The influence of sex on gene expression and protein evolution in *Drosophila melanogaster*

Lena Müller

München 2012

The influence of sex on gene expression and protein evolution in *Drosophila melanogaster*

Lena Müller

Dissertation
an der Fakultät der Biologie
der Ludwig-Maximilians-Universität
München

vorgelegt von
Lena Müller

München 2012

Erstgutachter:	Prof. John Parsch
Zweitgutachter:	Prof. Wolfgang Stephan
Tag der Abgabe:	01.02.2012
Tag der mündlichen Prüfung:	02.04.2012

ERKLÄRUNG

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

EIDESSTATTLICHE VERSICHERUNG

Ich versichere ferner hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt worden ist.

München, 31.01.2012 _____

DECLARATION OF AUTHOR'S CONTRIBUTION

In this thesis, I present the results of my doctoral studies conducted from November 2008 until January 2012. The results are shown in three chapters, all of which are the product of collaborations with other scientists. The work of this doctoral thesis has resulted in two publications. They constitute chapters 1 and 3 of this dissertation and are supplemented by appendices A and C. Chapter 2 is an unpublished manuscript.

In CHAPTER 1, I performed the microarray and qRT-PCR experiments. Stephan Hutter, Rayna Stamboliyska, and John Parsch analyzed the microarray data. Sarah Saminadin-Peter and myself performed the population genetic analysis of DNA sequences. Wolfgang Stephan and John Parsch conceived of the study, and participated in its design and coordination. I drafted the manuscript with assistance from John Parsch and input from all authors. This chapter has been published:

Müller, L., Hutter, S., Stamboliyska, R., Saminadin-Peter, S. S., Stephan, W., and Parsch, J. (2011). Population transcriptomics of *Drosophila melanogaster* females. BMC Genomics *12*, 81.

In CHAPTER 2, I chose the candidate genes that were subject to population genetic analyses. I performed all of the sequencing except for a 1 kb-fragment of the gene *Nipsnap*, which was sequenced by Korbinian von Heckel. I performed the population genetic analysis of DNA sequences and wrote the manuscript.

In CHAPTER 3, I performed the analyses of the *D. melanogaster* gene set, while the *D. ananassae* genes set was analyzed by Sonja Grath. Four of the newly sequenced genes of the *D. melanogaster* gene set were sequenced by myself, and two genes by Korbinian von Heckel. The manuscript was drafted by myself, Sonja Grath and John Parsch. This chapter has been published:

Müller, L., Grath, S., von Heckel, K., and Parsch, J. (2012). Inter- and Intraspecific Variation in *Drosophila* Genes with Sex-Biased Expression. International Journal of Evolutionary Biology *2012*, 1–10.

Lena Müller	John Parsch

Nothing in Biology Makes Sense Except in the Light of Evolution
Theodosius Dobzhansky

Meinen Eltern.

SUMMARY

A long-standing question in evolutionary biology concerns the molecular causes underlying adaptive evolution. These can either stem from structural changes in proteins or from changes in the expression patterns of proteins or mature RNAs. Over the last decade, many studies have shown that gene expression changes can have a huge impact on the phenotype of an organism and play an important role in adaptive evolution. A major prerequisite for adaptive evolution to occur at the gene expression level is the presence of expression variation among members of a population. This variation serves as the raw material for adaptive evolution.

The genetic causes underlying changes in expression patterns can either be located in *cis*-regulatory regions of the affected gene, such as transcription factor binding sites, or in *trans*-regulatory regions, such as transcription factors. Mutations in *cis*-regulatory elements have relatively few pleiotropic effects and their effects are often additive, thus, *cis*-regulatory changes are thought to be especially well-suited targets of selection.

A major factor influencing gene expression is the sex of an organism. The sex-bias of a gene also influences the pace at which proteins evolve, such that male-biased genes often show more rapid evolution than female-biased or unbiased genes between *Drosophila* species.

Here, we investigated genome-wide gene expression variation in adult females of two populations of *D. melanogaster*, one from the ancestral species range (Zimbabwe) and one from the derived species range (the Netherlands). We found relatively little expression polymorphism present within the populations and high expression divergence between the populations. More than 500 genes were expressed differentially between the populations. These are candidate genes for those that have undergone adaptive regulatory evolution to the new, derived environment. When comparing our study of female adults to a study investigating male adult flies of the same populations, we found that there is significantly less

expression polymorphism in females within the populations but significantly more expression divergence between the populations. Further, there was little overlap in genes that differ in expression between the populations in males and females. This suggests that general differences exist between the sexes in gene expression regulation and that regulatory evolution has been mainly sex-specific. Our findings show that extensive gene expression variation exists in *D. melanogaster* and further highlight the importance of accounting for sex when investigating gene expression.

In order to elucidate the genetic and evolutionary mechanisms that underlie differential gene expression between the populations, we employed a candidate gene approach. Analysis of molecular variation in the coding and upstream regions of several differentially expressed genes in both populations revealed evidence for a recent selective sweep in the European population for the gene *CG34330*. In the putative promoter region of the gene, there is one indel and one SNP where a derived variant is fixed in the European population, but at low frequency in the African population. These are candidates for those variants that control the expression level of the gene. For another gene, *Jon99Ciii*, we found evidence for recurrent structural protein evolution acting since the split of *D. melanogaster* from *D. simulans* and *D. sechellia*. However, no evidence for recent regulatory evolution could be found for this gene.

Motivated by findings that male-biased genes often evolve faster than both female- and unbiased genes between *Drosophila* species, we examined the molecular evolution of sex-biased genes and their contribution to within-population polymorphism, between-population divergence and between-species divergence in *D. melanogaster* and *D. ananassae*. This was studied on both the DNA-sequence level and the expression level. We found strong purifying selection limiting protein sequence variation within species. In contrast, a high proportion of divergence could be attributed to positive selection. In *D. melanogaster*, male-biased genes showed the highest fraction of adaptive substitutions, a pattern that was especially pronounced on the X chromosome. In contrast, male-biased genes did not show higher variation within or between populations, suggesting that inter-species divergence is not just a simple extension of inter-population divergence and intra-population variation. For *D. ananassae*, we did not observe a higher rate of adaptive evolution for male-biased genes, a finding that suggests that the type or strength of selection acting on sex-biased genes differs between lineages. Similarly, on the expression level, we found that sex-biased genes show high expression divergence between species, but low divergence between populations.

ZUSAMMENFASSUNG

Zentrale Fragen der Evolutionsbiologie beschäftigen sich mit den molekularen Ursachen adaptiver Evolution. Diese können entweder in strukturellen Veränderungen von Proteinen liegen, oder in Veränderungen der Expressionsmuster von Proteinen oder reifen RNA-Molekülen. Neuere Studien haben gezeigt, dass Änderungen der Genexpression starke Auswirkungen auf den Phänotyp eines Organismus haben können und darüber hinaus eine wichtige Rolle in adaptiven Prozessen spielen.

Eine wichtige Grundvoraussetzung, um adaptive Evolution auf dem Genexpressionslevel zu ermöglichen, ist das Vorhandensein von Genexpressionsvariation zwischen Individuen einer Population.

Die genetischen Ursachen für Änderungen von Expressionsmustern können auf verschiedenen Mechanismen basieren: Mutationen in *cis*-regulierenden Regionen, wie z.B in Transkriptionsfaktor-Bindestellen, oder Mutationen in *trans*-Faktoren, wie z.B in Transkriptionsfaktoren. Zwei Gründe machen letztere zu besonders guten Zielen von adaptiver Selektion: Mutationen in *cis*-regulatorischen Bereichen haben relativ geringe pleiotrope Effekte und sind häufig additiv in ihren Auswirkungen.

Das Geschlecht eines Organismus beeinflusst die Expression eines Genes in hohem Maße. Auch die Geschwindigkeit, mit welcher Proteine evolvieren, wird davon beeinflusst, ob ein Gen überwiegend in Männchen oder Weibchen exprimiert wird. So weisen überwiegend in Männchen exprimierte (*male-biased*) Gene häufig eine schnellere Proteinevolution auf als Gene, die überwiegend in Weibchen exprimiert werden (*female-biased*) oder Gene, die nicht geschlechtsspezifsch exprimiert werden (*unbiased*).

In der vorliegenden Arbeit untersuchten wir mittels genomweiter Microarrays, wie stark die Genexpression in adulten Weibchen zweier Populationen von *Drosophila melanogaster* variiert, wobei eine der beiden Populationen aus dem ursprünglichen afrikanischen Verbreitungsgebiet der Fliege stammt (Zimbabwe) und die andere aus dem neu erschlossenen Verbreitungsgebiet in Europa (Niederlande). Innerhalb der Populationen fanden wir relativ gering variierende Expressionshöhen der Gene, während zwischen den beiden Populationen

eine starke Divergenz der Expressionshöhen gefunden wurde; hier wiesen über 500 Gene Unterschiede in ihrer Expression auf. Die unterschiedliche Expressionshöhe dieser Gene könnte durch Adaption auf regulatorischer Ebene an die neue, abgeleitete Umwelt, verursacht worden sein könnte.

Vergleicht man die Ergebnisse unserer Studie mit einer vorausgegangenen Studie, welche adulte Männchen derselben Populationen untersucht hat, so fällt auf, dass signifikant weniger Expressionsvariation innerhalb der Populationen in Weibchen als in Männchen vorhanden ist. Dagegen fand sich in Weibchen zwischen den Populationen eine signifikant größere Divergenz der Expressionshöhen. Darüber hinaus gab es sehr geringe Überschneidung zwischen den differenziell exprimierten Genen in Weibchen und Männchen. Dies impliziert, dass allgemeine Unterschiede zwischen den Geschlechtern in Bezug auf die Expressionsregulation bestehen und dass regulatorische Evolution hauptsächlich geschlechtsspezifisch erfolgt ist.

Unsere Ergebnisse zeigen, dass eine große Variationsbreite innerhalb der Genexpression in *D. melanogaster* vorhanden ist und verdeutlichen, dass das Geschlecht eines Organismus einen starken Einfluss auf die Genexpression hat.

Um die der differenziellen Genexpression zugrunde liegenden genetischen und evolutionären Mechanismen aufzuklären, benutzten wir einen Kandidatengen-Ansatz. Die Analyse von DNS Variation in kodierenden und stromaufwärts gelegenen Regionen mehrerer differentiell exprimierter Gene zeigte für das Gen *CG34330* Hinweise auf einen kürzlich zurückliegenden *selective sweep* in der europäischen Population. Innerhalb der mutmaßlichen Promoterregion des Gens fanden wir ein Indel und einen SNP, wo eine abgeleitete Nukleotidvariante innerhalb der europäischen Population fixiert ist, diese hingegen in der afrikanischen Population in niedriger Frequenz vorkommt. Diese Stellen könnten also die Expression des Gens beeinflussen. Für das Gen *Jon99Ciii* haben wir Hinweise auf strukturelle, adaptive Proteinevolution gefunden, welche seit der Trennung von *D. melanogaster* von *D. simulans* und *D. sechellia* gewirkt hat.

Motiviert durch die Tatsache, dass spezifisch männlich exprimierte Gene zwischen verschiedenen Arten von *Drosophila* oftmals schneller evolvieren als spezifisch weiblich exprimierte Gene und Gene, die nicht geschlechtsspezifisch exprimiert werden, untersuchten wir die molekulare Evolution von geschlechtsspezifisch exprimierten Genen. In *D. melanogaster* und *D. ananassae* analysierten wir den Beitrag von geschlechtsspezifisch

exprimierten Genen zum Polymorphismuslevel innerhalb von Populationen, sowie deren Beitrag zur Divergenz zwischen Populationen und Arten. Dies untersuchten wir sowohl auf DNA-Ebene, als auch auf Expressions-Ebene. Wir fanden, dass starke negative Selektion Proteinsequenz-Variation innerhalb der Arten limitiert. Dagegen konnte ein großer Anteil der auf DNS Ebene beobachteten Divergenz auf positive Selektion zurückgeführt werden. In *D. melanogaster* zeigen *male-biased* Gene die meisten adaptiven Substitutionen. Dieses Muster ist auf dem X Chromosom besonders stark ausgeprägt. Im Gegensatz dazu zeigten *male-biased* Gene keine erhöhte Variation innerhalb oder zwischen den Populationen. Dies deutet darauf hin, dass die Divergenz zwischen Arten keine unmittelbare Fortführung von der Divergenz zwischen Populationen oder der Variation innerhalb von Populationen ist. Auch auf dem Genexpressionslevel fanden wir, dass geschlechtsspezifisch exprimierte Gene starke Divergenz zwischen den Arten aufweisen, aber geringe Divergenz zwischen den Populationen.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GENERAL INTRODUCTION

1.1 EVOLUTION AND GENETICS

In an abstract of his 1859 seminal work *On the Origin of Species*, Charles Darwin was the first to publish a theory of evolution driven by natural selection (Darwin 1859). Darwin's theory was based on the following observations that he had made during his 1831-1836 trip on board of the HMS Beagle: that individuals in species are variable; that at least part of this variability is heritable; that in each species, many more individuals are born than can survive which, as a consequence, elicits a "struggle for existence". It follows that individuals that have a better chance of survival, for any reason, will be selected for in nature. Since the advantageous trait can be passed on to the offspring, over time populations will adapt to the environment they are living in. Consequently, on a broader time scale, these same mechanisms can finally lead to the formation of new species. In his theory, Darwin recognized the importance of heritable variation being present in a population, which serves as the raw material for adaptive evolution.

Although Darwin's observations and ideas were conclusive, Darwin's theory suffered from one drawback: he could not explain how heritable information could be passed on from one generation to the next and what mechanisms could generate and maintain heritable variation. During the 19th century, the concept of blending inheritance was widespread among scientists. This concept states that the phenotypes of the parents blend to form an intermediate phenotype in the offspring. Thus, with each generation, variation would be removed from the population, resulting in equalization of all individuals. This in turn implies that no individual will have an advantage over another, such that natural selection would lack the basis to work on. In 1856, Gregor Mendel began his now famous crossing experiments with peas. His findings showed that traits are inherited as discrete factors that are passed on to the offspring in an unchanged form, so that variation is retained over generations (Mendel 1865). Ironically, Mendel's findings did not gain a lot of attention by the scientific community at that time and Charles Darwin was not aware of them. Darwin knew about the flaws of his theory: "Although much remains obscure, and will long remain obscure, I can entertain no doubt,

after the most deliberate study and dispassionate judgment of which I am capable, that the view which most naturalists entertain, and which I formerly entertained – namely, that each species has been independently created – is erroneous" (Darwin, 1859).

It was only in 1928 that DNA was proposed to be the carrier of heritable information (Griffith 1928), a finding proved in 1944 by Avery and colleagues (Avery *et al*. 1944). In 1952, an experiment by Hershey and Chase finally excluded proteins as hereditary material, a belief that was common at that time (Hershey and Chase 1952). In 1953, the unraveling of DNA's chemical structure (Watson and Crick 1953), along with the deciphering of the genetic code during the 1960s (reviewed in Nirenberg 2004) lay the foundation for modern genetics. Today it is clear that mutations occurring in the DNA of reproductive cells produce the variation that is necessary for adaptive evolution.

With these discoveries at hand, scientists today are able to investigate how and why DNA changes its informational content through time and how such changes correlate with changes in the makeup (the phenotype) of organisms. For example, why do individuals of the same species show differences in their outer appearance or why can some individuals cope better with certain environmental conditions and others not? What are the genetic differences between different species?

1.2 GENOMES AND GENETIC VARIATION

Modern techniques have enabled the rapid sequencing of genomic regions and – boosted by next generation sequencing techniques – even whole genomes. This genetic data represents a rich source to explore the function of different genomic regions and the genetic variation present within populations as well as between populations or species.

One of the first genomes to be sequenced was the genome of the fruitfly *Drosophila melanogaster*, the first draft of which was published in 2000 (Adams *et al.* 2000). The genome is about 180 megabases (Mb) in size, 120 Mb of which are euchromatic. It contains around 14,000 protein-coding genes. Most mutations arising in coding regions of the genome are recessive (Li 1997), implicating that natural selection can only effectively act on them after they have been raised to higher frequencies by drift. Once they appear in homozygotes, these mutations can render genes nonfunctional or lead to an altered amino acid sequence of

the affected protein, thus changing the structure or function of the protein. Since most proteins are involved in multiple cellular processes (Tomancak *et al.* 2002, Ohya *et al*. 2005), *i.e*. they are pleiotropic, mutations that change the characteristics of a protein most probably impact some or all of the traits that are influenced by this gene. This means that if a mutation in a gene is advantageous in one trait, it could still be disadvantageous in another, thus inhibiting the selection of this mutation. Therefore, evolution of proteins can be hampered by the pleiotropic effects they may be exerting (reviewed in Stern and Orgogozo 2008).

The biggest part of the genome of *Drosophila melanogaster*, though, does not encode proteins. Around 80% of the euchromatic DNA is non-protein-coding DNA (non-coding DNA). Although originally erroneously thought to be devoid of any function and consequently termed as 'junk-DNA', non-coding DNA is far from being useless: It is of functional importance for DNA replication, chromosome packaging, DNA secondary structure, and gene expression regulation.

1.3 REGULATION OF GENE EXPRESSION

How can non-coding DNA regulate gene expression? A gene *per se* is an inactive piece of DNA that stores information for the gene product. The majority of gene products are proteins, but they can also be mature RNA molecules (such as transfer RNA, ribosomal RNA, and microRNA). The process by which the gene information gets processed into the gene product is called gene expression and consists of two steps: The first step is transcription of the coding sequence DNA into messenger RNA (mRNA). In a next step, the mRNA gets translated into a chain of amino acids that together form a protein (in case the gene is a protein-coding gene). The expression of a gene can be regulated at different levels. These include chromatin state, transcriptional initiation, alternative RNA splicing, mRNA stability, control of translation, post-translational modification, and protein degradation (Wray *et al.* 2003). The most common way, though, is regulation at the initiation of transcription (Wray *et al.* 2003). This can be accomplished either by *cis*-regulatory or by *trans*-regulatory factors.

Cis-regulatory elements, such as transcription factor binding sites, lie in non-coding DNA regions and directly regulate the expression of genes lying on the same DNA strand. They can be located in enhancers, promoters, 5'UTRs, 3'UTRS or introns. In contrast, *trans*-regulatory elements are factors that are encoded elsewhere in the genome and interact with *cis*-regulatory

elements, such as transcription-factors or microRNAs. Transcription initiation of a gene is regulated in the promoter region: in eukaryotes, it consists of a core promoter region, where the basal transcriptional machinery (RNA polymerase II and general transcription factors) assembles. Nevertheless, the core promoter is not a common point of gene expression regulation. In addition to the core promoter, multiple transcription factor binding sites (TFBS) are spread in a module-like fashion along the promoter region of a gene, which facilitate fine-tuning of expression regulation (Figure 1.1). These are termed enhancers. A transcription factor (TF) bound to an enhancer can interact with the basal transcriptional machinery, thereby facilitating or inhibiting its association with the core promoter, resulting in an increase or decrease in transcription rates (reviewed in Wray *et al.* 2003). Mutations that occur in *cis*- or *trans* factors can lead to an altered expression of the gene concerning timing,

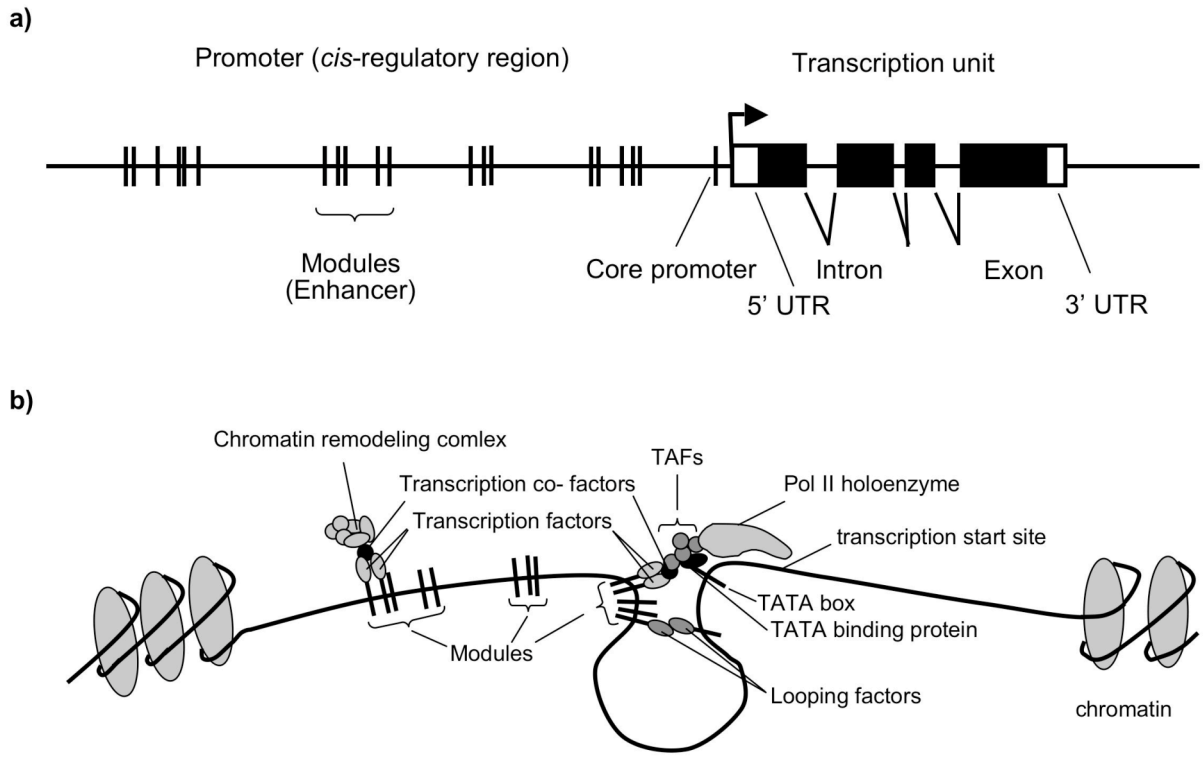


FIGURE 1.1 Promoter structure and function. (a) Organization of a generalized eukaryotic gene. Shown are the relative position of the transcription unit, basal promoter region (black box with bent arrow), and transcription factor binding sites (vertical bars). (b) Idealized promoter in operation. Many different proteins are required for initiating transcription, such as the RNA polymerase II holoenzyme complex (~15 proteins); TATA-binding protein (TBP; 1 protein); TAFs (TBP-associated factors, ~8 proteins); transcription factors (composition and number differs among loci and varies in space and time and according to environmental conditions); transcription cofactors; and chromatin remodeling complexes. (Adapted from Wray *et al*. 2003).

location or level of expression. Often, clusters of nearby TFBS operate as functionally coherent modules that control one single aspect of the expression profile (such as timing of expression or tissue-specificity) and act independently of other such modules (*e.g.*, DiLeone 1998, Jeong *et al*. 2008). Due to this modularity of *cis*-regulatory elements, mutations in these elements are predicted to be less pleiotropic than mutations in protein-coding regions and, therefore, their occurrence is thought to be less constraint. In addition, most mutations in *cis*-acting regions seem to be co-dominant in diploid organisms, meaning that each allele is transcribed independently (reviewed in Wray 2007). This has the advantage that natural selection can operate more efficiently on these mutations, since each new variant is immediately visible to natural selection in heterozygotes. Recessive mutations, in contrast, which make up the most part of coding mutations, first have to drift to higher frequencies in the population, until they appear as homozygotes, before selection can act on them.

Among the first authors to suggest that gene expression constitutes an important part in phenotypic change of organisms were Jacob and Monod in 1961 (Jacob and Monod 1961) and the first empirical evidence for this was provided in 1962 by Schwartz (Schwartz 1962). In 1975, King and Wilson (King and Wilson 1975) suggested that, due to the similarity of proteins and nucleic acids between humans and chimpanzees, the extensive phenotypic differences between the two species must be based on regulatory mutations. Ever since, more and more studies have investigated gene expression not only of single genes but have extended research to a genome-wide level. These studies were enabled to a large extent by the advent of microarray technologies during the 1990s (Figure 1.2). To date, numerous studies have shown that variation in gene expression is abundant in natural populations and also strongly contributes to divergence between species, *e.g.* in yeast (Cavalieri *et al.* 2000, Townsend *et al.* 2003, Fay *et al.* 2004), in fish (Oleksiak *et al.* 2002, Aubin-Horth *et al.* 2005, Whitehead and Crawford 2006), and in hominids (Enard *et al.* 2002, Stranger *et al.* 2005, Spielman *et al.* 2007, Storey *et al.* 2007). Gene expression has also been shown to be a heritable trait (Brem *et al.* 2002, Schadt *et al.* 2003, Monks *et al.* 2004). The abundance of expression variation in natural populations, together with the modularity of *cis*-acting elements, and the prevalence of co-dominant mutations in *cis*-regulatory regions make these especially well-suited targets for natural selection. Indeed, many studies have shown that adaptive evolution of gene expression is extensive in organisms and that mutations in promoter regions play an important role in adaptive evolution (*e.g.*, Tournamille *et al.* 1995,

Hamblin and Di Rienzo 2000, Gompel *et al.* 2005, Prud'homme *et al.* 2006, Wray *et al.* 2007, Fraser *et al.* 2010, Chan *et al.* 2010).

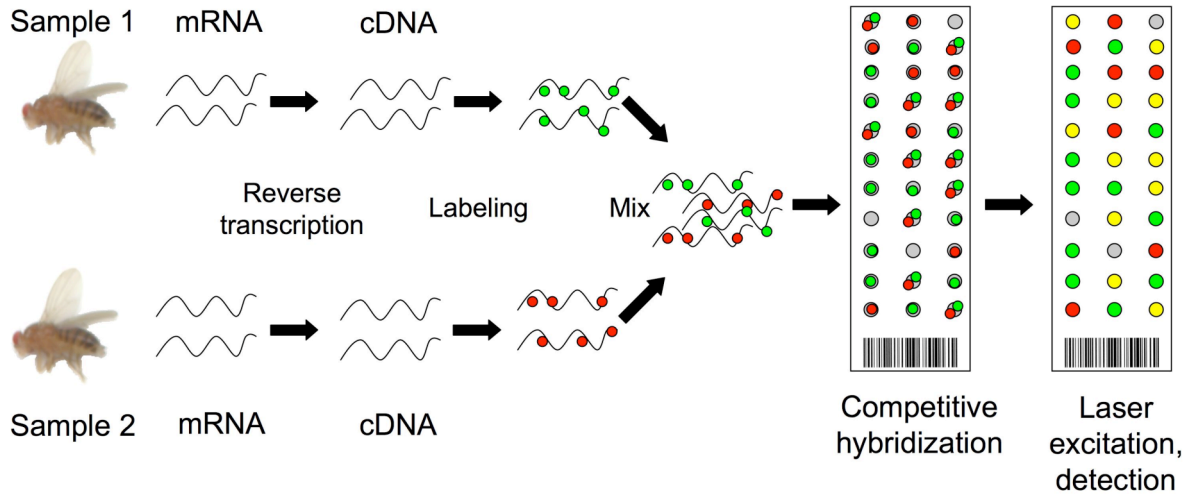


FIGURE 1.2 Principle of two-color microarray technology. mRNA is extracted from two samples, reverse transcribed into cDNA, and labeled with fluorescent dyes of two different colors. Samples are mixed and competitively hybridized onto the microarray slide, which features single stranded DNA probes that represent the *D. melanogaster* genome. Samples bind to their complementary probes, according to their frequency. Microarrays are exposed to laser excitation and fluorescence is detected.

Why is it important to be able to modulate gene expression patterns? Nearly all cells in a multicellular organism are genetically identical, irrespective of the function they are exerting or the tissue they are part of, which implies that most phenotypic differences between them are caused by differential regulation of gene expression. The same is true for male and female individuals of one species. Apart from the gene-poor sex-specific chromosome (*e.g.* the Y chromosome that is unique to males in humans and *Drosophila*), the genetic make-up of females and males is the same, nonetheless producing very different phenotypic outcomes. The key to producing different phenotypes from the same genetic background is the differential regulation of gene expression. In *Drosophila melanogaster*, up to 57% of all genes that are expressed in adult flies are expressed in a sex-specific manner (Jin 2001, Parisi *et al.* 2003, Ranz *et al.* 2003, Gibson *et al.* 2004, Ellegren and Parsch 2007) meaning they have a higher or exclusive expression in one sex. These genes are called sex-biased genes. Male-biased genes are higher expressed in males, whereas female-biased genes have a higher expression in females. Genes whose expression is equal in both sexes are termed unbiased. In

D. melanogaster, the majority of sex-biased genes are expressed in reproductive tissues (Parisi *et al.* 2003).

1.4 DEMOGRAPHIC HISTORY OF *DROSOPHILA MELANOGASTER*

Drosophila melanogaster has been used as a model organism in genetics research since the early 1900s. Several characteristics of the fly made early researchers, such as Thomas Hunt Morgan, select it as their model of research: A short lifecycle of about two weeks, small size, and a simple diet. In other words, they are easy and inexpensive to maintain in the laboratory. Over the years, more and more aspects of *Drosophila* genetics and development have been investigated, culminating in the sequencing of its genome in the year 2000 as one of the first complete available genome sequences. The thorough annotation of the genome, together with the availability of extensive genetic tools have made *D. melanogaster* also a valuable model organism for population geneticists. In 1988, two studies suggested that the origin of *D. melanogaster* lies in sub-Saharan Africa (David and Capy 1988, Lachaise *et al.* 1988), from where it expanded its species range and colonized Europe around 10,000-15,000 years ago, at the end of the last glaciation. This finding has been confirmed by extensive studies of microsatellites (*e.g.*, Kauer *et al.* 2002) and single nucleotide polymorphism (SNP) (*e.g.*, Ometto *et al.* 2005, Shapiro *et al.* 2007). There is evidence for a population bottleneck in populations residing in derived species ranges (Orengo and Aguade 2004, Ometto *et al.* 2005, Li and Stephan 2006, Thornton and Andolfatto 2006), which coincides with the colonization of these habitats. The out-of-Africa movement of this subtropical species presumably was accompanied by adaptation to the new European environment, where the flies were exposed to new conditions such as different temperatures, food sources, and pathogens.

The rediscovery of Mendel's studies (1856-1863) in the beginning of the 20th century allowed for a mathematical treatment of allele frequencies in populations, creating the field of population genetics. Mathematical frameworks were developed to allow the study of the forces that influence allele frequencies: mutation, natural selection, demography, drift, and recombination. In 1968 Kimura developed the neutral theory of molecular evolution (Kimura 1968), stating that most of the variation present in natural populations does not significantly affect the fitness of an organism and, therefore, the frequencies of the mutations are not determined by natural selection. In this theory, most mutations that arise in a population are

deleterious and thus purged from the population quickly. The remaining variation is selectively (nearly) neutral and will eventually be lost or driven to fixation by random processes, *i.e.* genetic drift: Since in populations of finite size, not all variation can be passed on from one generation to the next, eventually some of the variation will be randomly lost, and some will randomly go to fixation. This means that under the neutral theory, variation in a population depends only on its effective population size and the rate at which neutral mutations are introduced into the genome. Although there is ongoing debate regarding the extent to which selection influences the patterns of variation in natural populations, the predictions of the neutral theory provide a useful tool for population geneticists: the neutral theory can be used as a null hypothesis. Different tests of neutrality have been developed, which, when applied to empirical data can be used to either accept or discard neutral forces as causes for the observed molecular variation. When neutrality is rejected, other forces must have shaped the observed pattern of variation, and these can be of selective or demographic nature.

1.5 DETECTING SELECTION FROM DNA SEQUENCE DATA

How can selective events be detected from DNA sequence data? Selective events leave certain signatures in the genome. If by chance a mutation that confers a selective advantage arises in a population, this advantageous allele will spread through the population. Variation lying in the same chromosome that is physically linked to the mutation will be dragged along to fixation. This process is called genetic hitchhiking. Eventually all individuals of the population will be carriers of the advantageous allele as well as the polymorphism that hitchhiked along with it, resulting in a DNA region depleted of variation in the population (a selective sweep) (Maynard Smith and Haigh 1974) (Figure 1.3). Over time, some variation can be re-introduced into the region, which is characterized by an excess of mutations at low frequency. A selective sweep affects different properties of nucleotide variation: First, it causes a severe reduction of variation, second, extensive Linkage Disequilibrium (LD), third, a skew in the frequency spectrum to an excess of low-frequency polymorphisms.

Several statistics have been developed that summarize the level of polymorphism within DNA sequence sets. Two widely used statistics are π (Tajima 1983), which is the average number of pairwise differences between two sequences and θ (Watterson 1975), the number of segregating sites among the sequences. Tests are also available to look for an excess of low frequency polymorphism. Tajima's *D* (Tajima 1989), for example, compares the two

estimates for nucleotide variation, π and θ. Since π is influenced to a lesser extent by low-frequency variants than θ, their difference is negative when a recent selective sweep has taken place. In the presence of recombination, some variation is expected to remain in the region affected by the selective sweep, which is most likely segregating at low or high frequencies. Fay and Wu's *H* (Fay and Wu 2000) tests this prediction, by looking for an excess of high-frequency derived variants. A different approach to look for sites of potential selection in genomes is to look at genetic differentiation between populations. If an advantageous mutation spreads through one population, but not the other, one can look for regions of high population differentiation. A commonly used statistic is F_{ST}, the proportion of genetic diversity due to allele frequency differences among populations (Hudson *et al.* 1992). Another one is D_{XY}, the average number of pairwise sequence differences between alleles of the two populations (Nei 1987). It has to be noted, though, that all neutrality tests that exclusively rely on within-population data are susceptible to deviations from equilibrium demography. Population expansion, for example, can lead to an excess of low frequency variants just like positive selection would. This means that demographic causes must be excluded before one is able to infer selection.

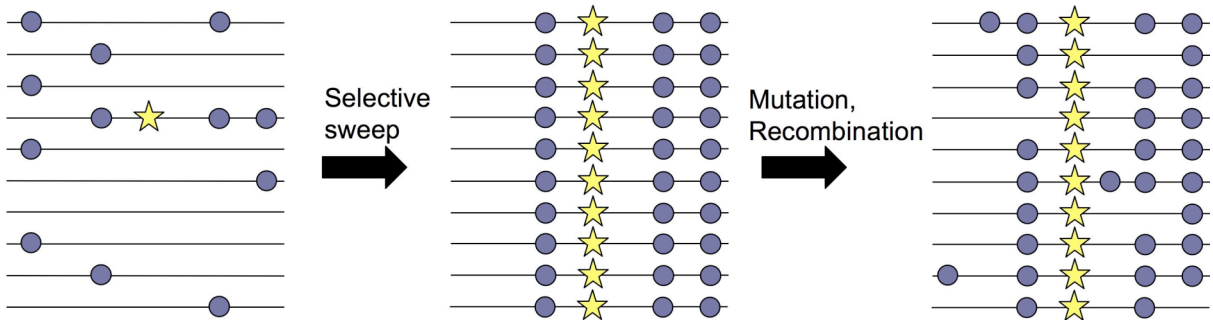


FIGURE 1.3 Selective sweep. When a beneficial mutation (indicated by star shape) is positively selected, the linked neutral variation (indicated by circles) of that allele hitchhikes along with the selected site. After completion of the selective sweep, new variation can be introduced again by mutation and recombination. This results in an excess of rare nucleotide variants. With time, this signal will be lost again due to accumulation of mutations and recombination breaking up the linkage between the selected site and neutral variants.

Other approaches to test for selective events are based on comparative approaches between species. The McDonald-Kreitman (MK) test, for example, compares within-species polymorphism (P) to between-species divergence (D) for synonymous (s) and

nonsynonymous (n) sites (McDonald and Kreitman 1991). Synonymous sites are assumed to evolve neutrally, while nonsynonymous sites are putative targets of selection. Under neutrality, the ratio of polymorphism to divergence is expected to be equal for both classes of sites, such that $P_s/P_n = D_s/D_n$. Deviations from this expectation suggest non-neutral processes for nonsynonymous sites. An excess of nonsynonymous divergence relative to polymorphism indicates positive selection favoring amino acid replacements, which is the result of recurrent selective sweeps since the split of the two species. The advantage of this test is that as both synonymous and nonsynonymous sites share the same genealogy and lie homogeneously interspersed in protein sequences, thus it is robust to assumptions about nonequilibrium demography (Nielsen 2001, Eyre-Walker 2002) and variation in recombination rates (Sawyer and Hartl 1992). The MK test can also be applied to non-coding regions of the genome, such as promoter regions (Andolfatto 2005). Here, synonymous sites from a neighboring gene serve as the neutral class of sites, while sites in the promoter region represent the putatively selected class.

What fraction of substitutions between species were driven by positive selection, as opposed to being fixed by genetic drift? Based on the logic of the MK test, Smith and Eyre-Walker (2002) developed a method to estimate the fraction of adaptive nucleotide substitutions between species, termed α. It can be estimated as $1- D_sP_n/D_nP_s$ and represents the excess of observed nonsynonymous substitutions over the expected number of substitutions under neutrality. One problem with this approach is that it is likely to be biased if there are slightly deleterious mutations segregating in the population. In *Drosophila*, there is evidence that some nonsynonymous mutations are slightly deleterious (Akashi 1996, Fay *et al.* 2002, Parsch *et al.* 2009). These mutations contribute proportionally more to polymorphism than to divergence, such that α will tend to be underestimated (McDonald and Kreitman1991, Smith and Eyre-Walker 2002). To account for this problem, other approaches to estimate α have been developed. A method by Eyre-Walker and Keightley (2009) estimates the proportion of new mutations that are slightly, intermediate, and strongly deleterious or neutral (the distribution of fitness effects, DFE) from the polymorphism data. Based on the inferred DFE, the fraction of substitutions stemming from neutral and slightly deleterious mutations between two species are estimated. If the observed number of substitutions is greater than the predicted number, the difference can be ascribed to adaptive substitutions.

1.6 SPECIATION GENETICS

When different populations of a species accumulate genetic differences that finally result in reproductive isolation of the populations, speciation occurs. This can happen due to geographic isolation between the populations (allopatric speciation) or even in the absence of such spatial isolation (sympatric speciation). When members of these diverged species come back into secondary contact, they are no longer able to interbreed. Mechanisms that cause reproductive isolation of two species can either act before or after mating. Pre-mating isolation includes ecological or behavioral factors, such as incompatible mating signals that prevent the species from mating. In contrast, post-mating isolation occurs when members of the two species do mate, but the hybrids are inviable, infertile or ecologically maladapted. Typically, the first step towards reproductive isolation is hybrid male sterility. In 1922, Haldane pointed out that when one sex is absent, rare or sterile in hybrids, it is generally the heterogametic sex, *i.e*. males in *Drosophila*. This observation is now known as Haldane's rule (Haldane 1922). A likely explanation for Haldane's rule is the dominance theory: If incompatibilities are due to recessive alleles on the X chromosome, their effect would be uncovered in the heterozygous sex, but not in the homozygous sex. Indeed, there is evidence that recessive incompatibility genes accumulate faster between *Drosophila* species (True *et al.* 1996, Masly *et al.* 2007). It has also been shown that the X chromosome has a disproportional large effect on hybrid sterility and inviability, the so-called large-X effect. There is evidence for a higher density of hybrid male sterility factors on the X chromosome compared to autosomes (Masly *et al.* 2007). Thus, genes involved in male reproduction should play an important role in speciation, especially on the X chromosome (Wu and Davis 1993). Concordant with this, sex-biased genes, particularly male-biased genes, tend to evolve faster than unbiased genes at the DNA sequence level and most of these genes are expressed in reproductive tissues (Parisi *et al*. 2003, Zhang *et al.* 2004, Haerty *et al.* 2007, Baines *et al.* 2008, Pröschel *et al*. 2006, Meisel 2011). Male-biased genes also show greater levels of expression variation within species than other classes of genes (Meiklejohn *et al.* 2003, Hutter *et al*. 2008), and more rapid expression divergence between species (Ranz *et al.* 2003). These findings indicate that sex-biased genes make a large contribution to between-species divergence, but it is not clear if there is a direct link between the evolution of sex-biased genes and speciation.

1.7 SCOPE OF THIS DISSERTATION

CHAPTER 1: To date, most studies of gene expression variation in *D. melanogaster* have focused on a small number of laboratory strains that were derived from non-African populations (Jin *et al.* 2001, Rifkin *et al.* 2003, Gibson *et al.* 2004). Since they did not include ancestral populations, they cannot provide a complete picture of expression variation present in the species. A study by Meiklejohn *et al.* (2003) did include derived as well as ancestral fly strains, but derived flies did not come from one single population and instead were a mixture of North American and Asian lab stocks. A study by Hutter and colleagues (Hutter *et al.* 2008) surveyed gene expression variation in a larger number of fly strains, including eight iso-female lines each from two natural populations of *D. melanogaster*. Besides studying a population from the derived species range in Europe (the Netherlands), they also included a population from the ancestral species range in sub-Saharan Africa (from Zimbabwe). This was the largest study of *D. melanogaster* gene expression variation to date, which provided a good idea about the variation present in the species. However, the study examined only adult males of these populations. This means that it provides only half the story regarding gene expression variation present in the species. In this chapter, I expand the gene expression survey of Hutter *et al.* (2008) to adult female flies of the same populations and strains.
In detail, I am addressing the following questions:

- How much expression variation is present within each population?
- How much expression divergence is there between the populations?
- How do the results of female flies compare to those of male flies?

CHAPTER 2: In the second chapter, I examine genes that show differential expression between the populations (as identified in Chapter 1) at the DNA sequence level. Since these genes show low expression polymorphism within the populations, but high expression divergence between the populations, they are good candidates for genes that have undergone adaptive regulatory evolution. I sequence the coding region as well as 1-2 kb of the upstream region (*i.e.* the putative promoter region) of these genes and perform population genetics analyses. DNA sequence polymorphism is surveyed for 12 fly strains from Europe and 11 fly strains from Africa, including the strains used in the expression analysis.
The questions I seek to answer are:

- Is there evidence for recurrent selective sweeps on these genes or their regulatory regions?
- Is there evidence for recent selection on these genes or regions linked to them?

- Are there DNA sequence differences in the putative promoter regions that are associated with expression differences?

CHAPTER 3: In the third chapter, I investigate the extent to which sex-biased genes contribute to within-population variation, between-population divergence, and between-species divergence. This is determined both at the DNA sequence-level and at the expression-level. To investigate this at the DNA sequence-level, I perform population genetics analyses of 143 genes from the African (Zimbabwe) and the European (the Netherlands) populations of *D. melanogaster*. In addition, a set of 43 genes is examined in *D. ananassae*. To investigate expression patterns, data from Chapter 1, as well as the published data of Hutter *et al*. (2008) and Ranz *et al*. (2003) are used.

The following questions are addressed:

- What are the contributions of sex-biased genes to within-population variation, between-population divergence, and between-species divergence at the DNA-sequence and at the expression level?
- Does a higher proportion of adaptive substitutions occur in sex-biased genes than in unbiased genes?
- Is there evidence for more adaptive substitutions on the X-chromosome than on the autosomes?

BMC Genomics

RESEARCH ARTICLE Open Access

Population transcriptomics of *Drosophila melanogaster* females

Lena Müller[1], Stephan Hutter[1], Rayna Stamboliyska[1], Sarah S Saminadin-Peter[1,2], Wolfgang Stephan[1], John Parsch[1*]

Abstract

Background: Variation at the level of gene expression is abundant in natural populations and is thought to contribute to the adaptive divergence of populations and species. Gene expression also differs considerably between males and females. Here we report a microarray analysis of gene expression variation among females of 16 *Drosophila melanogaster* strains derived from natural populations, including eight strains from the putative ancestral range in sub-Saharan Africa and eight strains from Europe. Gene expression variation among males of the same strains was reported previously.

Results: We detected relatively low levels of expression polymorphism within populations, but much higher expression divergence between populations. A total of 569 genes showed a significant expression difference between the African and European populations at a false discovery rate of 5%. Genes with significant over-expression in Europe included the insecticide resistance gene *Cyp6g1*, as well as genes involved in proteolysis and olfaction. Genes with functions in carbohydrate metabolism and vision were significantly over-expressed in the African population. There was little overlap between genes expressed differently between populations in females and males.

Conclusions: Our results suggest that adaptive changes in gene expression have accompanied the out-of-Africa migration of *D. melanogaster*. Comparison of female and male expression data indicates that the vast majority of genes differing in expression between populations do so in only one sex and suggests that most regulatory adaptation has been sex-specific.

Background

Over the past decade, microarray studies have shown that variation at the level of gene expression is abundant within natural populations [1,2]. Similar studies have also revealed extensive differences in gene expression between males and females [3]. Indeed, in the well-studied model organism *Drosophila melanogaster*, genes that differ in expression between the sexes (sex-biased genes) greatly outnumber those that differ in expression between individuals of the same sex [4-6]. Thus, it is important to account for sex when characterizing gene expression variation within species.

To date, most studies of gene expression variation within *Drosophila* species have been limited to a small number of laboratory strains, or to strains derived from a single non-African population [4-8]. These studies are useful for determining the amount and underlying genetic architecture of gene expression variation among individuals, but reveal little about the potential for gene expression levels to evolve adaptively in response to local environmental conditions. Studies of genomic and mitochondrial DNA variation suggest that *D. melanogaster* expanded from its ancestral range in sub-Saharan Africa and began to colonize Europe about 15,000 years ago [9-13], with a subsequent colonization of North America occurring within the past 500 years [14]. Presumably, the out-of-Africa expansion was accompanied by adaptation to the new, temperate environment, and several studies have provided evidence for genetic adaptation in derived *D. melanogaster* populations [11,15-17].

A previous microarray analysis of male gene expression variation in eight *D. melanogaster* strains from the ancestral species range (Zimbabwe, Africa) and eight

* Correspondence: parsch@zi.biologie.uni-muenchen.de
[1]Department of Biology II, University of Munich (LMU), 82152 Planegg-Martinsried, Germany
Full list of author information is available at the end of the article

strains from Europe (the Netherlands) identified 153 genes with a significant expression difference between the populations [18]. These genes represent candidates for those having undergone adaptive regulatory evolution in response to the local environment and were enriched for genes with functions in insecticide resistance, fatty acid metabolism, and flight [18]. The male expression data, however, provide only half of the story. Given the extent of sex-biased gene expression in *D. melanogaster* [19,20], the potential for differences in the mode of inheritance of gene expression between males and females [21], the impact of the Y chromosome on gene expression variation [22,23], and the proposed differences in effective population size between males and females of the African and European populations [24,25], it is desirable to investigate expression variation among females of the same populations.

Here we report a microarray survey of gene expression variation in adult females of the African and European *D. melanogaster* populations. Our analyses are performed on three levels. First, we use the new microarray data to determine levels of gene expression polymorphism among females of each population, as well as gene expression divergence between populations. Second, we examine the contribution of sex-biased genes to the observed patterns of expression polymorphism and divergence. Third, we compare the female results with previously published results from males in order to detect differences in expression variation between the sexes. We find that, in females, there is little gene expression polymorphism within populations, but a relatively large number of genes with a significant expression difference between populations. The latter represent candidates for population-specific gene regulatory evolution and several of these genes show evidence that positive selection has acted on linked, *cis*-regulatory sequences. We find that sex-biased genes do not make a disproportionate contribution to expression variation among females. A comparison of the female and male results suggests that substantial sex-specific adaptation of gene expression levels has occurred following the out-of-Africa migration of *D. melanogaster*.

Results and Discussion

Gene expression polymorphism

We analyzed gene expression variation among adult females of 16 strains of *D. melanogaster* (eight from Zimbabwe, Africa and eight from the Netherlands, Europe) using CDMC 14kv1 whole-genome microarrays (Figure 1). The microarray features 14,439 unique *D. melanogaster* probes corresponding to 13,688 unique protein-coding genes. After quality control, we detected expression of 6,578 probes corresponding to 6,308 unique genes in all 16 *D. melanogaster* strains. Of these, 1,536 (24%) showed a significant expression difference between at least two of the 16 strains at a nominal *P*-value of 0.001, which corresponds to a FDR of 30%. Overall, there was greater expression polymorphism among African strains than among European strains, but the greatest number of expression differences was found in comparisons between African and European strains (Table 1).

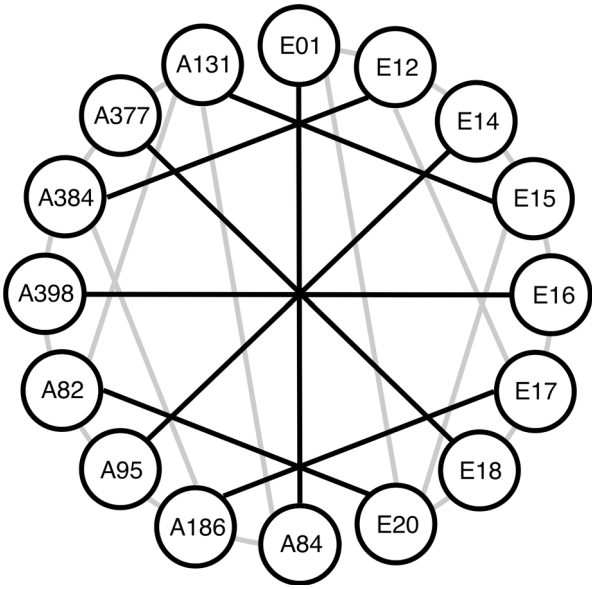


Figure 1 Microarray hybridization scheme. Each node represents one *D. melanogaster* strain, with 'E' indicating European and 'A' indicating African strains. Each line represents two microarray hybridizations (dye-swap replicates), with black indicating between-population and gray indicating within-population hybridizations.

Across all 16 *D. melanogaster* strains, we found significantly less expression polymorphism in females than what was previously reported for males of the same strains [18], with females having 1.7-fold fewer polymorphic genes (24% *vs*. 38%; χ^2 = 230, $P < 0.0001$), and 3.7-fold fewer significant pairwise differences per gene as males (0.89 *vs*. 3.28; Mann-Whitney test, $P < 0.0001$).

Table 1 Expression polymorphism within and between populations

Comparison	Number of polymorphic genes	Mean differences per pairwise comparison	Mean pairwise differences per gene (in %)
Among all strains	1536	49.8	0.74
Within Europe	305	22.6	0.33
Within Africa	547	37.5	0.57
Between populations	1364	65.7	0.99

Significant differences in expression between strains were determined using a *P*-value cut-off of 0.001 (FDR = 30%).

These comparisons are conservative, because they use a common *P*-value of 0.001 for both sexes, which corresponds to a FDR of 30% in females, but only 7% in males. Reducing the FDR in females would reduce the number of polymorphic genes even further. However, even using the minimal *P*-value possible in our analysis ($P = 0.0001$), the FDR does not drop below 20%. A contributing factor to the observed difference between the sexes may be that there is less statistical power to detect expression polymorphism in the female experiment. Townsend [26] proposed the statistic GEL_{50}, which is the fold-change difference at which there is a 50% chance of detecting a significant difference with $P < 0.05$, as a standard for comparing the power of microarray experiments. For the female experiment, the GEL_{50} was 1.85. This is higher than the GEL_{50} of 1.51 reported for the male experiment [18], but still within the range reported for similar surveys of expression polymorphism in *Drosophila* and other species [2]. However, it is possible that small differences in GEL_{50} can lead to large differences in the percentage of genes detected as differentially expressed [2].

To investigate the contribution of sex-biased genes to gene expression polymorphism among females, we classified all of the genes on our arrays as male-biased, female-biased, or unbiased using the 5% FDR meta-analysis of the Sebida database (release 2.0) [27]. Previous studies have shown that male-biased genes are the most polymorphic class of genes when assayed in males [18,28]. When assayed in females, there was no significant difference in the level of expression polymorphism among male-biased, female-biased, and unbiased genes (Table 2). However, the general pattern in females followed that in males, with male-biased genes showing the greatest expression polymorphism and female-biased genes showing the least (Table 2). As expected, there were significant differences in the proportion of genes of different classes that were detected as expressed in females, with 56% of the female-biased genes and 38% of the male-biased genes being detected (Table 2). It should be noted that the Sebida sex-bias classifications consider only adult flies raised under standard laboratory conditions and, thus, may overlook genes that show condition-dependent or transient sex-biased gene expression. Baker and Russell [29] identified over 3,500 genes that showed female-biased expression in adult female abdomens during at least one stage of egg development. However, levels of polymorphism in this set of female-biased genes were nearly identical to those in the Sebida female-biased gene set. Of the female-biased genes identified by Baker and Russell [29] that were detected as expressed in our experiment, 23.82% (470/1,973) were polymorphic. The corresponding number for the Sebida female-biased gene set was 23.79% (534/2,245).

Table 2 Expression polymorphism in sex-biased genes

	Sex-bias classification		
Feature	**Female**	**Male**	**Unbiased**
Number of genes on array	4002	2572	5988
Percentage of genes detected as expressed	56.1*	36.5*	44.8
Percentage of expressed genes:			
Polymorphic in Europe	5.1	5.1	4.6
Polymorphic in Africa	8.0	9.3	8.5
Polymorphic overall	23.8	24.4	24.7
Differentially expressed between populations	8.4†	10.9	11.6
Average percentage of pairwise differences:			
Within Europe	0.16	0.18	0.17
Within Africa	0.25	0.32	0.30
Overall	0.65	0.87	0.79

Genes were classified using the 5% FDR meta-analysis of the Sebida database [27]. *Significantly different from unbiased genes (FET, $P < 0.0001$).
†Significantly different from male-biased (FET, $P < 0.05$) and unbiased (FET, $P < 0.001$) genes.

It was previously found that, among males, genes residing on the X chromosome show less expression polymorphism than those residing on the autosomes [18]. This was attributed to the paucity of male-biased genes, which are the most polymorphic class in males, on the X chromosome [18]. Consistent with this interpretation, we found no significant difference in the level of expression polymorphism between X-linked and autosomal genes in females, where many fewer male-biased genes are expressed. The proportions of polymorphic X-linked and autosomal genes were 25.3% and 23.9%, respectively ($\chi^2 = 0.97$, $P = 0.33$). The ratio of X-linked to autosomal significant pairwise differences per gene was 0.96.

The above results suggest that the difference in expression polymorphism between males and females can be explained partly by sex-biased gene expression, as male-biased genes tend to show the greatest expression polymorphism whether assayed in males or in females [8,28] (Table 2) and make up a much greater proportion of the genes detected as expressed in males. However, when considering only unbiased genes (those expressed nearly equally in males and females), the percentage of polymorphic genes is still 1.6-fold lower in females than in males (24.7% *vs*. 39.2%; $\chi^2 = 230$, $P < 0.0001$). Similarly, unbiased genes show 3.9-fold fewer pairwise differences per gene in females than in males (0.95 *vs*. 3.74; Mann-Whitney test, $P < 0.0001$). This suggests that there are general differences between the sexes with respect to the regulation of gene expression and/or the level of purifying selection that restricts gene expression variation.

It has been observed that infection with sigma virus alters the expression of many more genes in males than

in females [30], which is consistent with male gene expression being more sensitive to genetic and/or environmental perturbations than female gene expression. It has also been shown that genetic variation on the Y chromosome can affect expression levels of many X-linked and autosomal genes [22,23]. Thus, one would expect there to be more expression variation among males, as this Y-linked source of expression variation is absent in females. Because our experiments used inbred strains that are homozygous over most of the genome, we are not able to detect gene expression variation caused by non-additive interactions between alleles in heterozygotes. Thus, the level of expression variation measured in our sample may be less than that observed among individuals sampled directly from natural populations. However, since the same inbred lines were used for both the male and female experiments, non-additivity cannot explain the difference observed between the sexes. Previous studies have shown, however, that non-additive interactions are more prevalent in females than in males [5,21], which suggests that the difference between male and female expression polymorphism might be smaller in natural populations than in comparisons of inbred lines.

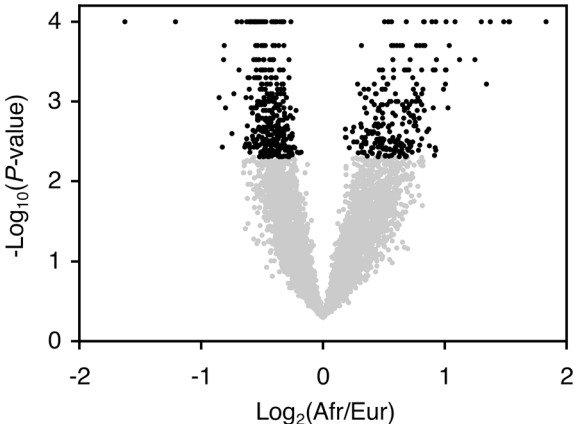


Figure 2 Volcano plot of the between-population analysis. Black points indicate genes with a significant expression difference between the African and European populations of *D. melanogaster* (FDR < 5%).

Gene expression divergence between populations

To identify genes that differ in expression between the European and African populations of *D. melanogaster*, we used data from the 16 microarray hybridizations that directly compared strains of the two populations (indicated by black lines in Figure 1). After quality control, we were able to compare hybridization intensities of 5,584 unique probes, corresponding to 5,370 genes, between the populations. Of these, 569 genes showed a significant inter-population expression difference with $P < 0.005$ (FDR = 5%; Figure 2; Additional file 1). More of the significant genes had higher expression in Europe than in Africa (330 *vs*. 239; $\chi^2 = 14.6$, $P < 0.0001$). However, the average magnitude of over-expression was greater in Africa than in Europe (1.56-fold *vs*. 1.37-fold; Mann-Whitney test, $P < 0.0001$). Only two genes showed greater than two-fold over-expression in Europe, while 16 showed greater than two-fold over-expression in Africa (Fisher's exact test (FET), $P < 0.0001$). Similarly, only 34 genes showed greater than 1.5-fold over-expression in Europe, while 115 showed greater than 1.5-fold over-expression in Africa (FET, $P < 0.0001$).

There was not an overrepresentation of sex-biased genes among those showing a significant expression difference between the African and European populations. In fact, there was a slight (but significant) under-representation of female-biased genes among the genes showing differential expression between the populations in females (Table 2). There was also no significant difference in the proportions of X-linked (10.0%) and autosomal (10.3%) genes that showed differential expression between the populations ($\chi^2 = 0.10$, $P = 0.76$).

The gene showing the strongest over-expression in the European population was *Cyp6g1*, a member of the cytochrome P450 gene family that is associated with insecticide resistance [31] (Figure 3). This gene was also found to have the greatest over-expression in male *D. melanogaster* [18]. Previous studies indicated that high levels of *Cyp6g1* expression, which provide increased resistance to DDT and other insecticides, are associated with the insertion of an *Accord* transposable element upstream of *Cyp6g1*, as well as with tandem duplication of the *Cyp6g1* gene [31-33]. The insertion and duplication are present at high frequency in many non-African populations of *D. melanogaster*, which has been suggested to be the result of selection for insecticide resistance [32,33]. To test for these features in our population samples, we performed PCR on all strains using the previously described diagnostic primers [31,33]. The *Accord* insertion was present in all European strains and in three of the eight African strains. All strains with the *Accord* insertion, but none of the others, had a tandem duplication of the *Cyp6g1* locus (Figure 4; Additional file 2). The three African strains with the insertion/duplication had 2.78-fold higher *Cyp6g1* expression than those without (Mann-Whitney test, $P < 0.05$). However, the expression level of the African strains with the insertion/duplication was still 1.57-fold lower than that of the European strains (Mann-Whitney test, $P < 0.05$). This suggests that other factors, either *cis*- or *trans*-acting, also contribute to the increased *Cyp6g1* expression observed for European strains.

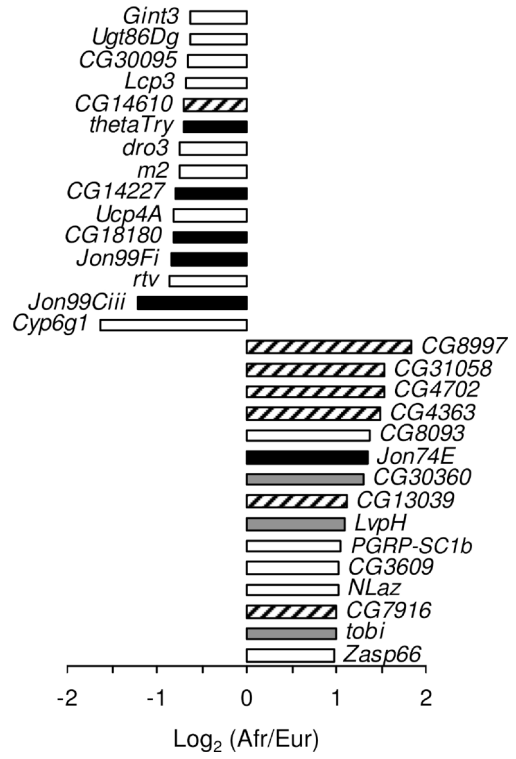


Figure 3 The top 15 over-expressed genes in each population. The horizontal bars indicate the ratio of African-to-European expression for each gene. Black bars represent genes that function in proteolysis, while gray bars indicate genes that function in carbohydrate metabolism. Genes of unknown function are indicated by hatched bars.

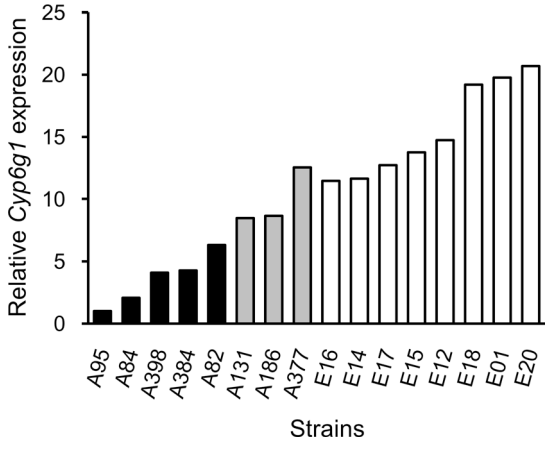


Figure 4 Cyp6g1 *expression levels*. Shown is the relative expression in adult females of all strains as determined by microarray experiments. All eight European strains contain the *Accord* insertion and a duplication of the *Cyp6g1* gene (white bars), whereas only three of the eight African strains have the insertion/duplication (gray bars). The remaining five African strains have neither the *Accord* insertion, nor the *Cyp6g1* duplication (black bars). Expression levels of all three groups are significantly different from each other (Mann-Whitney test, $P < 0.05$ for all comparisons).

Functional annotation of differentially expressed genes

Genes with proteolytic function, particularly serine-type endopeptidases, were consistently over-expressed in the European population (Table 3; Additional file 3). Among the 15 genes with the greatest over-expression in Europe, five function in proteolysis (Figure 3). Of these, the genes with the largest fold-change are members of the Jonah gene family, *Jon99Ciii* and *Jon99Fi*, which are serine-type peptidases expressed in the midgut of the adult fly. Other serine-type endopeptidases that were over-expressed in Europe include *CG18180*, *CG14227*, and *thetaTrypsin* (Figure 3). In contrast to the other proteases, one member of the Jonah gene family, *Jon74E*, showed significantly higher expression in Africa than in Europe.

Genes involved in sensory perception were enriched in both the Europe over-expressed and Africa over-expressed gene lists. However, the specific pathways differed between the two populations. In Europe, genes involved in olfaction and the detection of chemical stimulus were over-represented (Table 3), while in Africa genes involved in vision and the detection of light stimulus were over-represented (Table 4).

Genes involved in carbohydrate metabolism were also enriched among the genes over-expressed in the African population (Table 4) and several of these genes were among the most over-expressed, including the maltase *CG30360*, and two α- glucosidases *LvpH* and *tobi* (Figure 3). *tobi* has been shown to be a target of the insulin- and glucagon-like signaling system [34]. In this respect, it is noteworthy that the highly over-expressed gene *Nlaz*, which plays a role in stress response and

Table 3 GO-term enrichment of genes over-expressed in the European population

ID	Ontology	Term	*P*-value
GO:0004984	MF	Olfactory receptor activity	0.018
GO:0004252	MF	Serine-type endopeptidase activity	0.036
GO:0005337	MF	Nucleoside transmembrane transporter activity	0.039
GO:0035214	BP	Eye-antennal disc development	0.001
GO:0008052	BP	Sensory organ boundary specification	0.020
GO:0009593	BP	Detection of chemical stimulus	0.024
GO:0065004	BP	Protein-DNA complex assembly	0.027
GO:0007608	BP	Sensory perception of smell	0.027
GO:0009047	BP	Dosage compensation	0.033
GO:0001508	BP	Regulation of action potential	0.033
GO:0006544	BP	Glycine metabolic process	0.033
GO:0008380	BP	RNA splicing	0.036

In cases where multiple, related terms within a GO hierarchy were significant, only a single term is given. The complete list is provided in Additional file 3. Ontology abbreviations are: MF, molecular function; BP, biological process. *P*-values were determined by a hypergeometric test with Benjamini-Hochberg multiple-test correction.

Table 4 GO-term enrichment of genes over-expressed in the African population

ID	Ontology	Term	*P*-value
GO:0004558	MF	Alpha-glucosidase activity	0.006
GO:0004806	MF	Triglyceride lipase activity	0.021
GO:0003697	MF	Single-stranded DNA binding	0.021
GO:0019201	MF	Nucleotide kinase activity	0.021
GO:0004129	MF	Cytochrome-c oxidase activity	0.036
GO:0019318	BP	Hexose metabolic process	0.005
GO:0009586	BP	Rhodopsin mediated phototransduction	0.006
GO:0048814	BP	Regulation of dendrite morphogenesis	0.009
GO:0014866	BP	Skeletal myofibril assembly	0.015
GO:0048139	BP	Female germ-line cyst encapsulation	0.015
GO:0035075	BP	Response to ecdysone	0.025
GO:0006119	BP	Oxidative phosphorylation	0.028
GO:0012502	BP	Induction of programmed cell death	0.032
GO:0006631	BP	Fatty acid metabolic process	0.035
GO:0007015	BP	Actin filament organization	0.035
GO:0030713	BP	Ovarian follicle cell stalk formation	0.036
GO:0016028	CC	Rhabdomere	0.018
GO:0044429	CC	Mitochondrial part	0.025
GO:0016459	CC	Myosin complex	0.025
GO:0030425	CC	Dendrite	0.035

In cases where multiple, related terms within a GO hierarchy were significant, only a single term is given. The complete list is provided in Additional file 3. Ontology abbreviations are: MF, molecular function; BP, biological process; CC, cellular compartment. *P*-values were determined by a hypergeometric test with Benjamini-Hochberg multiple-test correction.

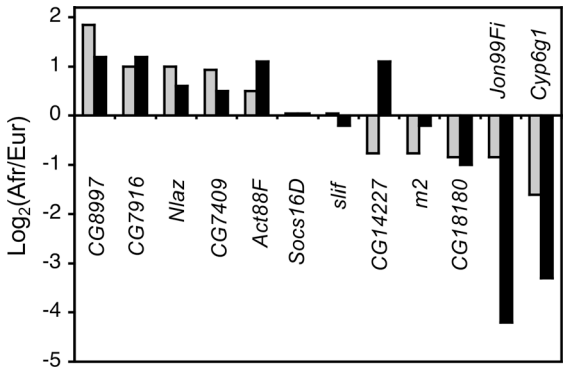


Figure 5 Comparison of microarray and qRT-PCR results. Shown is the relative expression difference between the eight African and eight European strains as measured by microarray (gray bars) or qRT-PCR (black bars) for 12 genes. Gene symbols are given below/above their corresponding values.

determination of adult lifespan, also functions in carbohydrate homeostasis and has been suggested to interfere with insulin signaling [35].

Other enriched functions among the Africa over-expressed genes included oxidative phosphorylation and muscle formation (Table 4). However, many of the Africa over-expressed genes are of unknown function, including six of the 15 genes with the greatest over-expression in Africa and the gene showing the highest overall difference in expression between the African and European populations, *CG8997* (Figure 3).

Validation of microarray results by qRT-PCR

In order to verify the between-population expression differences detected in our microarray analysis, we performed qRT-PCR on a subset of 12 genes, including five with over-expression in Africa, five with over-expression in Europe, and two control genes that showed no difference in expression between populations (Figure 5). For 10 of these genes, including all of those with Africa over-expression, the two control genes, and the three genes with the greatest Europe over-expression, the results were consistent with both methods. One of the genes (*m2*) showed strong (1.7-fold) Europe over-expression in the microarray experiment, but only slight (1.2-fold) Europe over-expression by qRT-PCR. Another gene (*CG14227*) showed over-expression in opposite populations when measured by the two methods (Figure 5). The reason for this discrepancy is unclear. It may be because the microarray and qRT-PCR probes match different regions of the *CG14227* transcript. However, there is only one annotated transcript for this gene in the current release of FlyBase (release 5.27). When considering all genes and strains, there was a good correlation between expression levels measured by microarray and by qRT-PCR (Pearson's $R = 0.5$, $P < 0.0001$; Additional file 4).

Comparison of inter-population gene expression divergence in males and females

There were many more genes that differed significantly in expression between the European and African populations in females than in males. In females, 10.6% (569/5370) of the genes analyzed showed a significant inter-population difference with a FDR of 5%. In males, 3.4% (153/4528) of the genes analyzed showed a significant inter-population difference with a FDR of 8.7% ($\chi^2 = 189$, $P < 0.0001$). The lower FDR of the female experiment indicates that this is a conservative comparison. Furthermore, the GEL_{50} values for the female and male experiments were 1.22 and 1.18, respectively, indicating that the female experiment had slightly less statistical power to detect differences. This suggests that the different amounts of inter-population gene expression divergence observed between females and males have a biological basis. At the protein level, it has been reported that autosomal female-biased genes show evidence for greater adaptive evolution in the European

population than in the African population [25]. If this is indicative of a general pattern of stronger selection on females to adapt to the European environment, it could explain the excess of between-population expression differences in females relative to males. A possible reason for this is that females may be under greater selection to survive through the winter, while males that do not survive the winter may still contribute genes to future generations if their sperm is stored in a surviving female [36]. The above hypothesis predicts that most expression differences between populations should be the result of changes occurring within the European population during colonization. At present, we do not have data that would allow us to infer the direction of inter-population expression changes and test this prediction.

Of the 569 genes identified as differentially expressed between the African and European populations in females and the 153 genes identified as differentially expressed between the same populations in males [18], only 14 genes overlapped (*i.e.*, were significant in both sexes; Table 5). Of these, 12 genes showed higher expression in the same population in both sexes, which is no more than would be expected by chance given the numbers of significant genes in each sex and the total number of genes analyzed in both sexes ($\chi^2 = 0.60$, $P = 0.44$). However, several of the overlapping genes (Table 5) represent good candidates for genes that have undergone adaptive regulatory evolution in response to changes in the environment. The gene showing the greatest over-expression in Europe in both males and females was the cytochrome P450 gene *Cyp6g1* (see above section, *Gene expression divergence between populations*). The gene *CG12262*, with an annotated function in oxidation/reduction and fatty acid metabolism, and the gene *CG17292*, which is also involved in fatty acid metabolism, both showed over-expression in Europe in both sexes. The gene *CG7409*, which has an annotated function in response to heat and unfolded protein binding, and the actin gene *Act88F*, which is a component of indirect flight muscle and also involved in the innate immune response, showed consistent African over-expression in both sexes. In addition, three genes of unknown function that are located in a cluster on chromosome arm 2L (*CG8997*, *CG7916*, and *CG7953*) showed significant African over-expression in both males and females (Figure 6). A fourth gene in this cluster, *CG33307*, showed significant African over-expression in males, but was not detected as expressed in females (Figure 6).

Table 5 Genes with a significant inter-population expression difference in both females and males

	Log_2 (Afr/Eur)		
Gene	Female	Male*	Function
CG8997	1.85	0.77	Unknown
CG7916	1.00	0.68	Unknown
CG34330	1.00	0.38	Unknown
CG7409	0.93	0.49	Unfolded protein binding; response to heat
retinin	0.85	0.58	Unknown; expressed in eye
CG7953	0.85	0.49	Unknown
Adk2	0.58	0.38	Adenylate kinase; ADP biosynthesis
Act88F	0.49	1.54	Actin filament; indirect flight muscle; immune response
Cyp6g1	-1.58	-2.14	Cytochrome P450; insecticide resistance
fau	-0.49	-0.38	Unknown; upregulated under anoxia
CG17292	-0.49	-0.19	Triglyceride lipase; lipid metabolism
CG12262	-0.38	-0.38	Acyl-CoA dehydrogenase; fatty acid beta-oxidation
for	0.51	-0.31	cGMP-dependent protein kinase; feeding behavior
CG11395	0.49	-0.23	Unknown

*Data from Hutter *et al.* [18].

The vast majority of genes detected as being differentially expressed between populations showed this pattern in only one sex. Of the 569 genes that differed in expression between females of the European and African populations, 557 showed this difference only in females. Of these, 310 genes were not detected as expressed in males, while 245 were detected as expressed but their expression did not differ significantly between the populations. Two other genes showed a significant expression difference between populations, but in opposite directions in the two sexes. The first, *CG11395*, is a gene of unknown function that had significant Africa over-expression in females, but significant Europe over-expression in males. The second is the *foraging* (*for*) gene, which encodes a cGMP-dependent protein kinase that influences larval and adult feeding behavior [37-39]. In females, *for* is significantly over-expressed in Africa, while in males it is significantly over-expressed in Europe.

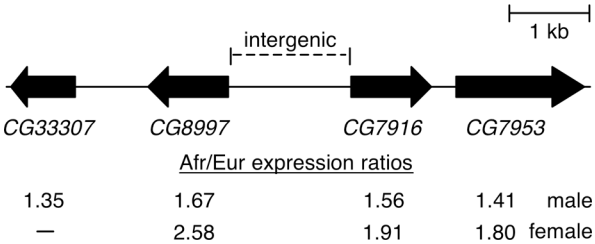


Figure 6 A cluster of genes on chromosome arm 2L with significant over-expression in the African population. Solid boxes represent transcriptional units, with the arrow indicating the direction of transcription. The African/European expression ratio of each gene, as determined by microarray experiments using males [18] and females (present study), is shown at the bottom. Expression of *CG33307* was not detected in females.

To further investigate the effect of sex on inter-population differences in gene expression, we performed a meta-analysis of the female and male expression data. For each of the 2,315 genes common to both experiments, we determined the difference in mean expression level between the African and European populations, as well as the standard deviation (SD) of this difference, in both females and males. We then calculated the statistic, *d*, by subtracting the Africa-Europe difference in males from that in females and dividing by the pooled SD of the difference in both sexes [40] (Additional file 5). We identified 94 genes for which the difference between males and females was greater than two SD units (FDR = 3.2%) and 209 genes for which the difference between males and females was greater than 1.7 SD units (FDR = 5%). Of these 209 genes, 176 (84%) showed enriched expression in opposite populations in the two sexes (Figure 7). There were a few cases in which a gene showed over-expression in the same population in both sexes, but the extent of over-expression was greater in one sex than the other (Figure 7).

Population genetics of a cluster of genes with African over-expression

The common expression pattern and genomic organization of *CG8997*, *CG7916*, and *CG7953* suggested that they might share a common regulatory element in the intergenic region between *CG8997* and *CG7916*. To investigate this further, we analyzed DNA sequence polymorphism in the coding and shared intergenic region of these two genes (Figure 6), which had previously been sequenced in the same strains used in our microarray experiments [41]. In both populations, levels of nucleotide polymorphism (π) were normal for a region of moderate recombination. In Africa, values of π at synonymous sites in *CG8997* and *CG7916* and all sites in the intergenic region were 0.028, 0.037, and 0.008, respectively. In Europe, the corresponding values were 0.017, 0.040, and 0.008. Furthermore, there was no haplotype structure or fixed sequence difference between the populations. Thus, there was no evidence for a recent selective sweep in this region of the genome. The most extreme difference in allele frequency was a single-nucleotide G/A polymorphism within the intergenic region in which the derived variant (G) was present in seven of the eight European strains, but only three of the eight African strains. We did, however, find evidence for non-neutral evolution in the coding regions of *CG8997* and *CG7916*, as well as the intergenic region, by the test of McDonald and Kreitman [42] (Table 6). Thus, this region of the genome appears to have been a target of both structural and regulatory adaptation in the past (*i.e.*, since the divergence of *D. melanogaster* and *D. sechellia*).

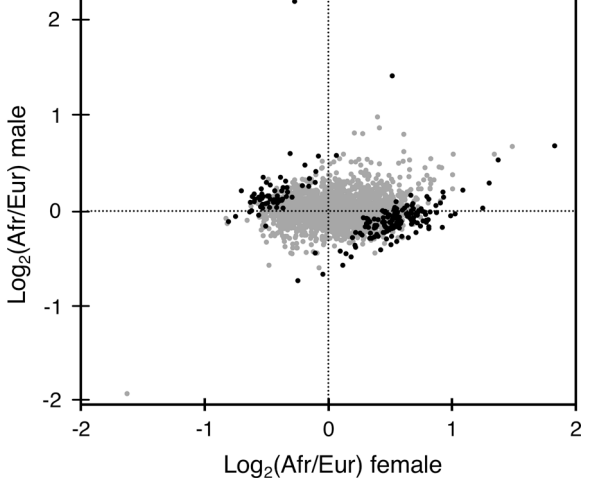


Figure 7 Comparison of inter-population differences in gene expression between males and females. Points represent African/European expression ratios of the 2315 genes that overlap between the present study and that of Hutter *et al*. [18]. Black points indicate the 209 genes showing a significant sex-by-population effect (FDR < 5%).

Expression and behavioral divergence between populations

Previous behavioral studies have shown that there is uni-directional mate-choice preference between *D. melanogaster* strains from Zimbabwe (Z) and cosmopolitan (M) strains, with Z females showing a preference for Z males [43,44]. Michalak *et al*. [45] investigated gene expression in female heads from flies of the two mating types and identified 45 candidate genes that might be involved in the behavioral difference. Only one of these genes (*CG7530*) was significant in our experiment using whole females. This gene had higher M expression in the experiment of Michalak *et al*. [45], but higher Zimbabwe expression in our experiment. Thus, there is no concordance between the putative mating-behavior genes expressed differently in Z and M female heads

Table 6 Results of McDonald-Kreitman (MK) tests

							MK-test *P*-value	
Gene	D_s	P_s	D_n	P_n	D_i	P_i	Nonsynonymous	Intergenic
CG8997	13	16	10	2	71	34	0.038	0.031
CG7916	13	24	6	1	71	34	0.032	0.008
Combined	26	40	16	3	71	34	<0.001	<0.001

Shown are the numbers of fixed differences (D) between *D. melanogaster* and *D. sechellia* and the number of polymorphic sites (P) within the African population of *D. melanogaster*. Subscripts indicate synonymous (s), nonsynonymous (n), and intergenic (i) sites. The intergenic region is shared between the two genes (see Figure 6).

and the genes expressed differently between the Zimbabwe and European populations in whole females. Two of the top candidate genes from Michalak *et al*. [45], *desaturase2* and *Odorant receptor 63a* were not detected as being consistently expressed in our experiment and, thus, were excluded from the analysis. However, *desaturase2* expression was detected in a higher proportion of Zimbabwe strains (7/8) than European strains (4/8) and, on average, showed two-fold higher expression in Zimbabwe strains in the hybridizations where it could be detected. This is consistent with the finding of Michalak *et al*. [45] that *desaturase2* shows over-expression in Z strains. A comparison of gene expression in male heads between a single Z strain and a single M strain uncovered 1216 genes that differed in expression between the mating types [46]. Although only 77 of these genes were detected as differentially expressed between the populations in our analysis, several of the overlapping genes were among those showing the greatest expression difference between Europe and Africa in both sexes, including *Cyp6g1*, *Act88F*, and the clustered genes *CG8997*, *CG7916*, and *CG7953*.

Conclusions

Our microarray survey identified over 500 genes showing low within-population expression polymorphism, but high between-population expression divergence in female *D. melanogaster* from Europe and Africa. The combination of low polymorphism and high divergence is a hallmark of positive selection and suggests that adaptive evolution at the gene regulatory level has occurred in conjunction with the recent colonization of non-African habitats. This is supported by the finding that *Cyp6g1*, whose expression is known to play an ecologically relevant role in insecticide resistance, was among the genes with the greatest inter-population expression difference. The functional basis for the inter-population divergence of the other genes is unknown, however there was an over-representation of genes involved in proteolysis, carbohydrate metabolism and sensory perception (both vision and olfaction). There was very little overlap between genes showing a significant expression difference between populations in females and in males. This suggests that most adaptive changes in gene expression are sex-specific and highlights the need for both sexes to be considered in studies of gene regulatory evolution.

Because our study focused on only one population from the ancestral species range (Zimbabwe) and one from the derived range (the Netherlands), it is not possible to distinguish global "out-of-Africa" adaptations from those that are specific to a local population. Surveys of nuclear DNA polymorphism indicate that there is little population structure within Europe, but more differentiation among some American, Asian, and African populations [47,48]. There is also evidence for adaptive evolution of pigmentation, a trait known to be influenced by gene-regulatory variation, among African populations [49]. Thus, there is likely to be gene expression divergence among various African and non-African populations. Further expression studies are needed to investigate this possibility.

Methods

Fly strains

Expression variation was surveyed for eight isofemale strains of both a European (Leiden, the Netherlands) and an African (Lake Kariba, Zimbabwe) population of *D. melanogaster*. The populations are as described in Glinka *et al*. [16]. The fly strains were the same as those used in the expression analysis of adult male flies by Hutter *et al*. [18]. Flies were maintained on standard cornmeal-molasses medium at 22° and constant lighting.

Microarray platform

The CDMC 14kv1 microarray (Canadian *Drosophila* Microarray Centre, Mississauga, Canada) was used for all hybridizations. This platform features a total of 32,448 oligonucleotide probes (65-69 bases), each spotted in duplicate. The probes represent 13,688 unique genes, which correspond to 92% of those in the current *D. melanogaster* genome annotation (FlyBase release 5.27). Since the transcript-specific probes were designed to release 4.1 of the genome, some genes in the current annotation are not represented on the array, whereas others are represented by more than one probe.

RNA extraction, hybridization, and scanning

For each strain, total RNA of 40 mated female flies, four-to-six days of age, was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) and samples were stored at -80°. Reverse transcription was conducted using 50 μg of total RNA per strain and anchored oligo (dT) primers. cDNA samples were labeled with Alexa Flour 555 and 647 dyes using the SuperScript Plus Indirect cDNA Labeling System (Invitrogen) and following the manufacturer's protocol.

To compare expression levels of all fly strains to each other, the hybridization scheme developed by Hutter *et al*. [18] was followed. This approach allows expression levels of all strains to be compared, while keeping the number of hybridizations at a practical level (Figure 1). Six or eight replicate hybridizations per strain were performed on a total 56 microarrays. For each strain, three or four competitive hybridizations with other strains, plus their respective dye-swap hybridizations were performed. For technical replicates (dye-swaps), RNA from the same extractions was used, whereas for biological

replicates (different pairwise hybridizations of strains), RNA extracted from a new set of flies was used. Arrays were pre-hybridized and washed using the Pronto! Universal Microarray Kit (Corning, Lowell, MA, USA) according to the manufacturer's protocol. Otherwise, hybridizations were conducted following the CDMC protocol. Arrays were scanned with an aQuire 2-laser microarray scanner and Qscan software (Genetix, New Milton, UK). All microarray data have been submitted to the Gene Expression Omnibus database under the accession numbers GSM580470-GSM580525 (platform GPL3603, series GSE23662).

Microarray data analysis

Raw fluorescence intensities were normalized using CARMAweb [50], which is a web-based interface to the 'limma' package of Bioconductor [51]. The default settings of 'minimum', 'printtiploess', and 'quantile' were used for background correction, within-array normalization, and between-array normalization, respectively. Between-array normalization was done using pairs of dye-swap hybridizations. As a quality control step to eliminate background noise from genes that are not expressed (or expressed only at very low levels) in adult females, we required that a spot have mean signal intensity at least one SD above local background in both channels to be included in the analysis. In cases where both replicate spots of a probe passed quality control, the arithmetic mean of their $\log_2$(red/green) intensities was used. Otherwise, only the red/green intensity of the spot passing quality control was used.

The resulting normalized red/green-ratios were used as input for BAGEL [52], a program that estimates relative expression levels for each gene in each of the 16 strains using a Bayesian framework. To determine the experiment-wide false discovery rate (FDR), we repeated the BAGEL analysis on a randomized version of our final data set. Randomization was performed by sampling- with-replacement within each hybridization (*i.e.*, randomizing within a column), thereby maintaining the underlying data structure (*e.g.*, excluded genes) within each hybridization. The resulting output was used to determine the FDR corresponding to a given *P*-value.

To identify genes that differ in expression between Africa and Europe on a population level, we repeated the BAGEL analysis using only the 16 hybridizations in which an African strain was compared directly to a European strain (black lines in Figure 1). All strains from the same population were combined into a single node and, thus, treated as biological replicates from within a population. To determine the FDR, BAGEL was run on a randomized data set that was created by permuting the expression ratios of the replicate hybridizations within each gene (*i.e.*, randomizing within a row). As an additional quality control step, we required that each gene be detected as expressed (by the criteria described above) in at least nine of the 16 replicate hybridizations.

qRT-PCR

For each strain, qRT-PCR of two biological replicates, representing two separate RNA extractions of 20 four-to-six day-old mated female flies, was performed. Following DNase I digestion, 5 μg of total RNA was reverse transcribed using Superscript II reverse transcriptase and random hexamer primers (Invitrogen). The resulting cDNA was diluted 1:40 and used for qRT-PCR with TaqMan probes and TaqMan Gene Expression Master Mix (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocol. To validate expression differences detected by our microarray analysis, qRT-PCR was performed on a Bio-Rad Real-Time thermal cycler CFX96 (Bio-Rad, Hercules, CA, USA) for the following target genes (TaqMan IDs are given in parentheses): *Cyp6g1* (Dm01819889_g1), *Jon99Fi* (Dm02146518_s1), *CG18180* (Dm01801887_s1), *CG14227* (Dm01845429_g1), *m2* (Dm02151465_s1), *CG8997* (Dm01791303_g1), *CG7916* (Dm01791305_g1), *Nlaz* (Dm01844577_g1), *CG7409* (Dm01840751_s1), *Act88F* (Dm02362815_s1), *CG18179* (Dm01801878_s1), *slif* (Dm01792789_g1), and *Socs16D* (Dm01813854_g1). Expression levels of all target genes were normalized to *Actin 5C* (Dm02361909_s1), which was used as an internal control. All assays were performed in three technical replicates, and for each gene the average threshold cycle (Ct) value over all biological and technical replicates was determined. ΔCt values were calculated by subtracting the control Ct from the target Ct value. The fold-change in expression between two samples was calculated as $2^{(-\Delta Ct1 - \Delta Ct2)}$. To determine the fold-change between the African and the European population, ΔCt values were averaged among strains within each population and the European value was used as ΔCt_2.

GO analysis

Enriched GO terms within the lists of differentially expressed genes were identified using the GOEAST web server [53]. Prior to analysis, the annotation of the CDMC microarrays was updated to match FlyBase release 5.27 by performing a BLAT search of all probe sequences with the UCSC genome browser [54]. Probes giving a unique hit to an annotated transcript were matched with their release 5.27 GO terms. Significant GO term enrichment was determined by the hypergeometric method with Hochberg FDR multiple-test correction [55], with the FDR set to 0.05. As a background for GO enrichment tests, we used all genes on the CDMC microarray that were detected as expressed in our experiments (*i.e.*, those passing the quality control steps described above).

DNA sequence analysis

DNA sequence polymorphism in the genomic region encompassing the genes *CG8997* and *CG7916* was previously reported [41]. These authors directly sequenced PCR-amplified genomic DNA from the same strains used in our microarray analysis, plus an additional four strains each from the Zimbabwe and Netherlands populations [16]. We used all of the available sequences for McDonald-Kreitman tests [42] of selection on nonsynonymous and intergenic sites using the DnaSP (v5) software [56]. The test compares ratios of divergence-to-polymorphism at the test sites (nonsynonymous or intergenic) to those at synonymous sites and provides evidence for adaptive evolution when there is a relative excess of divergence at the test sites, which is consistent with recurrent selective sweeps since the time of speciation.

To analyze sequence variation in the *Cyp6g1* region, we performed diagnostic PCR on the 16 strains used in our microarray analysis. The primers used to detect the *Accord* insertion were 5'-GAAAGCCGGTTGTGTTTAATTAT-3' and 5'-CTTTTTGTGTGCTATGGTTTAGTTAG-3', which flank the insertion site. An additional forward primer complementary to the *Accord* insertion (5'-GGGTGCAACAGAGTTTCAGGTA-3') was used to confirm its presence [31]. The primers used to detect tandem duplication of the *Cyp6g1* locus were 5'-CGAGTACGAGAGCGTGGAG-3' and 5'-ATTAAACACAACCGGCTTTCTCG-3' [33]. Following PCR, the products were sequenced to confirm that the expected target sequence was amplified.

Statistical analysis

For comparisons of categorical data (*e.g.*, numbers of polymorphic and non-polymorphic genes in males and females) we used standard 2 × 2 contingency table analyses. *P*-values were determined by Fisher's exact test or, when the sample sizes were large, by a chi-squared approximation. To test for differences between two samples (*e.g.*, *Cyp6g1* expression between strains with and without the *Accord* insertion) we used the non-parametric Mann-Whitney *U* test, which compares the rank-sums of the observed values of two samples. This approach was used to avoid making assumptions about the underlying distribution of gene expression levels among individuals or classes of genes. All tests were performed using R (version 2.10.1) [57].

Additional material

Additional file 1: Expression divergence between the African and European populations. Table of relative expression levels in the African and European populations of all genes used in the analysis.

Additional file 2: Diagnostic PCR for the *Accord* insertion and tandem duplication of the *Cyp6g1* gene. Agarose gel images of diagnostic PCR for the *Accord* element insertion and tandem duplication of the *Cyp6g1* gene.

Additional file 3: GO-term enrichment of genes over-expressed in the African and European populations. Table of all GO-terms with significant over-representation in each population.

Additional file 4: Correlation of fold-change expression differences as measured by microarray and qRT-PCR. Plot of 1,560 pairwise comparisons of all 16 *D. melanogaster* strains for 13 different genes.

Additional file 5: Meta-analysis of male and female between-population gene expression divergence. Table of differences in expression between the African and European populations for all genes common to the female and male experiments.

Acknowledgements

We thank Hedwig Gebhart and Carmen Iannitti for technical assistance and all members of the Munich population genetics group for comments and discussion. This work was carried out as part of the research unit "Natural selection in structured populations" (FOR 1078) funded by *Deutsche Forschungsgemeinschaft* grants STE 325/12-1 to WS and PA 903/5-1 to JP. RS was supported by a fellowship from the international research training program "RECESS: Regulation and evolution of cellular systems" (*Deutsche Forschungsgemeinschaft* GRK 1563).

Author details

[1]Department of Biology II, University of Munich (LMU), 82152 Planegg-Martinsried, Germany. [2]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA.

Authors' contributions

LM performed the microarray and qRT-PCR experiments. SH, RS and JP analyzed the microarray data. LM and SSS performed the population genetic analysis of DNA sequences. WS and JP conceived of the study, and participated in its design and coordination. LM and JP drafted the manuscript with input from all authors. All authors read and approved the final manuscript.

Received: 5 November 2010 Accepted: 28 January 2011
Published: 28 January 2011

References

1. Whitehead A, Crawford DL: **Variation within and among species in gene expression: raw material for evolution.** *Mol Ecol* 2006, **15**:1197-1211.
2. Clark TA, Townsend JP: **Quantifying variation in gene expression.** *Mol Ecol* 2007, **16**:2613-2616.
3. Ellegren H, Parsch J: **The evolution of sex-biased genes and sex-biased gene expression.** *Nat Rev Genet* 2007, **8**:689-698.
4. Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G: **The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*.** *Nat Genet* 2001, **29**:389-395.
5. Gibson G, Riley-Berger R, Harshman L, Kopp A, Vacha S, Nuzhdin S, Wayne M: **Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*.** *Genetics* 2004, **167**:1791-1799.
6. Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RRH, Mackay TF: **Systems genetics of complex traits in *Drosophila melanogaster*.** *Nat Genet* 2009, **41**:299-307.
7. Wayne ML, Pan YJ, Nuzhdin SV, McIntyre LM: **Additivity and trans-acting effects on gene expression in male *Drosophila simulans*.** *Genetics* 2004, **168**:1413-1420.
8. Baker DA, Meadows LA, Wang J, Dow JA, Russell S: **Variable sexually dimorphic gene expression in laboratory strains of *Drosophila melanogaster*.** *BMC Genomics* 2007, **8**:454.
9. David JR, Capy P: **Genetic variation of *Drosophila melanogaster* natural populations.** *Trends Genet* 1988, **4**:106-111.

10. Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P: **Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations.** *Genome Res* 2005, **15**:790-799.
11. Ometto L, Glinka S, De Lorenzo D, Stephan W: **Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation.** *Mol Biol Evol* 2005, **22**:2119-2130.
12. Thornton K, Andolfatto P: **Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*.** *Genetics* 2006, **172**:1607-1619.
13. Stephan W, Li H: **The recent demographic and adaptive history of *Drosophila melanogaster*.** *Heredity* 2007, **98**:65-68.
14. Townsend JP, Rand DM: **Mitochondrial genome size variation in New World and Old World populations of *Drosophila melanogaster*.** *Heredity* 2004, **93**:98-103.
15. Harr B, Kauer M, Schlötterer C: **Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 2002, **99**:12949-12954.
16. Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D: **Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach.** *Genetics* 2003, **165**:1269-1278.
17. Li H, Stephan W: **Inferring the demographic history and rate of adaptive substitution in *Drosophila*.** *PLoS Genet* 2006, **2**:e166.
18. Hutter S, Saminadin-Peter SS, Stephan W, Parsch J: **Gene expression variation in African and European populations of *Drosophila melanogaster*.** *Genome Biol* 2008, **9**:R12.
19. Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Oliver B: **Paucity of genes on the *Drosophila* X chromosome showing male-biased expression.** *Science* 2003, **299**:697-700.
20. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL: **Sex-dependent gene expression and evolution of the *Drosophila* transcriptome.** *Science* 2003, **300**:1742-1745.
21. Wayne ML, Telonis-Scott M, Bono LM, Harshman L, Kopp A, Nuzhdin SV, McIntyre LM: **Simpler mode of inheritance of transcriptional variation in male *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 2007, **104**:18577-18582.
22. Lemos B, Araripe LO, Hartl DL: **Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences.** *Science* 2008, **319**:91-93.
23. Jiang PP, Hartl DL, Lemos B: **Y not a dead end: epistatic interactions between Y-linked regulatory polymorphisms and genetic background affect global gene expression in *Drosophila melanogaster*.** *Genetics* 2010, **186**:109-118.
24. Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W: **Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data.** *Genetics* 2007, **177**:469-480.
25. Parsch J, Zhang Z, Baines JF: **The influence of demography and weak selection on the McDonald-Kreitman test: an empirical study in *Drosophila*.** *Mol Biol Evol* 2009, **26**:691-698.
26. Townsend JP: **Resolution of large and small differences in gene expression using models for the Bayesian analysis of gene expression levels and spotted DNA microarrays.** *BMC Bioinformatics* 2004, **13**:1-13.
27. Gnad F, Parsch J: **Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression.** *Bioinformatics* 2006, **22**:2577-2579.
28. Meiklejohn CD, Parsch J, Ranz JM, Hartl DL: **Rapid evolution of male-biased gene expression in *Drosophila*.** *Proc Natl Acad Sci USA* 2003, **100**:9894-9899.
29. Baker DA, Russell S: **Gene expression during *Drosophila melanogaster* egg development before and after reproductive diapause.** *BMC Genomics* 2009, **10**:242.
30. Carpenter J, Hutter S, Baines JF, Roller J, Saminadin-Peter SS, Parsch J, Jiggins FM: **The transcriptional response of *Drosophila melanogaster* to infection with the sigma virus (Rhabdoviridae).** *PLoS One* 2009, **4**:e6838.
31. Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Heckel D, Batterham P, Feyereisen R, Wilson TG, ffrench-Constant RH: **A single p450 allele associated with insecticide resistance in *Drosophila*.** *Science* 2002, **297**:2253-2256.
32. Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH, Schlotterer C: **World-wide survey of an *Accord* insertion and its association with DDT resistance in *Drosophila melanogaster*.** *Mol Ecol* 2004, **13**:2491-2504.
33. Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, Bogwitz MR, Martin J, Daborn PJ, Goddard ME, Batterham P, Robin C: **Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus.** *PLoS Genet* 2010, **6**:e1000998.
34. Buch S, Melcher C, Bauer M, Katzenberger J, Pankratz MJ: **Opposing effects of dietary protein and sugar regulate a transcriptional target of *Drosophila* insulin-like peptide signaling.** *Cell Metab* 2008, **7**:321-332.
35. Hull-Thompson J, Muffat J, Sanchez D, Walker DW, Benzer S, Ganfornina MD, Jasper H: **Control of metabolic homeostasis by stress signaling is mediated by the lipocalin NLaz.** *PLoS Genet* 2009, **5**: e1000460.
36. Collett JI, Jarman MG: **Adult female *Drosophila pseudoobscura* survive and carry fertile sperm through long periods in the cold: populations are unlikely to suffer substantial bottlenecks in overwintering.** *Evolution Int J Organic Evolution* 2001, **55**:840-845.
37. de Belle JS, Hilliker AJ, Sokolowski MB: **Genetic localization of *foraging* (*for*): a major gene for larval behavior in *Drosophila melanogaster*.** *Genetics* 1989, **123**:157-163.
38. Pereira HS, Sokolowski MB: **Mutations in the larval *foraging* gene affect adult locomotory behavior after feeding in *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 1993, **90**:5044-5046.
39. Shaver SA, Varnam CJ, Hilliker AJ, Sokolowski MB: **The *foraging* gene affects adult but not larval olfactory-related behavior in *Drosophila melanogaster*.** *Behav Brain Res* 1998, **95**:23-29.
40. Cohen J: *Statistical power analysis for the behavioral sciences*. 2 edition. Lawrence Erlbaum Associates, Hillsdale, NJ; 1988.
41. Parsch J, Novozhilov S, Saminadin-Peter SS, Wong KM, Andolfatto P: **On the utility of short intron sequences as a reference for the detection of positive and negative selection in *Drosophila*.** *Mol Biol Evol* 2010, **27**:1226-1234.
42. McDonald JH, Kreitman M: **Adaptive protein evolution at the *Adh* locus in *Drosophila*.** *Nature* 1991, **351**:652-654.
43. Wu CI, Hollocher H, Begun DJ, Aquadro CF, Xu Y, Wu ML: **Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation.** *Proc Natl Acad Sci USA* 1995, **92**:2519-2523.
44. Hollocher H, Ting CT, Wu ML, Wu CI: **Incipient speciation by sexual isolation in *Drosophila melanogaster*: extensive genetic divergence without reinforcement.** *Genetics* 1997, **147**:1191-1201.
45. Michalak P, Malone JH, Lee IT, Hoshino D, Ma D: **Gene expression polymorphism in *Drosophila* populations.** *Mol Ecol* 2007, **16**:1179-1189.
46. Wang HY, Fu Y, McPeek MS, Lu X, Nuzhdin S, Xu A, Lu J, Wu ML, Wu CI: **Complex genetic interactions underlying expression differences between *Drosophila* races: analysis of chromosome substitutions.** *Proc Natl Acad Sci USA* 2008, **105**:6362-6367.
47. Pool JE, Aquadro CF: **History and structure of sub-Saharan populations of *Drosophila melanogaster*.** *Genetics* 2006, **174**:915-929.
48. Nunes MD, Neumeier H, Schlötterer C: **Contrasting patterns of natural variation in global *Drosophila melanogaster* populations.** *Mol Ecol* 2008, **17**:4470-4479.
49. Pool JE, Aquadro CF: **The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*.** *Mol Ecol* 2007, **16**:2844-2851.
50. Rainer J, Sanchez-Cabo F, Stocker G, Sturn A, Trajanoski Z: **CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis.** *Nucleic Acids Res* 2006, **34**:W498-503.
51. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5**:R80.
52. Townsend JP, Hartl DL: **Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple strains or treatments.** *Genome Biol* 2002, **3**:RESEARCH0071.
53. Zheng Q, Wang XJ: **GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis.** *Nucleic Acids Res* 2008, **36**:W358-363.
54. Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, Fujita PA, Diekhans M, Smith KE, Rosenbloom KR, Raney BJ, Pohl A, Pheasant M, Meyer LR, Learned K, Hsu F, Hillman-Jackson J, Harte RA, Giardine B, Dreszer TR, Clawson H, Barber GP, Haussler D, Kent WJ: **The UCSC genome browser database: update 2010.** *Nucleic Acids Res* 2010, **38**:D613-619.

55. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Royal Stat Soc B* 1995, **57**:289-300.
56. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25**:1451-1452.
57. R Development Core Team: **R: A language and environment for statistical computing, reference index version 2.2.1.** Vienna, Austria: R Foundation for Statistical Computing; 2005.

doi:10.1186/1471-2164-12-81
Cite this article as: Müller *et al.*: Population transcriptomics of *Drosophila melanogaster* females. *BMC Genomics* 2011 **12**:81.

ELUCIDATING THE GENETIC AND EVOLUTIONARY MECHANISMS UNDERLYING DIFFERENTIAL GENE EXPRESSION IN TWO POPULATIONS OF *D. MELANOGASTER*

Lena Müller and John Parsch

(Unpublished manuscript)

Changes in gene expression have been shown to play an important role in adaptive evolution. Thus, they are expected to contribute to adaptation to new environmental conditions.

Here, we investigated several genes that show differential expression between an ancestral African and a derived European population of *D. melanogaster*. In order to elucidate the evolutionary and genetic mechanisms underlying the expression changes, molecular variation of coding and upstream regions of candidate genes was analyzed. For *CG34330*, a gene involved in neurogenesis, there is evidence for a recent selective sweep that has occurred in the European population. We identified a 1.64 kb region devoid of DNA sequence polymorphism in the European population. The monomorphic region spans along parts of the upstream region, the coding region, and the downstream region. Within the monomorphic upstream region there is one indel and one SNP where a derived variant is fixed in Europe, but at low frequency in Africa, suggesting that these variants are candidates for controlling the gene's expression level. For another gene, *Jon99Ciii*, evidence for recurrent adaptive structural protein evolution since the split of *D. melanogaster* from *D. simulans* and *D. sechellia* was found, while there was no evidence for recent adaptive regulatory evolution.

3.1 INTRODUCTION

Adaptive evolution can occur through the positive selection of mutations in coding regions that alter the function or the structure of a protein, or regulatory mutations that modify the timing, location or level of gene expression (Hoekstra and Coyne 2007, Wray 2007). Most studies to date have investigated the evolution of protein-coding regions, which is mainly a consequence of the relative ease with which one can distinguish neutral from non-neutral mutations in coding regions. Examples of adaptive protein evolution are plentiful (reviewed in Hoekstra and Coyne 2007) and studies investigating the proportion of amino acid replacements that are adaptive have yielded estimates as high as 45-95% among *Drosophila* species (Fay *et al*. 2002, Smith and Eyre-Walker 2002, Sawyer *et al*. 2007, Maside and Charlesworth 2007, Haddrill *et al*. 2008, Bachtrog 2008, Grath *et al*. 2009).

For non-coding regions, in contrast, it is difficult to predict functional regulatory nucleotides due to the lack of strict conservation of transcription factor binding sites (TFBS) (Martin *et al*. 1988, Magoulas *et al*. 1993, Tortiglione and Bownes 1997, Wolff *et al*. 1999, Ludwig *et al*. 2000, Dowell 2010) and the cumbersome work required to reveal their function. Thus, fewer examples of regulatory adaptation have been published to date. Still, studies have shown that the number of conserved non-coding nucleotides is roughly the same as the number of conserved coding nucleotides when comparing different taxa (Bergman and Kreitman 2001, Frazer *et al*. 2001, Shabalina *et al*. 2001, Keightley and Gaffney 2003) and there is evidence for selective constraints and positive selection acting on non-coding regions of the genome in *Drosophila* (Kohn *et al*. 2004, Andolfatto 2005, Andolfatto 2008, Haddrill *et al*. 2008). These findings imply that many noncoding nucleotides are indeed of functional importance, and might be involved in gene expression regulation. Although the location and physical extent of promoter regions varies greatly among genes, TFBS are often located within a few kb upstream of the basal promoter (Wray *et al*. 2003). Over the last decade, more and more examples of adaptive evolution in promoter regions have been identified (*e.g*., Tournamille *et al.* 1995, Hamblin and Di Rienzo 2000, Gompel *et al.* 2005, Prud'homme *et al.* 2006, Wray *et al.* 2007, Chan *et al.* 2010), suggesting that regulatory changes represent an important target of adaptive evolution.

Mutations altering the regulation of gene expression can occur in two general categories of regulatory DNA: *trans*-regulatory elements that are not physically linked to the gene they

regulate and *cis*-regulatory elements that are physically linked to the gene they regulate. *Trans*-regulatory elements typically encode transcription factors (TF) or microRNAs (miRNA), while *cis*-regulatory elements are often TFBS. Several TFBS can act together to form an enhancer. *Cis*-regulatory elements show several characteristics that make them good targets for natural selection. First, many enhancers act independently of one another, each controlling one aspect of the expression spectrum, such as tissue specificity or timing of expression (DiLeone 1998, Jeong *et al*. 2008). This modularity limits the adverse effects of pleiotropy, a phenomenon expected to be much more common for *trans*-regulatory and protein-coding mutations. Second, *cis*-regulatory mutations are more often co-dominant than *trans*-regulatory changes (Lemos *et al*. 2008, McManus *et al*. 2010). As a consequence of co-dominance, mutations in *cis*-regulatory elements are visible to selection immediately in heterozygotes. This implies that selection can act more efficiently on *cis*-regulatory than on *trans*-regulatory mutations or on mutations in protein-coding regions, which are mainly recessive (Li 1997). Indeed, it has been shown that positive selection acts on the turnover of TFBS, while purifying selection acts to maintain functional TFBS in *D. melanogaster* and *D. simulans* (He *et al*. 2011).

Drosophila melanogaster is a cosmopolitan species today, but its origin is thought to lie in sub-Saharan Africa (Lachaise *et al*. 1988, David and Capy 1988) from where it expanded its habitat to Europe around 10,000 - 15,000 years ago. Originally coming from a sub-tropical environment, the out-of-Africa movement of the species was presumably accompanied by adaptation to the new European environment, where derived populations are exposed to new biotic and abiotic factors, such as climate, food sources, and pathogens. Large scale genome scans of two *D. melanogaster* populations, one coming from the ancestral species range in Africa (Zimbabwe) and one from the derived species range in Europe (the Netherlands) have identified regions that may have been targets of positive selection (Ometto *et al*. 2005, Li and Stephan 2006) and there is evidence for recent selective sweeps in several genes and genomic regions in these populations (Beisswanger *et al*. 2006, Glinka *et al*. 2006, Beisswanger and Stephan 2008, Svetec *et al*. 2009).

Two studies, one by Hutter *et al*. (2008) and one by Müller *et al*. (2011), investigated genome-wide expression variation in these African and European populations of *D. melanogaster* using male and female flies, respectively. Their findings show that differences in gene expression are abundant between these populations. In males, 153 genes showed differential expression between the populations, while in females 569 genes showed such a

pattern. These genes are candidates for those that have undergone adaptive regulatory evolution in response to local environmental conditions. In both male and female flies, the gene showing the greatest over-expression in Europe was the insecticide-resistance gene *Cyp6g1*, a well-known example of adaptive evolution in *D. melanogaster*. It has been shown that the insertion of an *Accord* transposable element 291 bp upstream of the transcription start site of *Cyp6g1* is associated with over-expression of the gene and drives its expression in detoxification tissues (Daborn *et al*. 2002*,* Catania *et al*. 2004*,* Chung *et al*. 2007). This insertion was present in all European fly strains, but in only four of the eight African fly strains (Müller *et al*. 2011), suggesting that it can account for the observed expression differences between the populations. Evidence for recent positive selection in the European population has also been found for the gene *CG9509*, which showed the second highest over-expression in Europe in male flies. The target of selection could be narrowed down to a 1.2-kb region just upstream of the protein-coding region of the gene (Saminadin-Peter *et al*. 2011).

Here, a candidate gene approach and population genetic analyses are used to look for signatures of recent or recurrent selection in seven genes and their adjacent upstream regions. In addition, two promoter regions were investigated without sequencing their adjacent genes. Candidate genes were taken from the study by Müller *et al*. (2011) and belong to those genes that show differential expression between the ancestral African population and the derived European population. In order to look for mutations in the putative promoter regions that might be responsible for the observed changes in expression level, around 1-2 kb of the upstream region directly flanking the genes were sequenced. Sequence variants in the upstream regions that are private to one population or that are present at high frequency in only one population are good candidates for regulatory polymorphisms that control the expression of the target gene.

For *CG34330*, a gene involved in neurogenesis, we identified a 1.64 kb region devoid of DNA sequence polymorphism in the European population – a finding that is suggestive of a recent selective sweep in the derived population. The monomorphic region spans 499 bp of the upstream region of *CG34330*, the whole coding region, as well as 651 bp of the downstream region. In the monomorphic upstream region of *CG34330*, there is one SNP and one indel where a derived variant is fixed within the European population but at low frequency (18%) in the African population.

3.2 MATERIALS & METHODS

3.2.1 FLY STRAINS

Sequences were obtained from 12 isofemale fly strains of a European population (Leiden, the Netherlands) and 11 isofemale fly strains of an African population (Lake Kariba, Zimbabwe) of *D. melanogaster*. The populations are as described in Glinka *et al*. (2003). Fly strains were the same as those used in the expression analysis of Hutter *et al*. (2008) and Müller *et al*. (2011). Flies were kept at 22°C on standard cornmeal-molasses medium.

3.2.2 DNA EXTRACTION

Genomic DNA of single male flies was extracted using MasterPure™ DNA Purification Kit (Epicentre® Biotechnologies, Madison, WI, USA) following to the manufacturer's protocol for tissue samples. Quality and quantity of extracted DNA were subsequently assessed by gel electrophoresis and using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). If clear high-molecular weight bands were obtained on the gel that did not show any signs of degradation (smear), the ratio of sample absorbance at 260nm/280 nm was ~1.8, and the absorbance ratio at 260 nm /230 nm was 1.8-2.2, DNA was diluted with H_2O to a final concentration of 150 ng/µl and stored at -20°C until used.

3.2.3 PCR

Polymerase chain reaction (PCR) was performed for sequences using peqlab pegGOLD *Taq*-DNA-polymerase (PEQLAB Biotechnologie GMBH, Erlangen) and standard PCR conditions. Primers were designed based on the complete *D. melanogaster* genome sequence (www.flybase.org). For a complete list of primers, as well as standard PCR conditions, see supplemental material. PCR products were purified using Exo-SAP-IT® (United States Biochemical, Cleveland, OH, USA).

3.2.4 CLONING

For some of the autosomal genes that showed residual heterozygosity, PCR-products were

cloned prior to sequencing. This was done using the Invitrogen TOPO TA Cloning® Kit (Invitrogen, Carlsbad, CA, USA) and according to the manufacturer's protocol for chemical transformation. Colony PCR was performed with eight transformed colonies per cloning reaction and PCR products were subsequently sequenced using the standard sequencing protocol. One randomly chosen allele each was included in the final alignment.

3.2.5 SEQUENCING

Purified PCR products were sequenced using Big Dye version 1.1 chemistry and a 3730 automated sequencer (Applied Biosystems, Foster City, CA, USA). Primers used for sequencing were the same as those used for PCR. In some cases, additional sequencing primers were designed (see supplemental material). For the European population, the number of alleles sequenced ranged from 9–12 with a mean of 11.7. For the African population, the number of alleles sequenced ranged from 8–11, with a mean of 10.2.

For divergence calculations, a single allele from *D. simulans* was used (Drosophila 12 Genomes Consortium 2007), except for genes *CG34330* and *fau* where a complete *D. simulans* sequence was not available. Here, *D. sechellia* was used as an outgroup (Drosophila 12 Genomes Consortium 2007).

3.2.6 DATA ANALYSIS

Sequences were edited using DNAstar (Madison, WI). Multiple alignments were generated using MUSCLE (Edgar 2004) implemented in the MEGA5 software (Tamura *et al.* 2011) and adjusted manually.

3.2.7 STATISTICAL METHODS

All summary statistics were calculated using DnaSP v.5 (Librado and Rozas 2009). Nucleotide diversity was measured as π (Tajima 1983), the average number of pairwise differences per site, and Watterson's θ_W (Watterson 1975), the average number of segregating sites per nucleotide position. Between-population genetic differentiation was measured as D_{XY}, the mean number of pairwise sequence differences between alleles of the two populations (Nei 1987) and F_{ST}, the proportion of genetic diversity due to allele frequency differences among populations (Hudson *et al.* 1992). Divergence (K) was calculated as the

average number of nucleotide substitutions per site between species using the Jukes and Cantor correction (Nei 1987). To test for deviations from the standard neutral model, Tajima's *D* (Tajima 1989), Fay and Wu's *H* (Fay and Wu 2000) and the McDonald-Kreitman test (McDonald and Kreitman 1991) were performed. Tajima's *D* statistic contrasts low-frequency and intermediate-frequency sites in a sample. A significantly negative value of Tajima's *D* indicates an excess of rare variants, which is expected under directional selection or population size expansion. A significantly positive value of Tajima's *D* indicates an excess of intermediate-frequency variants, a scenario expected under balancing selection or population subdivision or weak/incomplete bottlenecks (Ometto *et al*. 2005). Fay and Wu's *H* statistic compares high-frequency derived variants to intermediate-frequency variants. Here, a significantly negative value indicates an excess of high-frequency derived variants, which is expected after positive selection. A significantly positive *H* statistic, in contrast, points to an overrepresentation of intermediate-frequency derived polymorphism, a scenario expected under balancing selection.

The McDonald-Kreitman (MK) test compares the polymorphism-to-divergence ratio at synonymous sites to the polymorphism-to-divergence ratio at nonsynonymous sites. If there is an excess of nonsynonymous divergence relative to polymorphism, recurrent positive selection for amino acid replacement since the split of the species can be inferred. A relative excess of nonsynonymous polymorphism is indicative of balancing selection or possibly weak purifying selection. For noncoding sequences, the MK test compares the polymorphism-to-divergence ratio of synonymous or intronic sites to the polymorphism-to-divergence ratio of noncoding sites (Andolfatto 2005).

3.2.8 CHOICE OF CANDIDATE GENES

Candidate genes that were chosen for sequencing were taken from the set of differentially expressed genes between the African and the European population in female flies (Müller *et al*. 2011). Some general criteria were applied when choosing the genes: priority was given to genes with great over-expression in one population, short gene spans, sufficiently long upstream regions without overlap with other genes, and known functions. However, not all of the above criteria could be satisfied by all genes.

Four of the genes were found to be differentially expressed in both female and male flies (*CG34330*, *CG12262*, *CG7409*, and *fau*). For the genes *Pbprp2*, *oaf*, and *fau*, alternatively

spliced transcripts were found to be over-expressed in different populations. *Nipsnap* and *CG31058* were over-expressed in Africa, and *Jon99Ciii* was over-expressed in Europe.

3.2.9 SEQUENCE DIVERGENCE IN UPSTREAM REGIONS

D_{XY}, F_{ST}, π, and divergence (*K*) were calculated for upstream regions, including 5'UTRs, for all of the differentially expressed genes above and compared to values of a set of genes that did not show a significant expression difference between populations, hereafter referred to as control genes. The set of control genes and values of D_{XY}, F_{ST}, π, and *K* for upstream regions of the control genes were taken from Saminadin-Peter (2008). Genes that showed differential expression in female flies and regions that did not blast to upstream regions due to changes in gene annotation were excluded from this gene set. The final set of control genes consisted of four genes that showed no significant differential expression between the populations in male flies and in female flies. Mann-Whitney *U* tests were performed using R (version 2.10.1) (R Development Core Team 2005).

3.3 RESULTS

All genes investigated in this study showed differential expression in adult female flies between a European and an African population of *D. melanogaster* in a genome-wide microarray study (Müller *et al*. 2011, chapter 1) (Table 3.1). Some of the genes were also found to be differentially expressed between the populations in male flies (Hutter *et al*. 2008).

In order to investigate if between-population expression differences could be caused by adaptive changes in *cis*-regulatory elements, the coding and upstream regions of seven candidate genes were sequenced in a sample of up to 12 fly strains from each population, including the strains used in the microarray expression analyses. The upstream regions of two additional candidate genes were also investigated, however the coding regions of these genes were not sequenced. (Table 3.1, Supplementary Figure 1).

TABLE 3.1 **All genes sequenced in this study**

				Expression ratio A/E		*P*-value		Nucleotides sequenced			Fly strains			
CG number	Symbol	Chr.	Transcript	Female [a]	Male [b]	Female [a]	Male [b]	Entire	Upstream	Gene region	EUR	AFR	Function	Biological process
CG34330		X		2.01	1.32	0.0004	0.0001	4351	2148	656	12	11	unknown	Neurogenesis
CG12262		3L		0.78	0.78	0.0013	0.0003	4017	1456	2190	12	10	Acyl-CoA dehydrogenase	Fatty acid beta-oxidation
CG7409		3L		1.89	1.4	0.0047	0.0006	1979	1017	788	12	9	Unfolded protein binding	Protein folding
CG6544	*fau*	3R						1126	1126		12	11	unknown	Response to anoxia
			RB	1.48	1.3	0	0.0016							
			RC	1.43		0.0221								
			RD	0.69		0.0027								
CG1668	*Pbprp2*	X						2185	662	1417	12	11	Pheromone/odorant binding	Sensory perception of chemical stimulus
			RA	0.75		0.0154								
			RB	1.89		0.0038								
CG9884	*oaf*	2L						1124	341		12	11	unknown	Female meiosis chromosome segregation
			RA	0.81		0.0036								
			RB	1.37		0.003								
CG9212	*Nipsnap*	X		1.46		0.0039		4915	1503	3438	12	11	unknown	unknown
CG31058		3R		2.89		0		2885	1523	1362	12	9	unknown	Lateral inhibition
CG31362	*Jon99Ciii*	3R		0.43		0		2376	1688	688	9	8	Endopeptidase	Digestion; Proteolysis

Shown are the chromosomal locations (Chr.), expression ratios of Africa/Europe and their corresponding significance, the number of nucleotides sequenced for the entire region (Entire), the upstream region (Upstream), and the gene region, the number of fly strains sequenced for each population, gene function and the biological process for each gene.

[a] Data from Müller *et al*. (2011)

[b] Data from Hutter *et al*. (2008)

3.3.1 GENES ANALYZED

Of the nine candidate genes, four were identified to be differentially expressed between Europe and Africa in male and female flies (Table 3.1). *CG34330*, which was over- expressed 2-fold in females and 1.3-fold in males in the African population has been shown to be involved in neurogenesis, although its specific function is unknown. *CG12262*, which was over-expressed 1.28-fold in Europe in females and 1.29-fold in Europe in males, has an annotated function in oxidation/reduction and fatty acid metabolism. *CG7409*, a gene with 1.89-fold higher expression in African female flies and 1.40-fold higher expression in African male flies has an annotated function in response to heat and unfolded protein binding. *fau*, which has an annotated function in recovery from anoxia (O_2 depletion), showed 1.48-fold over-expression for the transcript *fau-RB* in Africa in females and a 1.3-fold over-expression in Africa in males. In female flies, a second transcript, *fau-RD*, showed a 1.44-fold higher expression in Europe at a *P*-value cutoff of 0.005.

For the genes *Pbprp2*, which is predicted to function in pheromone and odorant binding, and *oaf*, which is involved in chromosome segregation during female meiosis, alternatively spliced transcripts were found to be over-expressed in different populations. *Pbprp2-RA* was 1.33-fold over-expressed in Europe ($P < 0.02$), while *Pbprp2-RB* showed a 1.89-fold over-expression in Africa. For *oaf*, the transcript *oaf-RA* was over-expressed 1.24-fold in Europe,

while *oaf-RB* showed 1.37-fold over-expression in Africa. These are candidates for those genes where alternatively spliced transcripts are preferentially expressed in the two populations.

CG31058, which is involved in lateral inhibition, and *Nipsnap*, a gene with unknown function, were over-expressed in Africa 2.89-fold and 1.46-fold, respectively. *Jon99Ciii*, a serine-type endopeptidase involved in digestion and proteolysis, was over-expressed 2.31-fold in Europe.

3.3.2 DNA SEQUENCE POLYMORPHISM AND NEUTRALITY TESTS

For all sequenced regions, the per-nucleotide estimate of sequence diversity, θ_W, was determined using a sliding-window approach (Supplementary Figure 1). Most of the genes showed average levels of nucleotide diversity (Hutter *et al*. 2007). Applying neutrality tests such as Tajima's *D*, Fay and Wu's *H* and the McDonald-Kreitman (MK) test did not show any deviations from neutrality for most genes (Tables 3.2, 3.3, 3.4) in coding or upstream regions.

TABLE 3.2 **Results of Tajima's *D* tests**

		Tajima's *D*				
Gene	Population	all sites	upstream	coding	syn.	nonsyn.
CG34330	Europe	-1.390	-1.567	n.a.	n.a.	n.a.
	Africa	-0.636	-0.498	n.a.	n.a.	n.a.
CG12262	Europe	-0.468	-1.154	-0.064	-0.064	n.a.
	Africa	0.412	1.137	-0.073	-0.073	n.a.
CG7409	Europe	0.463	0.334	0.229	0.229	n.a.
	Africa	-0.330	-0.744	-0.740	-0.740	n.a.
fau	Europe	0.756	0.756	n.a.	n.a.	n.a.
	Africa	0.217	0.217	n.a.	n.a.	n.a.
Pbprp2	Europe	-0.486	0.253	-1.291	-0.829	-1.451
	Africa	-1.111	-1.442	-0.159	-0.159	n.a.
oaf	Europe	-0.148	-1.067	n.a.	n.a.	n.a.
	Africa	0.180	0.696	n.a.	n.a.	n.a.
Nipsnap	Europe	1.714	0.563	2.196 *	2.196 *	n.a.
	Africa	-0.390	-0.819	-0.720	-0.720	n.a.
CG31058	Europe	0.064	0.162	0.095	0.276	-0.129
	Africa	-0.337	-1.040	0.155	0.133	0.176
Jon99Ciii	Europe	-0.928	-0.914	-0.896	-0.896	n.a.
	Africa	-0.197	-0.213	-0.075	-0.358	0.585

Shown are the results of Tajima's *D* tests for all, upstream, coding, synonymous (syn.) and nonsynonymous (nonsyn.) sites for each population. Upstream regions include 5'UTRs of genes, except for *fau* and *oaf*. If there was more than one variant of the 5'UTR, the longest one was included in the analysis.

*$P < 0.05$

An exception was *CG34330*. A 4.3 kb region of the X chromosome was sequenced, encompassing the entire coding region, ~ 2.1 kb of the 5' flanking region and ~ 1.5 kb of the 3' flanking region (Table 3.1). Along the upstream region, Fay and Wu's *H* statistic indicated

a significant excess of high frequency derived polymorphisms for the European population, a pattern in accordance with a recent selective sweep in this population (Table 3.3). Across the entire region, the estimate of sequence diversity per site, θ_W, in the European population was 0.0029, which is lower than the X-chromosomal mean value of θ_W (0.0047) in this population (Hutter *et al*. 2007). Interestingly, we found a 1.64 kb region completely depleted of polymorphism in the European population that spans 499 bp of the upstream region of *CG34330*, the whole coding region, as well as 651 bp of the downstream region (Figure 3.1). For the African population, no such reduction in nucleotide diversity could be found. Instead, this region shows normal levels of θ_W in Africa, as well as normal levels of divergence when comparing *D. melanogaster* to the outgroup *D. sechellia*. This indicates that neither extremely high selective constraint nor an exceptionally low mutation rate can account for the reduced nucleotide diversity in the European population. Such a finding is in accordance with a recent selective sweep in the European population.

TABLE 3.3 **Results of Fay and Wu's *H* tests**

		Fay and Wu's *H*				
Gene	Population	all sites	upstream	coding	syn.	nonsyn.
CG34330	Europe	-23.515 */**/*** [a]	-24.061 **/**/***	n.a.	n.a.	n.a.
	Africa	-5.327	1.509	n.a.	n.a.	n.a.
CG12262	Europe	-1.939	0.303	-2.848	-2.848 /**/*	n.a.
	Africa	18.489	14.667	-0.178	-0.178	n.a.
CG7409	Europe	1.727	0.697	0.242	0.242	n.a.
	Africa	4.528	1.306	1.417	1.417	n.a.
fau	Europe	-3.606	-3.606			
	Africa	-2.564	-2.564			
Pbprp2	Europe	-5.818	-1.848	-2.030 *	-2.333 */***/***	0.30303
	Africa	-0.655	-1.400	1.073	1.073	n.a.
oaf	Europe	-2.091	1.152			
	Africa	0.873	-0.400			
Nipsnap	Europe	-8.061	-1.758	0.303	0.303	n.a.
	Africa	-2.727	0.436	2.073	2.073	n.a.
CG31058	Europe	1.000	2.758	-1.758	-2.788 /**/**	0.727
	Africa	-5.056	0.444	-5.500	-3.694 /**/**	-0.528
Jon99Ciii	Europe	-7.250	-5.083	-2.167	-2.167 /**/**	n.a.
	Africa	0.286	2.143	-1.857	-2.5 /*/**	0.64286

Shown are the results of Fay and Wu's *H* tests for all, upstream, coding, synonymous (syn.) and nonsynonymous (nonsyn.) sites for each population. *P*-values were determined running coalescent simulations for a neutral infinite-sites model, assuming constant population size and no recombination. Upstream regions include 5'UTRs of genes, except for *fau* and *oaf*. If there was more than one variant of the 5'UTR, the longest one was included in the analysis.

[a] Different values of θ_W per site were used for simulations: average over the investigated region/ chromosomal average/ average over the whole sequenced region. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Within the monomorphic region in the European population, one indel and two SNPs could be identified that were fixed derived variants in the European population, but at low frequency in the African population (Table 3.5).

For another gene, *Jon99Ciii*, the MK test provided evidence for recurrent positive selection in Europe acting on nonsynonymous sites of the coding region (Table 3.4). This result was robust to whether *D. melanogaster* was compared to *D. simulans* or *D. sechellia*. In addition, a significant MK test was also found for upstream sites in the African population.

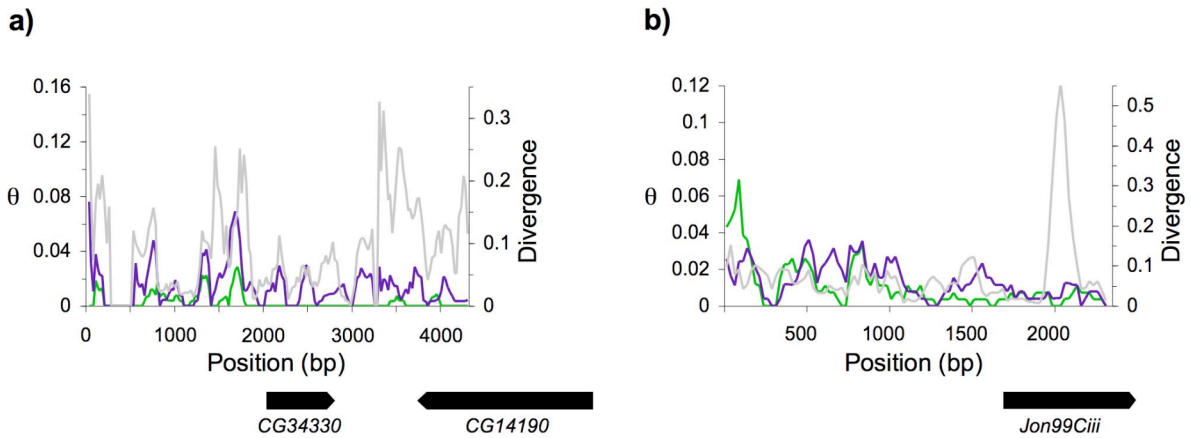


FIGURE 3.1 Polymorphism and divergence in the genomic regions of *CG34330* and *Jon99Ciii*

Sliding window analyses of polymorphism and divergence for a) *CG34330* and b) *Jon99Ciii* (window size = 100 bp, step size = 25 bp). Shown are nucleotide diversity (θ) in the European (green line) and the African (purple line) populations, as well as divergence between all *D. melanogaster* lines and an outgroup (gray line). *D. sechellia* was used as an outgroup for *CG34330*, *D. simulans* for *Jon99Ciii*. Positions of the genes are represented by arrows with the arrowhead indicating the direction of transcription.

3.3.3 SEQUENCE DIVERGENCE IN UPSTREAM REGIONS

Upstream regions of significantly differentially expressed genes were compared to those from genes that do not show any difference in expression between populations (control genes). If expression differences are due to DNA sequence variation within *cis*-regulatory elements, one might expect differentially expressed genes to harbor more sequence variation in their upstream regions. For these regions, the mean number of pairwise sequence differences between alleles of the two populations, D_{XY}, and the proportion of genetic diversity due to allele frequency differences among populations, F_{ST}, were examined. However, no significant differences between the two gene classes could be found (Figure 3.2).

Nucleotide diversity π and divergence (*K*) for each population were also investigated. In Africa, there were no significant differences for π and *K* among gene classes. In the European population, significantly higher values were found for differentially expressed genes versus control genes for π ($P < 0.05$, Mann-Whitney *U* test), while no significant differences could be found for divergence (*K*).

3.4 DISCUSSION

The sequence analysis of genes that showed differential expression between an African *D. melanogaster* population from Zimbabwe and a European population from the Netherlands did not show deviations from neutrality for most sequenced loci. However, there were two interesting exceptions.

TABLE 3.4 **Results of McDonald-Kreitman (MK) tests**

											MK test *P*-value		
Gene	Population	Outgroup	D_s	P_s	D_n	P_n	D_{int}	P_{int}	D_{up}	P_{up}	Nonsyn.	Upstream vs. syn.	Upstream vs. intronic
CG34330	Europe	*D. sec*	4	0	1	0			108	27	n.a.	0.587	
	Africa	*D. sec*	4	0	1	0			94	79	n.a.	0.129	
CG12262	Europe	*D. sim*	23	24	2	0	40	37	57	88	0.490	0.307	0.088
	Africa	*D. sim*	21	30	2	0	40	53	52	105	0.184	0.314	0.136
CG7409	Europe	*D. sim*	12	6	0	0			39	36	n.a.	0.302	
	Africa	*D. sim*	12	6	0	0			41	24	n.a.	1.000	
Pbprp2	Europe	*D. sim*	6	3	5	2	17	5	22	5	1	0.384	0.737
	Africa	*D. sim*	6	5	5	0	17	10	23	18	0.119	1.000	0.622
Nipsnap	Europe	*D. sim*	8	4	3	0	143	49	47	6	0.516	0.078	0.027 **
	Africa	*D. sim*	8	10	3	0	137	84	45	11	0.214	0.006 **	0.011 *
CG31058	Europe	*D. sim*	22	23	23	18			32	33	0.504	1.000	
	Africa	*D. sim*	22	24	24	17			33	29	0.391	0.697	
Jon99Ciii	Europe	*D. sim*	25	9	30	0			81	55	0.002 **	0.167	
	Africa	*D. sim*	24	8	30	3			76	75	0.108	0.012 *	
	Europe & Africa	*D. sim*	24	13	30	3			75	98	0.011 *	0.019 *	
Jon99Ciii	Europe	*D. sec*	9	9	8	0			85	51	0.023 *	0.441	
	Africa	*D. sec*	8	8	6	3			81	69	0.677	0.797	
	Europe & Africa	*D. sec*	8	13	30	3			80	92	0.236	0.048 *	

Shown are the number of fixed differences (D) between *D. melanogaster* and the respective outgroup and the number of polymorphic sites (P) within the populations. Subscripts indicate synonymous (s), nonsynonymous (n), intronic (int), and upstream (up) sites. Upstream regions include 5'UTRs of genes, except for *fau* and *oaf*. If there was more than one variant of the 5'UTR, the longest one was included in the analysis. Only constitutively spliced out introns were considered for intronic sites.

Nonsyn = nonsynonymous sites; syn = synonymous sites

* $P < 0.05$ (Fisher's exact test, FET), ** $P < 0.01$ (FET)

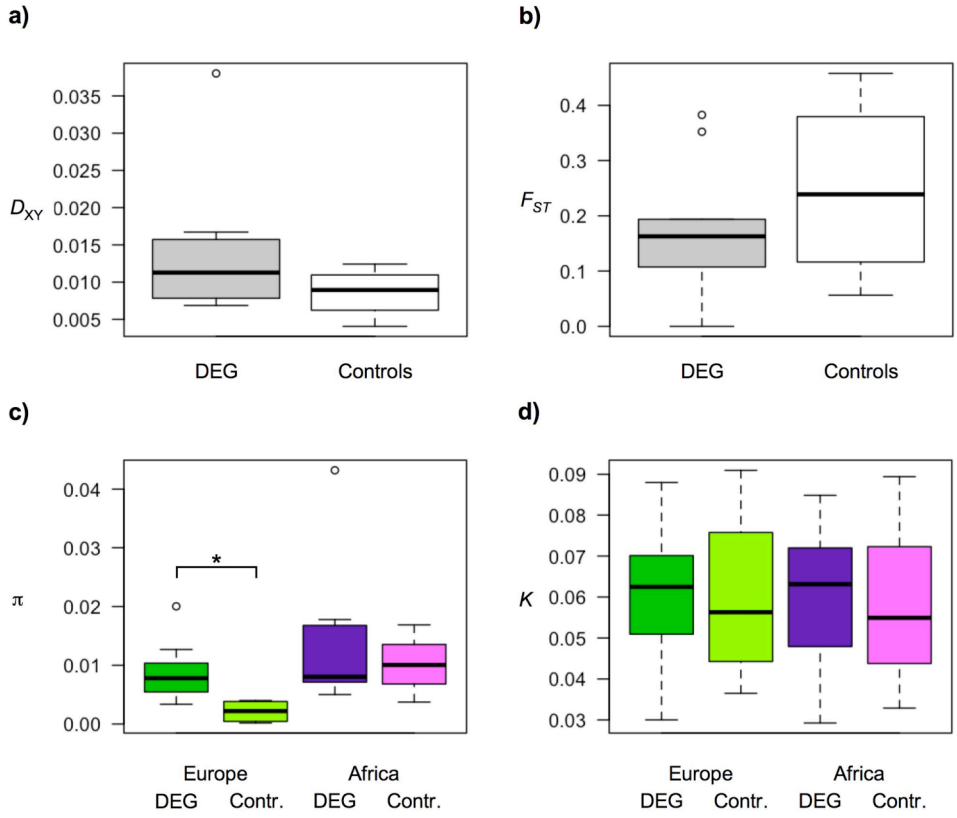


FIGURE 3.2 Comparison of upstream regions from differentially expressed genes (DEG) and genes without differential expression (controls = contr.)

a) D_{XY} between the European and the African population, b) F_{ST} between the European and the African population c) Nucleotide diversity π for each population, c) Divergence (K) for each population.

Brackets indicate the only significant difference which was found for π of differentially expressed genes vs. controls in Europe ($P < 0.05$, Mann-Whitney U test).

3.4.1 EVIDENCE FOR A SELECTIVE SWEEP IN THE GENOMIC REGION OF *CG34330*

The gene *CG34330*, which lies in the X chromosome and is involved in neurogenesis shows a significant over-expression in the African relative to the European population in male and female *D. melanogaster* (Hutter *et al*. 2008, Müller *et al*. 2011, Table 3.1). The extent of the over-expression, 1.3-fold in males and 2-fold in females, differs between the sexes, however this is not surprising given that the vast majority of all expression differences between these populations was found to be sex-specific. Several findings suggest that the gene has been a target of recent adaptive evolution in the European population. For the upstream region of *CG34330*, a highly significant negative value for Fay and Wu's H is found in Europe (Table 3.3), which implies an excess of derived polymorphism in high frequency

along this region. This finding is consistent with the recent action of positive selection, which has driven these variants to high frequencies in Europe. Since the coding region of the gene does not show any polymorphism in the European and the African population, the MK test could not be applied. Along the whole sequenced region as well as along the upstream region only, negative Tajima's *D* values point to an excess of low-frequency polymorphism (Table 3.2), a finding expected after a recent selective sweep.

Interestingly, there is a strong reduction in nucleotide diversity in the European population over the whole sequenced region when compared to the chromosomal average, which is especially pronounced along a 1.64 kb region where no variation is found among all derived European lines (Figure 3.1). This region includes 499 bp of the upstream region of *CG34330*, the whole coding region, as well as 651 bp of the downstream region. Two findings suggest that the reduced variation is neither a result of high selective constraints acting on this region nor due to an exceptionally low mutation rate: First, in the African population, this fragment shows normal levels of nucleotide diversity and second, divergence between *D. melanogaster* and *D. sechellia* is normal in this region.

According to the *Drosophila melanogaster* Recombination Rate Calculator (RRC) (Singh *et al*. 2005, Fiston-Lavier *et al*. 2010), the recombination rate for the chromosomal region in which *CG34330* is embedded is relatively low (1.83 cM/Mbp,) compared to the X-chromosomal average (3.09 cM/Mbp) or the genome-wide average (2.46 cM/Mbp), such that one might expect a larger area to be affected by the selective sweep. This suggests that the sweep either occurred relatively far in the past or the selection coefficient was relatively low. However, since nucleotide variation is still low in the sequenced portion of the neighboring gene *CG14190* in Europe (θ_W = 0.00115), footprints of the incomplete sweep might indeed extend further along *CG14190*. However, patterns of low polymorphism can also be caused by a reduction in population size. The European population used in this study is known to have undergone a bottleneck during the out-of-Africa movement (Ometto *et al.* 2005, Li and Stephan 2006). To be able to distinguish selection from demography, coalescent simulations with and without selection, explicitly modeling such a bottleneck, *i.e*. taking the demographic parameters from Li and Stephan (2006), should be performed.

If the gene's expression level is controlled by sequence variants adjacent to the gene, the upstream region is expected to harbor identifiable nucleotide differences between the populations. Within the monomorphic region detected in the European population, but nowhere else in the sequenced fragment, there are several variants that are fixed in Europe but

at low frequency in Africa (Table 3.5). Within the upstream region and the 5'UTR, there is one 6-bp indel and one SNP that have a derived variant fixed in the European population, but

TABLE 3.5 **Nearly fixed differences between the European and the African population for *CG34330***

	Position relative to start codon of *CG34330*						
Fly strains	-460	-126	-91	653	941	946	961
E01	**ATTCGC**	A	**C**	**C**	C	A	G
E02	**ATTCGC**	A	**C**	**C**	C	A	G
E11	**ATTCGC**	A	**C**	**C**	C	A	G
E12	**ATTCGC**	A	**C**	**C**	C	A	G
E13	**ATTCGC**	A	**C**	**C**	C	A	G
E14	**ATTCGC**	A	**C**	**C**	C	A	G
E15	**ATTCGC**	A	**C**	**C**	C	A	G
E16	**ATTCGC**	A	**C**	**C**	C	A	G
E17	**ATTCGC**	A	**C**	**C**	C	A	G
E18	**ATTCGC**	A	**C**	**C**	C	A	G
E19	**ATTCGC**	A	**C**	**C**	C	A	G
E20	**ATTCGC**	A	**C**	**C**	C	A	G
A84	**ATTCGC**	**C**	**C**	T	T	G	C
A95	**ATTCGC**	**C**	A	T	T	G	C
A131	-	**C**	A	T	T	G	C
A145	-	**C**	A	T	T	G	C
A157	-	**C**	A	T	T	G	C
A186	-	A	A	T	T	G	C
A191	-	**C**	A	T	T	G	C
A229	-	A	**C**	**C**	C	A	G
A377	-	**C**	A	T	T	A	G
A384	-	**C**	A	T	T	A	G
A398	-	**C**	A	T	T	G	C
D. sec	AT-	A	A	T	-	-	-
D. sim	-	A	A	T	-	-	-
D. yak	-	-	A	T	-	-	-
D. ere	-	A	A	T	-	-	-
D. ana	-	T	**C**	T	-	-	-
Portugal	**ATTCGC**	A	**C**	**C**	C	A	G
France	**ATTCGC**	A	**C**	**C**	**C**/T	**A**/G	**G**/C
Region	upstream	5'UTR		downstream			
Feature[a]		chinmo TFBS					

Shown are all differences that are fixed in Europe (E) but in low frequency in Africa (A). Derived variants are indicated in bold. Nucleotide variants for five outgroup species are also shown: *D. sechellia*, *D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae*. In addition, nucleotide states found in the population from Portugal (Pandey *et al*. 2011) and from France (S. Hutter, personal comm.) are shown. In France, the last three SNPs were not fixed. Instead, the predominant allele of the African population was found in two of the eight strains each.

[a] Data from Nègre *et al*. (2011)

they appear at low frequency (18%) in Africa. In addition, there is one SNP where the ancestral state is fixed in Europe, but occurs at low frequency in Africa (18%). These two SNPs are located in a TFBS of the TF chinmo, whose interaction has been verified in ChIP-

chip experiments during embryonic stages of *D. melanogaster* development (Nègre *et al*. 2011). However, whether this TF also binds during adult fly stages is unknown. It has been shown, though, that the emergence of new TFBS can result from changes in an already existing TFBS (Prud'homme *et al*. 2007), such that the presence of a TFBS provides a good opportunity for the formation of a new one. Downstream of the gene resides another derived SNP that is fixed in Europe, but has a low frequency in Africa (9%). Furthermore, there are three SNPs in the downstream region that are fixed in Europe and appear at low frequency in Africa (9%-27%), where the ancestral state could not be inferred since none of the outgroups aligned to the respective region. These sites, especially those with derived variants in Europe, are good candidates for sites that have been targets of positive selection and control the expression pattern of the gene.

Nevertheless, there is no clear correlation between the alternative states of the SNP variants and the expression level of the gene in each fly line as determined by the 16-node analyses of the microarray studies in male and female flies (Table 3.6).

Fly strain	Relative expression	
	Females[a]	Males[b]
E01	3.25	1.18
E12	2.19	1.00
E14	1.73	1.03
E15	1.00	1.35
E16	1.04	1.00
E17	2.20	1.33
E18	2.69	1.25
E20	1.48	1.23
A82	3.04	1.58
A84	4.91	1.35
A95	2.98	1.66
A131	4.97	1.48
A186	4.19	1.46
A377	3.64	1.77
A384	4.96	1.51
A398	2.71	1.55

TABLE 3.6 **Relative expression of *CG34330* for each fly strain as determined by microarray analysis**

[a] Data from Müller *et al*. (2011)

[b] Data from Hutter *et al*. (2008)

3.4.2 POLYMORPHISM PATTERNS IN OTHER NON-AFRICAN POPULATIONS

The region found to be monomorphic in the population from the Netherlands also shows low polymorphism in other non-African populations. The *D. melanogaster* reference sequence (Adams *et al*. 2000), which comes from a non-African lab strain is identical to the monomorphic region that was identified in our European population. In a pooled sample of 113 isofemale fly lines from a Portuguese population (Pandey *et al*. 2011), only four SNPs are present within the region. All of these SNPs are found at low frequencies (2.7%-7.1%),

suggesting that these represent new mutations that have occurred after the completion of the selective sweep. However, another European population from France (S. Hutter, personal comm.) consisting of eight lines shows 11 SNPs within this region, which corresponds to a value of θ_W of 0.0027. Still, a significantly negative Tajima's *D* in the upstream region of the gene, indicating an excess of low-frequency variants, suggests that this region has been a target of a recent selective sweep. This hypothesis is supported by a significantly negative Fay and Wu's *H* for upstream sites and all sites. Thus, these findings indicate that the selective sweep around *CG34330* has occurred in several non-African populations.

The indel and all of the SNPs that are fixed in the European population are also fixed in the population from Portugal. However, in the French population only those sites are fixed where the ancestral state could be inferred (Table 3.5). This indicates that these sites might play a crucial role for gene expression regulation of *CG34330* in derived populations.

3.4.3 STRUCTURAL ADAPTATION IN *JON99CIII*

In female *D. melanogaster*, the microarray study from chapter 1 has identified many genes over-expressed in the European population that are involved in proteolysis. Among these genes, four members of the *Jonah* gene family were identified. The gene with the second highest over-expression of all genes in Europe was *Jonah 99Ciii*, which is located in a cluster of three adjacent *Jonah* genes. For this gene, a significant MK test is found when comparing nonsynonymous to synonymous sites for the European population (Table 3.4). This holds true whether comparing *D. melanogaster* to *D. simulans* or *D. sechellia*. When comparing the combined data of both populations to *D. simulans*, the same pattern can be observed. In all cases, there is a significant excess of divergence relative to polymorphism at nonsynonymous sites, indicating that recurrent positive selection for amino acid substitutions has occurred since the split of the species before the out-of-Africa movement. It should be noted, however, that of the 893 nucleotide gene region, the last 206 are missing in the sequenced region. Since the gene occurs in a cluster with two other genes belonging to the same gene family, it was not possible to design unambiguously binding primers for this region. A significant MK test is also found for the upstream region of the African population and the combined data of both populations when compared to *D. simulans*. However, this is due to a relative excess of upstream polymorphism, indicating that either recurrent balancing selection or weak purifying selection has caused this pattern. No haplotype structure can be found along the upstream region, suggesting that weak purifying selection has caused the relative excess of

polymorphism in the upstream region. This is also supported by the finding of a negative Tajima's *D* along the upstream region (Table 3.2). All in all, levels of nucleotide diversity were normal over the whole sequenced region in both populations (Figure 3.1).

Together these findings indicate that *Jon99Ciii* has been a target of structural adaptation, altering the amino acid sequence of the gene since the divergence of *D. melanogaster* and *D. sechellia* or *D. simulans*, respectively. In contrast, no evidence for recent regulatory adaptation in any of the populations can be found, suggesting that other parts of the genome are responsible for the observed expression differences. This is in accordance with signatures of recurrent weak purifying selection in the upstream region of the gene, which might act to eliminate sequence changes in this region.

3.4.4 SEQUENCE DIVERGENCE IN UPSTREAM REGIONS

A recent study (McManus *et al*. 2010) has quantified allele-specific expression levels in *D. melanogaster* and *D. sechellia* and their F1 hybrids. For genes that show *cis*-regulatory divergence between species, the adjacent upstream regions harbored significantly more sequence changes than genes without significant *cis*-regulatory divergence. Similar patterns have also been found previously for other organisms (*e.g*. in yeast Tirosh *et al*. 2009, in *Arabidopsis* Zhang and Borevitz 2009). Thus, if the observed expression differences of the sequenced genes are caused by changes in *cis*, one could expect the upstream regions of these genes to show a greater D_{XY} or F_{ST} between the two populations than genes without significant differential expression. However, such a pattern cannot be found. In addition, there is no evidence for greater divergence between species in upstream regions of differentially versus non-differentially expressed genes (Figure 3.2). Only when looking at intrapopulation nucleotide diversity π, higher levels can be found in Europe for differentially expressed genes. Interestingly, several studies have shown that few changes in *cis*-regulatory elements are sufficient to account for changes in gene expression (*e.g*., Shirangi *et al*. 2009, Jeong *et al*. 2008, Prabhakar *et al*. 2008, Frankel *et al*. 2011). This suggests that upstream regions containing *cis*-regulatory elements that cause differential expression of a gene do not necessarily accumulate more changes than upstream regions of genes that do not show differential expression. However, we did not find any evidence for regulatory adaptation in *cis* for most of the sequenced genes. Thus, the absence of significant results for most of the summary statistics examined is most likely due to the absence of regulatory evolution in the genes' upstream region. The genetic basis for expression differences for the majority of the

genes examined in this study could reside in regions of the genome that are located further upstream. Although promoter sequences can be as short as 200-300 bp (Wei *et al*. 1995, Petronzelli *et al*. 1995), case studies have shown that *cis*-regulatory elements can reside as far as 800 kb upstream of the gene they regulate (DiLeone *et al*. 1998, Lettice *et al*. 2002). Another possibility is that the differential expression of these genes is caused by changes at the DNA-sequence- or the expression level of *trans*-factors, such as TFs or microRNAs. Such factors are encoded in other parts of the genome, and cannot be identified using the approach of the present study.

For one of the genes examined in this study, *CG34330*, evidence for adaptive evolution is found that possibly is caused by nucleotide variants located in the upstream region of the gene. In order to investigate if expression differences between the European and the African population are due to those changes that are fixed in Europe but at low frequency in Africa, reporter gene assays should be performed. This procedure has already successfully identified regulatory regions that control gene expression in many case studies (*e.g*., Gompel *et al*. 2005, Prud'homme *et al*. 2006, Jeong *et al*. 2008, Chan *et al*. 2010, Saminadin-Peter *et al*. 2011). By applying reporter gene assays, the effects of the alternative states of the nucleotide variants on gene expression can be tested in a common genetic background.

McManus *et al*. (2010) have identified significant *cis*- as well as *trans*- regulatory effects for *CG34330* when comparing *D. melanogaster* to *D. sechellia*. Since the two populations of *D. melanogaster* investigated here are less diverged, it is unclear if *trans*-regulatory factors also contribute to the observed expression difference. Performing reporter gene assays will also help to elucidate whether expression differences between the studied populations are caused by *cis*-regulatory factors, *trans*-regulatory factors, or a combination of *cis*- and *trans*-regulatory factors.

Hindawi Publishing Corporation
International Journal of Evolutionary Biology
Volume 2012, Article ID 963976, 10 pages
doi:10.1155/2012/963976

Research Article

Inter- and Intraspecific Variation in *Drosophila* Genes with Sex-Biased Expression

Lena Müller,[1] Sonja Grath,[2] Korbinian von Heckel,[1] and John Parsch[1]

[1] *Department of Biology II, University of Munich (LMU), 82152 Planegg-Martinsried, Germany*
[2] *Institute for Evolution and Biodiversity, University of Münster (WWU), 48149 Münster, Germany*

Correspondence should be addressed to John Parsch, parsch@zi.biologie.uni-muenchen.de

Received 13 July 2011; Revised 6 October 2011; Accepted 6 October 2011

Academic Editor: Artyom Kopp

Genes with sexually dimorphic expression (sex-biased genes) often evolve rapidly and are thought to make an important contribution to reproductive isolation between species. We examined the molecular evolution of sex-biased genes in *Drosophila melanogaster* and *D. ananassae*, which represent two independent lineages within the *melanogaster* group. We find that strong purifying selection limits protein sequence variation within species, but that a considerable fraction of divergence between species can be attributed to positive selection. In *D. melanogaster*, the proportion of adaptive substitutions between species is greatest for male-biased genes and is especially high for those on the X chromosome. In contrast, male-biased genes do not show unusually high variation within or between populations. A similar pattern is seen at the level of gene expression, where sex-biased genes show high expression divergence between species, but low divergence between populations. In *D. ananassae*, there is no increased rate of adaptation of male-biased genes, suggesting that the type or strength of selection acting on sex-biased genes differs between lineages.

1. Introduction

In sexually reproducing species, the evolution of reproductive isolation is closely coupled to the process of speciation. Indeed, the widely applied biological species concept defines species as "groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups" [1]. This definition has been of great utility to geneticists working with organisms like *Drosophila* that are separated into populations or species showing varying degrees of pre- and postzygotic reproductive isolation. The degree of isolation increases with the time since the species shared a common ancestor [2, 3].

Within species, prezygotic isolation is often observed as preferential mating of individuals (usually females) to other individuals from the same population. Such behavioral isolation has been observed for *Drosophila melanogaster* and *D. ananassae* populations that diverged within the past 15,000–20,000 thousand years [4, 5]. At the postzygotic level, it is often found that matings between closely related species produce hybrid offspring in which at least one sex (usually males) is either inviable or infertile. For example, species of the *D. simulans* complex, which diverged around 0.5–1.0 million years ago [6], produce viable hybrid offspring with only the males being infertile [7, 8]. Crosses between *D. melanogaster* and *D. simulans*, which diverged around 4-5 million years ago [6], produce viable, but sterile, offspring of only one sex (the sex of the *D. melanogaster* parent) [9, 10].

The observations from *Drosophila* suggest that the evolution of postzygotic reproductive isolation is a progressive process that involves the accumulation of incompatible alleles at many loci across the genome [11–13]. Since the first stage of isolation is typically hybrid male sterility, sequence divergence at genes involved in male reproduction is thought to be a major contributor to speciation [14]. Between the closely related species *D. simulans* and *D. mauritiana*, it is thought that ~60 loci contribute to hybrid male sterility [12]. To date, only a few of these loci have been mapped to the gene level [15–17]. For example, the first "speciation gene" identified between these two species, *OdsH*, encodes a homeodomain-containing transcription factor that is expressed in testis and shows extraordinary amino acid sequence divergence between *D. simulans* and *D. mauritiana* [15]. Within the homeodomain, 15 amino acids differ between these two

Drosophila species, while only 17 amino acids differ between mouse and common ancestor of the *D. melanogaster* clade [15]. These findings suggested that the early stages of speciation are driven by the rapid adaptive evolution of genes involved in male reproduction [18]. Consistent with this, it has been found that genes known to be involved in male reproduction, but not directly implicated in reproductive isolation between species, evolve at a faster rate than other classes of genes in the genome [19–21].

With the advent of transcriptomic technologies, such as microarrays, it became possible to examine gene expression differences between males and females on a genomic scale. In *Drosophila*, a large fraction of genes differ in expression between the sexes [22]. Such genes are referred to as "sex-biased." A meta-analysis over multiple experiments indicates that there are ~4,000 genes that show a large (greater than twofold) difference in expression between males and females of *D. melanogaster*, with ~2,000 showing male-biased expression and ~2,000 showing female-biased expression [23]. When statistical approaches are used to detect significant expression differences between the sexes, the number of sex-biased genes is even greater. For example, a meta-analysis with a false discovery rate of 5% classified 2,814 genes as male-biased and 4,056 genes as female-biased [23].

On average, male-biased genes display a faster rate of molecular evolution between species (as measured by the ratio of the nonsynonymous-to-synonymous substitution rates, d_N/d_S) than female-biased genes or genes with nearly equal expression in the two sexes ("unbiased genes") [23, 24]. By comparing levels of polymorphism within species to divergence between species, it could be shown that male-biased genes undergo more adaptive evolution than female-biased or unbiased genes [25]. This pattern was especially pronounced on the X chromosome, where X-linked, male-biased genes show exceptionally high d_N/d_S and the strongest signal of adaptive protein evolution [26]. Although species outside the *melanogaster* species subgroup have not been investigated as extensively, preliminary studies in *D. ananassae* and *D. pseudoobscura* suggest that male-biased expression does not have as much of an influence on evolutionary rate in these species as it does in *D. melanogaster* [27–29].

In this paper, we examine the molecular divergence of sex-biased genes within and between species using data from *D. melanogaster* and *D. ananassae*. We also investigate intra- and interspecific divergence at the level of gene expression. Our results indicate that much of the protein divergence observed between species is adaptive. Male-biased genes of *D. melanogaster*, especially those that reside on the X chromosome, show an exceptionally high rate of adaptation. However, these genes do not show unusually high sequence variation within or between populations. At the level of gene expression, we find that both male- and female-biased genes make a large contribution to expression differences between species but are underrepresented among genes that differ in expression between populations. These findings suggest that different selective forces contribute to interpopulation and interspecies divergence.

Table 1: Numbers of genes analyzed.

Species	Bias	Autosomal	X-linked
D. melanogaster	Male	35	18
	Unbiased	32	16
	Female	29	13
D. ananassae	Male	10	7
	Unbiased	9	5
	Female	10	2

2. Materials and Methods

2.1. D. melanogaster Genes and Populations. In total, we analyzed DNA sequence polymorphism in 143 *D. melanogaster* genes (see Table 1 in Supplementary Material available online at doi:10.1155/2012/963976), which were classified as male-, female-, or unbiased in their expression using the Sebida database [23]. The numbers of sex-biased genes, as well as the numbers of X-linked and autosomal genes, are given in Table 1. All of the genes were sequenced in a sample of isofemale lines from two populations, one from Europe (Leiden, the Netherlands) and one from Africa (Lake Kariba, Zimbabwe) [34]. The number of alleles sequenced per population ranged from 7 to 12, with a mean of 11. Sequences of 136 of these genes were reported previously [24, 25, 35] and are available from the GenBank/EMBL databases under accession numbers AM293861–AM294919, AM998825–AM999334, and FM244915–FM246454. In addition, seven genes were newly sequenced for the current study and are available under accession numbers JN252131–JN252193 and JN374903–JN374992. For divergence calculations, a single allele from *D. simulans* was used [30].

2.2. D. ananassae Genes and Populations. For *D. ananassae*, we surveyed polymorphism in 43 genes (Supplementary Table 1), which were classified as male-, female-, or unbiased in their expression using data from a custom amplicon microarray [29] and a whole genome microarray analysis [36]. The 43 genes were a subset of those analyzed in *D. melanogaster*. The numbers of sex-biased genes, as well as the numbers of X-linked and autosomal genes, are given in Table 1. All of the genes were sequenced in a sample of isofemale lines from Bangkok, Thailand [29], and the sequences are available from the GenBank/EMBL databases under accession numbers FN546265–FN546780. The number of alleles sequenced ranged from 8 to 12, with a mean of 11. To calculate divergence, a single allele from either *D. atripex* or *D. phaeopleura* was used. The phylogenetic relationship of the species is shown in Figure 1.

2.3. DNA Sequencing. Genomic DNA was purified from single male flies, and target genes were PCR-amplified using protocols, primers, and cycling conditions described previously [25, 26]. Following PCR, the amplified products were purified with ExoSAP-IT (USB, Cleveland, OH, USA), and both strands were sequenced using BigDye version 1.1

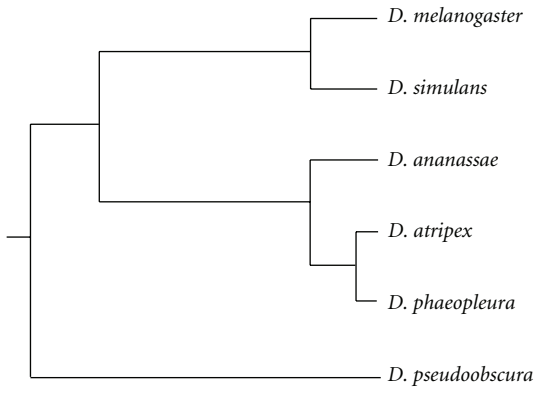


Figure 1: Phylogenetic relationship of the species used in this study [29, 30].

chemistry and a 3730 automated sequencer (Applied Biosystems, Foster City, CA, USA). Sequences were edited using DNASTAR (Madison, WI, USA) and multiple alignments were generated using MUSCLE [37].

2.4. Statistical Methods. Standard polymorphism and divergence statistics were calculated using DnaSP version 5 [38]. To assess the significance of differences between sex-bias classes, the Kruskal-Wallis tests and the Mann-Whitney *U* tests were performed using R version 2.12.2 [39].

The distribution of fitness effects of new nonsynonymous mutations and the proportion of adaptive amino acid replacements between species, α, were estimated using the DoFE software [40]. For this, the shape parameter was set to 0.5 and the number of repetitions for the MCMC chain was set to 5,000,000. Prior to running, a look-up table was generated, setting the upper limit of β to 1 and the number of steps to 200. Otherwise, the default parameters were used. Synonymous sites were used as the neutral reference. This method requires the same sample size (number of sequences) for all genes. For *D. melanogaster*, we used a common sample size of 10 sequences from the African population. When more than 10 sequences were available for a gene, we randomly excluded surplus sequences. Genes with fewer than 10 sequences were excluded from the analysis. For *D. ananassae*, the above procedure was followed, but a common sample size of eight sequences was used.

3. Results

3.1. Data. In total, we analyzed DNA sequence polymorphism and divergence in 143 *D. melanogaster* and 43 *D. ananassae* protein-coding genes. Within each species, the genes could be assigned to one of three expression classes (male-, female-, or unbiased) on the basis of microarray data (Table 1) [23, 29, 36]. The proportion of genes in each expression class was similar, although there was a slight over-representation of male-biased genes. The genes could further be separated into those residing on the X chromosome and those residing on the autosomes (Table 1). For *D. melanogaster*, approximately one-third of the genes within each sex-bias class were X-linked. This allowed us to perform additional analyses in which X-linked and autosomal genes were considered separately within each expression class. Because the *D. ananassae* sample size was much smaller, we did not analyze X-linked and autosomal genes separately.

3.2. Selective Constraint on Sex-Biased Genes. To infer selective constraints, we used the method of Eyre-Walker and Keightley [40], which estimates the distribution of fitness effects of nonsynonymous mutations. In both *D. melanogaster* and *D. ananassae*, we found evidence for strong constraint on male-, female-, and unbiased genes, with the vast majority (>85%) of new mutations having a strongly deleterious effect, in which the product of the effective population size and the selection coefficient ($N_e s$) is greater than 10 (Figure 2). Less than 10% of mutations fell within the neutral range ($0 < N_e s < 1$). The level of constraint was similar across all classes of genes and in both species.

When the X-linked and autosomal genes of *D. melanogaster* genes were analyzed separately, there was again evidence for the predominance of strong purifying selection in all classes of genes (Figure 3). For male-biased and unbiased genes, there was a trend towards less constraint on the X chromosome. This pattern was not seen for female-biased genes.

3.3. Adaptive Evolution of Sex-Biased Genes. In both the *melanogaster* and *ananassae* lineages, we found that positive selection has made an important contribution to protein sequence divergence between species. For all classes of genes, the estimated proportion of adaptive nonsynonymous substitutions, α, ranged from 0.29 to 0.83 (Figure 2). The 95% confidence interval of α excluded zero in all cases, except for the unbiased genes of *D. ananassae* where it was -0.04 to 0.56. In *D. melanogaster*, male-biased genes had the highest mean α and its 95% confidence interval did not overlap with that of female-biased or unbiased genes, indicating a significantly greater proportion of adaptive substitutions in male-biased genes. This pattern was not seen for *D. ananassae*, where α was highest for female-biased genes (Figure 2), but the 95% confidence intervals of α overlapped among all classes of genes.

Because the *D. ananassae* genes represented only a subset of those analyzed in *D. melanogaster*, it is possible that the observed differences in sex-biased gene evolution between species are a result of differences in gene composition or of limiting the *D. ananassae* genes to those that are well conserved and have identifiable orthologs in *D. melanogaster*. To examine these possibilities, we repeated our *D. melanogaster* analyses using only genes common to both species' gene sets (Figure 4(a)) or only genes with identifiable orthologs between species (Figure 4(b)). In both cases, we still observed higher values of α for male-biased genes than for female-biased or unbiased genes. For the set of common genes, which had a small sample size (37 genes total), the 95% confidence intervals of α overlapped among all classes of genes. However, for the set of genes with orthologs (108 genes total), the 95% confidence interval of α of male-biased genes did not overlap with that of female-biased or unbiased genes.

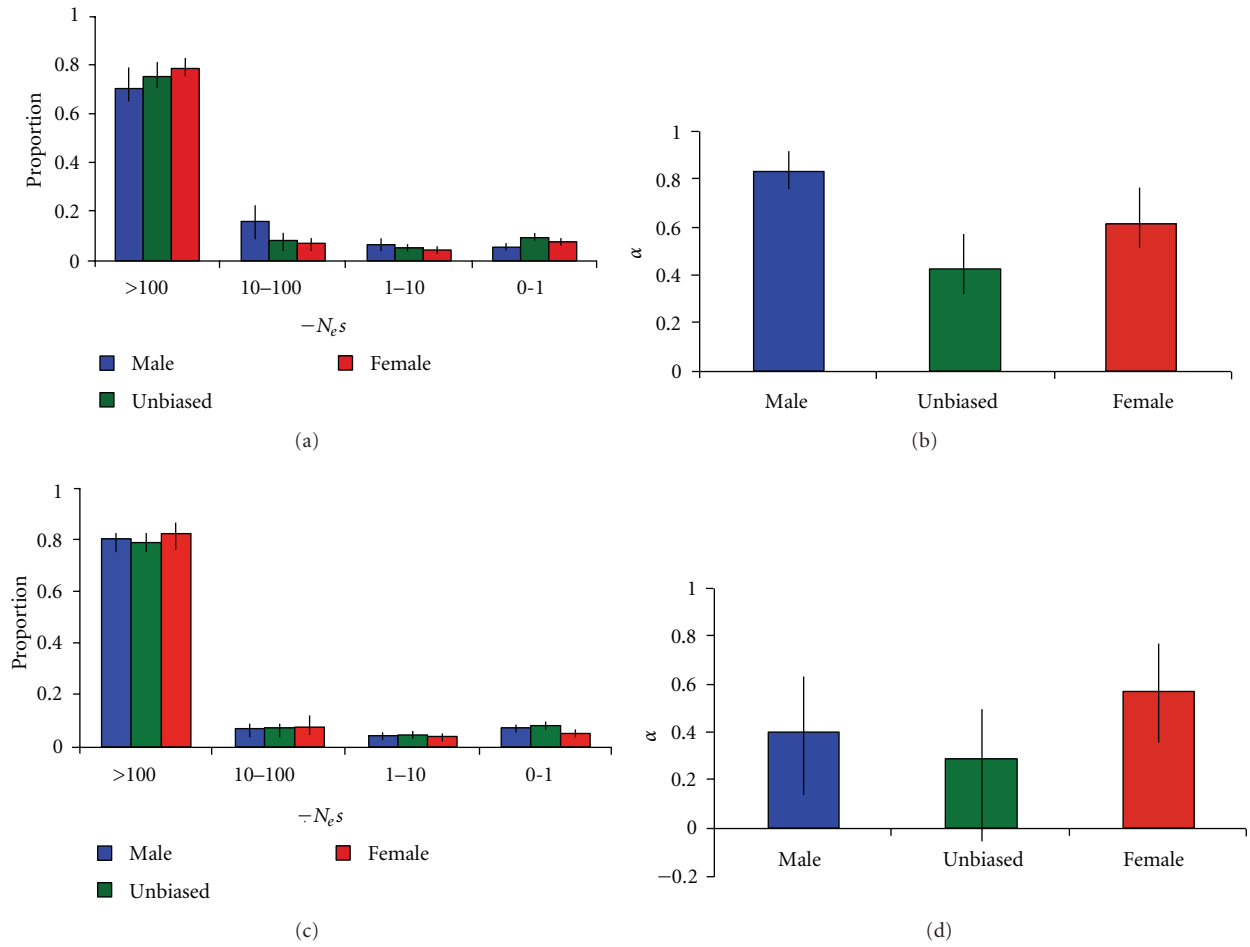


Figure 2: Distribution of fitness effects for nonsynonymous mutations within species and the proportion of adaptive nonsynonymous substitutions between species. Data for *D. melanogaster* are shown in (a) and (b), while those for *D. ananassae* are shown in (c) and (d).

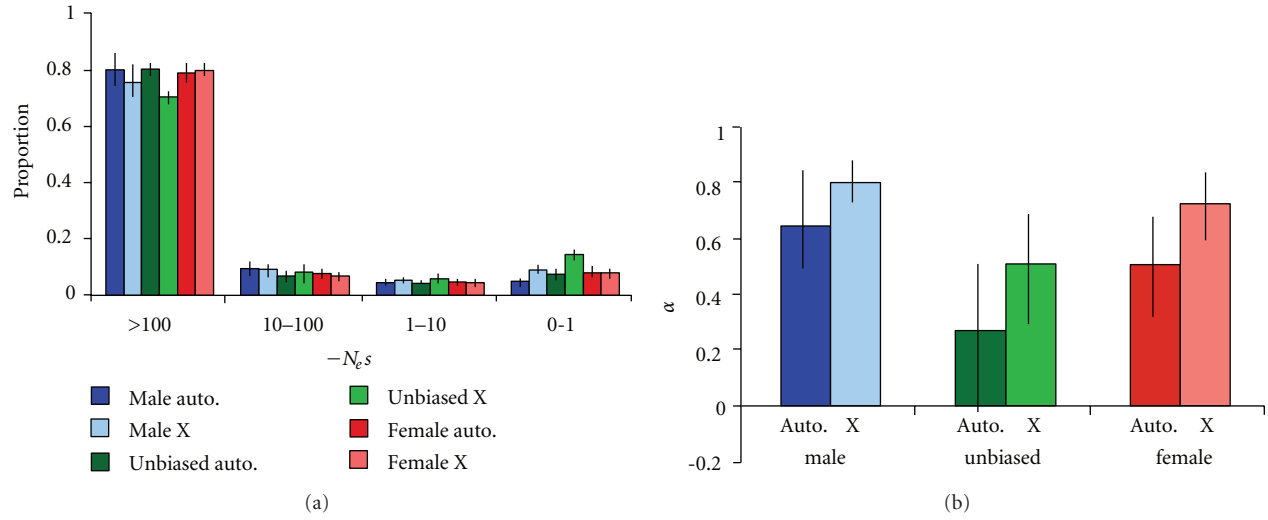


Figure 3: The distribution of fitness effects (a) and estimated proportion of adaptive substitutions (b) for autosomal and X-linked genes of *D. melanogaster*.

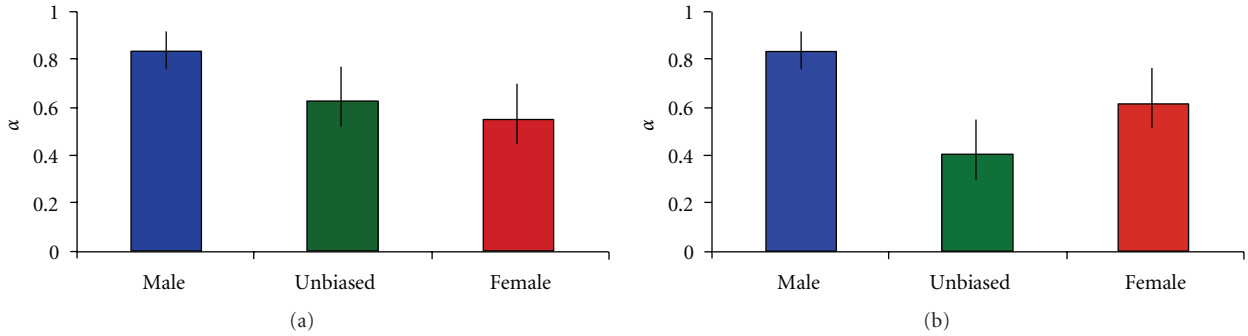


Figure 4: The estimated proportion of adaptive substitutions for *D. melanogaster* genes also present in the *D. ananassae* gene set (a) and for *D. melanogaster* genes that have an ortholog in *D. ananassae* (b).

Table 2: Rates of adaptive substitution in *D. melanogaster* and *D. ananassae*.

Species	Bias	Chrom.	d_N [a]	αd_N [b]
D. melanogaster	Male	Auto.	14.0	9.1
		X	60.8	49.8
	Unbiased	Auto.	8.3	2.3
		X	24.6	12.5
	Female	Auto.	16.1	8.3
		X	24.1	17.7
D. ananassae	Male	Auto. + X	17.7	7.1
	Unbiased	Auto. + X	15.9	4.6
	Female	Auto. + X	17.4	10.0

[a] Nonsynonymous substitutions per 1,000 nonsynonymous sites.
[b] Adaptive nonsynonymous substitutions per 1,000 nonsynonymous sites.

This indicates that the increased level of adaptive evolution of male-biased genes in *D. melanogaster* is not attributable to the rapid evolution of young, newly evolved genes that lack orthologs in *D. ananassae*.

When *D. melanogaster* autosomal and X-linked genes were considered separately, there was a consistent pattern of higher α for X-linked genes of all classes, with the highest value observed for male-biased, X-linked genes (Figure 3). This pattern was even more pronounced when the nonsynonymous substitution rate was taken into account, as X-linked genes showed greater nonsynonymous divergence (Table 2).

3.4. Sequence Variation of Sex-Biased Genes within Populations. Mean levels of nucleotide diversity (π) did not differ significantly among male-, female-, or unbiased genes in the Zimbabwe population of *D. melanogaster* or the Bangkok population of *D. ananassae* (Figure 5). This result held regardless of whether synonymous diversity (π_S), nonsynonymous diversity (π_N), or their ratio (π_N/π_S) was considered.

When *D. melanogaster* X-linked genes were considered separately, there was a significant difference in π_N among male-, female-, and unbiased genes (the Kruskal-Wallis test, $P = 0.03$). This was mainly a result of X-linked, unbiased genes having high nonsynonymous diversity (Figure 6). There were no significant differences in π_S, π_N, or π_N/π_S among autosomal male-, female-, or unbiased genes (the Kruskal-Wallis test, $P > 0.20$ in all cases). Within expression classes, there was consistently greater polymorphism at X-linked loci than at autosomal loci (Figure 6). This difference was significant only for unbiased genes, where π_N, and π_N/π_S were both greater on the X chromosome than the autosomes (the Mann-Whitney test, $P = 0.002$ and $P = 0.006$, resp.).

3.5. Sequence Divergence of Sex-Biased Genes between Populations. For *D. melanogaster*, we had sequence data for all 143 genes from both an African (Zimbabwe) and a European (the Netherlands) population, which allowed us to determine the contribution of sex-biased genes to interpopulation genetic differentiation. Two measures, F_{ST} and D_{XY} (the mean number of pairwise sequence differences between alleles of the two populations), indicated that there are similar levels of differentiation for male-, female-, and unbiased genes on both the X chromosome and the autosomes (Table 3). However, for all classes of genes, differentiation was greater at X-linked loci. For male- and female-biased genes, F_{ST} was significantly greater on the X chromosome when all sites or only synonymous sites were considered (Table 3). For unbiased genes, D_{XY} was significantly greater on the X chromosome for nonsynonymous sites (Table 3).

3.6. Intra- and Interspecific Divergence in Sex-Biased Gene Expression. To determine the contribution of sex-biased genes to variation within and between species at the level of gene expression, we analyzed data from published microarray studies that investigated expression polymorphism within *D. melanogaster* [31, 32] and expression divergence between *D. melanogaster* and *D. simulans* [33]. Three types of expression variation were examined (intrapopulation, interpopulation, and interspecies) using data from males and females separately (Figure 7). When expression was measured in males, male-biased genes showed the highest levels of intrapopulation and interspecies divergence. However, male-biased genes did not show increased expression divergence between populations. When measured in females, female-biased genes showed the least intrapopulation and interpopulation expression polymorphism, but the greatest interspecies expression divergence.

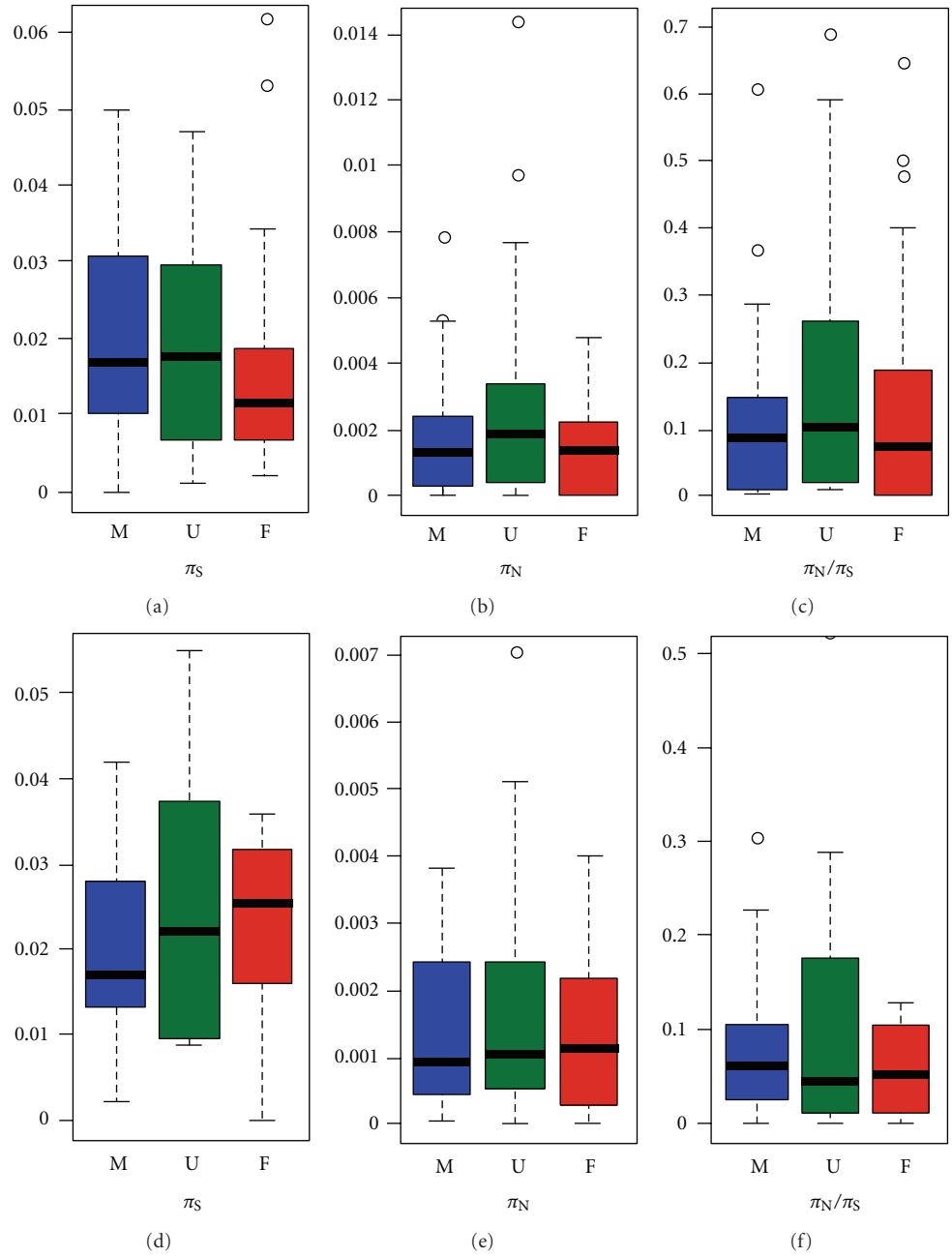


FIGURE 5: Intraspecies polymorphism in male-biased (M), unbiased (U), and female-biased (F) genes of *D. melanogaster* (a–c) and *D. ananassae* (d–f). Shown are distributions of synonymous nucleotide diversity (π_S), nonsynonymous nucleotide diversity (π_N), and their ratio (π_N/π_S). The *D. melanogaster* data are from the African (Zimbabwe) population. There were no significant differences among male-, female-, and unbiased genes in either species by any measure (the Kruskal-Wallis test, $P > 0.10$ in all cases).

4. Discussion

4.1. Selection on Sex-Biased Genes. Our analyses of polymorphism and divergence in *D. melanogaster* and *D. ananassae* uncovered several common patterns. First, there is strong purifying selection acting at the protein level in both species. We estimate that over 85% of all newly arising nonsynonymous mutations are deleterious. Second, a large proportion of amino acid substitutions that have become fixed between species can be attributed to positive selection. Our estimates of α range from 27 to 83% in *D. melanogaster* and 29–57% in *D. ananassae*. In *D. melanogaster*, male-biased genes showed the highest values of α (Figure 2), which is consistent with previous studies [25, 26]. In *D. ananassae*, there was no evidence for increased adaptive evolution of male-biased genes, which suggests that there are differences in sex-biased gene evolution among lineages[27, 29].

Our estimates of α are in line with previously published estimates and suggest that adaptive protein evolution is widespread across the *Drosophila* genus [41]. A recent study

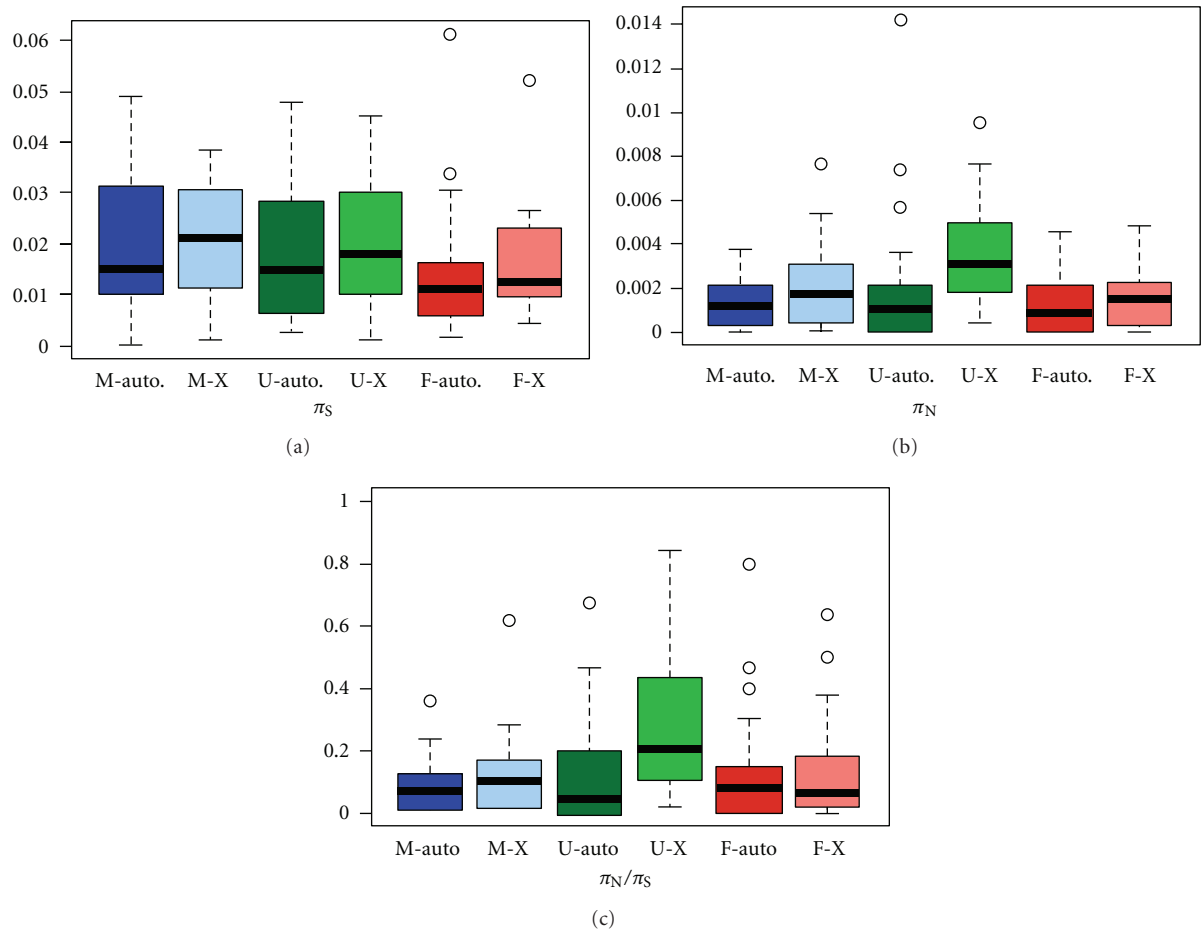


FIGURE 6: Intraspecies polymorphism in autosomal and X-linked genes of *D. melanogaster*. Shown are distributions of synonymous nucleotide diversity (π_S), nonsynonymous nucleotide diversity (π_N), and their ratio (π_N/π_S). The data are from the African (Zimbabwe) population. The only significant difference among expression classes was for π_N (the Kruskal-Wallis test, $P = 0.03$), where X-linked unbiased genes had significantly higher π_N than X-linked female-biased genes (the Mann-Whitney test, $P = 0.01$). Within expression classes, X-linked unbiased genes had significantly greater π_N and π_N/π_S than autosomal unbiased genes (the Mann-Whitney test, $P = 0.002$ and $P = 0.006$, resp.).

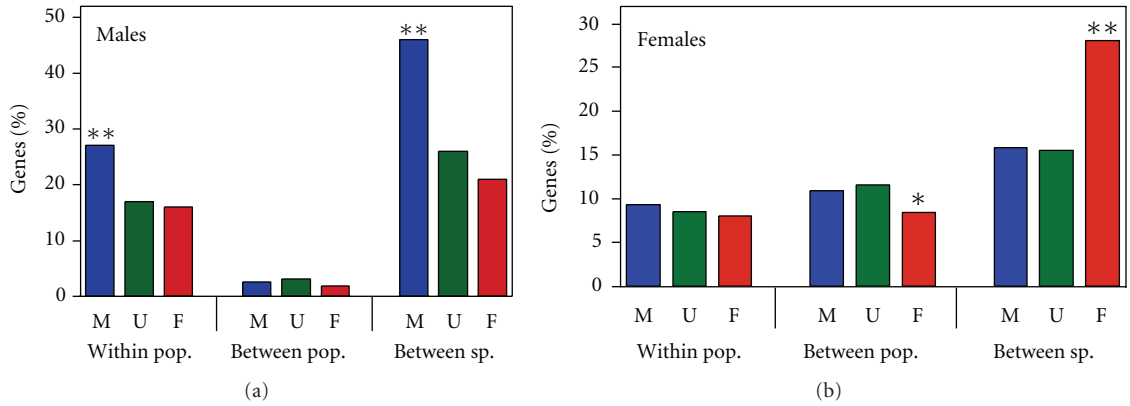


FIGURE 7: Gene expression variation within a population, between populations, and between species. Shown are the percentages of male-biased (M), unbiased (U), and female-biased (F) genes that show significant expression variation within a Zimbabwe population of *D. melanogaster*, between a Zimbabwe and a European population of *D. melanogaster* and between *D. melanogaster* and *D. simulans*. Expression variation was measured separately in males (a) and females (b). Data are from Hutter et al. [31], Müller et al. [32], and Ranz et al. [33]. Asterisks indicate significant differences from unbiased genes within the same comparison, as determined by Fisher's exact test. $^*P < 0.001$, $^{**}P < 0.0001$.

Table 3: Mean F_{ST} and D_{XY} between the African and European *D. melanogaster* populations.

Sites[a]	Bias[b]	Autosomal		X-linked	
		F_{ST} (SD)	D_{XY}[c] (SD)	F_{ST} (SD)[d]	D_{XY}[c] (SD)[d]
All	M	0.157 (0.142)	0.69 (0.35)	0.266 (0.141)**	0.72 (0.37)
	U	0.160 (0.125)	0.74 (0.60)	0.231 (0.231)	0.87 (0.48)
	F	0.195 (0.161)	0.56 (0.41)	0.343 (0.186)*	0.59 (0.30)
Syn	M	0.164 (0.164)	2.25 (1.87)	0.261 (0.151)*	2.68 (2.60)
	U	0.159 (0.139)	1.94 (1.63)	0.223 (0.020)	2.57 (1.56)
	F	0.185 (0.163)	1.54 (1.23)	0.319 (0.209)*	1.82 (1.27)
Non	M	0.090 (0.086)	0.14 (0.12)	0.185 (0.199)	0.27 (0.28)
	U	0.107 (0.123)	0.19 (0.27)	0.128 (0.143)	0.51 (0.80)**
	F	0.149 (0.171)	0.12 (0.15)	0.254 (0.305)	0.17 (0.17)

[a] "All," all sites (including introns); "Syn," synonymous sites; "Non," nonsynonymous sites.
[b] "M," male-biased; "U," unbiased; "F," female-biased.
[c] Mean pairwise differences between all African and all European sequences (in %).
[d] Asterisks indicate significant differences from autosomal genes by the Mann-Whitney test. $^*P < 0.05$, $^{**}P < 0.01$.

of *D. melanogaster* and *D. simulans* reported higher estimates of α for a randomly chosen (with respect to expression and function) set of genes [42]. However, this study was limited to X-linked genes, which tend to have higher values of α (Figure 3). This suggests that the use of only X chromosomal data may lead to an overestimate of the genome-wide proportion of adaptive substitutions.

4.2. Faster-X Evolution. Several factors could contribute to the increased rate of adaptive evolution of X-linked genes. For example, the X chromosome could have a larger effective population size (N_e) than the autosomes. Assuming an equal sex ratio, the number of X chromosomes in a population is expected to be 75% that of the number of autosomes. However, sexual selection acting on males can lead to a reduction in the N_e of the autosomes relative to the X chromosome, and this could accelerate X chromosome evolution [43, 44]. In our populations of *D. melanogaster* and *D. ananassae*, which are thought to come from the ancestral species ranges [45, 46], the X chromosome and the autosomes have nearly identical N_e [29, 35, 47], making this explanation unlikely. Furthermore, there is no evidence for increased purifying selection on the X chromosome (Figure 3), which would be expected if it had a larger N_e. This observation also argues against the possibility that an increased rate of recombination on the X chromosome leads to an increase in the efficacy of selection on X-linked loci by reducing interference among mutations [48, 49].

The accelerated rate of adaptive evolution on the X chromosome could also be explained by an increased rate of fixation of new, beneficial, recessive mutations due to their exposure to selection in hemizygous males [50, 51]. This would explain why the signal of adaptive evolution is strongest for male-biased genes, as they are expected to encounter selection mainly in the male (hemizygous) genetic background [26]. Although female-biased genes would be expected to receive the least benefit from selection in the male genetic background, a recent study found that mutations in female-biased genes often have phenotypic effects in males [52]. Thus, the increased rate of adaptive evolution seen for X-linked, female-biased genes could result from their exposure to selection in males.

4.3. Gene Expression Divergence. Our analyses of gene expression polymorphism and divergence revealed that the genes with the greatest expression divergence between species are those that are expressed predominantly in the sex that is used for comparison. When males are compared, male-biased genes show the greatest interspecific difference in expression (Figure 7). When females are compared, female-biased genes show the greatest interspecific difference in expression. These patterns are not seen for interpopulation expression divergence, where male- and female-biased genes consistently show less interpopulation expression divergence than unbiased genes, regardless of the sex that is compared (Figure 7). Thus, similar to protein divergence, gene expression divergence between species does not appear to be a simple extension of divergence between populations.

There are some caveats to our expression analysis. First, the experiments were performed by different groups, at different times, and with different microarray platforms. Thus, many factors may contribute to the differences seen among experiments. However, this problem will not apply to comparisons of male-, female-, and unbiased genes within each experiment, as all of the genes were measured together on the same microarrays. Thus, we expect comparisons of sex-biased genes within experiments to be robust. A second caveat is that the interspecies comparisons used only a single isofemale line of each species. This means that intraspecific polymorphism and interspecific divergence will be confounded. However, given the low level of expression polymorphism seen within species, it is unlikely that intraspecific gene expression polymorphism has much influence on measures of interspecific divergence. This is supported by the observation that, in females, there is no correspondence between the relative levels of expression polymorphism and divergence (Figure 7). However, for the experiments performed on males, we cannot exclude the possibility that the observed interspecific divergence of male-biased genes is inflated by intra-specific polymorphism.

4.4. Implications for Speciation. Although it is not possible to establish a direct link between sex-biased gene evolution and speciation in most cases, several of our observations are consistent with the rapid evolution of sex-biased genes (especially male-biased genes) contributing to reproductive isolation, at least for *D. melanogaster* and its close relatives. The male-biased genes examined here are expressed in reproductive tissues [26], and their rapid adaptive evolution could contribute to the male-hybrid sterility that is often seen as a first step in reproductive isolation. Furthermore, the rapid adaptive evolution of X-linked genes, especially those with male-biased expression, is consistent with the disproportionately large effect that the X chromosome has on hybrid breakdown [12, 13]. At the level of gene expression, male-biased genes make the largest contribution to the expression differences between species in males. Since the vast majority of male-biased genes are expressed in reproductive tissues [53], it is likely that expression differences also contribute to male hybrid sterility and other forms of hybrid breakdown.

Acknowledgments

This work was carried out as part of the research unit "Natural selection in structured populations" (FOR 1078) funded by the *Deutsche Forschungsgemeinschaft* grant PA 903/5.

References

[1] E. Mayr, *Systematics and the Origin of Species*, Columbia University Press, New York, NY, USA, 1942.

[2] J. A. Coyne and H. A. Orr, "Patterns of speciation in *Drosophila*," *Evolution*, vol. 43, no. 2, pp. 362–381, 1989.

[3] J. A. Coyne and H. A. Orr, ""Patterns of speciation in *Drosophila*" revisited," *Evolution*, vol. 51, no. 1, pp. 295–303, 1997.

[4] C. I. Wu, H. Hollocher, D. J. Begun, C. F. Aquadro, Y. Xu, and M. L. Wu, "Sexual isolation in *Drosophila* melanogaster: a possible case of incipient speciation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 92, no. 7, pp. 2519–2523, 1995.

[5] M. D. Schug, J. F. Baines, A. Killon-Atwood et al., "Evolution of mating isolation between populations of *Drosophila ananassae*," *Molecular Ecology*, vol. 17, no. 11, pp. 2706–2721, 2008.

[6] K. Tamura, S. Subramanian, and S. Kumar, "Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks," *Molecular Biology and Evolution*, vol. 21, no. 1, pp. 36–44, 2004.

[7] J. A. Coyne, "The genetic basis of Haldane's rule," *Nature*, vol. 314, no. 6013, pp. 736–738, 1985.

[8] L. W. Zeng and R. S. Singh, "The genetic basis of Haldane's rule and the nature of asymmetric hybrid male sterility among *Drosophila* simulans, *Drosophila* mauritiana and *Drosophila* sechellia," *Genetics*, vol. 134, no. 1, pp. 251–260, 1993.

[9] A. H. Sturtevant, "Genetic studies on *Drosophila* simulans. I. Introduction. Hybrids with *Drosophila* melanogaster," *Genetics*, vol. 5, no. 5, pp. 488–500, 1920.

[10] K. Sawamura, T. K. Watanabe, and M. T. Yamamoto, "Hybrid lethal systems in the *Drosophila* melanogaster species complex," *Genetica*, vol. 88, no. 2-3, pp. 175–185, 1993.

[11] J. R. True, B. S. Weir, and C. C. Laurie, "A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of *Drosophila* mauritiana chromosomes into *Drosophila* simulans," *Genetics*, vol. 142, no. 3, pp. 819–837, 1996.

[12] Y. Tao, S. Chen, D. L. Hartl, and C. C. Laurie, "Genetic dissection of hybrid incompatibilities between *Drosophila* simulans and D. mauritiana. I. Differential accumulation of hybrid male sterility effects on the X and autosomes," *Genetics*, vol. 164, no. 4, pp. 1383–1397, 2003.

[13] J. P. Masly and D. C. Presgraves, "High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*," *PLoS Biology*, vol. 5, no. 9, article e243, 2007.

[14] C. I. Wu and A. W. Davis, "Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases," *American Naturalist*, vol. 142, no. 2, pp. 187–212, 1993.

[15] C. T. Ting, S. C. Tsaur, M. L. Wu, and C. I. Wu, "A rapidly evolving homeobox at the site of a hybrid sterility gene," *Science*, vol. 282, no. 5393, pp. 1501–1504, 1998.

[16] Y. Tao, D. L. Hartl, and C. C. Laurie, "Sex-ratio segregation distortion associated with reproductive isolation in *Drosophila*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 23, pp. 13183–13188, 2001.

[17] L. O. Araripe, H. Montenegro, B. Lemos, and D. L. Hartl, "Fine-scale genetic mapping of a hybrid sterility factor between *Drosophila* simulans and *D. mauritiana*: the varied and elusive functions of "speciation genes"," *BMC Evolutionary Biology*, vol. 10, article 385, 2010.

[18] A. Civetta and R. S. Singh, "Broad-sense sexual selection, sex gene pool evolution, and speciation," *Genome*, vol. 42, no. 6, pp. 1033–1041, 1999.

[19] A. Civetta and R. S. Singh, "Sex-related genes, directional sexual selection, and speciation," *Molecular Biology and Evolution*, vol. 15, no. 7, pp. 901–909, 1998.

[20] J. Parsch, C. D. Meiklejohn, E. Hauschteck-Jungen, P. Hunziker, and D. L. Hartl, "Molecular evolution of the ocnus and janus genes in the *Drosophila* melanogaster species subgroup," *Molecular Biology and Evolution*, vol. 18, no. 5, pp. 801–811, 2001.

[21] W. Haerty, S. Jagadeeshan, R. J. Kulathinal et al., "Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*," *Genetics*, vol. 177, no. 3, pp. 1321–1335, 2007.

[22] H. Ellegren and J. Parsch, "The evolution of sex-biased genes and sex-biased gene expression," *Nature Reviews Genetics*, vol. 8, no. 9, pp. 689–698, 2007.

[23] F. Gnad and J. Parsch, "Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression," *Bioinformatics*, vol. 22, no. 20, pp. 2577–2579, 2006.

[24] Z. Zhang, T. M. Hambuch, and J. Parsch, "Molecular evolution of sex-biased genes in *Drosophila*," *Molecular Biology and Evolution*, vol. 21, no. 11, pp. 2130–2139, 2004.

[25] M. Pröschel, Z. Zhang, and J. Parsch, "Widespread adaptive evolution of *Drosophila* genes with sex-biased expression," *Genetics*, vol. 174, no. 2, pp. 893–900, 2006.

[26] J. F. Baines, S. A. Sawyer, D. L. Hartl, and J. Parsch, "Effects of X-linkage and sex-biased gene expression on the rate of adaptive protein evolution in *Drosophila*," *Molecular Biology and Evolution*, vol. 25, no. 8, pp. 1639–1650, 2008.

[27] M. Metta, R. Gudavalli, J. M. Gibert, and C. Schlötterer, "No accelerated rate of protein evolution in male-biased *Drosophila* pseudoobscura genes," *Genetics*, vol. 174, no. 1, pp. 411–420, 2006.

[28] Z. F. Jiang and C. A. Machado, "Evolution of sex-dependent gene expression in three recently diverged species of *Drosophila*," *Genetics*, vol. 183, no. 3, pp. 1175–1185, 2009.

[29] S. Grath, J. F. Baines, and J. Parsch, "Molecular evolution of sex-biased genes in the *Drosophila ananassae* subgroup," *BMC Evolutionary Biology*, vol. 9, no. 1, article 291, 2009.

[30] Drosophila 12 Genomes Consortium, "Evolution of genes and genomes on the *Drosophila* phylogeny," *Nature*, vol. 450, no. 7167, pp. 203–218, 2007.

[31] S. Hutter, S. S. Saminadin-Peter, W. Stephan, and J. Parsch, "Gene expression variation in African and European populations of *Drosophila* melanogaster," *Genome Biology*, vol. 9, no. 1, article R12, 2008.

[32] L. Müller, S. Hutter, R. Stamboliyska, S. S. Saminadin-Peter, W. Stephan, and J. Parsch, "Population transcriptomics of *Drosophila* melanogaster females," *BMC Genomics*, vol. 12, article 81, 2011.

[33] J. M. Ranz, C. I. Castillo-Davis, C. D. Meiklejohn, and D. L. Hartl, "Sex-dependent gene expression and evolution of the *Drosophila* transcriptome," *Science*, vol. 300, no. 5626, pp. 1059–1060, 2003.

[34] S. Glinka, L. Ometto, S. Mousset, W. Stephan, and D. de Lorenzo, "Demography and natural selection have shaped genetic variation in Drosophila melanogaster: a multi-locus approach," *Genetics*, vol. 165, no. 3, pp. 1269–1278, 2003.

[35] J. Parsch, Z. Zhang, and J. F. Baines, "The influence of demography and weak selection on the McDonald-Kreitman test: an empirical study in *Drosophila*," *Molecular Biology and Evolution*, vol. 26, no. 3, pp. 691–698, 2009.

[36] Y. Zhang, D. Sturgill, M. Parisi, S. Kumar, and B. Oliver, "Constraint and turnover in sex-biased gene expression in the genus *Drosophila*," *Nature*, vol. 450, no. 7167, pp. 233–237, 2007.

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

[38] P. Librado and J. Rozas, "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data," *Bioinformatics*, vol. 25, no. 11, pp. 1451–1452, 2009.

[39] R. Development Core Team, "R: a language and environment for statistical computing," Tech. Rep., R Foundation for Statistical Computing, Vienna, Austria, 2011.

[40] A. Eyre-Walker and P. D. Keightley, "Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change," *Molecular Biology and Evolution*, vol. 26, no. 9, pp. 2097–2108, 2009.

[41] G. Sella, D. A. Petrov, M. Przeworski, and P. Andolfatto, "Pervasive natural selection in the *Drosophila* genome?" *PLoS Genetics*, vol. 6, no. 6, article e1000495, 2009.

[42] P. Andolfatto, K. M. Wong, and D. Bachtrog, "Effective population size and the efficacy of selection on the X chromosomes of two closely related *Drosophila* species," *Genome Biology and Evolution*, vol. 3, pp. 114–128, 2011.

[43] B. Vicoso and B. Charlesworth, "Effective population size and the faster-X effect: an extended model," *Evolution*, vol. 63, no. 9, pp. 2413–2426, 2009.

[44] J. E. Mank, B. Vicoso, S. Berlin, and B. Charlesworth, "Effective population size and the Faster-X effect: empirical results and their interpretation," *Evolution*, vol. 64, no. 3, pp. 663–674, 2010.

[45] A. Das, S. Mohanty, and W. Stephan, "Inferring the population structure and demography of *Drosophila ananassae* from multilociis data," *Genetics*, vol. 168, no. 4, pp. 1975–1985, 2004.

[46] W. Stephan and H. Li, "The recent demographic and adaptive history of *Drosophila* melanogaster," *Heredity*, vol. 98, no. 2, pp. 65–68, 2007.

[47] S. Hutter, H. Li, S. Beisswanger, D. de Lorenzo, and W. Stephan, "Distinctly different sex ratios in African and European populations of *Drosophila* melanogaster inferred from chromosomewide single nucleotide polymorphism data," *Genetics*, vol. 177, no. 1, pp. 469–480, 2007.

[48] W. G. Hill and A. Robertson, "The effect of linkage on limits to artificial selection," *Genetical Research*, vol. 8, no. 3, pp. 269–294, 1966.

[49] G. Marais and B. Charlesworth, "Genome evolution: recombination speeds up adaptive evolution," *Current Biology*, vol. 13, no. 2, pp. R68–R70, 2003.

[50] B. Charlesworth, J. A. Coyne, and N. H. Barton, "The relative rates of evolution of sex chromosomes and autosomes," *American Naturalist*, vol. 130, no. 1, pp. 113–146, 1987.

[51] H. A. Orr and A. J. Betancourt, "Haldane's sieve and adaptation from the standing genetic variation," *Genetics*, vol. 157, no. 2, pp. 875–884, 2001.

[52] T. Connallon and A. G. Clark, "Association between sex-biased gene expression and mutations with sex-specific phenotypic consequences in *Drosophila*," *Genome Biology and Evolution*, vol. 3, no. 1, pp. 151–155, 2011.

[53] M. Parisi, R. Nuttall, D. Naiman et al., "Paucity of genes on the *Drosophila* X chromosome showing male-biased expression," *Science*, vol. 299, no. 5607, pp. 697–700, 2003.

GENERAL DISCUSSION

A long-standing debate in evolutionary biology has been if adaptive changes in organisms preferentially occur in protein-coding regions or in non-coding parts of the genome, such as promoter regions. Whereas an extensive body of work has investigated evolutionary processes in coding regions during the last 30 years, only relatively recently have people begun to investigate non-coding DNA evolution. Despite not coding for proteins, the majority of non-coding DNA is of functional relevance, due in large part to regulating gene expression. A major prerequisite enabling the occurrence of adaptive evolution on the gene expression level is the presence of heritable expression variation among members of a population. While variation at the DNA sequence level has been studied extensively in different populations of *D. melanogaster*, variation of gene expression in natural populations of *D. melanogaster* has not been examined exhaustively.

Today, it is widely accepted that gene expression regulation makes an important contribution to adaptive evolution, especially to phenotypic evolution. Nevertheless, more case studies are needed in order to gain a more profound understanding of the genetic, molecular, and evolutionary mechanisms that drive gene expression divergence.

A major factor influencing gene expression levels is the sex of an organism. Genes with sex-biased expression show faster rates of protein evolution between species than unbiased genes, suggesting that these genes can play important roles in reproductive isolation and speciation. However, it is not known if this also applies to between- and within population comparisons in *Drosophila*.

In this thesis, a combination of population transcriptomics and a candidate gene approach employing population genetic tools aimed at elucidating the extent and underlying genetic and evolutionary mechanisms of gene expression variation within and between natural populations of *D. melanogaster*. In addition, the role of sex-biased genes in intra- and inter-population divergence as well as inter-species divergence of *Drosophila* was studied on the DNA-sequence- and the gene expression level.

5.1 GENE EXPRESSION VARIATION WITHIN AND AMONG *D. MELANOGASTER* POPULATIONS

In Chapter 1 of this thesis, we investigated genome-wide gene expression levels in female adult flies from two natural populations of *Drosophila melanogaster*, one from the ancestral species range in sub-Saharan Africa and one from the derived species range in Europe. By comparing our results to those from a previous study examining whole-genome expression levels of adult male flies from the same populations, we contributed to the understanding of expression variation present within and among natural populations of the important and widely studied model organism *D. melanogaster*. Our main findings were threefold:

First, we found much lower levels of within-population variation in female compared to male flies, indicating that general differences exist between the sexes in gene expression regulation and /or there are stronger selective constraints acting on females. Second, there was a much higher number of genes (> 500) differentially expressed between the populations in females than in males. These genes are candidates for those that have undergone regulatory adaptation to the derived, European environment. Third, there was little overlap in genes that differ in expression between populations in males and females, suggesting that regulatory evolution has been mostly sex-specific.

Although gene expression variation has been studied previously in *Drosophila* (Jin *et al*. 2001, Rifkin *et al*. 2003, Meiklejohn *et al*. 2003, Gibson *et al*. 2004, Baker *et al*. 2007), these studies were based on a small number of lab strains or solely based on derived populations. Consequently, these studies neither provide a realistic picture of gene expression variation present in natural populations nor do they reveal the potential for gene expression levels to evolve adaptively in response to local environmental changes. An exception was a study by Hutter *et al*. (2008) that examined gene expression variation in male flies of natural populations of an ancestral and a derived *D. melanogaster* population. The same approach was applied to adult females flies of the same populations in our study of chapter 1, which provided three main improvements over existing studies. First, studying an ancestral and a derived population of *D. melanogaster* provides a fuller picture of gene expression variation present within the *D. melanogaster* species and allows for detection of genes that have been putative targets of regulatory evolution. Second, using strains derived from natural

populations instead of long-term lab strains gives a more realistic idea about gene expression variation present in the species. Third, the comparison of our data of female adults to that of male adults of the same strains (Hutter *et al*. 2008) enables to uncover sex-specific expression differences within and between the populations.

In female flies, we found significantly less gene expression variation within the populations than what has been found for male flies (Hutter *et al*. 2008). This is in agreement with a study by Baker *et al*. (2007), who detected less expression variation among females than among males in a study of five *D. melanogaster* long-term lab strains. There are several interesting findings that suggest that our results have a genetic basis: First, several studies have shown that variation present on the Y chromosome impacts the expression of hundreds of X-linked and autosomal genes in male flies (Lemos *et al*. 2008, Jiang *et al*. 2010, Lemos *et al*. 2010, Sackton *et al*. 2011). Since this source of variation is absent in female flies, one expects females to show lower variation of gene expression. Second, a recent study by Innocenti *et al*. (2011) showed that mitochondrial polymorphism influences gene expression of thousands of nuclear genes in male flies, whereas gene expression in female flies is left virtually unaffected.

Our findings of extensive differences in gene expression levels between the sexes fit in with previous findings (reviewed in Ellegren and Parsch 2007) and further highlight the importance of accounting for sex when investigating gene expression.

To be able to draw conclusions about sex-specific gene expression differences, we compared the data of our study to that of the study by Hutter *et al*. (2008). However, when comparing studies using different microarray platforms to each other, one has to be aware of biases introduced by platform differences, as has been pointed out by different studies (Irizarry *et al*. 2005, Larkin *et al*. 2005, Bammler *et al*. 2005). Still, these studies also showed that the biological treatment of the samples has a far greater impact on the measured expression levels than did the platform type and that when using standardized protocols for sample handling, hybridization, array processing, data acquisition and normalization, results of multiple platforms are comparable. Since all of these parameters were kept constant in both the male and the female array experiments, it is reasonable to assume that our comparison of the male and female experiments is reliable. This is also supported by concordant results of our microarray data and results from quantitative real-time PCR (qRT-PCR) that was performed on a subset of genes. The results indicated a good correlation between microarray data and qRT-PCR results for the female and the male experiment.

In our microarray analysis, we used RNA-extractions from whole-body adult flies raised under common lab conditions. This has several implications for the detection of expression variation. First, genes whose expression is limited to certain tissues can escape detection simply due to their low expression levels compared to other more broadly expressed genes. While this problem is inherent to the approach of studying whole-body gene expression, higher resolution of tissue-specific expression variation can be tackled by performing expression analyses of single tissues only. Second, using four-to-six day-old flies limits our analysis of expression to young adult flies, and does not allow for inferences about other developmental stages. Gene expression varies greatly between developmental stages, such that expression variation during other stages will differ from our findings for adult flies. In addition, immature fly stages can also be targets of selection (Sgrò *et al*. 2010, Frankel *et al*. 2011), and adaptation could at least in part occur through gene expression changes that occur over the course of development. Such expression changes could be missed when focusing on adult flies. Gene expression changes in specific developmental stages can be unraveled by analyzing expression within and between the populations for larval or pupal stages. Third, using flies raised under common lab conditions, we do not detect genotype-by-environment interactions that affect gene expression. The African and the European population are presumably adapted to their respective natural environments and gene expression results from a combination of genetic and environmental factors. Thus, gene expression levels might differ depending on the environment under which the flies have been raised. However, using common conditions for both populations ensures that all measured expression variation is due to genetic components. Measuring the environmental component of gene expression can be accomplished by analyzing the same strains kept under differing environmental conditions, such as different temperatures, levels of humidity, or exposure to different pathogens.

In our study, we detected more than 500 genes that were differentially expressed between the African and the European population. These expression changes can stem from *cis*-regulatory changes, *trans*-regulatory changes, or a combination thereof. However, this does not imply that these genes represent a complete list of genes where regulatory changes have occurred since the out-of-Africa movement. Compensatory interactions, which can occur among two or more sites in *cis*, two or more *trans*-factors, or in *cis* by *trans* interactions and that in sum leave the gene expression level unaffected, will go undetected using our approach. A recent meta-study of 230 enhancer swap experiments, *i.e*. experiments in which regulatory DNA from one species is used to drive gene expression in another species, found that regulatory changes accumulate with genetic divergence and that divergence in *cis* alone is

more common than divergence in *trans* alone. Furthermore, coevolution between *cis* and *trans* was also found to be common and its proportion increases with divergence time (Gordon and Ruvinsky 2012). Although the study focused on inter-species comparisons, this suggests that between our populations, most expression divergence is caused by changes exclusively in *cis* or *trans*.

5.2 EVOLUTIONARY AND GENETIC MECHANISMS UNDERLYING GENE EXPRESSION EVOLUTION

In Chapter 2, we applied population genetic analyses to upstream and coding regions of several differentially expressed genes (as identified in chapter 1) in order to shed light on the underlying genetic and evolutionary mechanisms that cause the genes' change in expression. For one of the genes, *CG34330*, evidence for recent adaptive evolution could be found. Low levels of polymorphism in the European but not the African population, most strikingly a 1.6 kb monomorphic region, as well as a significant Fay and Wu's *H* statistic suggest that this gene has been targeted by a recent selective sweep in the European population after the out-of-Africa migration. Within the monomorphic upstream region, one six bp indel and one SNP could be identified that are fixed in the European population but at low frequency in the African population, suggesting that the derived variants at these sites might be responsible for the differential expression of the gene between the populations. For another gene, *Jon99Ciii*, significant MK tests gave evidence for recurrent structural adaptation, *i.e.* changes in the amino acid sequence of the gene, since the split of *D. melanogaster* from *D. simulans* and *D. sechellia*.

Genomic studies have shown that *cis*-regulatory divergence is common between and within *Drosophila* species (Wittkopp *et al*. 2004, Genissel *et al*. 2008, Lemos *et al*. 2008, Wittkopp *et al*. 2009, McManus *et al*. 2010). However, the precise genetic basis of most *cis*-regulatory divergence is yet unidentified. Still, in recent years, more and more case studies have succeeded in pinpointing the exact location of functional regulatory changes. Although the overall number of such studies is still quite small, some general patterns of the genetic and evolutionary mechanisms underlying *cis*-regulatory divergence seem to evolve from them (Wittkopp and Kalay 2011). One striking finding is that often, few nucleotide substitutions in *cis*-regulatory elements can account for divergence in gene expression. A recent study by

Frankel *et al*. (2011) has identified five substitutions in an enhancer region of the gene *shavenbaby* that contribute to altered function of the enhancer in *D. sechellia* compared to *D. melanogaster*, resulting in strongly reduced differentiation of trichomes in larvae. Although the existence of additional loci contributing to the expression change cannot be excluded, these changes can explain a large proportion of the phenotypic change. Another recent study found that at least 11 substitutions control a novel gene expression pattern in the optic lobe of *D. santomea* (Rebeiz *et al*. 2011). Even one single point mutation was found to be sufficient to cause the expression of the gene *desatF* to change from a female-specific pattern to being expressed in both sexes in *D. takahashii*. This is due to the loss of a binding site for the transcription factor Dsx (Shirangi *et al*. 2009).

Although most studies of gene expression divergence have focused on between-species comparisons, *cis*-regulatory changes that account for expression differences have also been identified between populations of *D. melanogaster* (Rebeiz *et al*. 2009, Saminadin-Peter *et al*. 2011). Using the same approach and fly populations as we did, the study by Saminadin-Peter *et al*. (2011) identified a 1.2-kb *cis*-regulatory element of the gene *CG9509* that is associated with a selective sweep and can account for the differential expression of this gene between the populations. The ability of the 1.2-kb region to drive the altered expression level was confirmed using a reporter gene assay. The region contains only two fixed differences between the populations and six additional SNPs that are fixed in Europe, but appear at low frequency in Africa. Thus, these eight variants are candidates for those that control the gene's expression level.

In our study, we found one gene, *CG34330* that shows evidence for a recent selective sweep. In its upstream region, we identified one indel and one nucleotide change where a derived state is fixed in the European population, but occurs at low frequency in Africa. Thus, in light of the above findings from other studies, these changes can be sufficient to cause the observed expression difference between the populations. The study by Saminadin-Peter *et al*. (2011) has also proven the power of our candidate gene approach to detect *cis*-regulatory regions underlying adaptive gene expression differences.

5.3 THE CONTRIBUTION OF SEX-BIASED GENES TO WITHIN-POPULATION VARIATION, BETWEEN-POPULATION DIVERGENCE AND BETWEEN-SPECIES DIVERGENCE

In Chapter 3, we examined the molecular evolution of sex-biased genes and their contribution to within-population variation, between-population divergence, and between-species divergence in *D. melanogaster* and *D. ananassae*, two independent lineages within the *melanogaster* subgroup. We found strong purifying selection limiting protein sequence variation within species, whereas between species a considerable fraction of divergence could be attributed to positive selection. At the level of gene expression, sex-biased genes show high expression divergence between species, but low divergence between populations. In *D. melanogaster*, the proportion of adaptive substitutions between species is greatest for male-biased genes, especially for those lying on the X chromosome, whereas this pattern could not be detected in *D. ananassae*.

Despite many studies having identified male-biased genes as the fastest evolving class of genes (reviewed in Ellegren and Parsch 2007), some ambiguity has remained regarding the underlying causes of this finding. Most sex-biased genes are primarily expressed in sex-limited reproductive tissues and are involved in reproduction (Parisi *et al*. 2003, Meisel 2011). These genes often evolve rapidly because of sexual selection pressures. Additionally, it has been found that sex-biased genes typically have a narrow expression pattern, *i.e*. they are expressed in few tissues (Yang *et al*. 2006, Mank *et al*. 2008, Meisel 2011), and that narrowly expressed genes evolve faster than broadly expressed genes (Duret and Mouchiroud 2000, Zhang and Li 2004, Liao *et al*. 2006, Haerty *et al*. 2007). Consequently, the faster evolution of sex-biased genes could be due to several reasons: First, the genes' sexually dimorphic expression pattern, second, their expression in reproductive tissues and third, their narrow expression pattern. A recent study by Meisel (2011) shed light on this issue. By studying the rates of protein-coding sequence evolution of sex-biased genes in *D. melanogaster*, he found that the faster evolution of sex-biased versus unbiased genes that are expressed in somatic tissues of both sexes is mainly due to the genes' narrow expression pattern. However, sex-biased genes in sex-limited tissues evolve faster than other genes that show a narrow expression pattern. This implies that faster evolution for sex-biased genes applies only for those genes expressed in reproductive tissues, while other sex-biased genes benefit from their narrow expression pattern. The study also detected a positive correlation between the degree of sex-bias and the rate of protein evolution.

In our study, we found that male-biased genes show the highest proportion of adaptive substitutions between *D. melanogaster* and *D. simulans*, but such a pattern was not found when comparing *D. ananassae* to *D. atripex* or *D. phaeopleura*. However, as pointed out in Grath and Parsch (2012, under revision) this finding might be due to a sampling-effect. While the male-biased genes examined in *D. melanogaster* were highly male-biased, the male-biased genes examined in *D. ananassae* were not. These were taken from a study by Grath *et al*. (2009) and originally had been picked because their orthologs showed highly male-biased expression in *D. melanogaster*, not in *D.ananassae* (Pröschel *et al*. 2006, Baines *et al*. 2008). In *D. ananassae*, these genes did not show especially high levels of male-biased expression. As the degree of sex-bias positively correlates with the rate of protein evolution (Meisel 2011), the lack of this class of genes in the data set could explain why we did not detect strong adaptive evolution for male-biased genes in the *ananassae* lineage.

The study by Grath and Parsch (2012, under revision) that investigated the genome-wide rates of molecular evolution of sex-biased genes along the *D. melanogaster* and *D. obscura* lineage also found evidence for a fast-X effect in both lineages, *i.e*. X-linked genes evolve faster than autosomal genes. This effect was especially pronounced for male-biased genes. Concordant with these findings, our study revealed higher proportions of adaptive evolution for X-linked than for autosomal genes among all gene classes in *D. melanogaster*. The highest value was observed for male-biased, X-linked genes. A fast-X effect is expected if new beneficial mutations are, on average, recessive (Charlesworth *et al*. 1987). On the X-chromosome, these mutations are immediately exposed to selection in hemizygous males. As the fast-X effect is mainly driven by selection in males, male-biased genes, which primarily have fitness effects in males (Connallon and Clark 2011), are expected to be disproportionately affected by positive selection.

In our study, estimation of the rate of adaptive substitution, α, was performed using the method of Eyre-Walker and Keightley (2009), which is based on the logic of the MK test (McDonald and Kreitman 1991). α is estimated within a Bayesian framework using a Monte Carlo Markov Chain. This method accounts for the contribution of slightly deleterious mutations to polymorphism and divergence when inferring α. Previously, it has been suggested that this can be accomplished by excluding low-frequency variants from analysis (Fay *et al*. 2001). However, this approach will not remove all slightly deleterious mutations and also removes some effectively neutral mutations, leading to an underestimation of α (Charlesworth and Eyre-Walker 2008). In order to circumvent these problems, the method of

Eyre-Walker and Keightley (2009) estimates the distribution of fitness effects (DFE) of new deleterious mutations using the site frequency spectrum of the polymorphisms. It then uses the inferred DFE to predict the fraction of substitutions between species that stem from effectively neutral and slightly deleterious mutations and estimates the proportion of positively selected substitutions. One main assumption of the method is that strongly beneficial mutations do not contribute to the observed polymorphism, since these are rare and rapidly fixed. A new, similar method to infer the DFE and the proportion of positively selected substitutions (Wilson *et al*. 2011) also takes the occurrence of strongly beneficial mutations into account. Applying their method to X-linked data from *D. melanogaster*, they find that the vast majority (> 85%) of new, nonsynonymous mutations are very strongly or strongly deleterious, a finding in accordance with our study. Beyond that, they found that at least 0.7% of these mutations were moderately or strongly beneficial. Of the observed substitutions between *D. melanogaster* and *D. yakuba*, 80% were found to be moderately or strongly beneficial, a finding that is similar to our estimates of α on the X chromosome for male-biased and female-biased genes, but higher than our estimate for unbiased genes. However, the study by Wilson *et al*. (2011) did not consider the sex-bias of their genes chosen for analysis. An explanation for their high α value could lie in the gene set they analyzed. All of the loci were located in a region of the X chromosome that experiences very frequent recombination. Hence, their finding of a high proportion of adaptive substitutions is expected, since in regions with high recombination rates, selection can act more efficiently. Such regions have higher effective population sizes and thus experience higher rates of adaptive evolution and a higher efficacy of selection against deleterious mutations (Kimura 1983).

When applying methods that are based on the logic of the MK test to estimate the proportion of adaptive substitutions, one has to keep in mind that they all rely on certain assumptions. Synonymous changes are generally assumed to be unconstrained and serve as a neutral reference. However, there is evidence that in *Drosophila*, synonymous mutations are not necessarily neutral (Plotkin and Kudla 2011), as there can be selection between alternative codons encoding the same amino acid. Recently, Fay (2011) has particularly stressed the necessity to critically consider the assumption that all sites evolve independently of one another, as empirical studies suggest that extensive epistasis can occur between sites (*e.g.*, Malmberg and Mauricio 2005, Lunzer *et al*. 2010). Fay lists two scenarios that could lead to spurious signals of positive selection: First, epistasis in duplicated genes can lead to changes in selective constraint over time. Immediately after duplication, genes show low levels of constraint, which then increases over time. This implies that higher rates of protein divergence

contribute to divergence than to polymorphism, thus inflating the ratio of Dn/Ds over Pn/Ps. Second, positive selection can fix deleterious mutations if they hitchhike along with positively selected mutations. This can lead to overestimation of the frequency of positive selection. However, such a scenario is only realistic if the positive effect of the beneficial mutation outweighs the negative effect of the deleterious mutation. Hence, the extent to which these scenarios really influence estimates of positive selection based on the MK test must be investigated in more detail.

5.4 CONCLUSION

Changes in gene expression can have a strong impact on the phenotype of an organism and play an important role in adaptive evolution.

This dissertation has contributed to obtaining a fuller picture of gene expression variation that is present within and between populations of the important model organism *D. melanogaster*, underpinning the strong influence of sex on expression variation and adaptive evolution of regulatory changes. The genetic and evolutionary mechanisms underlying differential gene expression in one gene were partly elucidated, although functional tests are still needed to confirm the findings. In addition, it was shown that inter-species divergence is not just a simple extension of inter-population divergence and intra-population variation, but is influenced to a larger degree by selection acting on sex-biased genes, both with regard to protein sequence and gene expression.

APPENDIX A: SUPPLEMENTARY ONLINE MATERIAL MÜLLER *ET AL*. (2011)

ADDITIONAL FILE 1

Expression divergence between the African and European populations. Table of relative expression levels in the African and European populations of all genes used in the analysis.

see http://www.biomedcentral.com/1471-2164/12/81/additional

ADDITIONAL FILE 2

Diagnostic PCR for the Accord insertion and tandem duplication of the *Cyp6g1* gene. Agarose gel images of diagnostic PCR for the *Accord* element insertion and tandem duplication of the *Cyp6g1* gene.

ADDITIONAL FILE 3

GO-term enrichment of genes over-expressed in the African and European populations. Table of all GO-terms with significant over-representation in each population.

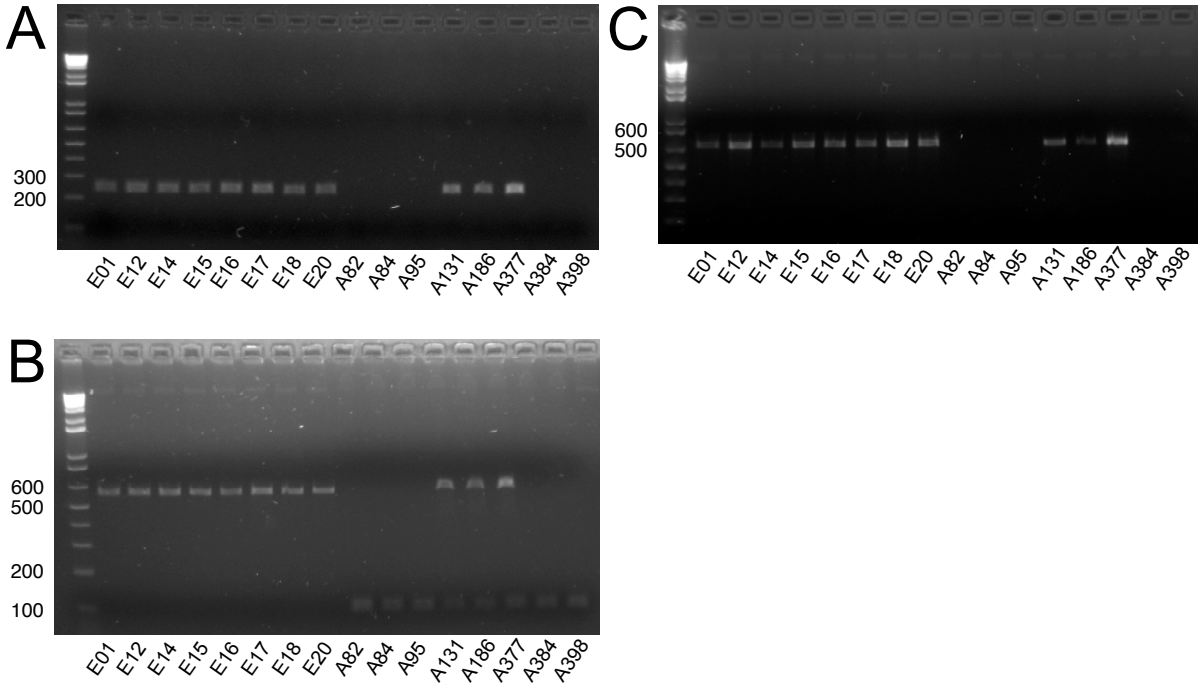
ADDITIONAL FILE 4

Correlation of fold-change expression differences as measured by microarray and qRT-PCR. Plot of 1,560 pairwise comparisons of all 16 *D. melanogaster* strains for 13 different genes.

ADDITIONAL FILE 5

Meta-analysis of male and female between-population gene expression divergence. Table of differences in expression between the African and European populations for all genes common to the female and male experiments.

see http://www.biomedcentral.com/1471-2164/12/81/additional



ADDITIONAL FILE 2 **Gel pictures of diagnostic PCR fragments for the *Accord* insertion and the *Cyp6g1* gene duplication**

Agarose gel (A) shows diagnostic fragments for the presence of the *Accord* insertion. Longer fragments in (B) demonstrate the presence of the *Accord* insertion, while short fragments demonstrate absence of the insertion. Picture (C) shows diagnostic fragments for the duplication of the *Cyp6g1* gene.

ADDITIONAL FILE 3

GO-term enrichment of genes over-expressed in the African and European populations. Table of all GO-terms with significant over-representation in each population.

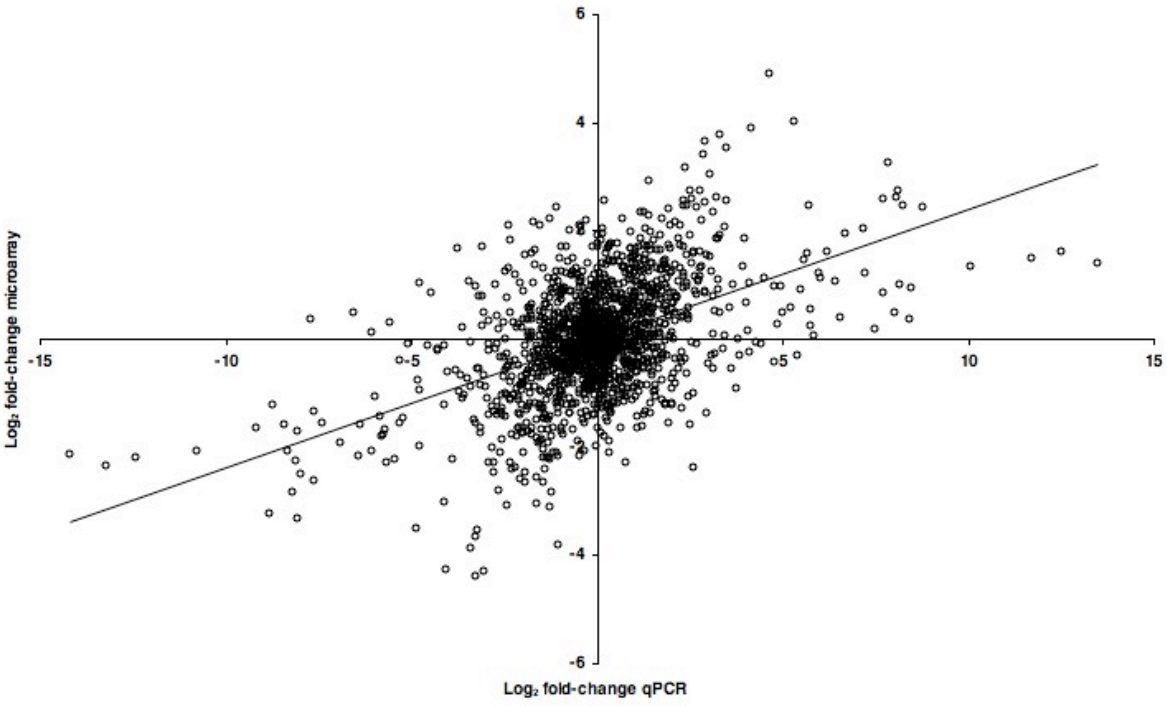
Additional file 3: GO-term enrichment of genes over-expressed in the African and European populations

Over-expression	GO ID	Ontology	Term	Adjusted P-value
Africa	GO:0004558	molecular_function	alpha-glucosidase activity	0.0055
Africa	GO:0015926	molecular_function	glucosidase activity	0.0181
Africa	GO:0004806	molecular_function	triglyceride lipase activity	0.0205
Africa	GO:0003697	molecular_function	single-stranded DNA binding	0.0205
Africa	GO:0019201	molecular_function	nucleotide kinase activity	0.0205
Africa	GO:0016701	molecular_function	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	0.0253
Africa	GO:0016776	molecular_function	phosphotransferase activity, phosphate group as acceptor	0.0354
Africa	GO:0004129	molecular_function	cytochrome-c oxidase activity	0.0356
Africa	GO:0015002	molecular_function	heme-copper terminal oxidase activity	0.0356
Africa	GO:0016675	molecular_function	oxidoreductase activity, acting on heme group of donors	0.0356
Africa	GO:0016676	molecular_function	oxidoreductase activity, acting on heme group of donors, oxygen as acceptor	0.0356
Africa	GO:0019205	molecular_function	nucleobase, nucleoside, nucleotide kinase activity	0.0467
Africa	GO:0019318	biological_process	hexose metabolic process	0.0053
Africa	GO:0009586	biological_process	rhodopsin mediated phototransduction	0.0055
Africa	GO:0016059	biological_process	deactivation of rhodopsin mediated signaling	0.0055
Africa	GO:0022400	biological_process	regulation of rhodopsin mediated signaling pathway	0.0055
Africa	GO:0006277	biological_process	DNA amplification	0.0055
Africa	GO:0022604	biological_process	regulation of cell morphogenesis	0.0088
Africa	GO:0048814	biological_process	regulation of dendrite morphogenesis	0.0088
Africa	GO:0050773	biological_process	regulation of dendrite development	0.0088
Africa	GO:0007603	biological_process	phototransduction, visible light	0.0088
Africa	GO:0005996	biological_process	monosaccharide metabolic process	0.0110
Africa	GO:0022603	biological_process	regulation of anatomical structure morphogenesis	0.0125
Africa	GO:0008277	biological_process	regulation of G-protein coupled receptor protein signaling pathway	0.0133
Africa	GO:0050908	biological_process	detection of light stimulus involved in visual perception	0.0133
Africa	GO:0050962	biological_process	detection of light stimulus involved in sensory perception	0.0133
Africa	GO:0050774	biological_process	negative regulation of dendrite morphogenesis	0.0154
Africa	GO:0035050	biological_process	embryonic heart tube development	0.0154
Africa	GO:0014866	biological_process	skeletal myofibril assembly	0.0154
Africa	GO:0030708	biological_process	germarium-derived female germ-line cyst encapsulation	0.0154
Africa	GO:0048138	biological_process	germ-line cyst encapsulation	0.0154
Africa	GO:0048139	biological_process	female germ-line cyst encapsulation	0.0154
Africa	GO:0009581	biological_process	detection of external stimulus	0.0205
Africa	GO:0008360	biological_process	regulation of cell shape	0.0205
Africa	GO:0031345	biological_process	negative regulation of cell projection organization	0.0205
Africa	GO:0009584	biological_process	detection of visible light	0.0205
Africa	GO:0006123	biological_process	mitochondrial electron transport, cytochrome c to oxygen	0.0205
Africa	GO:0006006	biological_process	glucose metabolic process	0.0212
Africa	GO:0042773	biological_process	ATP synthesis coupled electron transport	0.0212
Africa	GO:0042775	biological_process	mitochondrial ATP synthesis coupled electron transport	0.0212
Africa	GO:0007602	biological_process	phototransduction	0.0237
Africa	GO:0035075	biological_process	response to ecdysone	0.0253
Africa	GO:0048545	biological_process	response to steroid hormone stimulus	0.0253
Africa	GO:0007307	biological_process	eggshell chorion gene amplification	0.0253
Africa	GO:0048584	biological_process	positive regulation of response to stimulus	0.0253
Africa	GO:0035072	biological_process	ecdysone-mediated induction of salivary gland cell autophagic cell death	0.0253
Africa	GO:0035078	biological_process	induction of programmed cell death by ecdysone	0.0253
Africa	GO:0035081	biological_process	induction of programmed cell death by hormones	0.0253
Africa	GO:0022904	biological_process	respiratory electron transport chain	0.0273
Africa	GO:0006119	biological_process	oxidative phosphorylation	0.0281
Africa	GO:0010769	biological_process	regulation of cell morphogenesis involved in differentiation	0.0321
Africa	GO:0010975	biological_process	regulation of neuron projection development	0.0321
Africa	GO:0031344	biological_process	regulation of cell projection organization	0.0321
Africa	GO:0012502	biological_process	induction of programmed cell death	0.0321
Africa	GO:0044262	biological_process	cellular carbohydrate metabolic process	0.0348
Africa	GO:0030029	biological_process	actin filament-based process	0.0348
Africa	GO:0030036	biological_process	actin cytoskeleton organization	0.0348
Africa	GO:0022900	biological_process	electron transport chain	0.0353
Africa	GO:0006066	biological_process	alcohol metabolic process	0.0354
Africa	GO:0051606	biological_process	detection of stimulus	0.0354
Africa	GO:0006631	biological_process	fatty acid metabolic process	0.0354
Africa	GO:0007015	biological_process	actin filament organization	0.0354
Africa	GO:0045664	biological_process	regulation of neuron differentiation	0.0354
Africa	GO:0009583	biological_process	detection of light stimulus	0.0354
Africa	GO:0010927	biological_process	cellular component assembly involved in morphogenesis	0.0354
Africa	GO:0030713	biological_process	ovarian follicle cell stalk formation	0.0356
Africa	GO:0009582	biological_process	detection of abiotic stimulus	0.0357
Africa	GO:0009719	biological_process	response to endogenous stimulus	0.0417
Africa	GO:0009725	biological_process	response to hormone stimulus	0.0417
Africa	GO:0006091	biological_process	generation of precursor metabolites and energy	0.0424
Africa	GO:0015980	biological_process	energy derivation by oxidation of organic compounds	0.0442
Africa	GO:0032787	biological_process	monocarboxylic acid metabolic process	0.0452
Africa	GO:0008258	biological_process	head involution	0.0467
Africa	GO:0048134	biological_process	germ-line cyst formation	0.0467
Africa	GO:0030239	biological_process	myofibril assembly	0.0467
Africa	GO:0016203	biological_process	muscle attachment	0.0467
Africa	GO:0050793	biological_process	regulation of developmental process	0.0476
Africa	GO:0050906	biological_process	detection of stimulus involved in sensory perception	0.0490
Africa	GO:0042995	cellular_component	cell projection	0.0142
Africa	GO:0016028	cellular_component	rhabdomere	0.0181
Africa	GO:0044429	cellular_component	mitochondrial part	0.0253
Africa	GO:0016459	cellular_component	myosin complex	0.0253
Africa	GO:0019897	cellular_component	extrinsic to plasma membrane	0.0253
Africa	GO:0005740	cellular_component	mitochondrial envelope	0.0266
Africa	GO:0031966	cellular_component	mitochondrial membrane	0.0354
Africa	GO:0005743	cellular_component	mitochondrial inner membrane	0.0354
Africa	GO:0030425	cellular_component	dendrite	0.0354
Africa	GO:0043234	cellular_component	protein complex	0.0356

Additional file 3: GO-term enrichment of genes over-expressed in the African and European populations

Over-expression	GO ID	Ontology	Term	Adjusted P-value
Africa	GO:0005751	cellular_component	mitochondrial respiratory chain complex IV	0.0356
Africa	GO:0045277	cellular_component	respiratory chain complex IV	0.0356
Africa	GO:0005746	cellular_component	mitochondrial respiratory chain	0.0357
Africa	GO:0070469	cellular_component	respiratory chain	0.0357
Africa	GO:0016585	cellular_component	chromatin remodeling complex	0.0467
Africa	GO:0030867	cellular_component	rough endoplasmic reticulum membrane	0.0467
Europe	GO:0004930	molecular_function	G-protein coupled receptor activity	0.0092
Europe	GO:0004984	molecular_function	olfactory receptor activity	0.0179
Europe	GO:0004252	molecular_function	serine-type endopeptidase activity	0.0362
Europe	GO:0005337	molecular_function	nucleoside transmembrane transporter activity	0.0386
Europe	GO:0035214	biological_process	eye-antennal disc development	0.0014
Europe	GO:0010160	biological_process	formation of organ boundary	0.0074
Europe	GO:0050907	biological_process	detection of chemical stimulus involved in sensory perception	0.0083
Europe	GO:0050911	biological_process	detection of chemical stimulus involved in sensory perception of smell	0.0083
Europe	GO:0048645	biological_process	organ formation	0.0099
Europe	GO:0048859	biological_process	formation of anatomical boundary	0.0099
Europe	GO:0008052	biological_process	sensory organ boundary specification	0.0197
Europe	GO:0007600	biological_process	sensory perception	0.0233
Europe	GO:0007455	biological_process	eye-antennal disc morphogenesis	0.0242
Europe	GO:0009593	biological_process	detection of chemical stimulus	0.0242
Europe	GO:0065004	biological_process	protein-DNA complex assembly	0.0269
Europe	GO:0007606	biological_process	sensory perception of chemical stimulus	0.0288
Europe	GO:0007608	biological_process	sensory perception of smell	0.0288
Europe	GO:0000377	biological_process	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	0.0319
Europe	GO:0000398	biological_process	nuclear mRNA splicing, via spliceosome	0.0319
Europe	GO:0009047	biological_process	dosage compensation, by hyperactivation of X chromosome	0.0334
Europe	GO:0001508	biological_process	regulation of action potential	0.0334
Europe	GO:0006544	biological_process	glycine metabolic process	0.0334
Europe	GO:0000375	biological_process	RNA splicing, via transesterification reactions	0.0362
Europe	GO:0008380	biological_process	RNA splicing	0.0362
Europe	GO:0048800	biological_process	antennal morphogenesis	0.0386
Europe	GO:0050906	biological_process	detection of stimulus involved in sensory perception	0.0475

ADDITIONAL FILE 4



ADDITIONAL FILE 4 **Correlation of fold-change expression differences as measured by microarray and qRT-PCR**
Shown are 1,560 pairwise comparisons of all 16 *D. melanogaster* strains for 13 different genes (Pearson's $R = 0.5$, $P < 0.0001$). The genes are the same as those in Figure 5, with the addition of *CG18179*, which did not differ in expression between populations but showed high expression variation within each population.

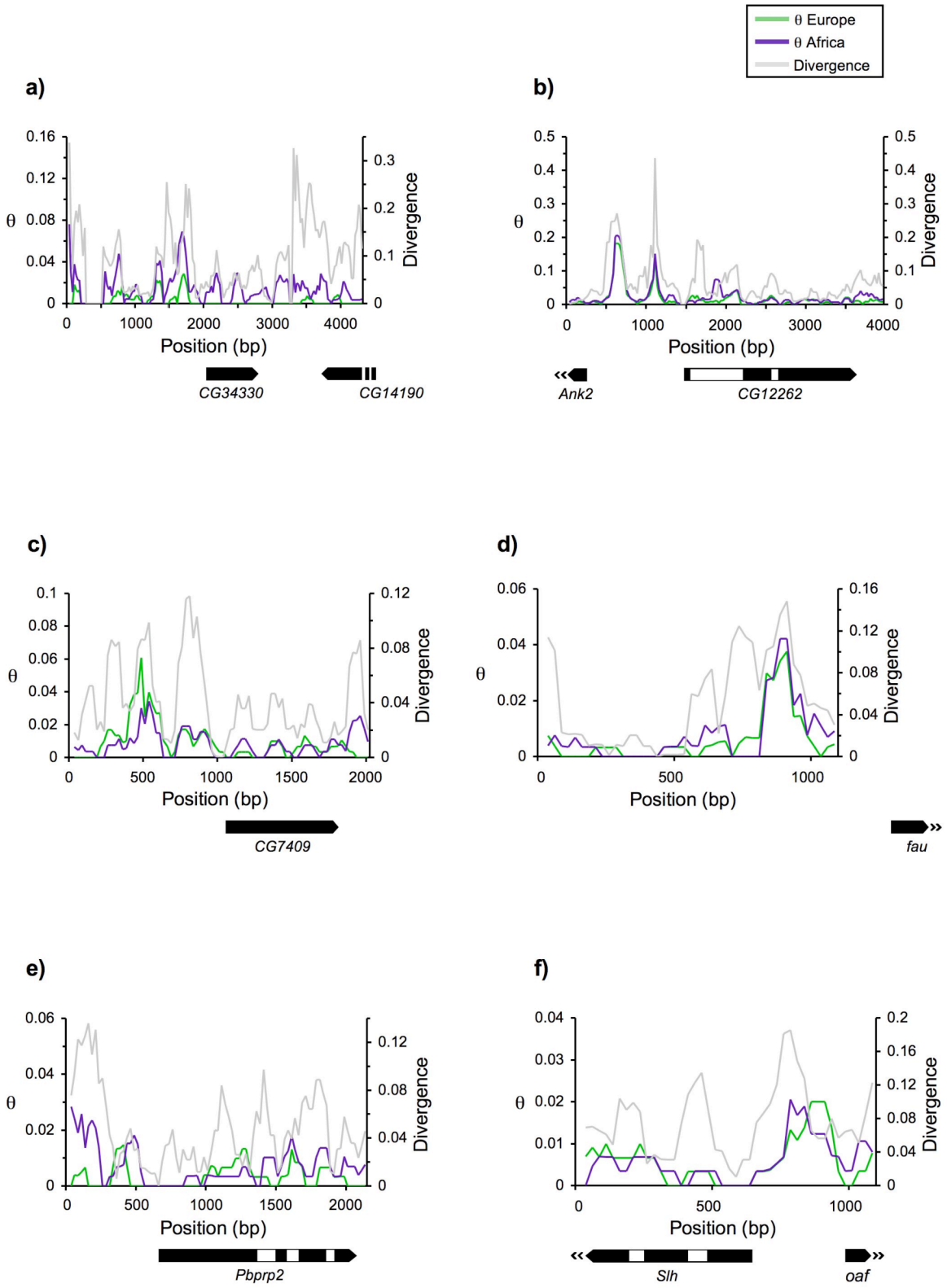
APPENDIX B: SUPPLEMENTARY MATERIAL MÜLLER AND PARSCH (UNPUBLISHED MANUSCRIPT)

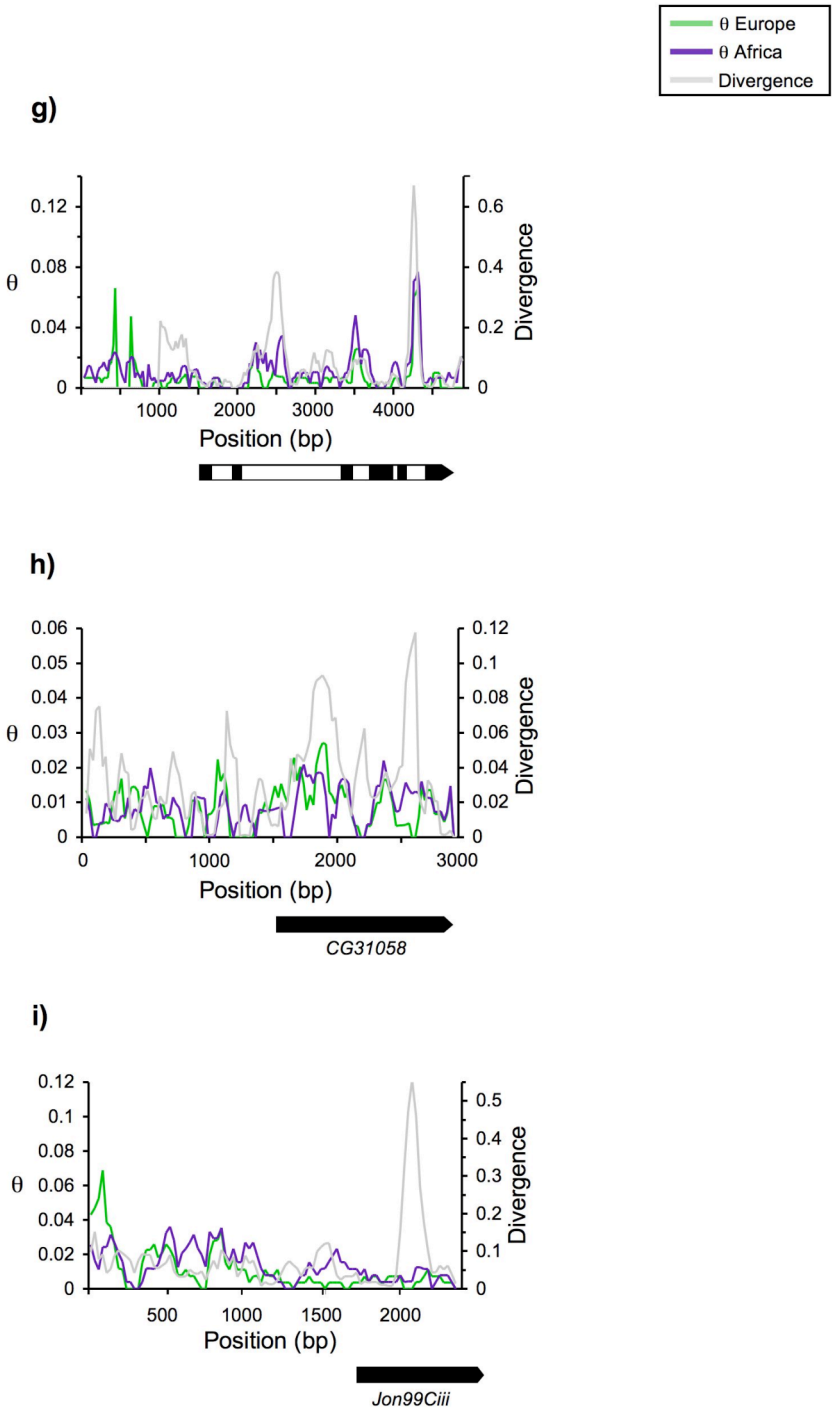
SUPPLEMENTARY FIGURE 1

Polymorphism and divergence for all genes analyzed

SUPPLEMENTARY TABLE 1

Primers used for PCR and sequencing reactions for all genes analyzed in this study


θ Europe
θ Africa
Divergence
a)
b)
c)
d)
e)
f)
θ
Divergence
Position (bp)
CG34330
CG14190
Ank2
CG12262
CG7409
fau
Pbprp2
Slh
oaf



SUPPLEMENTARY FIGURE 1 Polymorphism and divergence for all genes analyzed
Sliding window analyses of polymorphism and divergence for all genes analyzed in this study (window size = 100 bp, step size = 25 bp). Shown are nucleotide diversity (θ) in the European (green line) and the African (purple line) populations, as well as divergence between all *D. melanogaster* lines and an outgroup (gray line). *D. simulans* was used as an outgroup for all genes except for *CG34330* and *fau*, where *D. sechellia* was used. Positions of the genes are represented by arrows with the arrowhead indicating the direction of transcription, solid boxes representing exons and open boxes representing introns. Only constitutively spliced out introns are shown. Genes that extend beyond the sequenced region are indicated by double arrows or two vertical lines, respectively.

SUPPLEMENTARY TABLE 1 **Primers used for PCR and sequencing reactions for all genes analyzed in this study**

Gene	Forward primer	Primer sequence (5'-3')	Note	Reverse primer	Primer sequence (5'-3')	Note
CG34330	CG34330_F0	CGCTAGTATCTGGAAATCACCTG		CG34330_R0	GCAGCTGTGATACGCACATT	
	CG34330_F1	GGGGGCCAGTTTGTAATTTT		CG34330_R1	ACGTAGGCGAAGTTGTCCAC	
	CG34330_F2b	CACCCGGAGTAGTGTCCAGT		CG34330_R2b	GGCGAAAGAGTTTTGACTCG	
	CG34330_F3	TGATCAGCGCCCATCTGCGC		CG34330_R3	GCAGTTTCGAAGGTGCACTACGTT	
CG12262	CG12262_F1	GGGAGTGGGTAGCACAAAAA		CG12262_R1	GCTTGGACATTGGACCTTGT	
	CG12262_F2	GCTTTGGGAAAGTTCCATCA		CG12262_R2	GTTGTTCATCAGGCCCAGTT	
	CG12262_F3	TCCAATATCCGAAAATCCAG		CG12262_R3	TATCAGCCAGCATGAACTGC	
	CG12262_F4	GGTGTGGCGAACTGGTACTT		CG12262_R4	CAGCTGCAATTTGGAGACAA	
CG7409	CG7409_F1b	CAGGTCAGGCCCATAAATGT		CG7409_R1b	GAAACCATCCTTTCCGACAA	
	CG7409_F1seq	TACTCTCCGCTTCCGTTTGC	sequencing primer	CG7409_R1seq	TGCAAATCAACCCACCCAGC	sequencing primer
	CG7409_F2	AGTGGACGTCTATCAGAAAGC		CG7409_R2	TCGAGGGCGTAGTGTTTACC	
fau	fau_F1	AAACCTTTTCAAGTTGCAGACC		fau_R1	TATATCGTCGTTGCGTGTTGTT	
Pbprp2	Pbprp2_F1	GCAGTGCGTCTGCAGCAGGA		Pbprp2_R1	GACACCCCGCTGGAACTGCC	
	Pbprp2_F2	GGAAATAGGCGGACAGATGA		Pbprp2_R2	TGCTTGCTCATGACCTTGAC	
	Pbprp2_F3	GAGGAGATCAACAGGGACCA		Pbprp2_R3	CTGGAGAAACCCAGGTACGA	
oaf	oaf_F1	TACTGGTCGTAAACGCGATG		oaf_R1	GGCTGAGTGCCTTAAGATGC	
Nipsnap	Nipsnap_F0b	AATGAGTGATCGCACACTGC		Nipsnap_R0b	TGTTTTCTCCGCACTGTTGC	
	CG9212up_F	ACGGCCAATAGATTACGTAGC		CG9212up_R	TTAATCACCGGCGAAGGTAG	
	Nipsnap_F2b	ACGTGTAGGAGTGTGAATGC		Nipsnap_R2b	GTCCATATCGATAGCCAAGC	
	Nipsnap_F3	TCAATTTTTGGAAGGCCACT		Nipsnap_R3	CGCTGTCCGGTAAGAAAAAC	
	Nipsnap_F4	GGTTTGCCTTCTTCTCGTTG		Nipsnap_R4	CGACGGTGCGAATTTTAGTTAC	
	Nipsnap_F5	TGAGCAACTAAGCAGGCAAATC		Nipsnap_R5	CAGTCGAATGCTCAGCGATAAG	
				Nipsnap_R5seq	GGATTCTTCTCGCAGATCG	sequencing primer
CG31058	CG31058_F1	GGAGACGCAGCAGAGGTAAC		CG31058_R1	CAGAGATCCCAGCCAGCTAC	
	CG31058_F2	TTGCTCTTGCTGTCTGTGCT		CG31058_R2	GAACCCTTTTGCGAAACAAA	
	CG31058_F3	GCAAACTGGCCCATATCATT		CG31058_R3	TCAAGCCAAAGTGCAAAGTG	
	CG31058_seqF3	AGTCGGCACAAGTTGCAGTT	sequencing primer	CG31058_seqR3	TTACATTTGACGGGCGAATC	sequencing primer
Jon99Ciii	Jon99Ciii_F1	CGACTTTCGATTTTCGAAGC		Jon99Ciii_R1	GCTCTTTCGCTTTTCTCGTG	
	Jon99Ciii_F2	TTCTGCAGTGTACAACTGGG		Jon99Ciii_R2	AGTCCAGGTATCCGGTGACGC	

APPENDIX C: SUPPLEMENTARY ONLINE MATERIAL MÜLLER *ET AL*. (2012)

SUPPLEMENTARY TABLE 1

List of all genes used in this study, along with their chromosomal locations and their sex-biased expression classifications.

SUPPLEMENTARY TABLE 1 **List of all genes used in this study, along with their chromosomal locations and their sex-biased expression classifications.**

	D. melanogaster		*D. ananassae*	
Gene	Bias	Chrom	Bias	Chrom.
CG1980	Male	3R		
CG2140	Male	2R		
CG3085	Male	2R	Male	3L
CG3483	Male	2R		
CG5045	Male	2L		
CG5276	Male	3R		
CG5565	Male	2L		
CG6036	Male	3R	Male	2L
CG6130	Male	3R		
CG6255	Male	3R		
CG6332	Male	3R		
CG6971	Male	3R	Male	2L
CG6980	Male	3R	Male	2L
CG7251	Male	2L		
CG7387	Male	3L	Female	2R
CG7409	Male	3L		
CG7929	Male	3R		
CG7931	Male	3R		
CG8277	Male	3L	Male	2R
CG8564	Male	3L		
CG9314	Male	2L		
CG9531	Male	2L		
CG10252	Male	3R	Male	2L
CG10307	Male	2R		
CG10750	Male	2L	Unbiased	3R
CG11037	Male	3L		
CG11475	Male	2R		
CG13527	Male	2R		
CG14717	Male	3R	Unbiased	2L
CG14926	Male	2L		
CG15179	Male	3R		
CG17376	Male	2L		
CG17956	Male	3R		
CG18266	Male	2L	Male	3R
CG18418	Male	3L	Male	2R
CG1314	Male	X	Male	X
CG1503	Male	X		
CG1668	Male	X		
CG1950	Male	X		
CG2574	Male	X		
CG2577	Male	X	Male	X
CG5334	Male	X		
CG5662	Male	X		
CG6789	Male	X		
CG6999	Male	X		
CG7860	Male	X		
CG9156	Male	X		
CG10920	Male	X	Male	X
CG11379	Male	X	Male	X
CG11697	Male	X	Male	X
CG12395	Male	X		
CG12681	Male	X		
CG18341	Male	X	Unbiased	X
CG1600	Unbiased	2R		
CG3476	Unbiased	2L	Male	3R
CG3652	Unbiased	2L		
CG3683	Unbiased	2R		
CG5915	Unbiased	3R	Male	2L
CG5919	Unbiased	3R		
CG6094	Unbiased	2L		
CG6913	Unbiased	3R		
CG6981	Unbiased	3L	Unbiased	2R
CG7484	Unbiased	3L		
CG7508	Unbiased	3R	Unbiased	2L
CG7953	Unbiased	2L		
CG8392	Unbiased	2R		
CG8844	Unbiased	2L		
CG9283	Unbiased	3L		
CG9437	Unbiased	2R		
CG9617	Unbiased	3R		
CG9822	Unbiased	2R		

	D. melanogaster		*D. ananassae*	
Gene	**Bias**	**Chrom**	**Bias**	**Chrom.**
CG9893	Unbiased	2R		
CG10035	Unbiased	3R	Female	2L
CG10623	Unbiased	2L		
CG10853	Unbiased	3L	Unbiased	2R
CG11785	Unbiased	3R		
CG11981	Unbiased	3R	Female	2L
CG13189	Unbiased	2R	Unbiased	3L
CG13419	Unbiased	3R		
CG13845	Unbiased	3R		
CG13934	Unbiased	3L		
CG16985	Unbiased	3L		
CG17404	Unbiased	3R		
CG31058	Unbiased	3R		
CG33976	Unbiased	3R		
CG1397	Unbiased	X		
CG1751	Unbiased	X		
CG2555	Unbiased	X		
CG3603	Unbiased	X		
CG9164	Unbiased	X		
CG9571	Unbiased	X		
CG9723	Unbiased	X	Unbiased	X
CG9919	Unbiased	X		
CG11126	Unbiased	X		
CG14227	Unbiased	X		
CG14629	Unbiased	X		
CG14772	Unbiased	X		
CG14797	Unbiased	X		
CG15247	Unbiased	X		
CG15313	Unbiased	X		
CG15336	Unbiased	X	Unbiased	X
CG1239	Female	3R	Unbiased	2L
CG2867	Female	3R		
CG3509	Female	3R	Female	2L
CG3831	Female	2R		
CG3975	Female	2L		
CG4236	Female	3R		
CG4299	Female	3R		
CG4570	Female	3R		
CG4973	Female	3R	Unbiased	2L
CG5272	Female	3L	Female	2R
CG5363	Female	2L		
CG5499	Female	3R	Female	2L
CG5757	Female	2R		
CG6459	Female	2R	Female	3L
CG6554	Female	3R		
CG6874	Female	3L		
CG7840	Female	2L	Female	3R
CG9135	Female	2L	Female	3R
CG9273	Female	2L		
CG9383	Female	3L	Unbiased	2R
CG10206	Female	2L		
CG12262	Female	3L		
CG12276	Female	3R	Female	2L
CG12314	Female	2L		
CG12909	Female	2R		
CG13690	Female	2L		
CG17361	Female	3L		
CG17950	Female	2R		
CG32409	Female	3L		
CG1749	Female	X	Unbiased	X
CG2222	Female	X	Female	X
CG3004	Female	X	Unbiased	X
CG3024	Female	X	Male	X
CG3704	Female	X		
CG4593	Female	X	Male	X
CG8326	Female	X		
CG8675	Female	X		
CG9125	Female	X		
CG9915	Female	X		
CG12117	Female	X		
CG14434	Female	X		
CG15717	Female	X	Female	X

APPENDIX D: PROTOCOLS

1 MICROARRAYS

2 PCR

3 SEQUENCING

4 QRT-PCR

1 MICROARRAYS

I. RNA Extraction

2. For each "unit" of extraction (enough RNA for two hybridizations+) collect three sets of 13 females (40 females total), aged 4-6 days.

3. When ready for extraction, knock flies out and transfer each set of 13 females to individual 1.5 mL tubes on ice.

4. Add 200 uL Trizol to each tube and grind flies completely. Combine the 3 tubes of females into 1 tube (it may be necessary to first cut the pipette tip to avoid clogging with fly material).

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

6. Centrifuge 12,000 g at 4º C for 10 min., transfer supernatant to a clean tube.

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

8. Centrifuge 12,000 g at 4º C for 10 min., transfer the aqueous (upper) phase to a clean tube (must be Rnase-free at this stage!). Be careful not to transfer any of the interphase.

9. Add 500 uL isopropanol, incubate at room temp. for *exactly* 10 min.

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

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

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

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

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

15. Quantify RNA (dilution of 2uL sample + 98 uL Rnase-free water) with NanoDrop. Get concentration by A_{260} (1 abs = 40 ug/mL = 40 ng/uL) and sample purity by A_{260}/A_{280}.

16. Run 1 uL on a gel, rRNA bands should be visible, etc.

17. You should be left with ≈27 uL of sample at a concentration of 4-5 ug/uL. This is enough for two hybridizations (50-60ug per hyb required).

II. cDNA Synthesis and Amino Allyl Labeling

adapted from the Invitrogen SuperScriptTM Indirect cDNA Labeling System instruction manual

MATERIALS
This protocol makes use of the following Invitrogen packages: SuperScriptTM Indirect cDNA Labeling System, Alexa Fluor® 555 and Alexa Fluor® 647 Reactive Dye DecaPacks, and cDNA Labeling Purification Module ("SNAP" colums).

First-Strand cDNA Synthesis

Note: The following first-strand synthesis is to be performed in duplicate (RT for 2 X 25-30 ug) for hybridization to arrays

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

Component	Volume
25-30ug total RNA	X ul
Anchored Oligo(dT)B20B Primer (2.5ug/ul)	2 ul
DEPC treated H20	To 18 ul

*Note: the cocktail (without theSuperscript RT) in **step 3** should be prepared before starting, but wait until just before adding cocktail to add the Superscript RT enzyme).*

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

3. Add the following to each tube:

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

Note: if making a cocktail allow for some loss of volume when pipetting 12 ul of the above mix into each tube from step 1. For example, for 4 reactions, multiply each of the above numbers by 4.25 to make the cocktail.

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

5. Proceed directly to Hydrolysis and Neutralization.

Hydrolysis and Neutralization

This step is taken to degrade the original RNA

1. Add 15ul of 1N NaOH to each reaction tube from the First Strand cDNA synthesis reaction. Mix Thoroughly.

2. Incubate tube at 70ºC for 10 min.

3. Add 15ul of 1N HCl **immediately** after the 10min incubation to neutralize the pH and mix gently.

4. Proceed directly to Purifying First-Strand cDNA.

Purifying First-Strand cDNA

This step removes unincorporated dNTPs by ethanol precipitation.

Note: At this step, the two replicate tubes of first-strand cDNA should be combined for precipitation

1. Add 24 ul 3M Sodium Acetate, pH 5.2 and 2ul Glycogen to the combined neutralized reactions and mix.

2. Add 360 ul ice-cold 100% ethanol and mix by vortexing.

3. Place at –20ºC for at least 1 hr. Samples can also be incubated overnight at –20ºC, as this may increase yield, or stored for several days.

4. Centrifuge at 14,000g at 4º C for 20 min. Carefully remove and discard the supernatant.

Note: this is a good point to begin preparing the fluorescent dyes in the "Labeling with Fluorescent Dye" step.

5. Wash the pellet with 1 mL 75% ethanol and centrifuge at 14,000g at 4º C for 2 min. Carefully remove and discard the supernatant.

6. Centrifuge tube and carefully remove any supernatant that accumulated from the spin.

7. Air dry the samples to evaporate any ethanol that may still be on the sample. Sample will turn from white to clear and viscous (glass-like) when ready. Avoid over-drying, as it will be harder to resuspend the samples.

8. After samples have been appropriately dried, resuspend each sample in 5ul of 2X Coupling Buffer.

9. Proceed directly to Labeling with Fluorescent Dye.

Labeling with Fluorescent Dye

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

1. Remove the appropriate Alexa Fluor® dye vials from -20°C storage.

2. Add 2ul of DMSO directly to each dye vial and mix thoroughly.

Note: DMSO is hygroscopic and will absorb moisture from the air, which reacts with the dyes to reduce the coupling efficiency. So, warm DMSO to room temperature before use and keep the cap closed on the vial when not in use.

3. Centrifuge vials briefly.

4. Add the DMSO/dye solution to the tube from the Ethanol Precipitation step above. Add 3ul of DEPC-treated H2O to bring the final volume of the sample to 10ul.

5. Mix samples by vortexing, centrifuge briefly, and incubate at room temperature in the dark for 1-3 hours.

6. Go to Purification of Labeled cDNA after the dark incubation.

Purifying Labeled cDNA

Using the Invitrogen Purification Module

Before starting:

Prepare **Binding Buffer** with **Isopropanol** (if not already added)

-Add 6.5 ml Isopropanol to 18 ml (entire bottle) of Binding Buffer.

Prepare **Wash Buffer** with **Ethanol** (if not already added)

-Add 20 ml Ethanol to 5 ml (entire bottle) of Wash Buffer.

1. Add 700 ul of Binding Buffer to the reaction tube containing the labeled cDNA from Coupling. Vortex briefly to mix.

2. Each Low-Elution Volume Spin Cartridge is preinserted into a collection tube. For multiple reactions, clearly label each collection tube, and then load the cDNA/Binding Buffer solution directly onto the Spin Cartridge.

3. Centrifuge at 3,300g in a microcentrifuge for 1 minute. Remove the collection tube and discard the flow-through.

4. Place the Spin Cartridge in the same collection tube and add 600 ul of Wash Buffer to the column.

5. Centrifuge at maximum speed for 30 seconds. Remove the collection tube and discard the flow-through.

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

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

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

9. Centrifuge at maximum speed for 1 minute to collect the purified cDNA. The eluate contains your purified labeled cDNA.

> The sample can be stored at -20° C for up to one week prior to hybridization. Avoid freeze/thawing.

10. Next, it will be necessary to dry down the sample in a speed vac for ≈30 minutes in order to resuspend the sample in 12-24 ul of Pronto!™ Long Oligo/cDNA Hybridization Solution.

> *Note: the labeled cDNA resuspended in hybridization solution may be stored for up to one week at 4°C and protected from light.*

III. Prehybridization and Hybridization

The following protocols are from the Pronto!™ Microarray Hybridization Kit with small modification.

A. Preparation of Wash Solutions

Wash Solution 1

deionized water (Purelab from kitchen)	1,118.75ml
Universal Wash Reagent A	125ml
Universal Wash Reagent B	6.25ml

Wash Solution 2

deionized water (Purelab from kitchen)	3,562.5ml
Universal Wash Reagent A	187.5ml

Wash Solution 3

Wash Solution 2	750ml
deionized water (Purelab from kitchen)	3,000ml

B. Presoak and Prehybridization

1. Heat required volumes of both Pronto!™ Universal Pre-Soak Solution and Pronto!™ Universal Pre-Hybridization Solution to 42°C for at least 30 minutes.

2. Add 250 ul Sodium Borohydride Solution to 24.75 ml of 42°C Universal Pre-Soak Solution. Pipet several times to mix thoroughly. **DO NOT** add Sodium Borohydride to the Pre-Soak solution more than 15 minutes before use.

!!!!Do this in the hood, the Sodium Borohydride stock is very toxic!!!!

3. Immerse arrays in solution from Step 2 and incubate at 42°C for 20 minutes.

Note: DO NOT let arrays dry out between wash steps, because this will cause high backgrounds.

4. Transfer arrays to Wash Solution 2 and incubate at ambient temperature for 30 seconds.

5. Transfer to a fresh container of Wash Solution 2 for 30 seconds.

6. Transfer arrays to 42°C Universal Pre-Hybridization Solution (from Step 1) and incubate for 15 minutes.

7. Transfer arrays to a fresh container of Wash Solution 2 and incubate at ambient temperature for 1 minute.

8. Transfer arrays to Wash Solution 3 and incubate at ambient temperature for 30 seconds.

9. Transfer arrays to a fresh container of Wash Solution 3 and incubate at ambient temperature for 30 seconds.

10. Dip arrays in nuclease-free water at ambient temperature (22-25°C), and dry by centrifuging at 1,600 g for 2 minutes.

> *Note: This must be done immediately after taking the arrays out of the water. Prepare this step ahead of time by moving to the centrifuge at the beginning of Step 9!*

11. Store arrays in a dust-free environment (or in slide holder) until ready for use.

C. Hybridization

1. Wash the required number of pieces of cover glass (M-series lifter slips) with nuclease-free water, followed by ethanol. Dry cover glass by blowing high-purity compressed nitrogen gas or allow to air-dry in a dust-free environment.

2. Resuspend each of the two samples to be competitively hybridized in 40 ul of hybridization solution (for cDNA/long oligos). Combine the two samples into a PCR tube and mix well.

> *Note: ≈80 ul of total combined sample will fit under a 25x60 mm M-series lifter slip.*

3. Incubate the labeled cDNA solution at 95°C for 5 minutes, protecting samples from light.

4. Centrifuge the cDNA at 13,500g for 2 minutes to collect condensation. Do not place the solution on ice because this will cause precipitation of some of the components.

5. Place array in a Corning Hybridization Chamber (make sure to fill the two moisturizing wells with 10 ul dH_20 each). Pipet the labeled cDNA gently up and down and then transfer onto the surface of the printed side of the slide. Carefully place the cover glass on the array. Avoid trapping air bubbles between the array and the cover glass. Small air bubbles that do form usually dissipate during hybridization. Assemble the chamber.

6. Incubate the chamber-array assembly at 42°C for 14–20 hours using a water bath or a hybridization oven.

D. Post-Hybridization Washes

Note: Do not allow the arrays to dry out between washes, as this irreversibly increases background levels. Multiple containers are needed to perform the washes in the most efficient manner. Have all containers and the volumes of washing solutions ready before starting the procedure.

1. Heat required volume of Wash Solution 1 to 42°C for at least 30 minutes (note that Steps 3 and 4 both require prewarmed solutions).

2. Disassemble the hybridization chambers with the printed array side facing up.

3. Immerse arrays in Wash Solution 1 at 42°C for 1-2 minutes until the cover glass falls from the slide.

4. Transfer arrays to a fresh container of Wash Solution 1 at 42°C and incubate for 5 minutes.

5. Transfer arrays to Wash Solution 2 at ambient temperature (22-25°C) and incubate for 10 minutes.

6. Transfer arrays to Wash Solution 3 at ambient temperature and incubate for 2 minutes.

7. Repeat Step 6 twice.

8. Dry arrays by centrifugation at 1,600g for 2 minutes.

Note: This must be done immediately after taking the arrays out of the Wash Solution 3. Prepare this step ahead of time by moving to the centrifuge in the before the end of step 7!

9. Store arrays in light-proof container until ready to scan.

2 PCR

Setting up PCR Reaction

1. Create a mastermix consisting of the following ingredients for each sample:

ddH2O	15.70 µl
10x PCR-Buffer	2.0 µl
dNTP (10mM each)	0.50 µl
Primer F (10 pmol/µl)	0.40 µl
Primer R (10 pmol/µl)	0.40 µl
Taq Pol	0.20 µl
DNA (150 ng/µl)	0.80 µl

PCR Cycling conditions

1. Incubate at 94.0°C for 00:02:00
2. Incubate at 94.0°C for 00:00:45
3. Incubate at *°C for 00:00:45
4. Incubate at 72°C for **
5. Cycle to step 2 for 34 more times
6. Incubate at 72°C for 00:07:00
7. Incubate at 12°C forever

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

3 SEQUENCING

Protocol for Preparation of Sequencing Reactions with ABI 3730

1. Cleaning of PCR Product with ExoSAP

2. Create a mastermix consisting of the following ingredients for each sample:

ddH2O	3.5 μl
10x PCR-Buffer (blue)	0.5 μl
ExoSAP	1.0 μl

3. Put 5 μl of the mastermix in each sample, mix, put samples in thermocycler
4. Cycling conditions:
    a. Incubate at 37°C for 00:30:00
    b. Incubate at 80°C for 00:15:00
    c. Incubate at 12°C forever

2. Setting up Sequencing Reaction

1. For each sample use the following ingredients:

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

2. Use mastermix and automatic/multi-channel pipette when possible
3. Mix samples and put them into Thermocycler with sequencing cycling conditions

3. Sequencing Cycling Conditions

1. Incubate at 96°C for 00:01:00
2. Incubate at 96°C for 00:00:10
3. Incubate at 50°C for 00:00:15
4. Incubate at 60°C for 00:04:00
5. Cycle to step 2 for 34 more times
6. Incubate at 12°C forever

4 QRT-PCR

1. **RNA Extraction with Trizol**: see protocol microarrays

2. **DNAse I digestion:** according to Invitrogen DNase I protocol

3. **First-Strand cDNA Synthesis:** according to Invitrogen SuperScript™ II RT protocol

4. **qRT-PCR reaction mix**

20x TaqMan® Gene Expression Assays	1.0 μl
2x TaqMan® Gene Expression Master Mix	10.0 μl
cDNA template (1 to 100 ng)	4.0 μl
RNase-free water	5.0 μl

Cap the tube and invert several times
Centrifuge tube briefly
Transfer 20 μl of PCR reaction mix into each well of a 96-well reaction plate, cover with appropriate seal

Cycling conditions

1. Incubate at 95°C for 00:10:00
2. Incubate at 95°C for 00:00:15
3. Incubate at 60°C for 00:01:00
   Cycle to step 2 for 39 more times

BIBLIOGRAPHY

Adams, M. D. (2000). The genome sequence of *Drosophila melanogaster*. Science *287*, 2185–2195.

Akashi, H. (1996). Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. Genetics *144*, 1297–1307.

Andolfatto, P. (2005). Adaptive evolution of non-coding DNA in *Drosophila*. Nature *437*, 1149–1152.

Andolfatto, P. (2008). Controlling type-I error of the McDonald-Kreitman test in genomewide scans for selection on noncoding DNA. Genetics *180*, 1767–1771.

Aubin-Horth, N., Landry, C. R., Letcher, B. H., and Hofmann, H. A. (2005). Alternative life histories shape brain gene expression profiles in males of the same population. Proceedings Biological sciences / The Royal Society *272*, 1655–1662.

Avery, O. T., MacLeod, C. M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. Journal of Experimental Medicine, 137–157.

Bachtrog, D. (2008). Similar rates of protein adaptation in *Drosophila miranda* and *D. melanogaster*, two species with different current effective population sizes. BMC Evolutionary Biology *8*, 334.

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

Baker, D. A., Meadows, L. A., Wang, J., Dow, J. A., and Russell, S. (2007). Variable sexually dimorphic gene expression in laboratory strains of *Drosophila melanogaster*. BMC Genomics *8*, 454.

Bammler, T., Beyer, R. P., Bhattacharya, S., Boorman, G. A., Boyles, A., Bradford, B. U., Bumgarner, R. E., Bushel, P. R., Chaturvedi, K., Choi, D., *et al*. (2005). Standardizing global gene expression analysis between laboratories and across platforms. Nature Methods *2*, 351–356.

Beisswanger, S., Stephan, W., and De Lorenzo, D. (2006). Evidence for a selective sweep in the *wapl* region of *Drosophila melanogaster*. Genetics *172*, 265–274.

Beisswanger, S., and Stephan, W. (2008). Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America *105*, 5447-5452.

Bergman, C.M., and Kreitman, M. (2001). Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. Genome Research *11*, 1335-1345.

Brem, R. B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic dissection of transcriptional regulation in budding yeast. Science *296*, 752–755.

Catania, F., Kauer, M. O., Daborn, P. J., Yen, J. L., ffrench-Constant, R. H., and Schlötterer, C. (2004). World-wide survey of an *Accord* insertion and its association with DDT resistance in *Drosophila melanogaster*. Molecular Ecology *13*, 2491–2504.

Cavalieri, D., Townsend, J. P., and Hartl, D. L. (2000). Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. Proceedings of the National Academy of Sciences of the United States of America *97*, 12369–12374.

Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., Southwick, A. M., Absher, D. M., Grimwood, J., Schmutz, J., *et al*. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. Science *327*, 302–305.

Charlesworth, B., Coyne, J. A., and Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. The American Naturalist *130*, 113–146.

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

Chung, H., Bogwitz, M. R., McCart, C., Andrianopoulos, A., ffrench-Constant, R. H., Batterham, P., and Daborn, P. J. (2007). *Cis*-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. Genetics *175*, 1071–1077.

Connallon, T., and Clark, A. G. (2011). Association between sex-biased gene expression and mutations with sex-specific phenotypic consequences in *Drosophila*. Genome Biology and Evolution *3*, 151–155.

Daborn, P. J., Yen, J. L., Bogwitz, M. R., Le Goff, G., Feil, E., Jeffers, S., Tijet, N., Perry, T., Heckel, D., Batterham, P., *et al*. (2002). A single p450 allele associated with insecticide resistance in *Drosophila*. Science *297*, 2253–2256.

Darwin, C. (1859). On the origin of species by means of natural selection. John Murray. London.

David, J., and Capy, P. (1988). Genetic variation of *Drosophila melanogaster* natural populations. Trends in Genetics: TIG *4*, 106.

DiLeone, R. J., Russell, L. B., and Kingsley, D. M. (1998). An extensive 3' regulatory region controls expression of *Bmp5* in specific anatomical structures of the mouse embryo. Genetics *148*, 401–408.

Dowell, R. D. (2010). Transcription factor binding variation in the evolution of gene regulation. Trends in Genetics: TIG *26*, 468–475.

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

Duret, L., and Mouchiroud, D. (2000). Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Molecular Biology and Evolution *17*, 68–74.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research *32*, 1792–1797.

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

Enard, W., Khaitovich, P., Klose, J., Zöllner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., *et al.* (2002). Intra- and interspecific variation in primate gene expression patterns. Science *296*, 340–343.

Eyre-Walker, A. (2002). Changing effective population size and the McDonald-Kreitman test. Genetics *162*, 2017–2024.

Eyre-Walker, A., and Keightley, P. D. (2009). Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Molecular Biology and Evolution *26*, 2097–2108.

Fay, J. C. (2011). Weighing the evidence for adaptation at the molecular level. Trends in Genetics: TIG *27*, 343–349.

Fay, J. C., McCullough, H. L., Sniegowski, P. D., and Eisen, M. B. (2004). Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. Genome Biology *5*, R26.

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

Fay, J. C., Wyckoff, G. J., and Wu, C.-I. (2002). Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. Nature *415*, 1024–1026.

Fay, J., and Wu, C. (2000). Hitchhiking under positive darwinian selection. Genetics *155*, 1405–1413.

Fiston-Lavier, A.-S., Singh, N. D., Lipatov, M., and Petrov, D. A. (2010). *Drosophila melanogaster* recombination rate calculator. Gene *463*, 18–20.

Frankel, N., Erezyilmaz, D. F., McGregor, A. P., Wang, S., Payre, F., and Stern, D. L. (2011). Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. Nature *474*, 598–603.

Fraser, H. B., Moses, A. M., and Schadt, E. E. (2010). Evidence for widespread adaptive evolution of gene expression in budding yeast. Proceedings of the National Academy of Sciences of the United States of America *107*, 2977–2982.

Frazer, K. A., Sheehan, J. B., Stokowski, R. P., Chen, X. Y., Hosseini. R., Cheng, J.-F., Fodor, S. P. A., Cox, D. R., and Patil, N. (2001). Evolutionarily conserved sequences on human chromosome 21. Genome Research *11*, 1651–1659.

Genissel, A., McIntyre, L. M., Wayne, M. L., and Nuzhdin, S. V. (2008). *Cis* and *trans* regulatory effects contribute to natural variation in transcriptome of *Drosophila melanogaster*. Molecular Biology and Evolution *25*, 101–110.

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

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

Glinka, S., De Lorenzo, D., and Stephan, W. (2006). Evidence of gene conversion associated with a selective sweep in *Drosophila melanogaster*. Molecular Biology and Evolution *23*, 1869–1878.

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A, and Carroll, S. B. (2005). Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. Nature *433*, 481–487.

Gordon, K. L., and Ruvinsky, I. (2012). Tempo and mode in evolution of transcriptional regulation. PLoS Genetics *8*, e1002432.

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

Grath, S., and Parsch, J. (2012). Rate of amino acid substitution is influenced by the degree and conservation of male-biased transcription over 50 million years of *Drosophila* evolution. Genome Biology and Evolution, under revision.

Griffith, F. (1928). The significance of pneumococcal types. Journal of Hygiene *27*, 113–159.

Haddrill, P. R., Bachtrog, D., and Andolfatto, P. (2008). Positive and negative selection on noncoding DNA in *Drosophila simulans*. Molecular Biology and Evolution *25*, 1825–1834.

Haddrill, P. R., Waldron, F. M., and Charlesworth, B. (2008). Elevated levels of expression associated with regions of the *Drosophila* genome that lack crossing over. Biology Letters *4*, 758–761.

Haerty, W., Jagadeeshan, S., Kulathinal, R. J., Wong, A., Ravi Ram, K., Sirot, L. K., Levesque, L., Artieri, C. G., Wolfner, M. F., Civetta, A., *et al.* (2007). Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. Genetics *177*, 1321–1335.

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

Hamblin, M. T., and Di Rienzo, A. (2000). Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. American Journal of Human Genetics *66*, 1669–1679.

He, B. Z., Holloway, A. K., Maerkl, S. J., and Kreitman, M. (2011). Does positive selection drive transcription factor binding site turnover? A test with *Drosophila cis*-regulatory modules. PLoS genetics *7*, e1002053.

Hershey, A., and Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. Journal of General Physiology *36*, 39–56.

Hoekstra, H. E., and Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. Evolution *61*, 995–1016.

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

Hutter, S., Li, H., Beisswanger, S., De Lorenzo, D., and Stephan, W. (2007). Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. Genetics *177*, 469–480.

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

Innocenti, P., Morrow, E. H., and Dowling, D. K. (2011). Experimental evidence supports a sex-specific selective sieve in mitochondrial genome evolution. Science *332*, 845–848.

Irizarry, R. A., Warren, D., Spencer, F., Kim, I. F., Biswal, S., Frank, B. C., Gabrielson, E., Garcia, J. G. N., Geoghegan, J., Germino, G., *et al.* (2005). Multiple-laboratory comparison of microarray platforms. Nature Methods *2*, 345–349.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology *3*, 318–356.

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., and Carroll, S. B. (2008). The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. Cell *132*, 783–793.

Jiang, P.-P., Hartl, D. L., and Lemos, B. (2010). Y not a dead end: epistatic interactions between Y-linked regulatory polymorphisms and genetic background affect global gene expression in *Drosophila melanogaster*. Genetics *186*, 109–118.

Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., and Gibson, G. (2001). The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. Nature Genetics *29*, 389–395.

Kauer, M., Zangerl, B., Dieringer, D., and Schlötterer, C. (2002). Chromosomal patterns of microsatellite variability contrast sharply in African and non-African populations of *Drosophila melanogaster*. Genetics *160*, 247–256.

Keightley, P. D., and Gaffney, D. J. (2003). Functional constraints and frequency of deleterious mutations in noncoding DNA of rodents. Proceedings of the National Academy of Sciences of the United States of America *100*, 13402–13406.

Kimura, M. (1968). Evolutionary rate at the molecular level. Nature *217*, 624–626.

Kimura, M. (1983). The neutral theory of molecular evolution. Cambridge University Press. Cambridge, UK.

King, M.-C., and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107–116.

Kohn, M. H., Fang, S., and Wu, C.-I. (2004). Inference of positive and negative selection on the 5' regulatory regions of *Drosophila* genes. Molecular Biology and Evolution *21*, 374–383.

Lachaise, D., Cariou, M.-L., David, J. R., Lemeunier, F., Tsacas, L., and Ashburner, M. (1988). Historical biogeography of the *Drosophila melanogaster* species subgroup. Evolutionary Biology. *22*, 159–255.

Larkin, J. E., Frank, B. C., Gavras, H., Sultana, R., and Quackenbush, J. (2005). Independence and reproducibility across microarray platforms. Nature Methods *2*, 337–343.

Lemos, B., Araripe, L. O., Fontanillas, P., and Hartl, D. L. (2008). Dominance and the evolutionary accumulation of *cis*- and *trans*-effects on gene expression. Proceedings of the National Academy of Sciences of the United States of America *105*, 14471–14476.

Lemos, B., Araripe, L. O., and Hartl, D. L. (2008). Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. Science *319*, 91–93.

Lemos, B., Branco, A. T., and Hartl, D. L. (2010). Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. Proceedings of the National Academy of Sciences of the United States of America *107*, 15826–15831.

Lettice, L. A. (2003). A long-range *Shh* enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Human Molecular Genetics *12*, 1725–1735.

Li, W.-H. (1997). Molecular evolution. Sinauer Associates. Sunderland, Massachusetts.

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

Liao, B.-Y., Scott, N. M., and Zhang, J. (2006). Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. Molecular Biology and Evolution *23*, 2072–2080.

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

Ludwig, M. Z., Bergman, C., Patel, N. H., and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. Nature *403*, 564–567.

Lunzer, M., Golding, G. B., and Dean, A. M. (2010). Pervasive cryptic epistasis in molecular evolution. PLoS genetics *6*, e1001162.

Magoulas, C., Loverre-Chyurlia, A., Abukashawa, S., Bally-Cuif, L., and Hickey, D.A. (1993). Functional conservation of a glucoserepressible amylase gene promoter from *Drosophila virilis* in *Drosophila melanogaster*. Journal of Molecular Biology *36*, 234–242.

Martin, C. H., Mayeda, C. A., and Meyerowitz, E. M. (1988). Evolution and expression of the *Sgs-3* glue gene of *Drosophila*. Journal of Molecular Biology *201*, 273–287.

Malmberg, R. L., and Mauricio, R. (2005). QTL-based evidence for the role of epistasis in evolution. Genetical Research *86*, 89–95.

Mank, J. E., Hultin-Rosenberg, L., Zwahlen, M., and Ellegren, H. (2008). Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. The American Naturalist *171*, 35–43.

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

Masly, J. P., and Presgraves, D. C. (2007). High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. PLoS Biology *5*, e243.

Maynard Smith, J., and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genetical Research *23*, 23-35.

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

McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., and Wittkopp, P. J. (2010). Regulatory divergence in *Drosophila* revealed by mRNA-seq. Genome Research *20*, 816–825.

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

Meisel, R. P. (2011). Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein-coding sequence evolution. Molecular Biology and Evolution *28*, 1893–1900.

Mendel, G. (1865). Versuche über Pflanzen-Hybriden. Verhandlungen des Naturforschenden Vereins zu Brünn *4*, 3–47.

Monks, S. A., Leonardson, A., Zhu, H., Cundiff, P., Pietrusiak, P., Edwards, S., Phillips, J. W., Sachs, A., and Schadt, E. E. (2004). Genetic inheritance of gene expression in human cell lines. American Journal of Human Genetics *75*, 1094–1105.

Müller, L., Hutter, S., Stamboliyska, R., Saminadin-Peter, S. S., Stephan, W., and Parsch, J. (2011). Population transcriptomics of *Drosophila melanogaster* females. BMC Genomics *12*, 81.

Nei, M. (1987). Molecular evolutionary genetics. Columbia University Press. New York.

Nielsen, R. (2001). Statistical tests of selective neutrality in the age of genomics. Heredity *86*, 641–647.

Nirenberg, M. (2004). Historical review: Deciphering the genetic code – a personal account. Trends in Biochemical Sciences *29*, 46–54.

Nègre, N., Brown, C. D., Ma, L., Bristow, C. A., Miller, S. W., Wagner, U., Kheradpour, P., Eaton, M. L., Loriaux, P., Sealfon, R., *et al.* (2011). A *cis*-regulatory map of the *Drosophila* genome. Nature *471*, 527–531.

Ohya, Y., Sese, J., Yukawa, M., Sano, F., Nakatani, Y., Saito, T. L., Saka, A., Fukuda, T., Ishihara, S., Oka, S., *et al.* (2005). High-dimensional and large-scale phenotyping of yeast mutants.

Proceedings of the National Academy of Sciences of the United States of America *102*, 19015–19020.

Oleksiak, M. F., Churchill, G. A, and Crawford, D. L. (2002). Variation in gene expression within and among natural populations. Nature Genetics *32*, 261–266.

Ometto, L., Glinka, S., De Lorenzo, D., and Stephan, W. (2005). Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. Molecular Biology and Evolution *22*, 2119–2130.

Orengo, D. J., and Aguadé, M. (2004). Detecting the footprint of positive selection in a European population of *Drosophila melanogaster*: multilocus pattern of variation and distance to coding regions. Genetics *167*, 1759–1766.

Pandey, R. V., Kofler, R., Orozco ter Wengel, P., Nolte, V., and Schlötterer, C. (2011). PoPoolation DB: a user-friendly web-based database for the retrieval of natural polymorphisms in *Drosophila*. BMC Genetics *12*, 27.

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

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

Petronzelli, F., Kimura, A., Ferrante, P., and Mazzilli, M. C. (1995). Polymorphism in the upstream regulatory region of DQA1 gene in the Italian population. Tissue Antigens *45*, 258–263.

Plotkin, J. B., and Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. Nature Reviews Genetics *12*, 32–42.

Prabhakar, S., Visel, A., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., Plajzer-Frick, I., Morrison, H., Fitzpatrick, D. R., Afzal, V., *et al.* (2008). Human-specific gain of function in a developmental enhancer. Science *321*, 1346–1350.

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V. A, Williams, T. M., Yeh, S.-D., True, J. R., and Carroll, S. B. (2006). Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene. Nature *440*, 1050–1053.

Prud'homme, B., Gompel, N., and Carroll, S. B. (2007). Emerging principles of regulatory evolution. Proceedings of the National Academy of Sciences of the United States of America *104*, 8605–8612.

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

R Development Core Team (2005). R: A language and environment for statistical computing, reference index version 2.2.1. Vienna, Austria: R Foundation for Statistical Computing.

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

Rebeiz, M., Jikomes, N., Kassner, V. A, and Carroll, S. B. (2011). Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. Proceedings of the National Academy of Sciences of the United States of America *108*, 10036–10043.

Rebeiz, M., Pool, J. E., Kassner, V. A, Aquadro, C. F., and Carroll, S. B. (2009). Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. Science *326*, 1663–1667.

Rifkin, S. A, Kim, J., and White, K. P. (2003). Evolution of gene expression in the *Drosophila melanogaster* subgroup. Nature Genetics *33*, 138–144.

Sackton, T. B., Montenegro, H., Hartl, D. L., and Lemos, B. (2011). Interspecific Y chromosome introgressions disrupt testis-specific gene expression and male reproductive phenotypes in *Drosophila*. Proceedings of the National Academy of Sciences *108*, 17046–17051.

Saminadin-Peter, S. S. (2008). Evolution of gene expression and gene-regulatory sequences in *Drosophila melanogaster*. Dissertation, LMU München

Saminadin-Peter, S. S., Kemkemer, C., Pavlidis, P., and Parsch, J. (2011). Selective sweep of a *cis*-regulatory sequence in a non-African population of *Drosophila melanogaster*. Molecular Biology and Evolution, 1–8.

Sawyer, S. A, Parsch, J., Zhang, Z., and Hartl, D. L. (2007). Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America *104*, 6504–6510.

Sawyer, S. A, and Hartl, D. L. (1992). Population genetics of polymorphism and divergence. Genetics *132*, 1161–1176.

Schadt, E. E., Monks, S. A., Drake, T. A., Lusisk, A. J., Chek, N., Colinayok, V., Ruff, T. G., Milligan, S. B., Lamb, J. R., Cavet, G., *et al*. (2003). Genetics of gene expression surveyed in maize, mouse and man. Nature *422*, 297–302.

Schwartz, D. (1962). Genetic studies on mutant enzymes in maize. III. Control of gene action in the synthesis of ph 7.5 esterase. Genetics *47*, 1609–1615.

Sgrò, C. M., Overgaard, J., Kristensen, T. N., Mitchell, K. A, Cockerell, F. E., and Hoffmann, A. A. (2010). A comprehensive assessment of geographic variation in heat tolerance and hardening capacity in populations of *Drosophila melanogaster* from eastern Australia. Journal of Evolutionary Biology *23*, 2484–2493.

Shabalina, S. A., Ogurtsov, A. Y., Kondrashov, V. A., and Kondrashov. A. S. (2001). Selective constraint in intergenic regions of human and mouse genomes. Trends in Genetics *17*, 373–376.

Shapiro, J. A, Huang, W., Zhang, C., Hubisz, M. J., Lu, J., Turissini, D. A, Fang, S., Wang, H.-Y., Hudson, R. R., Nielsen, R., *et al*. (2007). Adaptive genic evolution in the *Drosophila* genomes. Proceedings of the National Academy of Sciences of the United States of America *104*, 2271–2276.

Shirangi, T. R., Dufour, H. D., Williams, T. M., and Carroll, S. B. (2009). Rapid evolution of sex pheromone-producing enzyme expression in *Drosophila*. PLoS Biology *7*, e1000168.

Singh, N. D., Arndt, P. F., and Petrov, D. A. (2005). Genomic heterogeneity of background substitutional patterns in *Drosophila melanogaster*. Genetics *169*, 709–722.

Smith, N. G. C., and Eyre-Walker, A. (2002). Adaptive protein evolution in *Drosophila*. Nature *415*, 1022–1024.

Spielman, R. S., Bastone, L. A., Burdick, J. T., Morley, M., Ewens, W. J., and Cheung, V. G. (2007). Common genetic variants account for differences in gene expression among ethnic groups. Nature Genetics *39*, 226–231.

Stern, D. L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? Evolution *62*, 2155–2177.

Storey, J. D., Madeoy, J., Strout, J. L., Wurfel, M., Ronald, J., and Akey, J. M. (2007). Gene-expression variation within and among human populations. American Journal of Human Genetics *80*, 502–509.

Stranger, B. E., Forrest, M. S., Clark, A. G., Minichiello, M. J., Deutsch, S., Lyle, R., Hunt, S., Kahl, B., Antonarakis, S. E., Tavaré, S., *et al.* (2005). Genome-wide associations of gene expression variation in humans. PLoS Genetics *1*, e78.

Svetec, N., Pavlidis, P., and Stephan, W. (2009). Recent strong positive selection on *Drosophila melanogaster HDAC6*, a gene encoding a stress surveillance factor, as revealed by population genomic analysis. Molecular Biology and Evolution *26*, 1549–1556.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. Genetics *105*, 437–460.

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

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution *28*, 2731–2739.

Thornton, K., and Andolfatto, P. (2006). Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics *172*, 1607–1619.

Tirosh, I., Barkai, N., and Verstrepen, K. J. (2009). Promoter architecture and the evolvability of gene expression. Journal of Biology *8*, 95.

Tirosh, I., Reikhav, S., Levy, A. A, and Barkai, N. (2009). A yeast hybrid provides insight into the evolution of gene expression regulation. Science *324*, 659–662.

Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E., *et al.* (2002). Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. Genome Biology *3*, research0088.1–0088.14.

Tortiglione, C., and Bownes, M. (1997). Conservation and divergence in the control of yolk protein genes in dipteran insects. Development Genes and Evolution *207*, 264–281.

Tournamille, C., Colin, Y., Cartron, J. P., and Le Van Kim, C. (1995). Disruption of a GATA motif in the *Duffy* gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nature Genetics *10*, 224–228.

Townsend, J. P., Cavalieri, D., and Hartl, D. L. (2003). Population genetic variation in genome-wide gene expression. Molecular Biology and Evolution *20*, 955–963.

True, J. R., Weir, B. S., and Laurie, C. C. (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.

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

Watson, J. D., and Crick, F. H. C. (1953). A structure for deoxyribose nucleic acid. Nature *171*, 737–738.

Wei, Z., Angerer, L. M., Gagnon, M. L., and Angerer, R. C. (1995). Characterization of the SpHE promoter that is spatially regulated along the animal-vegetal axis of the sea urchin embryo. Developmental Biology *171*, 195–211.

Whitehead, A., and Crawford, D. L. (2006). Neutral and adaptive variation in gene expression. Proceedings of the National Academy of Sciences of the United States of America *103*, 5425–5430.

Wilson, D. J., Hernandez, R. D., Andolfatto, P., and Przeworski, M. (2011). A population genetics-phylogenetics approach to inferring natural selection in coding sequences. PLoS Genetics *7*, e1002395.

Wittkopp, P. J., Haerum, B. K., and Clark, A. G. (2004). Evolutionary changes in *cis* and *trans* gene regulation. Nature *430*, 85–88.

Wittkopp, P. J., Stewart, E. E., Arnold, L. L., Neidert, A. H., Haerum, B. K., Thompson, E. M., Akhras, S., Smith-Winberry, G., and Shefner, L. (2009). Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. Science *326*, 540–544.

Wittkopp, P. J., and Kalay, G. (2011). *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics, 59–70.

Wolff, C., Pepling. M., Gergen, P., and Klingler, M. (1999). Structure and evolution of a pair-rule interaction element: runt regulatory sequences in *D. melanogaster* and *D. virilis*. Mechanisms of Development *80*, 87–99.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. A. (2003). The evolution of transcriptional regulation in eukaryotes. Molecular Biology and Evolution *20*, 1377–1419.

Wray, G. A (2007). The evolutionary significance of *cis*-regulatory mutations. Nature Reviews Genetics *8*, 206–216.

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

Yang, X., Schadt, E. E., Wang, S., Wang, H., Arnold, A. P., Ingram-Drake, L., Drake, T. A., and Lusis, A. J. (2006). Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Research *16*, 995–1004.

Zhang, L., and Li, W.-H. (2004). Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Molecular Biology and Evolution *21*, 236–239.

Zhang, X., and Borevitz, J. O. (2009). Global analysis of allele-specific expression in *Arabidopsis thaliana*. Genetics *182*, 943–954.

Zhang, Z., Hambuch, T. M., and Parsch, J. (2004). Molecular evolution of sex-biased genes in *Drosophila*. Molecular Biology and Evolution *21*, 2130–2139.

ACKNOWLEDGEMENTS

This thesis would not have been the same without the help of many people, whom I would like to express my gratitude.

First and foremost, my supervisor John Parsch who contributed considerably to making me a better scientist. I especially want to thank him for his accuracy, for sharing his vast knowledge, for improving my writing and presentation skills.

Desweiteren möchte ich Wolfgang Stephan danken, der mich mit der Evolutionsbiologie und der Populationsgenetik in Kontakt gebracht hat.

Vielen Dank auch an Dirk Metzler, welcher zur Lösung so mancher Probleme mit Statistik und R beigetragen hat.

I want to thank everybody of the Evolutionary Biology Department who contributed to the pleasant working atmosphere, nice lunch breaks and fruitful discussions!

Meiner langjährigen Mit-Büroinsassin Miri Linnenbrink danke ich für all die guten wissenschaftlichen Diskussionen, lustigen Unterhaltungen und hilfreichen Tipps, die sie mir über die Jahre hat angedeihen lassen. Sonja Grath danke ich für ihre permanente Bereitschaft, Fragen aller Art zu beantworten, für die gute Zusammenarbeit an unserem paper, für ihre ermutigende Art. Unvergessen bleibt auch, daß ich Miri und Sonja noch im Halbschlaf mit meinem Vortrag belästigen durfte! Lisha Naduvilezhath danke ich für ihre stete Hilfsbereitschaft, besonders ihre unendliche Hilfe bei R-Skripten, die guten Gespräche und die angenehme Zeit auf der Fraueninsel. Claus Kemkemer danke ich für seine hilfreiche und positive Art und seinen täglichen Lunch-Appell, Ana Catalán für die lustigen Unterhaltungen und die gute Atmosphäre im Büro. Thanks also to Amanda Glaser for contributing to the nice and pleasant atmosphere in the office. Dankeschön an Winfried Hense für das Ausleihen diverser Lehrbücher, das Teilen seines Wissens sowie seiner Schokoladenvorräte. Vielen Dank an Iris Fischer für die populationsgenetischen Diskussionen und daß sie mir so schön vorgemacht hat, wie man erfolgreich eine Dissertation einreicht. Danke auch an Anja Hörger für die netten Unterhaltungen, Danke an Sebastian Nagelmüller und Korbinian von Heckel für die gute Zusammenarbeit und Hilfe beim Sequenzieren. Martin Hutzenthaler danke ich für seine Hilfe mit R. Thanks to all the fly people for nice chats and discussions inside and outside the fly room. Stephan Hutter für die hochkompetente Beantwortung vielerlei Fragen. Ein riesiger Dank geht an Hedwig Gebhart, Hilde Lainer, Gisela Brinkmann, Ana Vrljic, Anne Steincke und Simone Lange für ihre Hilfe im Labor, das Teilen ihrer Reagenzien und Labortricks, das Fliegenfutter, die schönen Sequenzen und natürlich die guten Gespräche. Desweiteren danke ich Ingrid Kroiß, die auf alle Formalitätenfragen stets eine Antwort wußte.

Nicht zuletzt gebührt meiner Familie und meinen Freunden Dank für ihre stete Unterstützung und Ermutigung.