004) and substitution rates for 343 genes classified by male-biased, female-biased, and nonsex-biased (2-fold cutoff used).

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
Male-biased													
CG18284	2L	0.061	0.220	0.277	0.53	6.00	4.637	5.298	3.780	3.678	3.570	3.827	2.960
Kisir	2L	0.000	0.307	0.001	0.71	2.90	3.754	3.593	4.477	4.883	4.983	4.405	3.886
CG6488	2L	0.008	0.346	0.023	0.56	3.00	1.799	2.102	3.446	2.340	3.105	3.483	2.789
CG5171	2L	0.029	0.399	0.073	0.47	3.10	6.101	3.326	4.525	4.265	4.873	4.373	3.616
CG5390	2L	0.200	0.572	0.350	0.41	6.00	6.493	8.213	3.902	4.864	3.733	3.927	3.028
CG5375	2L	0.044	0.437	0.101	0.40	6.00	6.493	8.213	3.894	4.864	3.733	3.927	3.023
CG13770	2L	0.114	0.488	0.233	0.59	3.10	5.392	3.462	4.530	5.681	4.984	4.402	3.766
CG3214	2L	0.014	0.365	0.038	0.72	6.00	4.041	3.249	2.973	6.084	3.616	3.709	4.586
CG5879	2L	0.000	0.194	0.001	0.58	6.00	5.414	3.700	4.025	5.609	4.005	4.071	3.102
CG6115	2L	0.005	0.227	0.022	0.72	1.60	0.849	0.992	0.978	1.886	1.584	1.611	1.714
CG14043	2L	0.019	0.408	0.047	0.58	4.70	3.913	4.318	4.244	3.403	4.803	4.358	4.103
CG9336	2L	0.023	0.279	0.083	0.73	0.32	0.000	0.231	0.000	0.000	0.814	0.164	0.913
CG9259	2L	0.028	0.268	0.105	0.49	0.32	0.000	0.231	0.000	0.000	0.769	0.075	0.867
CG17022	2L	0.251	0.511	0.491	0.18	6.00	3.419	3.700	4.147	4.822	4.226	4.169	3.185
CG9188	2L	0.070	0.351	0.199	0.42	3.10	6.359	3.800	4.537	5.998	4.955	4.394	3.705
CG7778	2L	0.053	0.301	0.177	0.53	6.00	5.283	2.741	4.418	4.507	4.591	4.297	3.425
Alas	2R	0.035	0.562	0.062	0.67	1.70	2.867	3.410	2.836	2.750	2.740	0.880	4.440
CG9897	2R	0.079	0.441	0.179	0.48	4.20	3.759	3.001	3.266	3.488	2.773	1.257	4.274
Act57B	2R	0.033	0.534	0.062	0.85	4.20	3.120	4.886	3.807	3.579	3.088	1.154	3.926
PebIII	2R	0.025	0.444	0.056	0.84	1.70	2.867	3.410	2.805	2.750	2.740	0.880	na

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
CG4716	2R	0.154	0.377	0.408	0.50	2.50	1.026	3.383	2.837	0.406	2.398	2.220	2.630
JhI-26	2R	0.080	0.370	0.217	0.56	1.10	4.556	3.692	3.635	3.205	3.198	2.063	3.142
Mlf	2R	0.010	0.241	0.041	0.62	1.50	4.781	3.692	3.563	3.494	3.118	2.111	na
CG2249	2R	0.000	0.385	0.001	0.64	1.05	2.043	1.293	1.750	1.980	1.338	1.171	2.092
CG8701	2R	0.058	0.469	0.124	0.56	1.05	2.296	1.123	1.176	1.543	0.862	0.394	1.808
CG7163	3L	0.122	0.345	0.354	0.45	3.50	4.451	2.173	3.633	4.328	4.058	4.747	2.968
CG10472	3L	0.024	0.414	0.057	0.73	3.50	2.502	3.141	3.750	3.418	4.501	4.969	3.456
yip7	3L	0.073	0.337	0.216	0.68	3.50	2.502	3.141	3.748	3.418	4.501	4.969	na
kin17	3L	0.002	0.406	0.005	0.71	0.75	0.990	0.053	0.527	0.809	0.172	0.180	na
CG1316	3L	0.005	0.352	0.015	0.58	3.50	3.245	2.973	3.453	4.149	4.528	4.709	3.823
CG7118	3L	0.009	0.293	0.030	0.75	3.50	4.451	2.173	3.635	4.328	4.058	4.747	2.971
CG7542	3L	0.053	0.304	0.175	0.55	0.75	0.000	1.039	0.840	0.000	0.711	0.885	1.053
LysS	3L	0.006	0.593	0.010	0.76	1.70	1.563	1.617	1.799	1.449	3.204	1.525	4.463
fry	3L	0.005	0.188	0.025	0.54	6.20	3.469	3.098	3.380	4.212	3.645	4.348	2.710
CG6467	3L	0.104	0.414	0.250	0.74	3.50	2.502	3.141	3.748	3.418	4.501	4.969	3.461
CG6910	3L	0.006	0.357	0.016	0.60	4.40	2.481	1.991	2.594	3.099	2.781	3.390	2.176
Hn	3L	0.004	0.395	0.010	0.60	3.50	4.453	2.418	3.718	4.194	4.197	4.860	3.103
Pdh	3L	0.008	0.331	0.025	0.53	1.50	1.393	1.449	1.154	0.353	1.184	1.474	1.304
CG10589	3L	0.024	0.378	0.064	0.50	0.10	0.000	0.053	0.557	0.000	0.104	0.074	0.354
CG11892	3R	0.043	0.381	0.112	0.52	6.00	3.493	4.056	3.597	3.086	3.892	3.938	3.358
CG6921	3R	0.012	0.466	0.026	0.64	4.10	3.696	4.299	3.597	3.608	3.735	3.876	2.966
CG8417	3R	0.003	0.450	0.006	0.66	0.73	0.755	0.921	0.533	0.637	0.932	0.806	1.134
CG11876	3R	0.002	0.185	0.011	0.69	4.30	2.558	2.865	2.578	3.411	3.735	3.232	3.907

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
ocn	3R	0.030	0.435	0.068	0.46	2.00	2.033	2.328	2.114	2.200	3.651	2.851	na
CG8043	3R	0.025	0.246	0.101	0.56	0.42	0.272	0.736	0.365	0.297	0.632	0.502	0.979
CG10252	3R	0.002	0.409	0.005	0.46	3.00	4.833	2.804	3.658	3.753	3.837	3.969	3.131
CG5844	3R	0.002	0.369	0.006	0.62	1.40	1.389	1.165	1.141	1.182	1.591	1.495	1.517
CG9288	3R	0.070	0.355	0.198	0.57	2.40	1.562	1.763	1.507	2.020	1.923	1.850	1.705
Ser99Dc	3R	0.019	0.290	0.065	0.55	2.00	2.033	2.511	2.176	2.200	3.651	2.851	4.023
CG7956	3R	0.002	0.262	0.008	0.42	4.10	3.693	4.254	3.478	3.268	3.614	3.744	2.818
Tm2	3R	0.000	0.094	0.001	0.90	1.90	1.809	2.699	1.959	2.551	2.275	2.232	1.926
CG15188	3R	0.000	0.363	0.001	0.69	0.20	0.290	0.057	0.217	0.259	0.311	0.188	0.644
smp-30	3R	0.041	0.276	0.148	0.64	3.80	1.845	2.571	1.790	2.338	2.184	2.133	na
CG7584	3R	0.024	0.524	0.045	0.79	4.00	2.371	2.809	2.300	2.626	3.683	3.001	3.990
CG18111	3R	0.032	0.582	0.054	0.88	4.00	2.371	2.809	2.307	2.626	3.683	3.001	3.988
CG9920	3R	0.005	0.275	0.016	0.67	2.40	1.556	2.439	1.599	1.711	1.994	1.927	1.751
CG12163	3R	0.024	0.376	0.063	0.67	0.10	0.266	0.113	0.271	0.199	0.186	0.076	0.487
CG14683	3R	0.005	0.296	0.018	0.54	0.73	1.214	0.991	0.718	0.875	1.176	1.058	1.267
Mlc2	3R	0.000	0.025	0.000	0.70	2.00	1.795	2.328	2.037	1.998	3.634	2.770	4.059
chp	3R	0.000	0.245	0.001	0.63	0.00	1.603	2.328	1.379	2.878	3.541	2.273	4.208
to	3R	0.022	0.612	0.036	0.53	6.00	3.493	4.056	3.601	3.086	3.892	3.938	na
Spat	X	0.005	0.574	0.008	0.70	4.40	4.212	4.291	3.963	2.629	3.751	3.658	3.568
CG17333	X	0.038	0.389	0.097	0.62	3.60	2.957	3.786	4.270	2.053	3.087	2.402	3.534
CG15347	X	0.035	1.256	0.027	0.69	3.60	3.506	4.502	4.278	3.240	2.998	1.680	3.553
CG2254	X	0.046	0.484	0.095	0.70	3.40	3.405	4.502	4.241	3.400	3.054	1.486	3.556
CG8097	X	0.169	0.331	0.511	0.55	1.50	2.771	3.495	3.309	1.455	3.668	4.330	3.496

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
CG14619	X	0.004	0.249	0.016	0.46	2.10	0.838	0.459	1.246	1.267	4.149	1.341	3.450
Cyp4d2	X	0.023	0.490	0.046	0.54	3.10	1.822	0.365	1.859	1.054	2.018	0.585	3.604
Female-biased													
me31B	2L	0.003	0.034	0.096	0.40	6.00	7.553	4.693	3.924	6.452	3.917	4.028	3.042
Hel25E	2L	0.002	0.207	0.008	0.74	2.90	3.741	4.312	4.391	4.238	4.923	4.392	3.988
SmB	2L	0.005	0.184	0.026	0.54	6.00	5.616	5.298	3.860	5.058	3.634	3.868	3.004
Ip259	2L	0.000	0.222	0.001	0.62	6.00	5.616	8.161	3.866	5.058	3.634	3.868	3.007
eIF-4a	2L	0.006	0.178	0.035	0.77	2.90	3.672	4.282	4.469	5.276	4.968	4.402	3.899
Trip1	2L	0.003	0.164	0.015	0.68	4.70	3.913	4.318	4.242	3.403	4.803	4.358	4.103
Arc-p34	2L	0.002	0.213	0.011	0.54	0.32	0.000	0.231	0.000	0.000	0.871	0.278	1.001
CG12750	2L	0.093	0.381	0.244	0.33	0.90	0.652	0.449	0.400	1.406	1.287	1.089	1.399
ifc	2L	0.000	0.314	0.001	0.67	2.90	3.672	4.075	4.469	5.276	4.968	4.402	3.898
yip2	2L	0.007	0.262	0.026	0.66	6.00	4.423	3.700	4.031	5.652	4.073	4.103	na
CG12292	2L	0.019	0.231	0.083	0.57	1.90	1.909	1.852	2.909	3.059	2.709	3.113	2.546
CG10473	2L	0.010	0.186	0.054	0.64	0.41	0.566	0.546	0.225	0.764	1.165	0.860	1.278
CG16974	2L	0.018	0.199	0.090	0.47	1.30	2.071	1.852	2.682	3.714	2.550	2.943	2.448
CG7224	2L	0.018	0.478	0.038	0.59	6.00	5.071	3.326	4.470	4.069	4.732	4.330	3.496
RpL27A	2L	0.000	0.130	0.001	0.77	4.60	3.794	4.362	4.064	3.198	4.570	4.271	4.207
RpL7	2L	0.002	0.203	0.010	0.81	6.00	7.553	4.693	3.938	6.452	3.917	4.028	3.049
Rack1	2L	0.011	0.275	0.041	0.86	6.00	5.071	3.326	4.491	4.069	4.732	4.330	3.531
RpL13	2L	0.002	0.152	0.015	0.79	6.00	5.414	3.700	4.015	5.609	4.005	4.071	3.096
sop	2L	0.008	0.111	0.073	0.77	6.00	4.423	3.700	4.037	5.652	4.073	4.103	3.110
RpP2	2L	0.008	0.163	0.047	0.82	3.20	1.537	1.143	0.852	2.041	1.626	1.627	5.008

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
RpS27A	2L	0.000	0.056	0.002	0.69	6.00	5.616	8.234	3.865	5.058	3.634	3.868	3.007
RpL40	2L	0.000	0.147	0.001	0.82	5.50	3.135	4.362	3.967	2.941	4.413	4.201	4.254
oho23B	2L	0.000	0.189	0.001	0.84	6.00	3.811	3.110	3.178	5.659	3.746	3.803	4.531
RpS26	2L	0.000	0.435	0.001	0.86	0.90	0.652	0.449	0.401	1.406	1.287	1.089	1.400
RpS13	2L	0.006	0.135	0.041	0.78	6.00	5.283	2.741	4.414	4.507	4.591	4.297	3.420
CG7380	2L	0.014	0.578	0.024	0.64	6.00	5.330	3.326	4.466	4.478	4.681	4.323	3.489
CG7424	2L	0.000	0.036	0.000	0.81	6.00	5.330	3.326	4.464	4.478	4.681	4.323	3.486
porin	2L	0.000	0.040	0.000	0.73	6.00	3.388	5.298	3.704	2.731	3.452	3.749	2.919
CG11015	2L	0.008	0.120	0.063	0.70	3.10	4.724	3.462	4.524	5.414	4.988	4.404	3.787
CG8498	2L	0.014	0.333	0.041	0.63	6.00	5.330	2.741	4.448	4.478	4.681	4.323	3.462
CG7787	2L	0.013	0.439	0.030	0.65	6.00	5.658	2.741	4.423	4.674	4.618	4.305	3.430
CG5317	2L	0.016	0.164	0.097	0.68	3.00	1.691	2.102	3.204	2.305	2.977	3.372	2.676
CG9332	2L	0.043	0.236	0.184	0.41	0.32	0.000	0.231	0.000	0.000	0.814	0.164	0.919
CG17768	2L	0.006	0.251	0.024	0.52	6.00	5.616	8.161	3.868	5.058	3.634	3.868	3.008
Rpn11	2L	0.000	0.134	0.001	0.77	4.30	3.897	4.318	4.270	3.930	4.875	4.379	4.085
CG6249	2L	0.023	0.338	0.067	0.59	6.00	2.098	4.701	3.599	2.220	3.313	3.649	2.865
CG3227	2L	0.044	0.288	0.153	0.61	6.00	4.041	3.249	2.970	6.084	3.616	3.709	4.587
CG13089	2L	0.032	0.266	0.120	0.53	6.00	4.078	2.827	4.387	4.091	4.524	4.276	3.389
CG12253	2L	0.022	0.373	0.059	0.44	6.00	2.500	5.298	3.653	2.842	3.399	3.711	2.892
CG10084	2L	0.015	0.253	0.058	0.58	0.85	0.243	0.398	0.115	0.764	1.025	0.560	1.188
emb	3L	0.024	0.273	0.087	0.45	6.20	3.469	2.956	3.443	4.212	3.645	4.348	2.765
CG10702	2L	0.044	0.264	0.165	0.48	0.41	0.566	0.601	0.219	0.764	1.165	0.860	1.274
HmgD	2R	0.004	0.148	0.025	0.65	4.20	4.674	3.137	3.656	4.126	2.964	1.213	4.058

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
Nap1	2R	0.029	0.273	0.104	0.76	1.70	2.867	3.410	2.866	2.750	2.740	0.880	4.430
RpL29	2R	0.016	0.343	0.048	0.74	4.20	3.596	3.470	3.746	4.023	3.023	1.183	3.986
RpL19	2R	0.002	0.269	0.008	0.78	1.90	0.809	1.475	2.272	0.789	2.965	0.000	4.610
RpL46	2R	0.000	0.021	0.000	0.67	2.40	2.613	2.500	2.787	2.116	2.766	0.645	4.457
CG3751	2R	0.000	0.138	0.001	0.86	4.20	3.531	3.174	3.390	2.793	2.813	1.271	4.215
CG2852	2R	0.009	0.238	0.036	0.80	4.20	3.531	3.174	3.389	2.793	2.813	1.271	4.216
eIF-5A	2R	0.027	0.070	0.379	0.44	2.40	2.613	2.500	2.789	2.116	2.766	0.645	na
tsr	2R	0.000	0.167	0.001	0.57	2.40	2.613	3.108	2.796	2.116	2.766	0.645	4.454
Nxt1	2R	0.060	0.727	0.083	0.63	1.70	2.867	3.410	2.911	2.750	2.740	0.880	na
mago	2R	0.000	0.266	0.001	0.73	4.20	3.331	4.168	3.777	4.155	3.062	1.165	3.958
Fib	2R	0.000	0.231	0.001	0.75	4.20	3.789	3.103	3.232	3.299	2.758	1.239	4.289
CG4046	2R	0.000	0.044	0.000	0.76	4.20	3.777	3.174	3.402	3.662	2.833	1.270	4.209
CG17280	2R	0.015	0.248	0.059	0.85	2.60	3.783	3.638	3.033	3.217	2.734	1.146	4.370
CG3760	2R	0.016	0.157	0.099	0.69	1.90	0.809	1.576	2.291	0.789	2.965	0.000	na
mus209	2R	0.000	0.332	0.001	0.75	4.20	2.811	4.612	3.891	3.077	3.208	1.132	3.811
icln	2R	0.018	0.352	0.052	0.49	4.10	2.992	3.581	3.842	2.433	3.394	1.602	na
CG12391	2R	0.085	0.192	0.443	0.45	1.05	1.685	1.202	2.128	1.226	1.692	1.639	2.275
CG8415	2R	0.000	0.066	0.002	0.71	3.00	2.936	3.013	3.126	2.474	2.687	2.288	2.792
bic	2R	0.008	0.218	0.036	0.75	3.00	1.343	3.626	2.712	0.778	2.293	2.168	2.565
CG16868	2R	0.011	0.184	0.061	0.53	4.20	2.811	4.824	3.889	3.077	3.208	1.132	3.815
Aats-val	2R	0.006	0.189	0.034	0.68	2.50	1.026	3.383	2.819	0.406	2.398	2.220	2.621
CG8309	2R	0.002	0.166	0.011	0.69	3.00	2.936	3.013	3.111	2.474	2.687	2.288	2.783
RpS9	3L	0.000	0.077	0.001	0.84	6.20	3.469	3.098	3.411	4.212	3.645	4.348	2.737

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
CG5064	3L	0.027	0.326	0.082	0.63	6.20	4.428	2.752	3.529	4.659	3.871	4.574	2.849
asf1	3L	0.016	0.400	0.039	0.59	0.75	0.898	0.486	0.540	0.655	0.312	0.371	0.621
CG1240	3L	0.003	0.434	0.007	0.73	3.50	4.250	3.365	2.829	4.444	4.201	3.727	4.141
CG14103	3L	0.028	0.263	0.105	0.50	0.75	1.490	0.486	0.533	1.464	0.250	0.288	0.589
CG10424	3L	0.029	0.289	0.101	0.58	0.75	0.471	0.886	0.581	0.655	0.390	0.474	0.731
RpL23a	3L	0.003	0.139	0.019	0.86	1.70	1.889	2.344	2.138	1.932	3.461	2.067	4.371
CG10960	3L	0.052	0.343	0.150	0.62	2.70	2.052	1.899	2.365	1.456	2.444	2.996	2.040
SsRbeta	3L	0.014	0.210	0.066	0.75	1.50	1.124	1.449	1.221	0.353	1.237	1.539	1.351
RpS12	3L	0.000	0.098	0.001	0.78	2.70	2.112	1.899	2.294	1.946	2.378	2.918	2.000
RpP0	3L	0.002	0.233	0.007	0.78	0.10	0.000	0.027	0.738	0.000	0.071	0.006	0.115
RpS17	3L	0.003	0.073	0.043	0.84	6.20	3.469	3.062	3.430	4.212	3.645	4.348	2.754
CG6846	3L	0.000	0.240	0.001	0.90	0.75	0.357	0.989	0.604	0.655	0.427	0.522	0.772
RpL14	3L	0.002	0.078	0.031	0.76	4.60	4.496	2.752	3.600	4.402	3.932	4.632	2.928
CG7283	3L	0.002	0.162	0.012	0.79	4.40	2.919	1.566	2.720	3.545	2.939	3.571	2.252
NHP2	3L	0.008	0.228	0.036	0.66	1.30	1.229	1.818	1.661	1.534	1.831	2.263	na
CG11342	3L	0.057	0.407	0.140	0.49	3.50	3.245	2.973	3.476	4.149	4.528	4.709	3.805
Hsp67Bc	3L	0.011	0.228	0.048	0.50	6.20	3.469	2.956	3.443	4.212	3.645	4.348	2.766
CG10585	3L	0.000	0.240	0.001	0.56	0.10	0.000	0.053	0.566	0.000	0.104	0.074	0.333
CG8580	3L	0.011	0.328	0.034	0.59	3.50	4.101	2.453	3.751	4.208	4.332	4.948	3.183
CG11367	3L	0.015	0.278	0.055	0.50	0.10	0.000	0.027	0.922	0.000	0.081	0.000	-0.022
CG18543	3L	0.055	0.352	0.155	0.62	3.50	4.451	2.173	3.632	4.328	4.058	4.747	2.967
CG1475	3R	0.000	0.168	0.001	0.83	0.10	0.303	0.057	0.242	0.199	0.232	0.116	0.541
RpS3	3R	0.006	0.145	0.041	0.77	4.10	4.275	2.645	3.649	3.914	3.814	3.951	3.082

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
CG9796	3R	0.013	0.288	0.046	0.67	2.30	1.588	1.144	1.398	1.572	1.838	1.759	1.651
CG4800	3R	0.011	0.412	0.027	0.86	1.40	1.217	0.967	0.833	0.887	1.232	1.117	1.341
cdc2c	3R	0.000	0.412	0.001	0.53	3.20	2.731	3.331	3.353	2.710	3.465	3.574	2.706
CG6182	3R	0.021	0.534	0.039	0.56	1.70	4.310	3.742	3.659	3.296	3.872	3.982	3.197
CG7006	3R	0.002	0.363	0.006	0.70	1.70	3.726	4.056	3.632	3.080	3.891	3.953	3.296
wdn	3R	0.009	0.205	0.043	0.54	4.30	2.587	2.984	2.712	3.229	3.760	3.284	3.864
CSN5	3R	0.000	0.365	0.001	0.62	1.30	1.721	3.440	2.301	2.599	2.590	2.580	2.093
CG7834	3R	0.002	0.393	0.005	0.77	3.00	2.276	2.768	2.239	2.529	3.668	2.891	4.006
CG13603	3R	0.013	0.251	0.051	0.61	2.40	4.696	3.742	3.661	3.483	3.865	3.982	3.172
Pglym78	3R	0.007	0.289	0.025	0.84	4.30	2.559	2.865	2.562	3.411	3.708	3.112	3.912
CG10068	3R	0.026	0.270	0.098	0.52	0.44	0.177	0.020	0.230	0.259	0.429	0.301	0.758
CG8444	3R	0.000	0.320	0.001	0.76	0.73	0.639	0.201	0.449	0.297	0.849	0.721	1.062
CG15697	3R	0.006	0.055	0.116	0.78	3.20	2.731	3.331	3.373	2.710	3.465	3.574	2.723
CG7808	3R	0.006	0.185	0.032	0.78	3.00	2.276	2.511	2.206	2.529	3.668	2.891	4.015
CG10423	3R	0.000	0.092	0.001	0.72	6.00	3.493	4.056	3.601	3.086	3.892	3.938	3.353
CG9354	3R	0.002	0.145	0.017	0.78	0.73	0.639	0.201	0.471	0.297	0.849	0.721	1.082
CG8495	3R	0.000	0.161	0.001	0.74	0.73	0.755	0.921	0.536	0.637	0.932	0.806	1.136
RpL32	3R	0.000	0.107	0.001	0.84	2.00	2.033	2.328	2.110	2.200	3.651	2.851	4.040
CG4759	3R	0.000	0.117	0.001	0.82	7.30	3.053	4.056	3.552	2.708	3.889	3.903	3.416
RpS25	3R	0.000	0.074	0.001	0.84	1.40	1.217	0.967	0.835	0.887	1.232	1.117	1.341
PyK	3R	0.002	0.131	0.018	0.77	4.10	3.627	4.299	3.580	3.612	3.712	3.851	2.940
CG2099	3R	0.025	0.291	0.086	0.80	0.10	0.238	0.113	0.253	0.199	0.208	0.094	0.518
Dph5	3R	0.000	0.154	0.001	0.72	4.10	3.696	4.299	3.600	3.608	3.735	3.876	2.971

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
Hsc70-4	3R	0.027	0.178	0.150	0.72	1.90	1.809	2.668	1.938	2.551	2.275	2.232	1.916
CG18347	3R	0.003	0.211	0.013	0.64	1.40	1.504	1.051	1.020	1.187	1.478	1.375	1.450
RpS20	3R	0.003	0.213	0.015	0.75	3.20	2.731	3.331	3.368	2.710	3.465	3.574	2.719
CG3983	3R	0.000	0.233	0.001	0.58	1.00	1.556	3.795	2.474	2.692	2.730	2.737	2.180
CG2108	3R	0.012	0.190	0.064	0.64	0.10	0.303	0.057	0.239	0.199	0.232	0.116	0.549
CG17327	3R	0.001	0.561	0.001	0.74	1.90	1.459	1.191	1.285	1.572	1.761	1.677	1.593
GstD1	3R	0.016	0.220	0.070	0.79	1.40	1.389	1.051	1.124	1.182	1.591	1.495	1.507
CG7217	3R	0.005	0.465	0.011	0.75	3.50	2.759	2.905	2.775	2.501	2.936	2.968	2.337
AP-2sigma	3R	0.000	0.287	0.001	0.73	4.10	3.594	4.254	3.442	3.043	3.581	3.707	2.783
Vha13	3R	0.003	0.205	0.016	0.79	1.70	2.297	2.656	3.133	2.234	3.277	3.357	2.549
CG10217	3R	0.005	0.301	0.015	0.52	3.00	4.833	3.205	3.659	3.753	3.837	3.969	3.140
CG1866	3R	0.058	0.324	0.178	0.72	4.30	2.698	3.011	2.893	3.429	3.789	3.456	3.798
CG17202	3R	0.074	0.526	0.140	0.71	1.40	1.389	1.051	1.138	1.182	1.591	1.495	1.515
CG5220	3R	0.005	0.269	0.018	0.49	1.00	1.556	3.795	2.465	2.692	2.730	2.737	2.175
CG4338	3R	0.011	0.329	0.033	0.49	1.90	1.809	2.937	1.948	2.551	2.275	2.232	1.921
CG9926	3R	0.138	0.374	0.368	0.62	2.40	1.556	2.409	1.556	1.711	1.994	1.927	1.730
RpL1	3R	0.002	0.140	0.012	0.87	4.30	2.697	3.011	3.041	3.576	3.799	3.498	na
CG12054	3R	0.017	0.164	0.104	0.59	0.00	1.603	2.328	1.366	2.878	3.541	2.273	4.210
dhd	X	0.000	0.014	0.000	0.67	9.50	4.132	4.396	3.589	4.418	4.026	4.723	3.578
CG7033	X	0.021	0.229	0.092	0.76	8.00	4.676	5.306	4.316	3.893	2.892	3.324	3.548
Cyp1	X	0.003	0.219	0.012	0.79	4.20	3.387	4.279	3.081	2.073	3.365	3.667	3.490
Bap60	X	0.007	0.326	0.022	0.59	5.90	4.286	3.440	3.928	4.641	3.728	4.689	3.516
CG8326	X	0.023	0.246	0.095	0.54	1.20	3.198	1.919	2.645	2.584	2.484	2.591	3.480

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
CG11738	X	0.004	0.895	0.004	0.83	4.10	0.976	1.945	1.652	1.267	3.226	1.605	3.459
ran	X	0.000	0.351	0.001	0.71	5.20	3.279	3.878	4.242	2.589	3.156	2.545	na
CG1696	X	0.000	0.204	0.001	0.67	2.10	0.838	0.918	1.326	1.267	4.149	1.341	3.452
mxc	X	0.050	0.297	0.168	0.55	8.00	4.491	4.997	4.321	3.780	2.878	3.638	na
Scamp	X	0.006	0.397	0.015	0.67	1.50	2.500	3.495	3.350	2.113	3.749	4.560	na
CG9066	X	0.028	0.282	0.098	0.57	1.50	3.008	3.398	3.218	2.577	3.578	4.108	3.494
CG4094	X	0.001	0.529	0.001	0.63	5.00	3.726	4.286	3.999	2.787	3.598	2.997	3.567
S6kII	X	0.010	0.315	0.032	0.45	2.10	0.769	0.459	1.222	1.267	4.536	1.014	3.449
sqh	X	0.003	0.238	0.011	0.78	4.40	5.090	4.197	3.882	3.513	3.865	4.225	3.571
RpS6	X	0.003	0.169	0.015	0.79	3.80	3.236	2.482	4.219	2.822	3.167	1.460	3.557
RpL7A	X	0.000	0.244	0.001	0.81	4.40	4.212	4.291	3.963	2.629	3.751	3.658	3.568
RpL36	X	0.003	0.361	0.009	0.83	0.53	0.000	0.312	0.594	0.000	0.000	0.000	3.616
CG9091	X	0.000	0.148	0.001	0.83	1.50	2.635	3.764	3.440	3.039	3.798	4.718	3.500
CG4111	X	0.010	0.368	0.027	0.74	5.10	4.243	4.396	3.715	4.620	3.999	4.707	3.575
CG14206	X	0.006	0.112	0.057	0.77	3.10	2.611	2.437	1.946	2.354	2.240	2.500	3.464
sta	X	0.000	0.195	0.001	0.81	0.80	0.723	0.546	1.330	0.648	1.085	0.154	3.609
RpS19	X	0.003	0.111	0.024	0.74	3.40	4.125	2.725	2.979	2.974	3.158	3.318	3.488
Gip	X	0.007	0.434	0.017	0.58	2.90	3.427	2.634	4.305	3.290	2.937	2.813	3.538
CG9915	X	0.031	0.231	0.134	0.50	4.20	3.387	4.279	3.083	2.073	3.365	3.667	3.491
CG11752	X	0.027	0.275	0.099	0.63	5.20	3.866	4.202	4.220	3.097	3.293	3.000	3.530
CG18624	X	0.017	0.316	0.053	0.61	3.80	3.236	2.578	4.218	2.822	3.167	1.460	na
ctp	X	0.000	0.114	0.001	0.56	8.00	3.776	4.645	3.343	5.089	3.987	4.369	3.583
Gtp-bp	X	0.000	0.430	0.001	0.74	5.20	3.279	3.878	4.244	2.589	3.156	2.545	na

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
CG2998	X	0.000	0.098	0.001	0.71	7.40	4.340	4.052	4.325	3.780	2.876	3.855	3.544
Nonsex-biased													
CG6746	2L	0.004	0.301	0.014	0.67	3.00	1.691	2.102	3.229	2.305	2.977	3.372	2.687
SamDC	2L	0.019	0.388	0.049	0.40	6.00	6.493	8.213	3.890	4.864	3.733	3.927	3.021
Rapgap1	2L	0.003	0.195	0.014	0.52	3.00	5.694	3.326	4.518	4.805	4.846	4.366	3.591
Hs2st	2L	0.005	0.234	0.021	0.58	0.85	0.243	0.398	0.127	0.764	1.025	0.560	na
CG9324	2L	0.024	0.228	0.106	0.81	0.32	0.000	0.231	0.000	0.000	0.814	0.164	0.928
CG8869	2L	0.028	0.404	0.069	0.79	4.70	3.913	4.318	4.238	3.403	4.803	4.358	4.106
CG17347	2L	0.027	0.418	0.065	0.60	0.43	0.524	0.362	0.202	0.764	1.120	0.774	1.261
CG6724	2L	0.007	0.376	0.018	0.54	6.00	3.742	5.298	3.734	2.731	3.491	3.775	2.935
CG6922	2L	0.025	0.446	0.057	0.54	2.90	3.741	4.312	4.388	4.238	4.923	4.392	3.991
CG6583	2L	0.007	0.268	0.025	0.61	3.00	1.804	2.151	3.153	2.305	2.925	3.324	2.653
CG3609	2L	0.012	0.188	0.065	0.75	4.10	3.453	2.916	2.807	5.612	3.323	3.476	4.628
CG11024	2L	0.016	0.290	0.056	0.71	2.90	3.741	4.312	4.386	4.238	4.923	4.392	3.992
CG18661	2L	0.011	0.545	0.020	0.57	6.00	3.344	3.700	4.285	3.551	4.409	4.238	3.292
CG9306	2L	0.013	0.287	0.044	0.72	1.30	2.029	1.852	2.594	2.834	2.451	2.830	2.411
Pka-C1	2L	0.000	0.138	0.001	0.66	6.00	3.658	3.700	4.098	5.243	4.153	4.138	3.151
Ser4	2L	0.028	0.166	0.170	0.78	4.70	3.913	4.318	4.238	3.403	4.803	4.358	4.106
Traf1	2L	0.004	0.253	0.017	0.69	4.60	3.365	4.362	4.030	2.455	4.501	4.241	4.224
CG5885	2L	0.000	0.280	0.001	0.78	6.00	4.423	3.700	4.031	5.652	4.073	4.103	3.106
CG11455	2L	0.000	0.284	0.001	0.66	3.00	1.537	0.000	0.511	1.767	1.129	0.932	5.061
Os9	2L	0.069	0.379	0.182	0.55	0.89	0.000	0.462	0.020	0.000	0.943	0.422	na
ken	2R	0.006	0.303	0.020	0.69	1.70	2.867	3.410	2.881	2.750	2.740	0.880	4.424

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
CG10306	2R	0.009	0.231	0.037	0.79	4.20	4.674	3.137	3.663	4.126	2.964	1.213	4.053
Fkbp13	2R	0.008	0.348	0.022	0.77	4.20	4.050	3.470	3.706	4.023	2.999	1.195	4.020
CG9344	2R	0.000	0.285	0.001	0.75	4.20	3.120	4.886	3.802	3.579	3.088	1.154	3.932
HmgZ	2R	0.000	0.081	0.001	0.73	4.20	4.674	3.137	3.660	4.126	2.964	1.213	4.055
CG9849	2R	0.016	0.327	0.048	0.53	4.20	3.759	3.001	3.284	3.488	2.773	1.257	4.266
CG3683	2R	0.017	0.241	0.069	0.66	1.90	0.809	1.860	2.454	1.074	2.871	0.000	4.559
CG10320	2R	0.008	0.096	0.086	0.71	4.20	4.674	3.137	3.669	4.126	2.964	1.213	4.048
CG3907	2R	0.023	0.158	0.145	0.43	1.70	2.867	3.410	2.808	2.750	2.740	0.880	4.450
Mlp60A	2R	0.000	0.016	0.000	0.70	2.40	2.613	2.500	2.780	2.116	2.766	0.645	4.459
ProsMA5	2R	0.005	0.297	0.016	0.63	3.10	3.220	3.581	3.820	2.920	3.377	1.688	3.333
CG5770	2R	0.059	0.364	0.162	0.62	4.20	2.570	3.035	3.889	1.825	3.405	1.437	3.440
wal	2R	0.004	0.227	0.018	0.64	3.20	2.563	2.362	2.317	1.966	1.895	1.858	2.367
CG7461	2R	0.004	0.328	0.012	0.52	4.20	2.879	4.342	3.938	3.011	3.332	1.194	3.652
CG8707	2R	0.006	0.238	0.024	0.66	1.05	2.296	1.571	1.108	1.543	0.862	0.394	1.773
CG12384	2R	0.011	0.167	0.067	0.56	1.05	2.045	1.500	2.201	1.869	1.811	1.773	2.310
Sod2	2R	0.009	0.238	0.039	0.78	1.10	3.759	3.692	3.724	3.547	3.304	1.893	3.224
Syb	2R	0.000	0.280	0.001	0.73	1.05	1.927	1.025	1.821	1.334	1.421	1.291	2.127
LamC	2R	0.000	0.217	0.001	0.65	1.90	4.112	4.039	3.225	3.439	2.816	2.276	2.851
CG4692	2R	0.000	0.189	0.001	0.77	1.90	0.809	1.860	2.492	1.074	2.871	0.000	4.548
CG5134	2R	0.003	0.354	0.009	0.65	4.20	2.570	3.088	3.899	1.825	3.405	1.437	3.461
Iap2	2R	0.011	0.494	0.021	0.61	1.50	4.781	3.692	3.562	3.494	3.118	2.111	3.082
CG8586	2R	0.099	0.459	0.215	0.53	1.05	1.999	0.813	1.284	1.722	0.941	0.532	1.863
mbf1	3L	0.003	0.224	0.013	0.61	1.50	1.461	1.172	1.090	1.148	1.057	1.317	1.258

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
CG7375	3L	0.007	0.122	0.057	0.70	3.50	4.524	2.173	3.665	4.983	4.131	4.808	3.012
Tom	3L	0.000	0.107	0.001	0.70	1.60	1.144	1.818	1.602	1.534	1.772	2.192	1.595
CG6084	3L	0.005	0.168	0.029	0.63	3.40	2.965	1.566	2.788	4.167	3.006	3.648	2.294
CG18180	3L	0.031	0.404	0.076	0.73	6.20	3.469	3.098	3.379	4.212	3.645	4.348	2.709
Int6	3L	0.000	0.418	0.001	0.80	1.10	1.047	1.039	1.031	0.794	0.976	1.216	1.213
CG15877	3L	0.059	0.256	0.229	0.52	4.20	3.601	3.365	2.608	3.834	4.038	3.339	4.223
CG11357	3L	0.043	0.125	0.348	0.57	3.50	3.245	2.973	3.526	4.149	4.528	4.709	3.767
CG15012	3L	0.018	0.237	0.076	0.68	3.50	3.245	2.973	3.455	4.149	4.528	4.709	3.821
LanB2	3L	0.019	0.278	0.067	0.69	6.20	3.469	3.098	3.384	4.212	3.645	4.348	2.714
CG8583	3L	0.015	0.241	0.060	0.63	3.50	4.101	2.453	3.751	4.208	4.332	4.948	3.185
CG13298	3L	0.000	0.256	0.001	0.64	3.50	2.502	3.077	3.759	3.418	4.501	4.969	3.429
CG13043	3L	0.004	0.144	0.026	0.62	1.50	1.393	1.449	1.168	0.353	1.184	1.474	1.314
CG18081	3L	0.005	0.412	0.013	0.70	1.60	1.340	1.501	1.313	1.084	1.401	1.740	1.413
CG7770	3L	0.044	0.296	0.148	0.75	0.75	1.490	0.064	0.529	1.464	0.250	0.288	0.564
CG7170	3L	0.021	0.414	0.050	0.72	3.50	4.451	2.173	3.635	4.328	4.058	4.747	2.971
CG12091	3L	0.002	0.271	0.007	0.66	1.70	1.889	2.344	2.040	1.932	3.461	2.067	4.399
CG5582	3L	0.016	0.279	0.055	0.50	0.75	0.000	1.039	0.784	0.000	0.585	0.725	0.998
CG18178	3L	0.054	0.471	0.115	0.59	6.20	3.469	3.098	3.359	4.212	3.645	4.348	2.693
CG3344	3L	0.025	0.245	0.102	0.61	0.80	0.000	0.847	1.329	0.316	2.797	0.683	4.577
fln	3L	0.003	0.164	0.016	0.42	0.75	1.556	0.064	0.527	1.464	0.231	0.263	0.550
CG7706	3R	0.051	0.278	0.182	0.48	3.30	3.618	2.820	2.907	2.975	3.072	3.123	2.411
CG5902	3R	0.006	0.284	0.020	0.73	2.40	4.696	3.742	3.661	3.483	3.865	3.982	3.172
CG6171	3R	0.161	0.249	0.646	0.36	1.90	1.810	0.561	1.974	2.551	2.331	2.294	na

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
Rpn7	3R	0.005	0.475	0.010	0.68	4.10	3.696	4.299	3.589	3.608	3.735	3.876	2.953
CG14639	3R	0.037	0.476	0.077	0.65	0.10	0.000	0.056	0.402	0.053	0.118	0.021	0.343
CG2943	3R	0.007	0.249	0.028	0.49	0.38	0.071	0.063	0.280	0.259	0.525	0.395	0.868
CG8031	3R	0.014	0.189	0.073	0.54	1.90	1.459	1.191	1.288	1.572	1.761	1.677	1.595
CG7048	3R	0.032	0.255	0.126	0.69	4.10	3.696	4.299	3.591	3.608	3.735	3.876	2.957
CG2046	3R	0.061	0.310	0.196	0.56	0.10	0.334	0.057	0.233	0.458	0.256	0.138	0.565
sds22	3R	0.027	0.391	0.068	0.64	3.50	1.709	2.567	2.564	3.103	2.787	2.801	na
CG9795	3R	0.088	0.353	0.248	0.39	0.10	0.000	0.056	0.389	0.053	0.118	0.021	0.353
CG6666	3R	0.003	0.271	0.011	0.68	1.40	1.217	0.967	0.833	0.887	1.232	1.117	1.341
CG1939	3R	0.012	0.478	0.025	0.70	0.44	0.177	0.031	0.229	0.259	0.429	0.301	0.751
CG6364	3R	0.000	0.311	0.001	0.65	1.70	4.131	3.742	3.657	2.769	3.881	3.980	3.215
Rab1	3R	0.000	0.160	0.001	0.49	4.10	3.594	4.254	3.441	3.043	3.581	3.707	2.783
CG11069	3R	0.000	0.274	0.001	0.48	1.70	3.726	4.056	3.629	3.080	3.891	3.953	3.303
CG15189	3R	0.035	0.198	0.178	0.73	0.20	0.290	0.057	0.217	0.259	0.311	0.188	0.643
CG10219	3R	0.018	0.242	0.075	0.68	3.00	4.833	3.205	3.659	3.753	3.837	3.969	3.140
CG9645	3R	0.072	0.270	0.266	0.36	3.30	1.735	2.439	1.635	2.227	2.038	1.974	1.769
CG14558	3R	0.009	0.365	0.024	0.73	2.60	2.925	2.663	3.476	2.558	3.874	3.817	3.488
CG9471	3R	0.013	0.365	0.036	0.69	0.73	0.755	0.921	0.535	0.637	0.932	0.806	1.135
CG13618	3R	0.009	0.115	0.078	0.61	1.70	4.056	4.204	3.649	3.179	3.890	3.961	3.249
mira	3R	0.013	0.312	0.042	0.77	1.90	2.187	2.501	3.196	2.638	3.320	3.407	2.591
CG7718	3R	0.014	0.311	0.046	0.56	3.30	3.563	2.884	2.923	2.975	3.101	3.156	2.421
CG4413	3R	0.029	0.424	0.068	0.65	3.20	2.150	2.501	3.216	1.967	3.349	3.440	2.605
Mlc1	3R	0.008	0.068	0.115	0.77	4.30	2.622	3.011	3.126	3.374	3.811	3.548	3.700

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
Tpi	3R	0.005	0.295	0.016	0.88	2.00	1.795	2.328	2.060	1.998	3.634	2.770	4.053
CG6503	3R	0.019	0.324	0.058	0.54	2.60	2.995	2.814	3.433	2.606	3.868	3.787	3.522
Cys	3R	0.084	0.577	0.145	0.53	3.80	1.721	2.483	1.738	2.338	2.112	2.055	1.819
eIF-1A	3R	0.000	0.123	0.001	0.74	3.50	2.974	2.905	2.804	3.172	2.974	3.012	na
CG1746	3R	0.007	0.023	0.305	0.61	0.00	1.603	2.328	1.371	2.878	3.541	2.273	4.209
tacc	3R	0.090	0.304	0.295	0.39	0.10	0.299	0.112	0.327	0.252	0.148	0.044	0.415
CG6398	X	0.000	0.169	0.001	0.66	1.20	2.682	1.919	2.573	1.939	2.299	2.487	3.478
CG5703	X	0.000	0.274	0.001	0.71	1.20	3.198	1.919	2.644	2.584	2.484	2.591	3.480
CG1532	X	0.007	0.366	0.020	0.51	3.50	0.901	0.918	1.425	1.267	3.815	1.605	3.454
CG3040	X	0.017	0.283	0.060	0.64	4.70	3.431	4.286	4.064	2.494	3.476	2.447	3.565
CG4645	X	0.024	0.338	0.071	0.57	5.90	4.312	3.440	3.979	5.021	3.677	4.512	3.518
CG2471	X	0.011	0.411	0.027	0.44	3.50	4.057	4.385	4.148	4.065	3.497	3.813	3.525
CG10992	X	0.021	0.337	0.062	0.62	5.60	4.799	4.421	3.791	5.025	3.847	5.023	3.511
CG15027	X	0.017	0.371	0.046	0.63	11.01	2.883	3.764	3.455	3.448	3.829	4.833	3.500
crl	X	0.008	0.152	0.055	0.55	3.40	4.125	2.725	2.993	2.974	3.158	3.318	na
fh	X	0.071	0.273	0.259	0.61	8.00	4.676	5.306	4.318	3.893	2.892	3.324	3.547
C3G	X	0.000	0.189	0.001	0.64	5.00	3.726	4.286	4.024	2.787	3.598	2.997	3.566
CG12203	X	0.010	0.362	0.028	0.72	3.70	3.013	2.903	1.984	3.105	2.073	2.528	3.465
CG5548	X	0.006	0.189	0.034	0.86	11.01	2.883	3.764	3.461	3.448	3.829	4.833	3.501
CG13356	X	0.038	0.320	0.119	0.58	2.10	0.656	0.000	1.073	1.267	4.862	0.722	3.446
CG3446	X	0.003	0.253	0.011	0.72	4.00	4.791	4.223	3.916	3.466	3.829	4.068	3.570
CG4666	X	0.017	0.453	0.038	0.71	4.40	5.090	4.347	3.840	3.513	3.865	4.225	3.572
CG15043	X	0.091	0.316	0.287	0.68	4.50	2.418	2.838	2.343	2.161	1.954	2.434	3.473

(Continues...)

Table S5. *(Cont.)*

Gene	Loc	d_N	d_S	d_N/d_S	F_{op}	ACE	RTE	HKw	HKp	CC99	KH93	CK00	MMD01
ade5	X	0.009	0.459	0.019	0.72	5.90	4.207	3.440	4.008	4.640	3.656	4.436	na
Ag5r	X	0.036	0.299	0.121	0.53	9.50	3.073	4.092	3.539	3.584	3.851	4.934	na

Table S6. List of sequenced genes.

Gene	CGnum	Loc.	M/F	Sites	Syn	Non	n	D_S	P_S	D_N	P_N	P-value	TDs	TDn
Male-biased														
janB	CG7931	3R	9.20	607	97	322	12	12	7	6	6	0.470	-1.30	-1.63
ocn	CG7929	3R	6.64	515	71	265	12	5	7	2	2	0.772	-0.51	-1.25
T4	CG10252	3R	35.16	880	173	514	12	19	3	0	0	na	-0.11	na
T5	CG14926	2L	25.19	685	116	406	10	15	4	7	3	0.600	1.26	-0.50
T7	CG17956	3R	5.03	192	44	124	12	4	0	1	1	0.333	na	-1.14
T9	CG17376	2L	12.50	594	51	150	11	1	3	0	1	1.000	-0.36	-1.13
T10	CG1980	3R	6.95	853	132	603	11	28	2	12	1	0.904	-1.41	-1.11
T21	CG5565	2L	20.83	796	168	552	9	11	15	13	5	0.047*	-0.07	-1.30
T22	CG18418	3L	16.06	942	222	711	12	27	33	13	5	0.040*	-0.01	0.25
T24	CG6980	3R	8.08	860	149	580	12	16	5	13	4	0.984	0.43	0.18
T25	CG6255	3R	17.96	1026	254	772	12	28	25	10	1	0.011*	0.57	1.83
T26	CG6332	3R	23.88	1267	256	802	7	34	24	4	0	0.151	-0.46	na
T27	CG3483	2R	10.77	1185	276	897	11	33	5	6	4	0.070	-1.15	-1.32
T28	CG10307	2R	7.94	1145	243	780	12	24	27	9	7	0.521	0.45	-0.66
T29	CG10750	2L	12.73	1176	222	771	9	21	20	10	0	0.004*	-0.33	na
T30	CG3085	2R	19.81	1417	298	998	11	25	41	5	1	0.028*	-0.45	-0.07
T31	CG8564	3L	35.21	1585	371	1141	12	30	23	27	7	0.026*	-0.45	-0.69
T32	CG11475	2R	14.59	1392	316	1076	11	31	39	22	4	0.000*	0.06	-1.04
T33	CG7387	3L	19.63	1451	326	1021	9	45	12	18	6	0.772	-0.69	-1.07
T34	CG18266	2L	13.68	1461	323	1135	9	35	11	55	9	0.189	-0.83	-0.69
T35	CG9314	2L	24.06	1608	351	1167	9	33	36	2	1	0.520	0.05	-1.13
T36	CG6971	3R	9.72	806	170	577	12	26	7	0	0	na	-1.33	na
T37	CG7251	2L	11.14	1026	213	717	10	26	10	20	9	0.774	-0.36	-1.10
T38	CG9531	2L	3.48	1087	230	739	11	14	15	6	6	0.920	-0.97	-1.29
T39	CG6130	3R	8.22	768	162	606	10	18	6	23	2	0.101	-0.88	-1.41
T40	CG5045	2L	19.81	876	187	572	11	17	8	2	1	0.963	-0.39	-1.13
T41	CG5276	3R	5.67	1314	301	956	12	46	5	16	5	0.133	-0.62	-1.28
T50	CG15179	3R	14.09	776	146	514	12	10	2	6	1	0.890	1.82	0.57
T51	CG8277	3L	19.91	802	140	556	9	14	5	24	3	0.180	0.12	-0.38
T52	CG14717	3R	16.29	921	228	690	12	23	11	7	6	0.380	-0.18	-1.43
T53	CG11037	3L	18.48	955	217	659	11	24	3	5	0	1.000	-1.60	na
T54	CG6036	3R	13.66	1194	251	862	10	26	21	11	5	0.341	0.32	-1.70
T55	CG13527	2R	15.87	1046	210	660	9	23	12	15	6	0.656	0.57	0.46
Female-biased														
O9	CG2867	3R	0.30	1638	402	1232	11	46	5	5	0	1.000	0.08	na
O22	CG9383	3L	0.28	657	148	506	11	14	1	2	0	1.000	-0.09	na
O23	CG9273	2L	0.22	946	175	563	12	18	9	2	6	0.035§	-0.19	-1.92
O24	CG4973	3R	0.32	1196	233	835	11	41	6	9	12	0.000§	-0.66	-1.30
O25	CG1239	3R	0.34	1134	204	696	12	23	5	18	2	0.439	-0.99	-0.03

(Continues...)

Table S6. *(Cont.)*

Gene	CGnum	Loc.	M/F	Sites	Syn	Non	n	D_S	P_S	D_N	P_N	P-value	TDs	TDn
O26	CG6554	3R	0.24	1395	250	878	11	29	2	3	1	0.285	-0.77	-1.12
O27	CG7840	2L	0.18	978	230	748	11	18	17	12	4	0.106	-0.04	0.42
O28	CG9135	2L	0.16	1586	345	1115	10	28	19	3	0	0.279	-0.78	na
O29	CG5363	2L	0.21	1357	203	688	11	22	7	13	0	0.079	0.70	na
O31	CG3831	2R	0.32	1698	324	975	9	27	13	3	2	0.741	0.49	-1.32
O32	CG12276	3R	0.27	1153	250	761	12	41	3	8	5	0.008§	0.02	-0.91
O33	CG17950	2R	0.09	420	68	262	11	4	1	0	0	na	1.15	na
O34	CG5499	3R	0.24	1612	110	313	8	7	3	0	3	0.070	0.20	-1.45
O35	CG10206	2L	0.18	1714	354	1176	8	19	18	1	2	0.545	-0.30	-1.31
O37	CG4299	3R	0.24	1377	167	640	10	20	4	3	0	1.000	0.02	na
O38	CG4236	3R	0.32	1475	287	1003	12	26	6	0	0	na	0.09	na
O41	CG5272	3L	0.08	781	184	536	12	17	4	14	3	0.911	-0.90	-0.42
O42	CG5757	2R	0.26	750	148	485	12	17	8	11	2	0.255	-0.78	0.34
O43	CG12314	2L	0.28	759	177	582	11	26	1	30	3	0.392	1.44	1.44
O44	CG4570	3R	0.10	825	184	641	11	19	3	11	0	0.534	-0.50	na
O45	CG12909	2R	0.26	846	174	669	12	29	7	14	5	0.560	-0.64	-1.82
O46	CG6874	3L	0.18	663	159	501	12	17	4	22	0	0.048*	0.11	na
O47	CG13690	2L	0.18	1044	234	807	12	19	4	15	5	0.540	1.43	0.32
O48	CG3509	3R	0.13	1141	234	810	10	26	13	35	6	0.048*	0.36	-0.24
O49	CG17361	3L	0.32	546	117	420	11	17	8	27	6	0.225	-0.50	-0.45
O51	CG32409	3L	0.26	781	149	514	11	18	25	2	3	0.936	0.29	0.10
O52	CG6459	2R	0.18	969	179	610	11	21	7	5	3	0.465	0.0001	-0.0122
O53	CG3975	2L	0.25	1299	292	1001	11	22	30	31	17	0.029*	-0.0002	0.0054
Nonsex-biased														
U1	CG11981	3R	1.03	615	142	473	8	15	1	0	0	na	0.33	na
U2	CG5919	3R	0.98	824	179	589	11	21	8	7	1	0.353	-0.94	-1.12
U3	CG5915	3R	1.00	747	140	481	11	9	3	1	0	1.000	1.19	na
U4	CG9893	2R	0.95	576	139	437	10	7	17	3	4	0.502	0.51	-1.14
U5	CG13189	2R	1.03	1023	258	765	12	26	22	3	0	0.249	-0.22	na
U6	CG10853	3L	0.99	668	128	328	12	6	7	2	4	0.600	0.00	-0.78
U7	CG6913	3R	1.00	651	145	440	12	20	2	0	0	na	0.15	na
U8	CG9437	2R	0.95	942	223	716	11	11	22	6	10	0.770	-1.20	0.72
U9	CG8392	2R	0.95	795	168	504	11	11	9	2	0	0.490	0.12	na
U10	CG7953	2L	1.07	988	210	681	11	20	7	12	2	0.380	1.00	-1.43
U11	CG13419	3R	0.98	653	125	394	12	8	8	0	0	na	0.76	na
U12	CG7484	3L	0.95	935	122	412	12	20	3	3	3	0.064	-0.38	-0.58
U13	CG9283	3L	0.98	627	122	490	12	16	4	10	7	0.159	-0.26	-0.94
U14	CG13934	3L	0.92	779	106	323	10	6	14	2	12	0.277	-0.03	0.36
U15	CG17404	3R	0.93	944	226	659	11	32	3	7	3	0.104	-1.13	-1.61
U16	CG10623	2L	0.95	1139	241	752	12	21	13	11	5	0.629	-0.03	-0.72
U17	CG3652	2L	0.98	914	174	498	10	25	9	4	1	0.752	-1.51	-1.12

(Continues...)

Table S6. *(Cont.)*

Gene	CGnum	Loc.	M/F	Sites	Syn	Non	n	D_S	P_S	D_N	P_N	P-value	TDs	TDn
U18	CG9617	3R	0.98	537	122	409	12	11	4	6	0	0.281	-1.39	na
U19	CG6981	3L	1.03	486	118	368	12	13	4	0	0	na	0.10	na
U20	CG9822	2R	1.02	920	181	608	12	9	23	5	6	0.298	0.51	-0.29
U21	CG13845	3R	1.02	894	151	548	11	16	11	4	0	0.269	0.15	na
U22	CG3683	2R	1.04	822	121	405	12	11	2	0	0	na	-0.85	na
U23	CG3476	2L	0.98	1094	235	662	11	17	16	0	6	0.027§	-0.99	-1.56
U24	CG6094	2L	1.00	673	152	457	12	14	22	4	4	0.566	-0.15	-1.75
U25	CG8844	2L	1.04	942	101	376	12	5	9	0	1	1.000	0.38	1.38
U26	CG7508	3R	1.11	939	219	711	12	23	4	9	3	0.454	-0.37	-0.42
U27	CG16985	3L	1.10	776	104	343	11	12	8	2	0	0.515	0.66	na
U28	CG11785	3R	1.07	880	137	481	11	3	6	2	3	0.800	0.12	-0.39
U29	CG10035	3R	1.06	930	233	625	12	25	5	9	4	0.310	-0.99	0.50
U30	CG18553	3R	1.06	903	164	487	11	3	1	7	4	0.670	1.18	-0.94

NOTE:
1. Gene = Symbol of genes defined in my thesis
2. CGnum= CG number (*D. melanogaster* genome release 4.0)
3. Loc = Location of gene (chromosome arm)
4. M/F = Average male/female expression ratio
5. Sites = Total number of sites sequenced, including introns
6. Syn = Number of synonymous sites
7. Non = Number of nonsynonymous sites
8. n = Sample size (number of *D. melanogaster* alleles sequenced)
9. Ds = number of synonymous fixed differences between species
10. Ps = number of synonymous polymorphisms within *D. melanogaster*
11. Dn = number of nonsynonymous fixed differences between species
12. Pn = number of nonsynonymous polymorphisms within *D. melanogaster*
13. P = P-value from MK test.
14. TDs = Tajima's *D* for synonymous sites
15. TDn = Tajima's *D* for nonsynonymous sites

*= sig. MK test (positive selection)
§ = sig. MK test (balancing/purifying selection,)

Table S7. Nucleotide polymorphism of genes with significant McDonald-Kreitman tests. Grey color represents intron regions. Bold in box represents nonsynonymous divergence compared with *D. simulans*. Bold with underline represents nonsynonymous polymorphism within *Zimbabwe* population.

T21 CG5565

	4	7	9	3	7	7	8	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	4
				6	3	7	6	1	0	2	2	2	2	2	2	2	3	3	3	3	4	5	5	6	8	8	1	2	2	3	4	8	9	2	4	4	6	7	7	3
									9	1	2	3	4	5	6	7	1	2	3	4	1	4	7	3	1	2	3	4	6	9	8	7	6	1	6	9	1	0	6	3
ZB82	C	A	C	T	C	C	A	G	C	C	-	-	-	T	T	C	T	G	C	G	C	C	A	T	A	T	T	C	G	C	A	T	**C**	C	T	G	A	G	G	G
ZB84	**T**	.	.	.	.	A	.	.	.	.	-	-	T	.	.	T	.	T	T	.	.	.	C	C	.	.	.	.	.	.	.	.	G	.	.	.	.	A	.	A
ZB95	.	.	.	.	.	A	.	C	.	.	-	-	T	.	.	T	.	T	T	.	.	.	.	.	G	.	.	.	.	.	.	**A**	G	.	.	A	.	.	.	A
ZB145	.	.	.	.	.	.	.	.	.	.	-	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**C**	.	.	.	.	.	.	.
ZB157	.	.	T	.	.	A	.	C	.	.	T	T	T	.	.	T	.	T	T	.	.	.	C	C	.	.	.	.	.	.	.	.	G	**T**	.	.	.	.	A	A
ZB186	.	.	.	.	.	A	.	C	.	.	-	-	T	.	.	T	.	T	T	.	.	.	C	C	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
ZB191	.	.	.	.	.	.	.	C	.	.	-	T	T	.	.	T	.	T	T	.	.	.	C	C	G	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
ZB229	.	.	.	.	.	A	.	C	.	.	-	-	T	.	.	T	.	T	T	.	.	.	.	.	G	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
ZB384	.	.	.	.	.	A	.	C	.	.	T	T	T	.	.	T	.	T	T	.	.	.	C	.	G	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
D. sim	.	**T**	.	C	T	.	G	C	-	-	-	-	-	-	-	T	C	T	T	C	T	T	C	C	G	C	**C**	**A**	**C**	**G**	**G**	.	G	.	**A**	.	G	.	.	A

	4	4	4	4	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7	7	7
	3	7	8	8	1	1	3	4	0	0	2	4	4	4	7	7	8	2	4	6	7	9
	4	2	1	7	3	7	9	4	1	7	8	0	3	9	3	4	0	1	2	0	3	0
ZB82	A	A	C	T	C	T	T	A	T	T	C	A	G	G	A	A	A	C	**C**	A	A	A
ZB84	.	.	T	C	.	.	.	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.
ZB95	.	.	.	C	.	G	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	G
ZB145	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
ZB157	.	.	.	C	.	G	.	.	.	.	T	G	.	T	.	.	.	.	G	.	.	.
ZB186	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	G
ZB191	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
ZB229	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
ZB384	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.
D. sim	**C**	C	.	C	**A**	G	**A**	G	C	A	.	.	A	.	T	**C**	**C**	.	G	C	**G**	G

(Continues...)

Table S7. *(Cont.)*

T22 CG18418

	9	21	36	42	51	67	87	96	111	144	150	163	168	180	187	191	204	207	210	228	240	253	255	270	276	291	330	342	348	354	360	379	396	408	411	432	435	492	507	529
Zim82	G	T	A	C	A	A	A	C	T	A	A	A	A	A	G	T	G	A	A	G	A	A	G	G	T	G	T	C	G	C	G	A	G	G	C	C	A	A	G	T
Zim84	.	.	.	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.	.	.	G	.	.
Zim95	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.
Zim131	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	A	.	.	A	.	.	C	.	.	A	.	.	.	A	T	.	.	.	.	.	G	.	.
Zim145	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
Zim157	.	.	.	.	G	.	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
Zim186	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	A	.	.	A	.	.	C	.	.	A	.	.	.	A	T	.	.	.	.	.	G	.	.
Zim191	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	T	T	.	.	G	.	.
Zim229	.	.	.	.	G	**G**	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.
Zim377	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim384	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	T	T	.	.	G	.	.
Zim398	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
D. sim	A	G	G	T	G	G	C	.	C	G	C	**C**	C	T	**A**	**G**	.	**G**	G	.	G	**T**	.	.	C	.	C	.	C	.	.	**T**	A	.	.	.	G	G	C	C

	534	573	582	885	595	606	618	624	657	660	708	711	722	723	724	749	753	773	783	786	792	798	816	831	843	861	766	782	822	892	901	907	913
Zim82	G	T	T	A	A	T	T	A	T	T	C	A	A	G	G	G	T	A	C	A	T	C	T	C	G	T	A	T	C	T	T	A	A	A	T	T
Zim84	.	.	.	.	.	**A**	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim95	.	.	.	.	.	**A**	.	T	C	.	A	C	**G**	.	.	**A**	C	.	.	.	.	.	.	.	.	.	.	C	T	C	.	.	.	**G**	.	.
Zim131	.	.	.	.	.	**A**	.	C	C	.	A	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	T	C	.	.	.	.	.	.
Zim145	.	.	C	.	.	.	.	T	C	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	C	.	C	.	.	.	**G**	.	.
Zim157	.	.	C	.	.	.	.	T	C	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	G	C	T	C	.	.	.	.	.	.
Zim186	.	.	.	.	.	**A**	.	C	C	.	A	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	T	C	.	.	.	.	.	.
Zim191	A	.	.	.	.	**A**	.	T	C	.	.	.	**G**	.	.	.	.	.	.	.	.	.	.	T	.	.	.	C	.	C	.	.	.	**G**	.	.
Zim229	.	.	.	.	.	.	.	T	C	.	.	.	.	A	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	**G**	.	.
Zim377	.	.	.	.	.	.	.	.	.	.	A	C	**G**	.	.	.	C	.	.	G	.	.	.	.	.	.	.	C	.	C	.	.	.	.	.	.
Zim384	.	.	.	.	.	**A**	.	T	C	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	C	T	C	.	.	.	.	.	.
Zim398	.	.	C	.	.	.	.	T	C	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.	.	C	.	C	.	.	.	.	.	.
D. sim	.	G	.	G	**G**	.	C	.	C	C	.	C	.	.	**C**	.	C	**G**	**G**	G	C	.	C	.	**C**	C	.	C	.	C	C	**G**	**G**	G	C	C

(Continues...)

Table S7. *(Cont.)*

T25 CG6255

	6	11	12	19	40	42	43	52	108	112	121	126	144	147	201	216	222	228	240	300	303	309	339	342	345	348	360	378	397	409	420	429	465	477	531	546	552	562	609	624
ZB82	A	A	T	A	**A**	G	A	A	C	G	G	G	C	G	C	A	G	A	C	G	C	A	C	G	C	A	T	C	G	C	A	T	T	C	A	G	C	C	G	C
ZB84	.	.	.	.	T	A	.	.	.	.	.	A	.	.	.	.	A	G	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	C	G	.	.	A
ZB95	T	.	.	.	**A**	.	.	.	.	.	.	A	.	.	.	.	A	.	.	A	.	.	.	.	T	C	.	.	.	.	.	.	C	A	.	.	.	.	.	.
ZB131	.	.	.	.	T	A	.	.	.	.	.	A	.	.	.	.	A	.	.	.	A	.	.	.	T	C	.	.	.	.	G	.	C	A	.	.	.	.	.	A
ZB145	.	.	.	.	T	A	.	.	.	.	.	A	.	.	.	.	A	G	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	C	G	.	.	A
ZB157	.	.	.	.	**A**	.	.	.	.	.	.	A	.	.	.	.	A	G	.	.	.	.	.	.	T	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.
ZB186	.	.	.	.	T	A	.	.	.	.	.	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	C	A	.	C	.	.	.	.
ZB191	.	.	.	.	T	A	.	.	.	.	.	A	.	.	.	.	A	G	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	C	G	.	.	A
ZB229	.	.	.	.	T	A	.	.	.	.	.	A	.	.	.	.	A	G	.	A	.	.	.	.	T	C	.	.	.	.	G	.	C	A	.	.	.	.	.	.
ZB377	.	.	.	.	T	A	.	.	.	.	.	A	.	.	.	.	A	G	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	C	G	.	.	A
ZB384	.	.	.	.	T	A	.	.	.	.	.	A	.	.	.	.	A	G	.	.	.	.	.	.	T	C	.	.	.	.	G	.	C	.	.	C	G	.	.	.
ZB398	.	.	.	.	T	.	.	.	.	.	.	A	.	.	T	.	A	G	.	.	.	.	.	.	T	C	.	.	.	.	G	.	.	.	.	C	.	.	.	.
D. sim	C	**G**	C	**C**	.	.	**G**	**G**	T	**T**	**C**	A	.	A	.	G	C	G	T	.	.	C	T	A	.	C	A	T	**T**	A	G	C	C	.	C	C	.	T	A	.

	633	651	684	720	744	759	795	798	807	834	843	882	901	918	921	936	996	998	1000	1001	1100	1101	1107	1120	1023
ZB82	G	T	C	G	G	C	G	T	T	T	T	C	C	C	T	C	G	A	A	A	T	C	C	G	
ZB84	.	.	.	.	.	T	A	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	G	.	
ZB95	.	.	.	.	.	.	.	.	.	.	.	T	.	A	C	.	.	.	.	.	.	.	G	.	
ZB131	.	.	.	.	.	T	A	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	
ZB145	.	.	.	.	.	T	A	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	G	.	
ZB157	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
ZB186	.	.	.	.	.	T	.	.	.	.	.	T	.	A	C	.	.	.	.	.	.	.	.	A	
ZB191	.	.	.	.	.	T	A	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	G	.	
ZB229	.	.	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
ZB377	.	.	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	
ZB384	.	.	.	.	.	T	A	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	
ZB398	.	.	.	.	.	.	.	.	.	A	.	T	.	A	C	.	.	.	.	.	.	.	.	.	
D. sim	A	C	G	A	**T**	.	.	G	C	.	C	.	T	.	C	A	C	**G**	**G**	**C**	C	T	.	.	

(Continues...)

Table S7. *(Cont.)*

T29 CG10750

	9	4	4	5	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	3	3	3	3	4	4	5
		2	5	6	9	0	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	6	7	7	9	9	0	1	5	8	0	1	7	9	0	8	7	2
						2	3	4	5	6	1	2	3	4	5	5	6	7	8	9	1	2	8	9	0	1	5	0	3	5	6	3	3	1	3	0	1	9	7	
ZB82	A	G	T	G	G	C	A	G	A	G	-	-	-	-	-	T	C	A	T	A	T	T	T	G	A	A	G	G	C	G	A	C	A	G	A	C	T	A	G	A
ZB84	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	T	C	.	.	.
ZB131	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	C	.	A	.
ZB145	C	.	.	.	.	.	.	.	.	.	-	-	-	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
ZB157	.	.	.	.	A	.	.	A	G	A	C	T	A	C	A	.	.	.	.	.	C	.	.	.	.	T	A	.	.	.	.	.	.	.	.	G	.	C	.	.	.
ZB186	.	.	.	.	.	.	.	A	G	A	C	T	A	C	A	.	.	.	.	.	C	.	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	C	.	.	.
ZB229	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	G	.	C	.	A	.
ZB384	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
ZB398	.	.	.	.	.	.	.	A	G	A	C	T	A	C	A	.	.	.	.	.	C	.	.	.	.	T	A	.	.	.	.	.	.	.	.	G	.	C	.	A	.
D. sim	.	C	C	**A**	.	T	G	A	G	T	-	-	-	-	-	A	G	-	-	-	C	C	C	C	T	T	A	A	A	T	**C**	T	C	.	G	.	C	G	.	G

	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7	7	8	8	8	8	8	8	9	9	9	9	1	1	1	1	1	1
	1	4	5	7	8	0	1	1	4	6	6	7	7	7	8	8	8	8	9	0	2	6	8	9	4	5	5	5	9	9	1	3	5	9	0	0	0	0	0	0
	9	1	3	7	1	5	0	9	8	1	4	1	5	6	3	4	5	6	7	5	4	1	6	8	6	2	5	6	1	7	2	9	1	6	0	1	3	3	3	4
																																			7	6	1	2	5	1
ZB82	A	T	A	G	G	A	A	T	T	A	T	C	A	T	T	T	G	-	T	G	G	T	G	C	C	G	G	G	C	G	C	T	A	A	T	A	T	C	T	T
ZB84	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.
ZB131	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	G	C	.	C	C
ZB145	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB157	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	A	G	.	.	.	.	.	.	.	.	.	.
ZB186	.	.	.	A	.	.	.	C	C	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB229	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	C	G	C	.	C	C
ZB384	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB398	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	C	G	C	.	C	C
D. sim	**C**	C	T	.	**A**	**T**	G	.	C	T	A	G	T	G	-	-	-	C	C	A	**A**	**A**	**T**	A	.	.	T	**T**	T	.	.	C	G	.	G	.	C	A	C	C

(Continues...)

Table S7. *(Cont.)*

T29 CG10750							
	10771	10074	10804	11042	11228	11285	11555
ZB82	C	C	C	A	T	G	A
ZB84	.	.	.	.	.	.	.
ZB131	T	.	.	.	C	.	.
ZB145	.	.	.	.	.	.	.
ZB157	.	.	T	.	.	.	.
ZB186	.	.	.	.	.	A	.
ZB229	T	.	.	.	G	.	.
ZB384	.	.	.	.	.	A	.
ZB398	T	.	.	.	.	.	.
D. sim	.	T	.	G	C	.	G

(Continues...)

Table S7. *(Cont.)*

T30 CG3085

	66	72	75	78	103	108	109	111	114	120	145	147	204	215	230	231	235	236	242	243	245	247	251	253	261	268	277	322	340	343	358	361	364	433	506	523	559	595	629	640
ZB84	T	T	G	C	C	G	T	G	C	C	T	G	G	G	A	T	A	A	A	C	G	A	T	A	G	A	C	C	G	A	G	G	G	A	T	C	G	C	C	A
ZB95	.	G	C	T	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	C	.	A	.	.	.
ZB131	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	A	.	.	.
ZB145	.	G	C	T	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	A	.	.	.
ZB157	.	G	C	T	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	A	.	.	.
ZB186	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	C	.	A	.	.	.
ZB191	.	.	.	.	.	.	.	.	.	.	C	A	.	T	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	A	.	C	C	G	A	.	.	.
ZB229	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	C	.	.	.	.	.	.	.	T	.	.	C	.	.	.	C	C	.	A	.	.	.
ZB377	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.	.	.
ZB384	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	C	C	.	.	T	.	.
ZB398	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	A	.	.	.
D. sim	C	G	.	.	A	A	**G**	T	T	T	C	.	.	.	G	G	G	G	.	T	T	C	C	T	A	G	.	T	.	C	.	.	A	.	C	.	.	.	**A**	G

	64	65	67	67	69	69	70	71	72	72	73	73	73	73	73	75	76	66	67	78	84	83	85	85	88	93	93	96	100	100	100	100	100	100	100	101	111	116	117
	14	48	70	06	47	97	03	16	23	40	02	23	34	47	50	60	06	62	24	50	56	66	64	37	71	00	08	12	27	34	64	69	93	29	02	59	65	74
ZB84	C	C	T	A	G	A	C	A	G	A	C	A	G	T	C	C	A	C	T	T	C	C	T	G	T	A	G	A	G	A	C	C	C	C	G	T	G	G	G	G
ZB95	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	C	G	.	.	.	.	.	.	.	.	A	.	.	.	.	.
ZB131	T	.	.	C	.	.	.	G	A	.	.	.	.	A	.	.	.	.	.	.	.	.	C	.	C	G	A	.	A	.	.	.	.	T	.	.	.	.	.	.
ZB145	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	C	G	.	.	.	.	A	.	.	.	.	.	.	.	.	.
ZB157	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	C	G	.	.	.	.	A	.	.	.	.	.	.	.	.	.
ZB186	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
ZB191	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	C	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.
ZB229	T	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	C	.	C	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.
ZB377	.	T	.	C	.	.	.	G	A	.	.	.	.	A	.	.	.	.	.	.	.	.	C	.	C	G	A	.	A	.	.	.	.	.	.	.	T	.	.	.
ZB384	.	.	.	.	.	.	.	G	A	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	C	G	A	.	.	.	T	.	.	.	.	.	.	.	A	.
ZB398	T	.	.	C	.	.	.	G	A	.	.	.	.	A	.	.	.	.	.	.	.	.	C	.	C	G	A	.	A	.	.	.	.	.	.	.	T	A	.	.
D. sim	.	.	C	C	A	G	G	.	.	T	T	T	-	.	.	.	C	T	C	**A**	G	A	C	.	C	G	.	G	.	**G**	G	.	T	.	.	C	.	.	.	.

(Continues...)

Table S7. *(Cont.)*

T30 CG3085															
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	2	2	2	2	2	3	3	3	3	3	3	3	3
	8	9	0	1	2	4	5	0	0	1	1	5	8	9	0
	9	4	7	6	2	3	5	6	9	0	5	6	4	0	
ZB84	G	T	G	C	T	C	C	G	C	T	C	A	G	T	
ZB95	.	.	.	.	.	T	.	.	.	.	.	.	.	.	
ZB131	.	.	A	.	C	.	.	A	.	.	.	**G**	.	.	
ZB145	.	.	.	A	C	.	.	.	.	.	.	.	.	.	
ZB157	.	.	.	A	C	.	.	.	.	.	.	.	.	.	
ZB186	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
ZB191	.	.	.	.	.	.	.	.	.	.	.	**G**	.	.	
ZB229	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
ZB377	.	.	.	.	C	.	.	.	.	.	.	.	.	.	
ZB384	A	.	.	.	.	.	T	A	.	.	.	.	.	.	
ZB398	.	.	.	.	C	.	.	.	.	.	.	.	.	.	
D. sim	.	C	.	A	C	.	.	.	T	C	T	.	C	C	

(Continues...)

Table S7. *(Cont.)*

T31CG8564

	15	33	34	45	58	61	66	68	84	105	120	130	153	155	166	168	224	226	246	255	276	279	287	303	309	339	360	375	384	396	409	471	483	501	504	519	531	555	567	579
ZB82	G	A	G	G	G	A	C	**T**	T	G	T	G	C	C	G	C	C	C	C	T	G	G	C	T	G	T	A	A	C	C	G	T	G	T	G	T	T	C	A	T
ZB84	C	G	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	.	G	.	G	.	.	.	.	.	.
ZB95	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	T	G	.	.	.	.	.	.
ZB131	C	.	.	.	**A**	.	.	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
ZB145	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	G	.	.	.	.	.	.
ZB157	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	G	.	.	.	.	.	.
ZB186	C	G	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
ZB191	C	G	**A**	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
ZB229	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
ZB377	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
ZB384	C	G	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	.	.	.
ZB398	C	.	.	T	.	.	.	C	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	G	.	.	.	.	.	.
D. sim	C	T	.	.	.	**G**	G	C	C	A	**G**	**A**	G	**A**	**C**	A	**G**	.	.	C	**A**	.	**T**	**A**	**C**	A	G	T	.	.	A	G	.	G	T	C	C	T	T	C

	598	602	628	647	667	679	703	724	760	775	883	907	909	925	970	997	1012	1033	1055	1057	1108	1114	1132	1135	1160	1208	1220	1246	1250	1254	1265	1283	1303	1313	1333	1338	1349	1373	1449
ZB82	T	G	C	G	T	T	A	T	A	A	T	G	A	G	T	T	C	G	T	A	T	C	A	A	C	T	C	A	G	A	**C**	A	C	T	G	A	C	A	A	G
ZB84	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	G	C	.	.	.	.	.	.	.	.	.	**C**	.	.	.	.	.	.	.	.	.
ZB95	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	G	C	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
ZB131	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	C	G	C	.	.	.	.	.	.	.	.	.	A	.	G	.	.	.	.	.	.	.
ZB145	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	G	C	.	.	.	.	.	.	.	.	.	**C**	.	.	.	.	.	.	.	.	.
ZB157	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	.	G	C	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
ZB186	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
ZB191	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**C**	.	G	.	.	.	.	.	.	.
ZB229	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	C	G	C	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
ZB377	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.
ZB384	.	.	T	.	.	.	.	.	.	.	.	A	.	A	C	.	.	.	C	G	C	.	.	.	.	.	.	.	.	.	**C**	.	G	.	.	.	.	.	.	.
ZB398	.	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	.	.	C	G	C	.	.	.	.	.	.	.	.	.	**C**	.	.	.	.	.	.	.	.	.
D. sim	C	-	.	**A**	**A**	C	G	C	G	G	C	.	**G**	.	C	A	T	.	C	G	.	.	C	G	T	**G**	**A**	T	**C**	**G**	.	**G**	.	**A**	C	-	-	-	**T**	**T**

(Continues...)

Table S7. *(Cont.)*

T31 CG8564																	
	13666	13677	13880	14044	14434	14453	14665	14668	14669	14885	14995	15527	15546	15550	15552	15554	15567
ZB82	C	T	G	A	C	A	T	G	T	A	C	**C**	C	C	T	**A**	G
ZB84	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	C	A
ZB95	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	C	A
ZB131	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	C	A
ZB145	.	.	.	.	.	.	.	.	.	.	T	**C**	.	.	.	C	A
ZB157	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	C	A
ZB186	.	.	.	.	.	.	.	.	.	.	T	**C**	T	.	.	C	A
ZB191	.	.	.	.	.	.	.	.	**C**	.	T	**C**	.	.	.	C	A
ZB229	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	C	A
ZB377	.	.	.	.	.	.	.	.	.	.	T	**C**	.	.	A	C	A
ZB384	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	C	A
ZB398	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.	C	A
D. sim	T	**G**	**A**	.	**A**	G	G	**T**	.	**G**	.	T	.	**T**	.	C	A

(Continues...)

Table S7. *(Cont.)*

T32CG11475

	4	15	18	20	26	36	60	69	70	84	93	100	108	120	132	138	162	192	199	202	210	215	222	235	240	243	249	251	255	261	288	336	346	358	360	363	371	388	394	397
ZB82	C	C	A	C	A	G	C	G	C	G	G	T	T	T	C	T	G	T	C	A	G	C	G	G	C	A	A	A	C	C	C	A	T	T	G	T	A	C	G	C
ZB84	.	.	.	.	.	.	.	.	.	.	A	C	A	.	.	.	.	C	T	.	.	.	.	.	T	.	.	.	T	.	T	.	.	.	A	.	**G**	.	.	.
ZB95	.	.	.	.	.	.	.	.	T	A	.	C	.	.	.	.	.	C	T	.	A	.	.	.	T	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.
ZB131	.	.	.	.	.	.	.	.	T	.	A	C	A	.	.	.	.	C	T	.	.	.	A	.	T	.	.	.	.	.	.	G	**A**	.	.	G	.	.	.	.
ZB145	.	.	.	.	.	.	.	.	T	A	.	.	.	.	.	.	.	C	T	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB157	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	T	.	.	.	.	.	T	.	.	.	T	.	.	G	.	.	.	.	.	.	.	.
ZB186	.	.	.	.	.	.	.	.	T	A	.	C	A	.	.	.	.	.	T	.	.	.	A	.	T	.	.	.	.	.	.	.	**A**	.	.	G	.	.	.	.
ZB191	.	.	.	.	.	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB377	.	.	.	.	.	.	.	.	T	A	.	C	.	.	.	.	A	C	T	.	.	.	.	.	T	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.
ZB384	.	.	.	.	.	.	.	.	T	A	.	.	.	.	.	.	.	C	T	.	A	.	.	.	T	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.
ZB398	.	.	.	.	.	.	.	.	T	.	A	C	A	.	.	.	.	C	T	.	.	.	A	.	T	.	.	.	.	.	.	G	**A**	.	.	G	.	.	.	.
D. sim	**T**	**A**	**T**	.	**G**	C	.	.	.	.	.	C	A	C	T	G	.	C	T	C	.	**T**	.	**A**	T	C	G	**T**	.	**G**	.	**T**	.	**G**	.	.	.	**A**	.	**A**

	423	427	429	430	494	516	537	540	555	558	579	591	593	600	621	624	634	657	696	714	729	740	765	807	813	825	829	900	912	933	942	943	945	948	951	954	987	1002	1004	1005
ZB82	C	G	A	A	C	C	T	T	G	G	A	A	A	C	C	G	A	T	A	C	G	T	G	G	T	G	C	G	G	T	A	G	A	G	T	A	T	T	G	C
ZB84	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	A	.	.	.	.	.	A	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB95	.	.	.	.	.	.	.	.	A	.	.	T	.	.	.	.	.	A	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	C	.	.	.
ZB131	T	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB145	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
ZB157	.	.	.	.	**G**	.	.	.	.	A	.	T	.	.	.	.	.	A	.	.	.	.	.	.	T	.	C	.	.	.	.	A	.	.	C	.	.	.
ZB186	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB191	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
ZB377	.	.	.	.	.	.	.	.	.	A	.	T	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
ZB384	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A	.	.	.	**C**	.	.	.	.	.	A	C	.	.	.	.	.	.	.	C	.	.	.
ZB398	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	T	A	.	.	.	.	.	.	.	.	C	.	.	.
D. sim	.	.	G	**G**	.	T	.	C	.	.	C	.	**G**	T	A	.	**G**	A	G	.	C	C	.	.	C	.	.	.	.	C	G	**C**	G	.	C	.	.	G	**C**	T

(Continues...)

Table S7. *(Cont.)*

T32 CG11475																							
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	0	0	0	0	0	0	0	1	1	1	1	2	2	2	2	2	3	3	3	3	3	3	
	0	2	6	7	8	9	9	1	6	8	8	2	2	4	6	8	0	0	1	4	5	8	
	8	3	2	0	6	2	9	3	7	0	5	7	8	5	0	2	3	9	7	7	3	0	
ZB82	T	C	T	A	G	A	C	C	C	A	A	A	T	G	C	A	A	G	C	C	G	T	
ZB84	C	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	
ZB95	C	.	C	.	.	.	.	T	.	.	.	.	.	.	G	.	.	.	A	.	.	.	
ZB131	C	.	.	.	.	.	.	T	T	.	.	.	.	.	G	.	.	.	A	.	.	.	
ZB145	C	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	
ZB157	C	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	T	.	.	
ZB186	C	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	
ZB191	C	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	
ZB377	C	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	
ZB384	C	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
ZB398	C	.	.	.	.	.	.	T	.	.	.	.	.	.	A	.	.	.	.	.	.	.	
D. sim	C	.	C	**G**	C	G	T	.	.	**T**	G	C	**G**	T	G	**C**	.	**C**	G	.	A	C	

(Continues...)

Table S7. *(Cont.)*

O23 CG9273

	21	22	25	26	35	41	55	56	67	105	1069	1290	1401	1441	1442	1443	1510	1606	1681	1686	1888	1901	2037	2466	2618	2888	3099	3224	3277	3422	3455	3666	3877	4177	4207	4777	4899	5071	5311
Zim82	T	T	C	C	G	T	A	G	-	A	A	G	G	T	G	A	G	G	G	C	T	A	C	T	A	A	A	G	C	G	C	C	G	T	A	A	G	C	T	C
Zim84	.	A	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	A	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Zim95	.	A	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	A	.	.	.	.	G	.	G	.	.	T	.	.	.	.	.	.	.	A	.	.	.
Zim131	.	A	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	A	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Zim145	.	A	.	.	.	.	.	.	-	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim157	.	A	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	A	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Zim186	.	.	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	A	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Zim191	.	A	.	.	.	.	.	A	-	.	.	.	.	.	.	.	C	.	A	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Zim229	.	A	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	A	.	.	.	.	G	.	G	.	.	.	.	.	.	.	T	.	.	A	.	.	.
Zim377	.	A	.	.	.	.	.	A	-	.	.	.	.	.	.	.	.	.	A	.	.	.	.	G	.	G	.	.	.	.	.	.	.	.	.	.	A	.	.	.
Zim384	.	A	.	.	.	.	.	.	-	.	.	.	.	.	.	.	.	.	A	.	.	.	.	G	.	G	.	.	T	.	.	.	.	.	.	.	A	.	.	.
Zim398	.	A	.	.	.	.	.	A	-	.	.	.	.	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. sim	G	A	T	A	A	G	G	.	T	C	G	.	-	-	-	-	.	A	A	T	C	T	A	G	G	G	C	A	.	A	T	A	.	G	G	C	A	A	C	G

	6007	6099	6255	6626	6679	6788	6881	6899	7012	7129	7235	7341	7480	7586	7679	8000	8039	8404	8707	8783	8836	8879	8890	8994	9042	9771
Zim82	G	G	G	C	C	A	G	C	A	G	C	C	C	G	T	T	A	T	A	T	T	A	T	G	C	A	-
Zim84	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Zim95	**T**	.	**C**	**G**	.	G	.	A	**C**	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	T
Zim131	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Zim145	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T
Zim157	.	.	.	.	.	G	.	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Zim186	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Zim191	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Zim229	.	.	**T**	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Zim377	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Zim384	.	.	.	.	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
Zim398	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	T
D. sim	.	C	.	.	T	.	T	.	.	A	.	T	T	.	C	G	G	G	C	.	G	T	G	T	T	C	T

(Continues...)

Table S7. *(Cont.)*

O24 CG4973

	30	48	63	66	76	131	141	161	178	208	226	294	296	297	298	301	304	307	316	319	325	338	361	382	388	391	430	463	464	478	490	496	511	520	541	568	586	619	622	628
Zim82	G	G	C	T	C	C	G	C	T	G	C	G	G	T	T	T	T	T	C	A	C	C	A	A	A	T	G	C	T	C	A	C	G	T	T	C	T	A	T	G	T
Zim84	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim95	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
Zim131	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim157	C	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim186	C	.	.	.	**G**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim191	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim229	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim377	C	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim384	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
Zim398	C	.	.	.	**G**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. sim	C	A	A	C	.	T	A	A	G	A	T	C	A	C	G	C	A	C	A	C	A	T	G	T	G	C	A	A	A	T	G	A	C	C	C	.	C	G	G	A	G

	64	65	66	68	67	68	73	73	74	76	78	80	80	87	90	92	92	93	94	94	94	95	95	95	97	100	100	101	101	101	101	101	101	101	101	101	101	101	101	101
	35	54	64	86	76	82	31	61	45	63	85	03	04	71	04	22	29	36	40	45	81	29	40	51	52	57	85	05	59	11	12	16	17	18	22	28	36	53	55	72	87
Zim82	T	A	T	A	A	T	A	A	A	A	G	A	**T**	T	T	A	G	T	A	T	C	T	C	C	G	G	T	T	G	T	G	A	G	G	G	A	G	G	G	T
Zim84	.	.	.	.	.	.	.	G	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**C**	.	.	.	.	.	.
Zim95	.	.	.	.	.	.	**C**	.	.	.	**A**	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim131	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim157	.	.	.	.	.	.	.	G	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	**C**	.	.	.	.	.	.	T	.	A	.	.	.	.	.
Zim186	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim191	.	.	.	.	.	.	.	G	.	.	.	.	C	C	.	G	.	.	C	.	.	.	.	.	.	.	**C**	.	**A**	.	.	.	.	T	.	A	.	.	A	.	.
Zim229	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim377	.	.	.	.	.	.	.	G	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim384	.	.	.	**C**	.	.	.	.	**T**	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim398	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	**C**	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. sim	C	C	C	.	G	C	.	T	.	G	.	G	G	C	C	.	A	C	.	A	T	C	T	T	C	A	C	C	.	C	-	-	-	.	T	T	.	C	A	.

(Continues...)

Table S7. *(Cont.)*

O32 CG12276

	45	1121	1212	1626	1672	1772	1778	1996	2002	2041	2711	2777	2884	3006	3101	3212	3228	3315	3357	3499	3525	3588	3595	4211	4311	4436	4487	4560	5117	6077	6170	6212	6313	6414	6616	6717	6828
ZB82	T	A	A	C	G	A	T	C	T	C	G	T	G	G	A	T	T	A	C	G	G	C	A	C	G	G	A	T	T	C	T	G	G	T	C	C	-	A	C	C
ZB84	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	-	.	.	.
ZB95	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.
ZB131	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	-	.	.	.
ZB145	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.
ZB157	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	-	.	.	.
ZB186	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	-	.	.	.
ZB191	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	-	.	.	.
ZB229	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.
ZB377	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	-	.	.	.
ZB384	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.
ZB398	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	-	.	.	.
D. sim	G	C	G	T	A	C	C	T	C	T	A	C	**A**	**C**	G	**C**	G	G	**G**	A	T	A	G	T	**A**	A	**G**	G	C	T	G	C	A	C	A	A	C	G	T	T

	631	633	634	635	637	638	639	640	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	668	705	706	708	763	766	769	823	841
ZB82	T	T	G	C	T	A	C	-	-	T	G	T	C	C	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	C	C	G	G	T	C	A	C
ZB84	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.	.
ZB95	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.	.
ZB131	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.	.
ZB145	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.	.
ZB157	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	**A**	.	.	.	.	.	.
ZB186	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.	.
ZB191	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	.	.	.	.	.	.	.
ZB229	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.	.
ZB377	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	.	.	**T**	.	.	.	.
ZB384	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.	.
ZB398	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.	.
D. sim	G	A	A	A	C	G	A	A	T	G	T	C	T	T	T	G	C	T	A	A	C	T	T	A	A	T	G	T	C	C	A	C	.	G	.	C	**A**	T	T	G

(Continues...)

Table S7. *(Cont.)*

O32 CG12276

	8 5 6	8 6 0	8 6 1	8 6 5	8 6 8	8 9 5	9 1 3	9 9 1	9 9 7	1 0 0 6	1 0 1 2	1 0 3 3	1 0 4 8	1 0 8 1	1 0 8 7	1 1 1 1	1 1 1 4	1 1 1 7	1 1 4 6
ZB82	A	A	G	C	G	G	C	A	G	A	C	C	A	G	A	C	A	G	G
ZB84	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB95	.	**T**	**T**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB131	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB145	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB157	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB186	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB191	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	**A**
ZB229	.	**T**	**T**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB377	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**A**
ZB384	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB398	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. sim	G	.	**A**	G	C	C	T	G	T	G	.	T	C	A	T	T	C	A	A

(Continues...)

Table S7. *(Cont.)*

O46 CG6874

	49	56	78	81	1117	1188	1132	1140	2258	2738	2888	2889	2900	2901	2912	2923	2944	2955	2966	2977	2988	2999	3000	3004	3013	3114	3115	3225	3226	3227	3228	3229	3230	3333	3335	3336	3340	3545	3767	3777
Zim84	G	A	T	C	T	T	A	C	C	A	T	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	-	-	-	-	-	-	T	T	C	G	A	G	G
Zim95	.	.	.	.	C	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.
Zim131	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.
Zim145	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.
Zim157	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.
Zim186	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.
Zim191	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.
Zim229	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.
Zim377	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.
Zim384	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.
Zim398	.	.	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-	-	-	-	-	.	-	-	-	-	-	-	-	-	-	.	.	.	.	.	.	.
D. sim	C	G	C	T	C	C	G	T	A	G	G	T	C	C	A	C	T	T	C	G	T	C	T	T	C	G	G	G	A	A	T	T	G	G	C	T	A	G	-	-

	378	391	399	401	414	415	429	441	444	447	494	496	503	510	520	528	541	547	549	552	564	567	600	609	613	626	635
Zim84	G	C	A	A	C	A	A	T	G	T	A	C	T	A	C	G	C	T	T	C	C	C	G	G	C	C	A
Zim95	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Zim131	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim145	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim157	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim186	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	A	.	.	.	.	.	.
Zim191	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Zim229	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Zim377	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Zim384	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
Zim398	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. sim	-	G	C	C	G	G	C	C	A	C	C	G	G	G	G	T	T	A	C	T	.	T	T	C	T	T	C

(Continues...)

Table S7. *(Cont.)*

O48 CG3509

	4	21	24	39	73	76	87	102	114	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	167	185	190	198	218	222	245	253	264	267	282	319	339
ZB82	A	A	C	T	C	G	C	A	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	T	A	T	C	T	G	C	C	C	G	C	G
ZB84	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	G	A	T	G	A	G	G	A	G	G	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB95	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	G	A	T	G	A	G	G	A	G	G	A	G	.	.	.	.	.	.	.	.	.	T	.	.	.
ZB131	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	G	A	T	G	A	G	G	A	G	G	A	G	.	.	.	.	.	.	.	.	.	T	.	.	.
ZB145	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	G	A	T	G	A	G	G	A	G	G	A	G	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB186	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	G	A	T	G	A	G	G	A	G	G	A	G	.	.	.	.	.	.	.	.	.	T	.	.	.
ZB191	.	.	T	.	.	.	T	.	.	-	-	-	-	-	-	G	A	T	G	A	G	G	A	G	G	A	G	.	.	**G**	.	.	.	.	.	.	T	.	.	.
ZB229	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	G	A	T	G	A	G	G	A	G	G	A	G	.	.	.	.	.	.	.	.	.	T	.	.	.
ZB377	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	G	A	T	G	A	G	G	A	G	G	A	G	.	.	.	.	.	.	.	.	.	T	.	.	.
ZB384	.	.	.	.	.	.	.	.	.	-	-	-	-	-	-	G	A	T	G	A	G	G	A	G	G	A	G	.	.	.	.	.	.	.	.	.	T	.	.	.
D. sim	**C**	**T**	T	**A**	**T**	**C**	.	G	A	A	T	G	G	A	G	G	A	G	G	A	G	G	A	G	G	A	G	**A**	**C**	.	**G**	**T**	C	**A**	**A**	A	T	A	T	A

	429	439	451	453	483	489	492	493	495	513	519	592	593	594	595	596	611	613	614	616	618	623	642	645	662	668	680	701	722	725	726	728	729	730	731	733	734	735	736
ZB82	C	C	T	A	G	T	G	C	A	T	**G**	G	A	C	T	A	T	C	A	G	G	G	C	T	C	C	A	A	A	C	T	A	-	-	-	-	-	-	-	-
ZB84	T	.	C	G	.	.	.	.	.	.	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	A	.	-	-	-	-	-	-	-	-
ZB95	T	.	C	G	.	.	.	.	.	.	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	A	.	-	-	-	-	-	-	-	-
ZB131	T	.	C	G	.	.	.	.	.	.	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	A	.	-	-	-	-	-	-	-	-
ZB145	T	.	C	G	.	.	.	.	.	.	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	A	.	-	-	-	-	-	-	-	-
ZB186	T	.	C	G	.	.	.	.	.	.	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	A	.	-	-	-	-	-	-	-	-
ZB191	.	.	.	.	.	.	.	.	.	.	**G**	.	.	**T**	.	.	.	.	.	A	.	.	.	.	G	.	.	.	.	.	.	.	-	-	-	-	-	-	-	-
ZB229	.	.	.	.	.	.	.	.	.	.	**G**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	A	.	-	-	-	-	-	-	-	-
ZB377	.	.	.	.	.	A	.	.	.	.	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	A	.	-	-	-	-	-	-	-	-
ZB384	.	.	.	.	.	A	.	.	.	.	T	A	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	T	T	A	.	-	-	-	-	-	-	-	-
D. sim	.	**A**	C	.	A	.	A	A	G	**G**	.	.	C	.	C	**G**	**C**	**G**	G	.	**A**	**A**	T	**G**	T	.	G	.	.	.	.	G	A	A	C	C	T	A	T	A

(Continues...)

Table S7. *(Cont.)*

O48 CG3509

	7 3 7	7 3 8	7 3 9	7 4 0	7 4 1	7 4 2	7 4 3	7 4 8	7 4 9	7 5 9	7 6 1	7 8 0	7 8 1	7 8 4	7 8 5	7 8 6	7 8 7	7 8 8	7 9 0	7 9 1	7 9 2	7 9 3	7 9 4	7 9 5	7 9 7	7 9 8	7 9 9	8 0 5	8 0 6	8 1 5	8 1 7	9 1 4	9 2 2	9 2 3	9 2 4	9 2 8	9 4 0	9 4 5	9 4 6	9 6 3
ZB82	-	-	-	-	-	-	C	C	A	T	G	A	A	T	G	A	T	A	T	T	A	A	T	A	T	-	-	G	A	C	A	T	A	A	G	C	A	G	G	T
ZB84	-	-	-	-	G	T	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB95	-	-	-	-	G	T	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	**A**	.	.	.	.
ZB131	-	-	-	-	G	T	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB145	-	-	-	-	G	T	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB186	-	-	-	-	G	T	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	C	.	.	.	.	.	.	.	.	.	.	.
ZB191	-	-	-	-	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	C	.	.	.	.	.	.	.	.	.	.	.
ZB229	-	-	-	-	G	T	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB377	-	-	-	-	G	T	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB384	-	-	-	-	G	T	A	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.
D. sim	T	A	A	A	G	T	A	.	C	A	A	T	T	C	C	T	A	T	A	A	T	G	A	T	G	T	A	A	.	T	C	**G**	-	-	-	.	**T**	A	**A**	C

	9 6 6	9 9 0	1 0 0 5	1 0 0 8	1 0 1 0	1 0 0 3	1 0 0 9	1 0 1 3	1 0 0 1	1 0 3 8	1 0 0 5	1 0 0 0	1 0 6 3	1 0 0 6	1 0 6 6	1 0 6 7	1 0 7 2	1 0 8 2	1 0 9 9	1 1 3	1 1 7 1	1 2 1 1	1 2 2 3	1 2 3 1	1 2 7 2	1 2 7 3
ZB82	A	A	A	C	A	C	T	G	G	T	C	C	G	C	T	C	**T**	G	C	T	A	G	A	A	G
ZB84	.	G	.	.	.	.	.	.	.	.	.	.	**T**	.	.	.	C	.	.	.	.	.	.	.	.
ZB95	.	G	.	.	.	.	.	.	.	.	.	.	**T**	.	.	.	C	.	.	.	.	.	.	.	.
ZB131	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**T**	.	.	.	.	.	.	.	.
ZB145	.	G	.	.	.	.	.	.	.	.	.	.	**T**	.	.	.	C	.	.	.	.	.	.	.	.
ZB186	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
ZB191	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	A	.	.	.	.	.	.	.
ZB229	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**T**	.	.	.	.	.	.	.	.
ZB377	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
ZB384	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
D. sim	G	G	T	A	**T**	**T**	**G**	**A**	A	**C**	**A**	**A**	.	**A**	**G**	**T**	C	.	G	**C**	C	C	**C**	G	**C**

(Continues...)

Table S7. *(Cont.)*

O53 CG3975

	8	24	28	36	43	47	69	84	110	120	123	159	179	214	218	241	279	309	325	336	349	354	366	368	397	403	423	425	427	463	468	476	479	483	502	515	519	529	531	545
ZB82	T	C	T	C	G	G	T	A	T	C	C	G	**G**	T	**C**	A	T	A	G	G	T	G	A	A	A	C	T	C	C	C	A	C	T	A	T	A	A	A	G	T
ZB84	.	.	.	.	.	.	.	.	.	.	T	.	A	C	A	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
ZB95	.	.	.	.	.	.	.	.	.	.	.	.	A	C	A	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB131	**C**	.	.	.	.	.	.	.	.	.	.	.	A	C	A	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
ZB145	.	.	.	.	.	.	.	.	.	.	.	.	A	C	A	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
ZB157	.	.	.	.	.	.	.	.	.	.	T	A	A	C	A	.	.	C	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB191	.	.	.	.	.	.	.	.	.	.	.	.	A	C	A	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB229	.	T	.	.	.	.	.	G	.	.	.	.	A	C	A	.	.	C	.	.	.	A	C	**C**	.	.	C	**T**	**A**	.	.	.	.	.	.	.	.	.	.	.	.
ZB377	.	.	.	.	.	.	.	.	.	.	.	.	A	C	A	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB384	.	.	.	.	.	.	.	.	.	.	.	.	A	C	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB398	.	.	.	.	.	.	.	.	.	.	.	.	A	C	A	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
D. sim	.	G	A	T	A	A	C	G	C	A	.	.	A	C	A	G	C	C	A	A	C	.	C	.	G	G	C	A	A	T	G	T	C	G	G	T	C	G	C	C

	548	556	557	558	559	560	561	562	621	681	718	756	776	786	789	799	800	828	837	852	858	859	863	873	882	883	900	909	912	942	985	1048	1063	1066	1067	1069	1072	1077	1099	1110
ZB82	C	-	-	-	-	-	-	T	A	**G**	A	A	A	G	G	T	C	G	C	A	A	A	C	T	A	C	T	G	C	G	G	T	A	A	T	G	T	T	G	C
ZB84	.	-	-	-	-	-	-	.	.	T	.	G	**G**	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB95	.	-	-	-	-	-	-	.	.	T	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	**A**	.	.	.	.	.	.	.	.	.
ZB131	.	-	-	-	-	-	-	**G**	.	T	**G**	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**A**	**G**	.	.	.	.	.
ZB145	.	-	-	-	-	-	-	.	T	**G**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**A**	.	.	.
ZB157	.	-	-	-	-	-	-	.	.	T	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	**C**	.	.	.	.	.	.	.	.	.	.
ZB191	.	-	-	-	-	-	-	.	T	**G**	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	**A**	**G**	C	.	.	.	.
ZB229	.	-	-	-	-	-	-	.	T	T	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB377	.	-	-	-	-	-	-	.	.	T	**G**	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C	T	.	.	.	.	.	.	**A**	**G**	C	**T**	.	.	.
ZB384	.	-	-	-	-	-	-	.	.	**G**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB398	.	-	-	-	-	-	-	.	T	**G**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.
D. sim	G	G	G	C	A	A	G	.	.	T	.	G	.	C	A	G	T	.	T	G	C	T	A	C	G	.	C	C	.	.	.	.	T	G	C	.	G	C	.	G

(Continues...)

Table S7. *(Cont.)*

O53 CG3975

	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2
	2	2	3	3	4	5	5	6	6	6	7	7	8	9	0	2	3	3	5	7
	1	2	4	7	6	2	5	1	5	7	2	3	0	7	0	4	3	6	7	2
ZB82	**G**	A	C	T	T	A	A	C	C	T	A	C	G	T	T	G	C	G	G	T
ZB84	A	.	T	.	A	G	.	.	.	C	.	.	.	A	A	A	.	C	.	.
ZB95	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB131	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.
ZB145	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB157	A	.	.	.	.	.	.	.	.	.	.	.	.	A	C	.	T	.	.	.
ZB191	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB229	A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.
ZB377	A	.	T	.	.	G	.	.	.	.	.	.	.	A	A	A	.	A	.	.
ZB384	**G**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB398	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. sim	A	**T**	T	C	A	G	G	T	**T**	C	**T**	**G**	**A**	A	.	A	.	C	A	C

(Continues...)

Table S7. *(Cont.)*

U23 CG3476

	24	60	75	84	102	120	125	140	149	163	180	206	216	234	255	309	321	327	384	444	450	484	486	501	561	569	570	583	591	597	645	646	647	648	649	651	652	653	654	659
ZB82	A	A	C	T	T	G	T	C	A	T	C	**T**	G	T	G	T	G	C	C	T	C	C	G	G	G	T	T	A	A	G	T	A	T	A	T	T	A	T	A	A
ZB84	.	G	G	.	.	.	.	.	.	.	.	C	.	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.
ZB95	.	G	.	.	A	.	.	.	.	.	.	C	.	A	.	.	A	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	-	-	-	T
ZB131	.	G	.	.	A	.	.	.	.	.	.	C	.	A	.	.	A	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	-	-	-	T
ZB145	.	G	G	.	.	.	.	.	.	.	.	C	.	A	.	.	A	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	-	-	-	T
ZB157	.	.	.	.	C	.	.	.	.	.	.	C	.	A	.	.	A	.	.	.	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB191	.	G	.	.	A	.	.	.	.	.	.	C	.	A	.	.	A	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	A	-	-	-	T
ZB229	.	G	G	.	A	.	.	.	.	.	.	C	.	A	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB377	.	G	.	.	A	.	.	.	.	.	.	.	.	A	.	.	A	.	T	.	.	**A**	.	T	.	**A**	.	**C**	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB384	.	G	.	.	.	.	.	.	.	.	.	C	.	A	.	.	A	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB398	.	G	.	.	A	.	.	.	.	.	.	C	.	A	.	.	A	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. sim	G	G	.	C	.	A	C	G	T	A	A	.	A	A	A	C	A	T	.	C	T	.	T	.	.	.	A	.	C	.	A	T	A	T	A	-	-	-	-	.

	66	66	67	67	69	68	71	72	75	75	76	80	82	86	88	89	90	92	93	94	97	98	98	100
	67	73	36	91	81	18	43	73	76	38	93	89	63	99	26	28	45	43	9	0	1			
ZB82	G	G	A	A	A	C	A	A	C	G	A	G	G	G	T	G	T	T	A	C	C	A	C	A
ZB84	.	.	.	.	T	.	.	.	.	.	.	.	.	.	G	.	.	C	C	.	T	.	.	.
ZB95	-	A	.	.	T	.	.	.	.	.	.	.	.	.	G	.	.	C	.	**T**	.	.	.	.
ZB131	-	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
ZB145	-	A	.	.	T	.	.	.	**A**	A	T	.	.	A	G	.	.	C	.	.	.	.	.	.
ZB157	.	.	.	.	T	.	.	.	.	.	.	.	.	.	G	.	.	C	.	.	T	.	.	.
ZB191	-	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
ZB229	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	C	.	.	.	.	.	.
ZB377	.	.	.	.	T	.	.	.	.	.	.	.	.	A	G	.	.	C	C	.	.	G	.	.
ZB384	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ZB398	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D. sim	-	A	G	-	C	T	G	C	A	.	T	A	A	.	G	C	C	C	.	.	.	.	T	.

Table S8. PCR and sequencing primers.

Gene	Type	Forward primer (5' - 3')	Reverse primer (5' - 3')	Temp
Male-biased				
CG7931	PCR	GCAGTTAGTCTGTAGCTTTGC	CCGAAAAGAAACTGGTATGAACGG	54
CG7929	PCR	CCGTTCATACCAGTTTCTTTTCGG	GGCAAGATGATGTTGTAATGCTGG	51
CG10252	PCR	CGGCCAACACCTAATATTGG	CCATAAACAATGAAACTTGG	55
CG14926	PCR	CGCATCTCCAGGTGATCCATTCG	GGTTCGGATCATTGAGAC	58
	Int1	GAGTCTATGGGATTCGATGGAG	TCAAGACCAGGCTCTGAGC	
CG17956	PCR	TATAAGCTTTAGCCAATGCT	GAAGACTGCGGATCGATTCT	54
CG17376	PCR	CGGACCAGTCGCTTCGACTT	TTATATGGGTCGTCCGTAGT	53
CG1980	PCR	GACAAATAGTCTCCAGCTGT	AACAGAGCTGAATGAATT	52
CG5565	PCR	AATTAAGCACCGTTCTGGCC	GAATACTAATTTTGACGGATC	53
CG18418	PCR	CATACGAATTTGGCTCGTTTCC	CAGAACAAGCGCGCTTTACAC	53
CG6980	PCR	GCAGTCGAACCAGTCGGTATC	CATGCTTCACAGATTTGGGCAG	53
CG6255	PCR	AACTCCATTCTAAATCAGGCC	GAGCACTAAGTTAGTTTATTGG	53
CG6332	PCR	CAAAATTCGGAAATCCACAG	GATACCAAATGTAGCCGAAA	53
	Int1	CGCTACAGCTATCTATCGCT	GATCTCCTCCAAATAGGCCT	
CG3483	PCR	CTCACAACTCACATTTCGGG	CCAGGCACATAAAGTCGC	52
	Int1	GGCCAACATCATGCGGATGA	CAGGTAGCGCTCCTCGAACA	
CG10307	PCR	GGCGACAGTTCGAAACTACA	ATAAACCACAGATTTATGTG	52
	Int1	GTGAACACCTCGAGACGCTG	CCCCTTGAGATTCAGATGCA	
CG10750	PCR	ATCCTTGACTGTGCTGTCTG	GTTCCATTACTTAAAGGATT	50
	Int1	TGCGCTCAAGGCCACGGTGG	TGTGGAGAACCCACTTAG	
CG3085	PCR	GAGTTCTAGTTGGGAGCACA	AATTTGGACAAATTACAACG	53
	Int1	CTGGTCGTGCTGGAGAACAA	CTGGCATCGATCGGCCAGCA	
	Int2	GGATCTTCGAGACGCAGCGG	TTCAGTGCATCCGTCTTGTC	
CG8564	PCR	CTCAAGTGATAAATCGTTTTC	TCACGTTGCCACACTAATCAC	56
	Int1	ACATACAAGGGTGAAAGTCC	ACATGATACTTTGGCCATAC	
CG11475	PCR	TTCCTGGCAAAGAGGGCTTCG	AAATATGGGCTGTTAATTGGG	53
	Int1	CACGGTGACCTAAACATCAC	CAAACATCGAACGACTGATC	
CG7387	PCR	CCCCATAAAATAAACAGAAAC	GCAAGGCAACCCAAGCAAATC	53
	Int1	AACACGATCCTTGAATCGCC	GCTTCTTGCATCCCACTGTC	
CG18266	PCR	CCGACACGACATATCGTCTCG	ATGGCGATGGACTAACGAATG	53
	Int1	ACCTTCTACAAGCAGGCCTC	GGTACATGTCCGACCAGCTG	
CG9314	PCR	CAAGCAAAATCACTCACAACGC	ATATTTCAACAGCCCACAGTC	53
	Int1	TGATGTCCCACGATCCGGAC	GACCTATATAATTCCTACCTG	
CG6971	PCR	CCACAGGCATTTCTGATTTTC	GTCGGCTCCAGCTGAAGAGTC	53
CG7251	PCR	ATTATTTCGTTTAACAGTGTC	CTGTACATTTGGAGTCGCCAA	53
CG9531	PCR	TGGGTCGTCAAAATTTTGTTG	CAGCTAAATGCATGGGCAACG	53
CG6130	PCR	CCCAACTGCCCTGAGAAGTCC	CACGTTCCTGCCCCACTGCAA	53

(Continues...)

Table S8. *(Cont.)*

Gene	Type	Forward primer (5' - 3')	Reverse primer (5' - 3')	Temp
CG5045	PCR	AACATTGCGAATCCGGTTGA	TTGGAAAGTGTATATTCTGT	53
CG5276	PCR	AACGAGATCTTGCTGACGTG	CTGAAGCACCTGGATCGAAG	53
	Int1	GTGAATGACAAGCCCATTCC	TTGCCCATGGAGCCCACATA	
CG15179	PCR	GGCCAGAGCCTCATAAGTTAG	CAACACTTGCGCTATTTTATG	54
CG8277	PCR	GTAAGCAATTCCCACTCGAATC	CACACTTACGAATCTGGTGTAC	54
CG14717	PCR	CATGGATAACAAAAGAAAACCC	TGCCAATTTTATATAAAAGTCAC	53
CG11037	PCR	GATACTTATTGGACTGTGTCGC	CTATAATTAAGCAAGCCAAATCCG	63
CG6036	PCR	ATTGAATTGTCATTGATTAA	CACAGCTGTTGCTACATAGT	50
	Int1	ACGCTGGCAGTCAAATATCA	TCATCCAACTGCAGGAAGCC	
	Int2	GAGGTTTGTGAATTTATTAG	CCCTGCTGCCTTTATGCAGG	
CG13527	PCR	CACTCAGCTGCCAGATGAAC	AGCTGAATTCGTGTATCCAT	54
	Int1	CCTCGAGTTGGCCAAGTACG	CAGCAGCCATCGAGCCTTGT	
	Int2	GTGCAGCGGCGACATCGGAT	AACCGCAGCCAATCGGATAG	
Female-biased				
CG2867	PCR	CAACAACAACAGTCGCAGCAG	GTGACAGTAATTCCGAGCCCTCT	54
CG9383	PCR	GTCTACGCTCTGCAATCTC	CATTGTTCGGACTGCCGGCAG	55
CG9273	PCR	CCTTGTGAAATTATATAGAAT	AATAGTGGTTACACGATTAGA	50
CG4973	PCR	GTGAATTACTAGATAAGTACA	ACTTTCACAACCTGTATAAATGAT	59
CG1239	PCR	TATACTTTGGATTTCACA	GAAGCGAATGAAGTAGTA	54
CG6554	PCR	TAGCGTGCTGTAGAAACAGT	CCTTACGTCTAGTCATTG	55*
	Int1	TTGAGCTTCCAAATGGTATT	CTCGATGGCCGTGATATACA	
CG7840	PCR	GGCTGCATAGTGCCGAATCA	GAAATGTTTCCAGTCTTCGG	55
CG9135	PCR	CAGACCAGTCCAGTCCACAA	CAGTTCTGAGAATTTCCTAC	54
	Int1	CGCCAATTAACTACCGAGGT	AAAGTACTTGGCGTCAGTAT	
CG5363	PCR	TTTAAAAGTCGGTGGCTTGC	GCTTTAGATTAGATACCAAG	55*
	Int1	GAATTCCTATCGATGGACCT	CGATTAGTAAGTTCTGCGGC	
CG3831	PCR	TTTCCACCCTCTATCACCCAC	GGTCGCCAAGTTGTAATCTAG	53
	Int1	GTGTCGGCAGTGTTGGTATG	GGAAAACCCTCTCGATGAGG	
CG12276	PCR	TTGTTGCGGCGTCTCACTG	CAACTGGTGTTATTTGATATTCAG	53
	Int1	CCTGGGCGTCAAGTTCATAG	GCTGTGCTTCTGCAGACTGG	
CG17950	PCR	CAGTGTAACACTTGTTAAATAGTTTC	GCTGGTTGTGTGTGACTTGG	53
CG5499	PCR	GGCGCAGTTTACAGGCAG	CTGCTGAAATCTGAATTAAACATTAG	53
	Int1	TTCCGCCGAGTTTTCCCGC	CTGTACACAATAAGAAAAGTGG	
CG10206	PCR	GAAGGCACACGTGCTTGCAC	TTATATACCACGCCGTTGAC	54
	Int1	TGGCGTCCAGGAGCTGATGC	GATCTTGTCGGGCGAGAACT	
	Int2	AATGATGGCCATGGCTCCGA	CGGCGCCCAGAATTTGTACA	
	Int3	CGATAAACCTTGGCTGTCTG	GGTTGGTAGGTGAACACCTC	
CG4299	PCR	AAAATCCGCTTCGCTTCTCT	TATGTAAATGTTGCATAACC	54*

(Continues...)

Table S8. *(Cont.)*

Gene	Type	Forward primer (5' - 3')	Reverse primer (5' - 3')	Temp
	Int1	ACACATGTACAAGAAATTTT	CACGATTTTTCCGGGAGGGG	
	Int2	TGCTCACCAAAGAGTTCCAC	AAGCAGCTTTAATTGAAA	
CG4236	PCR	TGCACACGTTTTGCCGCTTC	GCCATTTAGGCTGGATAGCT	58*
	Int1	TCGAGATTAAGATCAACCAC	GACGGCGTCTTGGTGGCGAT	
	Int2	TCGGATCGGTGGCTGACGAC	CAGAATGAACTCCGAGTAGG	
CG5272	PCR	AGCAGTTAATTACAGTTAG	CATACAAATTCAAATCCACATG	53
CG5757	PCR	CAGATTTGTTCAGCCCTGGTC	CGTAAATTAAAAGTTAAGCGGC	53
CG12314	PCR	CGAATGGGCAAGCGCATGGC	CTGAGGAACTGTTTGGTAGTC	53
CG4570	PCR	CCGCCAATTGCAGCTTGCTCG	CTAATCAGAGTGTACCATGTAC	54
CG12909	PCR	CACGTGCGTTTTGTTTGGCCTC	TGTATGCAGTATCTAAAGTAC	53
CG6874	PCR	CGTTTGGAAGTTAAAATGCCAG	GGCTCCTGCCAAGCTAACATC	53
CG13690	PCR	AATGACGAATGAGTCACACTG	CACGAGATTGGCGGAAACAC	53
CG3509	PCR	CGTGTCTTAAAAGTTGTTGAG	ACAAAGTTTCACATGTACTG	53
CG17361	PCR	CGGGGTCTTTGCCAT T	GTCCACACCCGGACTCA	53
	Int1	ACATCTACGA CAACAAGAGC	CAATCCTCGGCCTCCAGGAG	
CG32409	PCR	CATTGTGGCCAAGCTGCCGG	TAAGCGGCACACGTCACCAG	55.5
	Int1	CCGATCCCAATACGGACATG	AGACCACGTGCCTCCTTCTC	
CG6459	PCR	GGAGGATCAAGAACCAGAGAA	GTGATGAAACTGCGACAGCA	56
	Int1	CACCCGCAGCCTGTGGCACA	GTGTGCACGTTGCAGCCGCA	
	Int2	AGCAAGCCCCAGTTCGA	AGACATCGTCTGAAAGGTAA	
CG3975	PCR	GTTGGCACGCCCGCAATACG	ATCAGACACGCGAATCTATA	54
	Int1	AATTGTCGAAATGTCGCTTA	AGATCGGTGACCAGCACA	
	Int2	TGGAGCAGCGTCCTGGAGCG	TCCTTGGGCGAGGTCTTCT	
	Int3	GACGAGGAGGAAAAGCTGGA	GTCCTCAGAGTCGCTGATTC	
Nonsex-biased				
CG11981	PCR	AGTCTGGCAACGCAGAGCAG	CGCACATGTTGTACCTAAAC	53
CG5919	PCR	GTTGCAAACAGTTCGCGCAG	GTCCGCACACAGACACACAC	53
CG5915	PCR	GAAATAACCCCGAGTAAGCC	GAATCTGAACTTTGCATCCTATC	53
CG9893	PCR	GACATCGATTGTTTTGCAGACTG	CCTTCAGCTTAAGTGGAACC	50
CG13189	PCR	GGTAAACAATTGCTGCGCAAC	GGCTGGTTCTGAGATTTGTAC	53
CG10853	PCR	AAACACCTGTGGGATCTGGAA	TAATGTGAGCACATTTTCGCC	54
CG6913	PCR	GCAACACAGTCTTCCAAGCAG	CTAGGACAAACTATGATTATAGG	54
CG9437	PCR	TAGCATTGTGTCGATGACAA	CTTTGTATATTTGTACTGCATT	53
	Int1	TCGCTATCGGCGGTGTCGATC	ACTATGTTAAAACCTATGTAC	
CG8392	PCR	CAAGCCAGTCATTTCGTGTTC	GCATGAGACGAGTTTGCAACTC	53
CG7953	PCR	CTTCTACGCGTTATCGTCTGC	CAATAATTGAATATATCGACAAG	53
	Int1	GATTTTGGTATATAAGACTCTG	GCAGGAAATGAGAGTCCAATGG	
CG13419	PCR	ATTTACGCTGTCGAGCGGCC	TGTGTACAAGCAGAAACGGC	56.5

(Continues...)

Table S8. *(Cont.)*

Gene	Type	Forward primer (5' - 3')	Reverse primer (5' - 3')	Temp
	Int1	GTCGCTGTGCCAGTTATA	GACTCCTGGCAGCACATGCA	
CG7484	PCR	CAGTCCGTATCGATTGTAGG	TGGCGCGTAGAAAAGCTC	53
	Int1	GTAAACAACGACGTGGACGT	AACCCACTTGATGGTATCCA	
	Int2	CTGGAGGTGTGCACCTGCAA	CTTACGTATTTGATCTG	
CG9283	PCR	CATTTGAGCGATGTTCACGC	TCGCTTGCAGTCAGTCAG	56.5
	Int1	ACATCCCTCATGGTGGATAT	TCCACCCATTGGTGGTGA	
CG13934	PCR	TTTCAACTGCTGCCGTAACA	ATTTTGCAATTTACGGCTGCG	54
	Int1	TTGTGTGTTTATCTACCA	CCACGTCGGAGATGTGGTAC	
CG17404	PCR	GCTAAGCTCTTATTTAAGAA	GGAACTGTGCCGCCATCT	53
	Int1	GTGAACTATCCCAATGTTCT	AGATCTCCGTATCCGTAA	
CG10623	PCR	ATATCACGTATCTGTCAGAT	TAGGTCTAGGTCTAGTCACA	52
	Int1	ATCATGGGCTCAATAGGTCC	TTACTCACCATGCACTGCAG	
CG3652	PCR	CGATAACCGTTATCGATGGGTG	ACGCTATGGAACGTGCACTG	53
CG9617	PCR	CATTTTATATCTGGTAGCACTGCC	CTCTGTTAATCTCTCTCGCTG	53
CG6981	PCR	GAAAAGATTGTGAGTGCCCC	GCGGATAATAACTAATTTGCAC	45
CG9822	PCR	ACCTGCAAATCAAAATTTCG	CTGAATTGGTAATTAGCAAC	53
	Int1	AGTACCTGGCCACGTTGAAC	ATGTAACTGCTGTCGATGAG	
CG13845	PCR	GGCCCTGTGCGTATATCAAT	ATGCTCTAGCAGATTGTACG	53
	Int1	CACACCTTCGGCATTCAGTA	CTTGCGCAGGTGCTGACGTA	
CG3683	PCR	CGACTGCCAGTTGTCGATGA	TACTGCGGATGCTGCCTGAA	53
	Int1	GCGAGCAGGCCAATAATGTT	TTCACCTTGCGGAAGAAGTC	
CG3476	PCR	TTATCATAGGTCAGCAACCG	TCGGATACCAATTATTGGCG	53
	Int1	ACAGTGCCCACCGATCGTAT	TCAGAATGCACGCGCATGTT	
CG6094	PCR	CAACTCACCTCGAACAGCTG	CAATGGCGTCCATAGCATAG	53
CG8844	PCR	GTATTTTCCGTCAAAGGAAAGCG	GGAATATCGTTTTACGGCGTG	53
CG7508	PCR	AGACCAGTCAGAAACCCGCC	GCGGTAATTCACTACTGGGC	53
CG16985	PCR	CGCCATGTGAAATTTCTCCC	GGGCAAATCGATATGCAAAGTC	53
CG11785	PCR	ACATCGGTCTTGTGCCAGCTC	CGTGGTAGTTGCATTTATTTCAG	51
CG10035	PCR	GGCTTTAAAAGAATCGATCGTG	GCATTCCACCTAATTTACGAGC	55
CG18553	PCR	GGTGTTTCAACCTAAAAGTGCG	GTAAATACTGCAGTCTCATGC	55

NOTE:
1. PCR primers were used for both PCR and sequencing.
2. For some genes, additional internal (Int) primers were used for sequencing.
3. PCR used 25 cycles of (95º for 1:00, Temp for 0:30, 72º for 1:20), where "Temp" represents the annealing temperature given below for each primer.
4. Temps with an asterisk indicate that 6 "touch down" cycles were run prior to the above program, starting with a temp 3º above the given temp and reducing it by 0.5º each cycle.

Curriculum Vitae

Dr. Zhi Zhang

Address: Home
App. K30, Connollystr. 7, Munich
Germany 80809
Phone number:
+49-017622629441 (private)
+49-89 2180 74109 (work)
E-Mail Address:
zhang@zi.biologie.uni-muenchen.de
http://www.zi.biologie.uni-muenchen.de/evol/EvoGen.html

EDUCATION

2002-present: University of Munich, Munich, Germany, Ph.D. student

1990-1997: Bethune University of Medical Science, China, Master's Degree in Medicine

WORK EXPERIENCE

1997-2002: PuNan Hospital of ShangHai, China, Physician

SKILLS

- Techniques of molecular biology: PCR, TA-clone, gel eletrophoresis, ABI sequencing.
- Techniques of functional genomics: microarray design, lab work (mRNA extracting, cDNA synthesis, labelling, hybridization, scanning of arrays) and analysis (image processing, normalization, detecting of different gene expression, ontology analysis of gene expression).
- Standard Software: Word, Endnote, Excel, PowerPoint, Internet, Mac and PC.
- Analysis of population genetics: DNasp, Seqman, BLAST, Sequencer, Sequence analysis.
- Bioinformatics tools: perl, bio-perl, mysql, php, database design, bioconductor.

LANGUAGES

Chinese (Native language), English (Fluent).

REFERENCES

1. Prof. Dr. John Parsch
Department Biologie II
University of Munich (LMU)
Grosshaderner Str. 2
82152 Planegg-Martinsried
Germany
Tel: +49 89 2180 74 107
Fax: +49 89 2180 74 104
parsch@zi.biologie.uni-muenchen.de

2. PD Dr. Beate Nurnberger
Department Biologie II
University of Munich (LMU)
Grosshaderner Str. 2
82152 Planegg-Martinsried
Germany
Tel: +49 89 2180 74205
Fax: +49 89 2180 74204
nurnbb@zi.biologie.uni-muenchen.de

3. Dr. John Baines
Department Biologie II
University of Munich (LMU)
Grosshaderner Str. 2
82152 Planegg-Martinsried
Germany
Tel: +49 89 2180 74 110
Fax: +49 89 2180 74 104
baines@zi.biologie.uni-muenchen.de

ANNEX

PUBLICATIONS

1. Proeschel, M., **Z. Zhang**, and J. Parsch. 2006. Widespread adaptive evolution of Drosophila genes with sex-biased expression. Genetics. (in press)
2. **Zhang, Z.**, and J. Parsch. 2005. Positive correlation between evolutionary rate and recombination rate in Drosophila genes with male-biased expression. Mol Biol Evol 22:1945-1947.
3. **Zhang, Z.**, T. M. Hambuch, and J. Parsch. 2004. Molecular evolution of sex-biased genes in Drosophila. Mol Biol Evol 21:2130-2139.

POSTER PRESENTATIONS:

1. Molecular evolution of sex biased genes in Drosophila, **Zhi Zhang** and John Parsch, poster at 10th Congress of ESEB, held in Krakow Poland in August 2005.

Acknowledgements:

Upon the successful completion of this dissertation, I look back with a great sense of gratitude. This thesis is the result of four years of work, whereby I have been accompanied and supported by many people. I am pleased to have the opportunity to express my gratitude to all of them.

The first person I would like to thank is my academic supervisor Prof. Dr. John Parsch for his continuous supervision, encouragement, constructive comments, continual patience and confidence in me and my work. I would like to express my gratitude to Dr. Tina Hambuch and Dr. John Baines for the instruction for my lab work. Thanks to my collaborator Matthias Proeschel for his nice PCR and sequencing work, which helped me a lot for my thesis. Thanks to Winfied Hense for his warm-hearted and kind help for my questions on German language. Thanks to Sarah Peter and Xiao Liu for the lively chat. A special thank to PD Dr. Beate Nürnberger and Dr. Laura Rose for their helpful comments and suggestions in my committee meetings.

I give my thanks to all my colleagues: Wolfgang Stephan, David De Lorenzo, Joachim Hermisson, Pleuni Pennings, Haipeng Li, Peter Pfaffelhuber, Sylvain Mousset, Thomas Städler, Kerstin Roselius, Aparup Das, Ying Chen, Ann Arunyawat, Nina Stoletzki, Aura Navarro-Quezada, Jose Alvarez-Castro, Daven Pregraves, Michael Kopp and Lukasz Grzeskowiak: thank you for the good working atmosphere. I'm very grateful to all for all their cooperation and help.

I also express a lot of thanks to our super technicians: Hedwig Gebhart and Yvonne Cämmerer for their excellent work.

Many thanks go to my friends, especially the friends in Munich like

Ming Zhao and also to all the other Chinese friends and German friends I knew in Munich and all the members in my wife's group for their friendship that let me feel in Munich as at home.

I express my heartfelt gratitude to my family in China, to my parents for continual love and support and encouragement, to my sister, to my uncles and aunts, to other relatives, and to my wife's family.

I would like to give my special thanks to my wife for her love, support, encouragement and patience during the Ph. D. period.

Finally, I will dedicate this thesis to my son for the pleasure he gives to me during the time in Munich.