Local adaptation of gene regulation in natural populations of *Drosophila melanogaster*

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

vorgelegt von
Ana Gabriela Catalán Ramírez
aus Guatemala, Guatemala

München, 16 Dezember 2013

1. Gutachter: Prof. Dr. John Parsch

2. Gutachter: Prof. Dr. Nicolas Gompel

Dissertation eingereicht am: 16. Dezember 2013

Datum der Disputation: 2. July 2014

Erklärung:

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

Eidesstattliche Erklärung:

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

München, 16.12.2013

_____

Ana Gabriela Catalán Ramírez

Declaration of author's contribution

In this thesis, I present the results of my doctoral studies conducted from October 2010 until December 2013. The results are shown in four chapters, all of which are collaborations with other scientists. The work of this doctoral thesis has resulted in three publications. They constitute chapters 1, 3 and 4 of this dissertation. Chapter 2 is an unpublished manuscript.

In CHAPTER 1, I and John Parsch conceived the study and its design. I performed the brain dissections, RNA extractions and qRT-PCR. Stephan Hutter, John Parsch and I analyzed the RNA-seq data. I and Stephan Hutter performed the population genetic analyses. I and John Parsch wrote the manuscript with input of Stephan Hutter. John Parsch, Stephan Hutter and I read and approved the final manuscript, which was published in:

CATALÁN A., HUTTER S., PARSCH J., 2012 Population and sex differences in *Drosophila melanogaster* brain gene expression. BMC Genomics **13**: 654.

In CHAPTER 2, I and John Parsch conceived the study and its design. I performed the brain dissections, RNA extractions, qRT-PCR, genotype-gene expression study, copy number variation assays, correlation analysis, molecular cloning, brain confocal imaging, and microRNA-3' UTR prediction. I with the support of Angélica Cuevas performed the sequencing. John Parsch analyzed the RNA-seq data. John Parsch, Pablo Duchén and I did the population genetics analyses. I and John Parsch wrote the manuscript.

In CHAPTER 3, I maintained fly stocks, performed PCR, sequencing, editing and alignment of the sequences. I performed population genetic analyses (calculations of pi, theta, Fst, etc.), constructed the neighbour-joining tree and did the bootstrap analysis. I wrote the parts of the methods and results sections corresponding to the above. I also wrote the section of discussion about possible functions of CG9509. I prepared figure 3 and tables 1 and S1 and revised the manuscript.

GLASER-SCHMITT A*., **CATALÁN A***., PARSCH J., 2013 Adaptive divergence of a transcriptional enhancer between populations of *Drosophila melanogaster*. Phil. Trans. R. Soc. B **368**: 20130024. * Equal contributions.

In CHAPTER 4, I maintained fly stocks and performed crosses. I designed and optimized *in-situ* hybridization on fly testis from six transformant flies and one control line. I performed the imaging of the mounted testis and prepared figure 2 of the manuscript. I wrote the methodology part of the testis *in-situ* hybridization assays and revised the manuscript.

KEMKEMER C., **CATALÁN A.**, PARSCH J., 2013 "Escaping" the X chromosome leads to increased gene expression in the male germline of *Drosophila melanogaster*. Heredity (in press), doi: 10.1038/hdy.2013.86.

Table of Contents

List of Figures

List of Tables

"Quetzal, imán del sol, Tecún, imán
del tún, Quezaltecún, sol y tún, tún-bo
del lago, tún-bo del monte, tún-bo del
verde, tún-bo del cielo, tún, tún, tún, tun-
bo del verde corazón del tún, palpitación de
la primavera, en la primera primavera
tún-bo de flores que bañó la tierra viva."

Miguel Ángel Asturias (1899-1974)

A mi abuelita,
María Elba Catalán

Abstract

THE central goal of this dissertation is to understand the genetic and functional aspects of how populations adapt to new or changing environments. Genetic variation within a population, either at protein coding genes or at regulatory elements, provides the substrate upon which natural selection can act to drive adaptation. There is considerable evidence that changes in gene expression account for a large proportion of morphological, physiological and behavioral variation between and within species that can contribute to adaptation and speciation. Due to the major role that gene expression changes can have in shaping phenotypes, the first three chapters of this dissertation deal with the study of how changes in gene expression can facilitate adaptation. I use *Drosophila melanogaster* from ancestral and derived regions of the species' range as a model system for studying local adaptation.

In chapter 1, I perform high-throughput RNA-sequencing (RNA-seq) of brain tissue of flies from an ancestral (Zimbabwe) and a derived (the Netherlands) population. The whole brain transcriptome was assayed for differences in gene expression between African and European flies in order to understand how differences in brain expression may lead to local adaptation. I found over 300 candidate genes that differed significantly in expression between the populations, including a cluster of genes on chromosome arm 3R that showed reduced expression in Europe and genetic evidence for positive selection. Other candidate genes involved in stress response, olfaction and detoxification were also identified. Additionally, I compared brain gene expression between males and females and found an enrichment of sex-biased genes on the X chromosome.

Chapter 2 presents a detailed study of one of the candidate genes identified in chapter 1. The metallothionein gene, *MtnA*, shows over four-fold higher expression in the brain of European flies than of African flies. I found a derived deletion in the 3' untranslated region (UTR) of *MtnA* that segregates at high frequency within the Dutch population, but is absent from the Zimbabwean population. The presence of the deletion was perfectly associated with the observed variation in *MtnA* expression. When additional populations of *D. melanogaster* were screened for the presence of the deletion, I found that it showed a clinal distribution that was significantly correlated with latitude and temperature. Furthermore, using population genetic data and a selective sweep analysis I show that the *MtnA* locus is evolving under positive selection.

In Chapter 3 I report a population genetic analysis of the enhancer region of *CG9505*, a gene that shows significantly higher expression in European than in African populations of *D. melanogaster*. A previous study found that there was very low nucleotide polymorphism in the enhancer region of *CG9509* in flies from the Netherlands, a pattern that is consistent with a selective sweep. I analyzed an additional set of five populations from Zambia, Egypt, Malaysia, France and Germany in order to gain a better understanding of how selection has affected the evolution of this enhancer. I found that there is a depletion of nucleotide diversity in all of the non-sub-Saharan populations (Egypt, Malaysia, France and Germany), which share a common high-frequency derived haplotype. Population genetic analyses suggest that a selective sweep took place in the enhancer region of *CG9509* just after *D. melanogaster* migrated out of sub-Saharan Africa.

Finally, in chapter 4 I performed *in situ* hybridizations to examine the expression of tissue-specific reporter genes in the *D. melanogaster* testis. In the male germline of *D. melanogaster*, reporter genes that reside on the X chromosome show a reduction in expression relative to those located on the autosomes. This phenomenon was demonstrated by randomly inserting reporter gene constructs on the X chromosome and the autosomes. By doing *in situ* hybridizations on testis of flies having reporter gene insertions on the X chromosome and autosomes, I could show that the expression difference mainly occurs in meiotic and post-meiotic cells. For most

constructs, expression was very low or absent in the testis apex, which is enriched with pre-meiotic cells. These results suggest that the suppression of X-linked gene expression in the *Drosophila* male germline occurs through a different mechanism than the MSCI (meiotic sex chromosome inactivation) known to occur in mammals.

General introduction

"Natural selection is a mechanism for generating an exceedingly high degree of improbability"

Ronald Fisher (1890 - 1962)

CONSIDER a species that is undergoing population expansion. As its population grows, individuals of this species will start colonizing new habitats. The newly encountered habitat might differ from the habitat from which the original or ancestral population comes from. From an abiotic perspective, this new habitat might be different in factors like temperature, humidity, salinity or exposure to daylight. On the other hand, the newly encountered environment might also differ in various biotic factors, including exposure to novel parasites, predators and competitors for natural resources. The individuals of this species have to undergo adaptation in order to establish a new population in the new habitat successfully. Not only do species have to adapt when they colonize new habitats, but also to the continuous changes to their ecosystem that occur over time.

Through changes in morphological, physiological or behavioral traits, a species' fitness can go towards its optimum in a particular habitat or at a certain time point. Understanding the process of adaptation is one of the major goals of the study of evolutionary biology. In this general introduction I want to give an overview of the efforts that have been undertaken to understand how species adapt to novel habitats, the theoretical tools that we have to detect adaptation through natural selection, and the advantage of using *Drosophila melanogaster* as a model organism for the study of adaptation.

Natural selection acts on phenotypic variation present in a population. Errors made during DNA replication, as DNA polymerase synthesizes a new DNA strand, give rise to new point mutations in a population. Indels, inversions, transposable elements, translocations and gene conversion also account for genetic variability in genomes. These genetic variations might have an effect on a particular phenotype and thus be the target of natural selection. Changes in both coding genes and regulatory sequences can serve as the basis for adaptive evolution. One example of adaptation at the protein level is the amino acid variation segregating at the *Agouti* locus in different populations of the deer mouse, *Peromyscus maniculatus* (Vignieri et al. 2010, Linnen et al. 2013). Variation in the *Agouti* locus has been linked with adaptive melanism in deer mouse, conferring a better fur camouflage within a specific habitat. Another classical example of a fitness advantage caused by variation in a protein-coding gene is the heterozygous advantage (malaria resistance) due to an amino acid replacement in the beta chain of hemoglobin in human populations where malaria has high prevalence (Anastasi 1984).

It has been suggested that the variation observed at the protein level cannot explain all of the phenotypic variation that is observed between and within populations, and that changes in gene expression might account for a considerable proportion of phenotypic variation (Caroll 2005). Changes in gene expression could occur through changes in *trans*-regulatory elements, like transcription factors, or due to changes in *cis*-regulatory elements, like promoters or enhancers. When a mutation occurs in a coding region, it has the potential of having a large effect on the phenotype and fitness, and thus it is predicted that mutations in coding genes mostly have deleterious effects. This assumption might be especially true for highly pleiotropic proteins that are ubiquitously expressed and that are part of complex gene networks. Gene

duplications can also provide genetic variation for natural selection to act upon. Susumu Ohno (1970) proposed that duplication events are a form of genetic variation in which changes at the protein level could be facilitated. While the original copy of the duplicated gene can maintain its original function, the new gene copy can accumulate variation due to relaxation of purifying selection, which could eventually lead to neofunctionalization of the paralog (Zhang 2003). One example of neofunctionalization is the duplication event that occurred in the eosinophil-derived neurotoxin gene (EDN), which is part of the RNase A gene superfamily (Zhang 2003). It has been proposed that this duplication event happened some 30 million years ago in the hominoid and Old World monkey branch. The tandem duplication of EDN allowed its paralog to accumulate substitutions that led to the paralog becoming a new gene with antibacterial activity (Zhang et al. 1998).

Variation occurring in *cis*-regulatory elements can circumvent the possible harmful effects that might arise from amino acid changes in highly pleiotropic genes, since changes in *cis*-regulatory elements can alter gene expression patterns specifically in time and space (Carroll 2008). Numerous studies have identified variation in *cis*-regulatory elements with functionally significant effects on morphological, physiological and behavioral traits. For example, evidence of adaptation through changes in gene expression patterns has been found at the *Cyp6g1* locus in *D. melanogaster*. Insertion of an *Accord* transposable element in the 5' upstream region of *Cyp6g1* leads to a higher *Cyp6g1* gene expression level, which also translates into higher resistance to pesticides (Schmidt et al. 2010). Similarly, an indel polymorphism in the promoter region of *Dca* (*Drosophila* cold acclimation) causes differences in the expression of *Dca,* which has been associated with adaptive variation in wing size across a latitudinal cline (McKechnie et al. 2009).

Interrogating the whole genome for differences in gene expression between populations of the same species or between different species is one of the approaches used to detect regulatory adaptation. By using microarrays or RNA-sequencing to detect genes differentially expressed between an ancestral sub-Saharan population of *D. melanogaster* and a derived northern European population, hundreds of candidate genes for regulatory adaptation have been found (Hutter et al. 2008, Müller et al. 2011, Catalán et al. 2012).

By implementing population genetic tools, molecular biology methods and functional assays, some of the genes showing differences in gene expression between populations could be validated as being targets of natural selection. In the first, second and third chapter of this dissertation, evidence is presented that shows how differences in gene expression can be maintained in a species through the action of positive selection in one or more local populations. When a mutation arises in a population, it can either be lost or fixed solely by genetic drift. However, a beneficial mutation can also be fixed by positive selection if it confers a fitness advantage. The size of the population determines how quickly a mutation can be lost or fixed by genetic drift. The smaller the population size, the stronger the influence of genetic drift and the more likely it is that a mutation will be lost of fixed just by chance.

Another goal in population genetics is to discern between DNA patterns left by genetic drift or demographic events and those caused by natural selection. The neutral theory of molecular evolution developed by Motoo Kimura (1984) is widely used as a null model for the detection of natural selection. The neutral theory postulates that if a DNA sequence is evolving neutrality, the nucleotide variation occurring in it depends only on the mutation rate and the population size. Factors like natural selection (positive, negative, balancing selection or background selection), changes in population size, population structure or compensatory evolution can leave polymorphism patterns that differ from the neutral expectations (Hamilton 2009). Having a first overview of the nucleotide diversity levels at many loci in a population can already give information about the evolutionary forces acting in a population. Watterson's theta (θ_w) (Watterson 1975) and the nucleotide diversity estimate π (Nei and Li 1979) are two indices commonly used for the estimation of nucleotide diversity in a population. Watterson's theta uses the number of segregating sites present in a population to estimate nucleotide diversity and π is the average number of pairwise differences present in a sample of DNA sequences. Patterns of nucleotide diversity can give us information about demographic or selective forces acting at particular loci in the genome. For example, the test developed by Tajima, the *D* statistic (Tajima 1989), uses θ_w and π estimates obtained from DNA polymorphism data to test for deviations from the neutral model. Tajima's *D* statistic makes the assumption that, under neutrality, the two estimators θ_w and π should be equal and the *D* statistic should not deviate significantly from zero. A significantly negative value of Tajima's *D* corresponds to an excess of rare variants, as is expected under positive or negative

selection, or when a population has undergone expansion. On the other hand, a significantly positive value of Tajima's *D* indicates an excess of intermediate frequency variants, which can occur when a locus is under balancing selection or when the samples come from a structured, subdivided population.

Another test that is commonly used for the detection of natural selection at the protein level is the McDonald–Kreitman test (McDonald and Kreitman 1991). This test compares the ratio of nonsynonymous to synonymous polymorphism to the ratio of nonsynonymous to synonymous divergence. Since nucleotide changes at a specific locus within and between species depend on the mutation rate (assuming a constant population size and random mating), these two ratios should be equal under neutral expectations. If significant deviations are found in the ratio of fixed differences to polymorphic changes, natural selection might be acting on the protein sequence. For example, if a higher ratio of nonsynonymous polymorphism is found in comparison to nonsynonymous divergence, balancing selection could be maintaining a high level of nonsynonymous variation within a population. The McDonald-Kreitman test can circumvent the potential problems caused by a population's demographic history, since drift or a demographic event is expected to affect synonymous and nonsynonymous sites in an equal manner. One drawback of the McDonald-Kreitman test is that it can be applied only to coding regions where synonymous and nonsynonymous changes can be detected. However, the McDonald-Kreitman test can be modified to use other types of sites as the test and/or neutral classes (Andolfatto 2005).

Another way to look for adaptation through positive selection is searching for DNA polymorphism patterns that resemble a selective sweep. When an adaptive mutation arises in the genome, its fitness advantage causes it to spread and increase in frequency in the population. When this occurs, the linked neutral variants flanking the selective site also increase in frequency until they reach fixation (Smith and Haigh 1974). Thus, a depletion of DNA polymorphism is then observed in the genomic region surrounding the selected site. As one moves away from the positively selected site, recombination starts to break up the liked genomic regions and DNA polymorphism starts to increase (Kim and Stephan 2002, Kim and Stephan 2003). The signature of a selective sweep can be used to detect regions of the genome under positive selection (Nielsen et al. 2005, Pavlidis et. al 2013). The extent of the swept

region, and thus the power to detect a selective sweep, depends on the recombination rate and the strength of selection (Kim and Nielsen 2004). A demographic event like a bottleneck could also result in a similar DNA polymorphism pattern to that produced by a selective sweep (Jensen et al. 2005). One way to circumvent this problem is to test for the independency of linkage between the flanking sites of the positively selected site, as is done by the omega statistic (Alachiotis et al. 2012). This test is based on the expectation that there should be strong linkage disequilibrium within the flanking regions of the beneficial mutation, but not between the two flanking regions. By performing neutral coalescent simulations that take into account the demographic changes occurring through time in a population as well as the recombination rates of the loci in question, it is possible to establish a significance threshold to discriminate between DNA polymorphism patterns left by chance and those resulting from selection (Hudson 2002).

Using *Drosophila melanogaster* as a model organism to study evolutionary biology has many advantages. One of the most important advantages for the study of adaptation is the geographic distribution of *D. melanogaster*. From biogeographic, morphological and DNA polymorphism data it has been concluded that the origin of *D. melanogaster* lies in sub-Saharan Africa, most probably in Zimbabwe or Zambia (Lachaise et al. 1988, Begun and Aquadro 1993, Stephan and Li 2007, Pool et al. 2012). After a population expansion that occurred about 60,000 years ago, *D. melanogaster* started colonizing new habitats and now has a cosmopolitan geographical distribution (Stephan and Li 2007). It has been hypothesized that the colonization of new habitats by *D. melanogaster* has been aided by human migrations (Keller 2007). It has also been estimated that the out-of-Africa migration occurred between 13,000–19,000 years ago and that about 2,500 years ago the European and Asian populations of *D. melanogaster* split from a common ancestral population (Laurent et al. 2011, Duchen et al. 2012, Werzner et al. 2013). A more recent colonization has occurred in North America, where *D. melanogaster* successfully established itself about 250 years ago (Johnson 1913, Keller 2007, Duchen et al. 2012). Recent SNP data suggest that the *D. melanogaster* population residing in North America has 85% European ancestry and 15% African ancestry. The colonization of the Americas by *D. melanogaster* is also tightly linked with human migrations, as throughout the nineteenth century many Europeans migrated to North America bringing fruit flies with them. Flies coming along with the slave trade from Africa may explain the 15% of

African ancestry (Johnson 1913, Duchen et al. 2012). The fact that *D. melanogaster* has a cosmopolitan geographical distribution, where different populations encounter very different habitats, offers an ideal system to study local adaptation. With a well-annotated reference genome, the availability of many genetic and molecular tools and resources, its easy maintenance in the laboratory and short life cycle, *D. melanogaster* is a very convenient and powerful model organism for the study of evolution.

For most of the work presented in this dissertation I focused on understanding how populations colonizing novel habitats undergo adaptation. I mostly worked with two populations of *D. melanogaster*, one ancestral population from Zimbabwe and one derived population from the Netherlands. The underlying hypothesis is that the derived *D. melanogaster* population from the Netherlands has undergone adaptation to the new biotic and abiotic factors encountered in its new habitat. In CHAPTER 1, I aim to understand how changes in gene expression in the brain can be the basis for adaptation to new environments. Previous studies of gene expression variation have found hundreds of genes that are differentially expressed between females and males from an ancestral and a derived population of *D. melanogaster* (Hutter et al. 2008, Müller et al. 2011). Although these studies uncovered a lot of candidate genes for adaptation, most of the time it is challenging to find a direct link between gene expression variation, the evolutionary forces maintaining this variation, and how the gene expression variation affects a specific phenotype that potentially affects fitness. In the studies by Hutter et al. (2008) and Müller et al. (2011), gene expression variation was surveyed using whole bodies. By using whole bodies, the signal detected from the genes that are differentially expressed is likely to come from those genes that are expressed ubiquitously or those that are expressed in organs that occupy a large percentage of the fly's body. In order to acquire a more precise understanding of how populations adapt to a novel environment, I assessed gene expression differences in the brain of Zimbabwean and Dutch flies. In animals with a central nervous system, the brain is the organ receiving all of the information collected from the environment by the sensorial organs (in insects: antennae, proboscis, setae, eyes, ovipositor). All of this information is processed by the different brain centers and a behavioral or physiological response is generated as the output from a particular stimulus. The stimuli received from two different types of environments can result in different types of outputs. One of the goals in chapter 1 is to find genes that are differentially expressed between African and

European flies that could be under positive selection and involved in behaviors or physiological responses controlled by the brain that might be adaptive in their local environment.

Over three hundred genes were identified as candidates for adaptation in the survey performed by RNA-seq of *D. melanogaster* brains from an African and a European population (Catalán et al. 2012). In CHAPTER 2, I present evidence for adaptation in one of these genes, *Metallotheionein A* (*MtnA*), in northern European populations of *D. melanogaster*. In European flies (the Netherlands), *MtnA* shows four-fold higher expression in comparison with flies from Zimbabwe. An indel polymorphism in the 3' untranslated region (UTR) of *MtnA* was identified as the most probable cause of the difference in expression observed between Europe and Africa, since all the flies having an ancestral 49-bp fragment had significantly lower *MtnA* expression than the flies having the derived deletion in the 3' UTR. A survey of additional populations revealed that the deletion showed a clinal distribution and was in high frequency in northern European populations. Population genetic analysis also indicated that the *MtnA* locus was a target of positive selection in European populations.

Another gene that shows a robust difference in expression between African and European populations of *D. melanogaster* is *CG9509*. The European (the Netherlands) population shows two- to three-fold higher expression of *CG9509* when compared with African (Zimbabwe) flies (Saminadin-Peter et al. 2012). The enhancer region of *CG9509* has several derived fixed or nearly-fixed differences between the African and the European population and in the European population this region shows a depletion of nucleotide diversity, which is consistent with a selective sweep. Furthermore, the European version of the *CG9509* enhancer region is capable of driving three-fold higher expression of a reporter gene than the African version (Saminadin-Peter et al. 2012). In CHAPTER 3, I perform an extensive population genetics analysis of four other *D. melanogaster* populations coming from the ancestral species range (Zambia) and the derived species range (Egypt, Malaysia and Germany). From this analysis, I present evidence of positive selection in non-sub-Saharan populations (Egypt, Malaysia and Germany), which also show a strong reduction of nucleotide diversity and the same high-frequency derived haplotype present in the Netherlands. These results suggest that a selective

sweep in the enhancer region of *CG9509* took place just after the out-of-Africa migration of the species.

The last chapter of this dissertation, CHAPTER 4, deals with the observation that the expression of X-linked genes is suppressed in the male germline of *D. melanogaster*. A previous study by Hense et al. (2007) showed that when a reporter gene driven by a testis-specific promoter is inserted on the X chromosome, it shows significantly lower expression than when it is inserted on an autosome. To test the generality of this result, a reporter gene driven by three different promoters from X-linked, testis-specific genes was used. The constructs were randomly inserted on the X chromosome and the autosomes, with the result being that the reporter gene expression of the constructs inserted on the X chromosome had significantly lower expression levels than the constructs inserted on the autosomes. I performed *in situ* hybridizations on the testes of transformed flies that targeted the reporter gene. The results confirmed that the reporter genes inserted on the X chromosome have lower expression than those inserted on the autosomes. Additionally, the *in situ* hybridizations could localize reporter gene expression to the meiotic and post-meiotic stages of spermatogenesis. My results suggest that X chromosome inactivation in the *Drosophila* male germline occurs through a different mechanism than previously predicted and one that differs from the meiotic sex chromosome inactivation (MSCI) that occurs in mammals.

Chapter 1

Population and sex differences in *Drosophila melanogaster* brain gene expression

Ana Catalán, Stephan Hutter and John Parsch

BMC Genomics, 2012, **13**:654.

BMC Genomics

RESEARCH ARTICLE Open Access

Population and sex differences in *Drosophila melanogaster* brain gene expression

Ana Catalán, Stephan Hutter and John Parsch*

Abstract

Background: Changes in gene regulation are thought to be crucial for the adaptation of organisms to their environment. Transcriptome analyses can be used to identify candidate genes for ecological adaptation, but can be complicated by variation in gene expression between tissues, sexes, or individuals. Here we use high-throughput RNA sequencing of a single *Drosophila melanogaster* tissue to detect brain-specific differences in gene expression between the sexes and between two populations, one from the ancestral species range in sub-Saharan Africa and one from the recently colonized species range in Europe.

Results: Relatively few genes (<100) displayed sexually dimorphic expression in the brain, but there was an enrichment of sex-biased genes, especially male-biased genes, on the X chromosome. Over 340 genes differed in brain expression between flies from the African and European populations, with the inter-population divergence being highly correlated between males and females. The differentially expressed genes included those involved in stress response, olfaction, and detoxification. Expression differences were associated with transposable element insertions at two genes implicated in insecticide resistance (*Cyp6g1* and *CHKov1*).

Conclusions: Analysis of the brain transcriptome revealed many genes differing in expression between populations that were not detected in previous studies using whole flies. There was little evidence for sex-specific regulatory adaptation in the brain, as most expression differences between populations were observed in both males and females. The enrichment of genes with sexually dimorphic expression on the X chromosome is consistent with dosage compensation mechanisms affecting sex-biased expression in somatic tissues.

Keywords: Transcriptomics, Adaptation, Population genetics, Insecticide resistance, Sexual dimorphism

Background

The successful colonization of new habitats requires populations to adapt to novel biotic and abiotic conditions. Understanding the basis of this ecological adaptation is a major goal of evolutionary genetics. Because of its demographic history, the fruit fly *Drosophila melanogaster* offers an opportunity to address this fundamental issue in a well-established model system. Presently, *D. melanogaster* has a worldwide distribution spanning a wide variety of habitats. However, biogeographic and population genetic studies indicate that the species has its origin in sub-Saharan Africa and only began to colonize non-African regions about 15,000 years ago [1-6]. The expansion of the species to new, non-tropical environments is thought to have been accompanied by extensive genetic adaptation [4,7-12], although the identification of ecologically adapted genes and the characterization of their functions have proven difficult. Because changes in gene expression are expected to play an important role in adaptation [13-16], transcriptomic studies offer the possibility to bridge the gap between genotypic and phenotypic evolution and identify candidate genes that may have been the targets of regulatory adaptation. With this aim in mind, several microarray studies have been performed to identify gene expression differences between African and non-African *D. melanogaster* [17-19]. A comparable study has been carried out in *D. simulans*, which has a similar demographic history to *D. melanogaster* [20].

Although previous studies identified genes that differ in expression between African and European *Drosophila* populations [18-20], they suffered from some limitations. For example, these studies used mRNA extracted from

* Correspondence: parsch@bio.lmu.de
Department of Biology II, University of Munich (LMU), Grosshaderner Str. 2, Planegg-Martinsried 82152, Germany

whole flies. This approach provides a general picture of gene expression averaged over all tissues, but it is biased towards highly expressed genes and those expressed in many (or large) tissues. The use of whole flies typically does not provide the resolution to detect expression changes that occur only in a single tissue. A second limitation to the previous population studies is that they examined flies of only one sex per experiment [18-20]. Because gene expression is highly sexually dimorphic [21], especially when using whole flies or gonads [22], expression profiles can differ greatly between males and females. Consistent with this, there was very little overlap among the genes differing in expression between populations that were identified separately in males and females from whole-fly microarray studies [19], which suggests that most of the between-population expression divergence is sex-specific.

In order to get a more detailed picture of gene expression divergence between African and non-African *D. melanogaster*, we performed high-throughput RNA sequencing (RNA-seq) of mRNA isolated from dissected brains of adult males and females from two populations, one from the ancestral species range in sub-Saharan Africa (Zimbabwe) and one from the derived species range in Europe (the Netherlands). We chose to study gene expression variation in the brain because it plays a critical role in processing sensorial input from the environment. The visual, olfactory, and tactile stimuli coming from biotic sources, such as predators and food resources, as well as environmental conditions, such as temperature and humidity, differ greatly between these populations. Many of these environmental stimuli are detected by the sensorial organs of the fly's head (eyes, antennae, and proboscis) and are then processed by the brain, which produces a specific output that results in a behavioral and/or physiological response. Previous studies have shown that differences in gene expression in the brain can affect traits such as learning, memory, reproductive diapause, lifespan, and foraging behavior [23-26]. Furthermore, many behaviors that vary between strains or populations, including courtship, mating, aggression, and olfactory response, also exhibit sexual dimorphism [27,28].

The goal of this work is to identify genes that differ in their basal levels of brain gene expression between *D. melanogaster* strains originating from Africa and Europe. To this end, we use a 'common garden' experimental design in which flies from both populations are reared under identical laboratory conditions. This approach detects expression differences that have a genetic basis, but it cannot detect the effects of environment or gene-by-environment interactions. In total we identify 343 genes that differ in expression between the populations and 91 genes that differ in expression between the sexes. Our study represents the first brain-specific comparison of gene expression between African and non-African *D. melanogaster* and uncovers many genes that may play a role in ecological adaptation.

Results

The *D. melanogaster* brain transcriptome

To investigate population differences in gene expression, brains were dissected from 11–12 inbred lines each of an African and a European population (Figure 1). Total RNA was isolated from pooled brains within each population and used to generate cDNA libraries for RNA-seq. In total, we obtained over 270 million short sequence reads from eight cDNA libraries, which included two biological replicates of each sex and population (Table 1). On average, 71% of the reads could be mapped to annotated transcripts. Of the remaining reads, a large proportion (9–20% depending on the library) mapped to ribosomal RNA (rRNA). Because all samples were

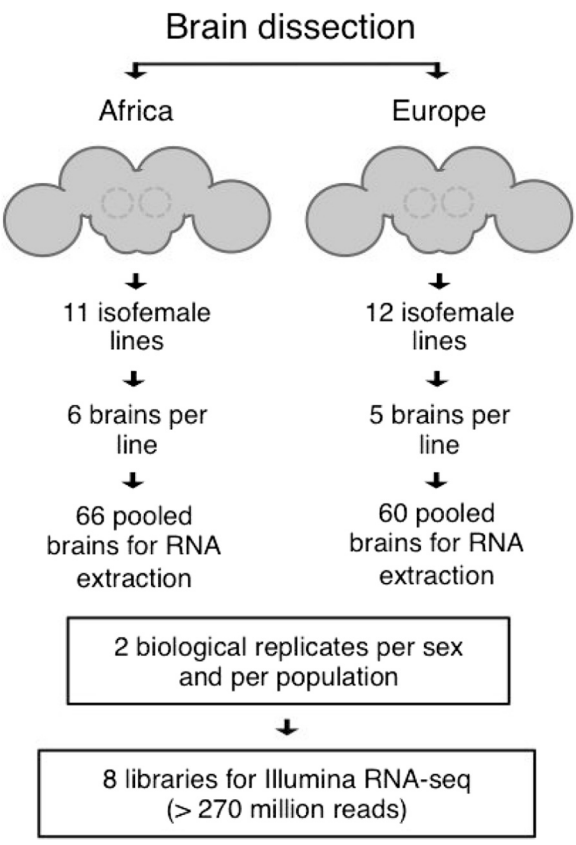


Figure 1 Overview of RNA sample preparation. Whole brains were dissected from isofemale lines derived from a European (the Netherlands) and an African (Zimbabwe) population. Brains were pooled within each population for RNA extraction. The dissection and pooling procedures were performed for two biological replicates of each population and sex, resulting in a total of eight samples that were used for cDNA library construction and high-throughput sequencing.

Table 1 Number of total and mapped reads (in millions) per sample

		Mapped reads (%)			
Sample	Reads	Transcripts	rRNA	Other*	Unmapped (%)
AfrFemale-R1	24.6	18.7 (75.8)	2.3 (9.2)	3.5 (14.3)	0.20 (0.81)
AfrFemale-R2	44.4	31.6 (71.1)	7.5 (17.0)	5.0 (11.1)	0.37 (0.82)
AfrMale-R1	29.4	20.0 (71.3)	5.0 (16.9)	3.3 (11.1)	0.23 (0.80)
AfrMale-R2	28.6	21.3 (74.5)	3.8 (13.4)	3.0 (10.5)	0.46 (1.60)
EurFemale-R1	27.2	18.3 (67.2)	5.7 (20.7)	2.7 (10.1)	0.54 (1.97)
EurFemale-R2	23.5	16.8 (71.3)	3.8 (16.0)	2.8 (11.8)	0.22 (0.94)
EurMale-R1	48.4	32.3 (66.6)	8.2 (16.9)	4.8 (10.0)	3.16 (6.53)
EurMale-R2	47.3	34.4 (72.7)	6.9 (14.6)	5.2 (11.1)	0.78 (1.64)

*Includes intergenic regions, introns, transposable elements, non-coding RNA (excluding rRNA), and pseudogenes.

enriched for poly(A) mRNA before cDNA synthesis, differences in the proportion of rRNA among libraries are likely to reflect differences in mRNA enrichment efficiency. Most of the reads that did not map to transcripts or rRNA could be mapped to intergenic regions (9%) or introns (2%). These may represent unannotated genes or transcript isoforms, but could also result from spurious transcription or intron retention. Around 2% of all reads could not be mapped to the genome (Table 1).

Of the 13,920 protein-encoding genes annotated in FlyBase release 5.43 [29], 13,575 had at least one mapped read in at least one of the libraries, while 10,873 had at least one mapped read in every library. A total of 11,531 genes had at least 16 reads when summed over all libraries (Figure 2), which was the minimum needed to detect significant differential expression given our experimental design and replication scheme. This set of genes was used for subsequent statistical analyses. Read counts per gene were highly correlated between the biological replicates, with Pearson's correlation coefficient, *R*, ranging from 0.93 to 0.99.

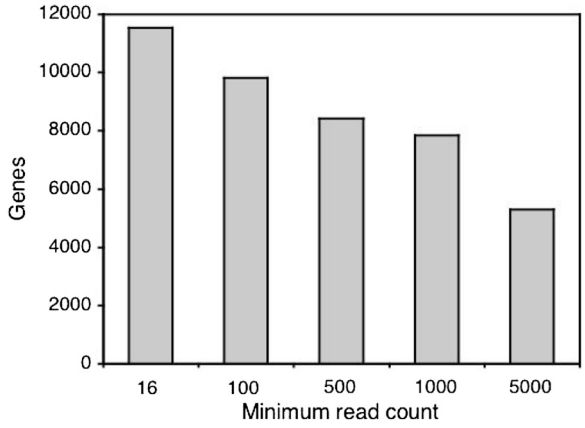


Figure 2 Number of genes meeting various read-count thresholds. The Y-axis indicates the number of genes that have the minimum number of mapped reads given on the X-axis. Read counts are summed over all libraries. The set of genes with at least 16 mapped reads was used for analysis of differential expression between sexes and populations.

Expression differences between the sexes

We identified sex-biased genes as those whose expression showed a significant effect of sex in a two-factor analysis that accounted for both sex and population (Figure 3; Additional file 1). Overall, the amount of sexually dimorphic expression was low, with 91 genes showing a significant difference in expression between the sexes at a false discovery rate (FDR) of 5% (Table 2). There was a slight tendency for genes with male-biased expression in the brain to show the same bias in whole

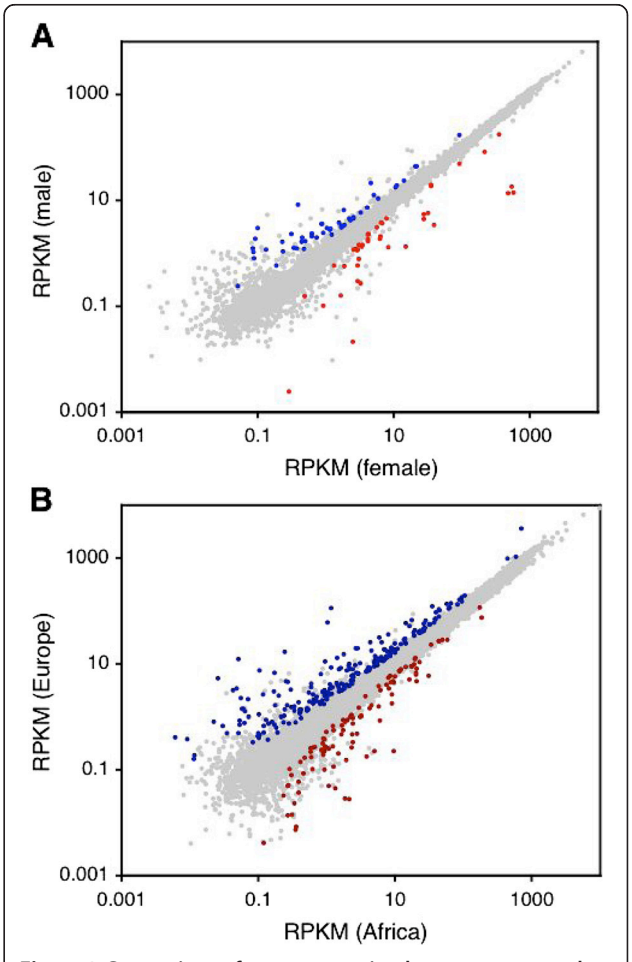


Figure 3 Comparison of gene expression between sexes and populations. (**A**) Dot plot of reads per kilobase per million mapped reads (RPKM) from female and male libraries. Genes with significant (FDR<5%) female-biased expression are shown in red. Genes with significant male-biased expression are shown in blue. (**B**) Dot plot of RPKM values from African and European libraries. Genes with significant (FDR<5%) African-biased expression are shown in dark red. Genes with significant European-biased expression are shown in dark blue.

Table 2 Number of sex-biased genes

Expression	Total genes	X-linked (%)	*P*
Sex-biased	91	52 (57)	$2x10^{-16}$
Male-biased	49	39 (80)	$2x10^{-16}$
Female-biased	42	13 (31)	0.017

Enrichment on the X chromosome was tested by Fisher's exact test.

flies (Additional file 2). For example, 25 of the 49 genes with male-biased expression in brain also had male-biased expression in whole flies [30,31]. However, of the 24 other genes with male-biased expression in the brain, 16 had female-biased expression and eight had unbiased expression in whole flies. Of the 42 genes with female-biased expression in the brain, 11 had female-biased expression, 10 had male-biased expression, and 22 had unbiased expression in whole flies (Additional file 2).

There was an enrichment of sex-biased genes on the X chromosome, which was significant for both male- and female-biased genes, but much stronger for male-biased genes (Table 2). A previous RNA-seq study using *D. melanogaster* heads found a similar enrichment of X-linked sex-based genes and suggested that it was related the mechanism of dosage compensation that occurs on the male X chromosome [32]. To test for a possible influence of dosage compensation on sex-biased expression in the brain, we examined the correlation between the $\log_2$(male/female) expression ratio of all X-linked genes with at least 100 mapped reads in each sex and the distance to the nearest male-specific lethal (MSL) binding site [33], which represents the assembly point for the dosage compensation complex (DCC). The correlation was significantly negative (Spearman's $\rho = -0.11$; $P < 0.001$), indicating that genes with relatively high expression in males tend to be close to MSL binding sites. This result held when the minimum read count was increased to 200 or 500 reads per sex. When genes of the different sex-bias classes were compared, male-biased genes were found to be significantly closer to MSL binding sites than female-biased or unbiased genes (Table 3).

Because most RNA-seq reads could not be mapped unambiguously to individual transcripts of genes with alternatively spliced isoforms, we had little power to detect sexually dimorphic expression among transcript isoforms. Nonetheless, we did detect significant sexual dimorphism in the expression of *transformer* and *doublesex* isoforms in the brain (Additional file 3). We also detected transcripts of the ribosomal protein genes *RpL17* and *RpS6* that had highly female-biased expression.

Expression differences between populations

We identified genes that differed in expression between the African and European populations as those with a significant effect of population in a two-factor analysis that accounted for both population and sex (Figure 3). This revealed a total of 343 differentially expressed genes at an FDR of 5% (Additional file 1). There were 16 genes that showed a significant effect of both sex and population on their expression (Additional file 1). In all of these cases, the direction of the population bias (European or African) was the same in both sexes. In general, the ratio of European-to-African expression per gene was highly correlated between males and females (Spearman's $\rho = 0.63$, $P < 0.0001$), indicating that there is little sex-dependent divergence in brain expression between populations. To further investigate this, we analyzed between-population expression divergence separately in females and males using a one-factor (population) analysis within each sex. This revealed 48 genes that were differentially expressed between the populations in one sex, but not the other, and were not detected in the two-factor analysis. The vast majority of these genes (42 out of 48) showed differential expression only in males, which is in contrast to the results previously reported for whole flies [19]. However, even among these genes there was a strong correlation between the European-to-African expression ratios observed in males and females (Spearman's $\rho = 0.63$, $P < 0.0001$; Figure 4), which again indicates that there is little sex-dependent gene expression divergence in brain between the populations.

There were more genes that showed relative over-expression in the European population (232) than in the African population (111; sign test, $P < 0.0001$). This could result from the mapping of RNA-seq reads to the reference genome (which was generated from a non-African lab strain) being more efficient for the European sample than for the African sample. To test for such a bias in mapping efficiency, we applied our read mapping

Table 3 Distance (*d*) to nearest MSL binding site for X-linked genes

			Number of reads with:		
Expression	Genes	Median *d* (bp)	*d* = 0 bp (%)	*d* < 3 kb (%)	*d* < 10 kb (%)
Male-biased	39	157*	19 (49)	31 (79)**	35 (90)**
Female-biased	13	4,795**	1 (8)	5 (38)	7 (62)
Unbiased	2,089	1,593	771 (37)	1,164 (56)	1,446 (69)

Differences in *d* between male-biased (or female-biased) and unbiased genes were tested by a Wilcoxon test. Differences in the proportion of genes in each category were tested by Fisher's exact test. *$P < 0.01$, **$P < 0.001$.

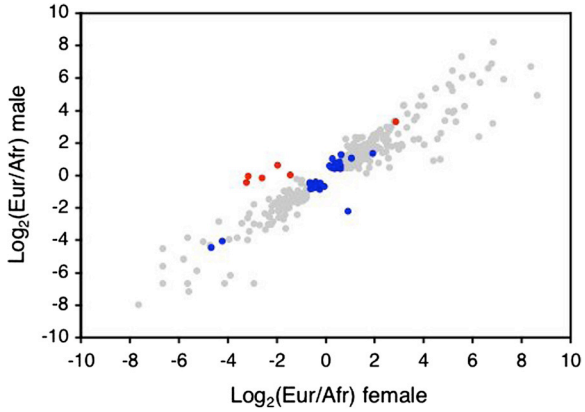


Figure 4 Between-population expression divergence in females and males. The ratio of European to African expression (on a $\log_2$ scale) is plotted for all genes that showed a significant (FDR < 5%) difference in expression between the populations. Gray points represent genes showing a significant effect of population across both sexes. Blue points represent genes that were significant only in males, while red points represent genes that were significant only in females.

procedure to simulated RNA-seq reads from each population. Overall, the mapping efficiency was very high with ~98% of all simulated reads from both populations being mapped to the correct gene in the reference sequence (Table 4). The remaining reads either could not be mapped to the transcriptome (~0.02%) or were mapped to an incorrect gene (~2%). Incorrect mapping occurred mostly when the exonic content of a gene showed overlap with another gene or, in rare cases, when gene families consisted of closely related paralogs. Across all genes, European reads showed slightly, but not significantly, higher mapping efficiency (Table 4). A similar result was observed for the subsets of genes with significant over-expression in either Africa or Europe (Table 4). Given that the observed median difference in expression of significant genes between populations was 2.7-fold, the contribution of mapping bias to the observed expression differences is expected to be negligible.

At the transcript level, we were able to identify 63 individual transcripts of multiple-transcript genes that differed in expression between the populations at an FDR of 5% (Additional file 4). The vast majority of these were cases where one transcript of a gene showed a significant bias towards one population and the other transcripts of that gene were either biased towards the same population or were not detected. Two transcripts of the gene *CHKov1* that are associated with a polymorphic transposable element insertion [34] showed significant over-expression in Europe (see below).

Table 4 Mapping efficiency of simulated RNA-seq reads to the reference *D. melanogaster* transcriptome

		Mean mapping efficiency in %		
Data set	**Genes**	**Africa**	**Europe**	***P***
All genes	13,520	97.63 (9.73)	97.65 (9.45)	0.07
Africa over-expressed	110	98.33 (6.45)	98.40 (6.74)	0.48
Europe over-expressed	218	97.93 (8.00)	98.17 (6.49)	0.78

Standard deviations are given in parentheses. *P*-values are from Wilcoxon signed-ranks tests.

Validation of RNA-seq results by qRT-PCR

For a subset of the genes analyzed by RNA-seq, we attempted to confirm the observed expression difference between populations using RNA extracted from new biological replicates and quantitative reverse-transcription PCR (qRT-PCR). The genes tested included five that were over-expressed in Europe (*CG31157*, *Cyp6a23*, *dsf*, *Hspc70-2*, and *TotA*), five that were over-expressed in Africa (*CG13331*, *CG16772*, *Est-Q*, *GstD3*, and *mtg*), and two that showed no difference in expression between the populations (*Ace* and *Robo3*). Overall, the expression ratios measured by the two methods were highly correlated (Pearson's $R = 0.79$, $P = 0.002$). Qualitatively, all of the genes gave consistent results with the two methods, with the exception of *mtg*, which showed high over-expression in the African population by RNA-seq, but weak over-expression in the European population by qRT-PCR (Figure 5). The gene *Hsc70-2* showed European over-expression by both methods, however the magnitude of over-expression was much greater in the RNA-seq data (Figure 5). Otherwise, there was good agreement in the expression levels detected by RNA-seq and qRT-PCR (Additional file 5).

Functional classification of differentially expressed genes

Of the genes expressed differentially between the European and African populations, six are known to be involved in the response to heat stress. These include the heat shock protein *Hsc70-2*, which showed 70-fold higher expression in European flies. A role for *Hsc70-2*

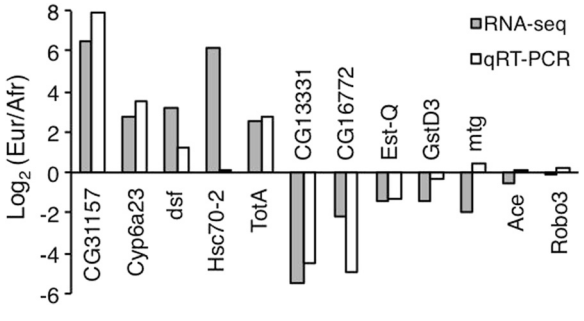


Figure 5 Comparison of RNA-seq and qRT-PCR results. The $\log_2$ ratio of European-to-African expression is shown for a subset of 12 genes that were measured by both RNA-seq (gray bars) and qRT-PCR (open bars).

in ecological adaptation is supported by parallel clines in genetic variation in both Australia and North America [35]. Another heat shock protein, *Hsp23*, showed two-fold over-expression in European flies. *Hsp23* is induced by both high and low temperatures [36,37], and its expression is associated with faster chill coma recovery [38], a phenotype known to differ between the African and European lines used in our analysis [39]. Two other genes that showed high over-expression in Europe, *TotA* (6-fold) and *TotC* (8-fold), are known to be involved in general stress response and show an induction in expression under both high and low temperatures [36,40].

A set of genes encoding chemosensory receptors was found to differ in expression between the populations. These included ionotropic glutamate receptors (*Ir93a* and *GluRIIA*), odorant-binding proteins (*Obp18a* and *Obp49a*), gustatory receptors (*Gr61a*), and olfactory receptors (*Or45b, Or63a, Or67d*, and *Or88a*). The four olfactory receptors were all expressed at higher levels (1.6–3.9-fold) in Europe than in Africa. *Or67d* binds to 11-*cis*-vaccenyl acetate, which is a volatile male-specific pheromone known to trigger aggregation and mating behavior in both sexes as well as male-male aggressive behavior [41-43]. *Or88a* is activated when flies are exposed to odors from virgin or mated females, although its exact ligand has not been identified [44].

The differentially expressed genes also included six glutathione S-transferase and seven cytochrome P450 genes, which are known to be important for detoxification. Notably, these included the cytochrome P450 gene *Cyp6g1*, whose over-expression is associated with resistance to DDT and related insecticides [45]. Previous studies of the same populations identified *Cyp6g1* as the gene with the greatest European over-expression when whole flies were examined [18,19]. In brain, *Cyp6g1* also shows strong over-expression (>4-fold) in European flies.

A cluster of differentially expressed genes on chromosome arm 3R

Three genes that showed significant over-expression in the African population (*CG10560*, *CG10562*, and *CHKov2*) are located in a cluster on chromosome arm 3R. This cluster also contains the gene *CHKov1*, which is known to produce different transcript isoforms due to the presence/absence of a polymorphic *Doc* transposable element insertion [34]. All of these genes are predicted to encode choline kinases. We found that the *Doc* element insertion, which promotes transcription of *CHKov1* isoforms that exclude the choline kinase domain, was present in all 12 of our European lines, but only one of the 11 African lines (Additional file 6). This region of the genome shows a strong reduction in nucleotide polymorphism that is limited to the European population (Figure 6), which is consistent with a recent selective sweep. Furthermore, there are blocks of strong linkage disequilibrium (LD) on either side of the region of reduced polymorphism (Figure 6), as is expected in the

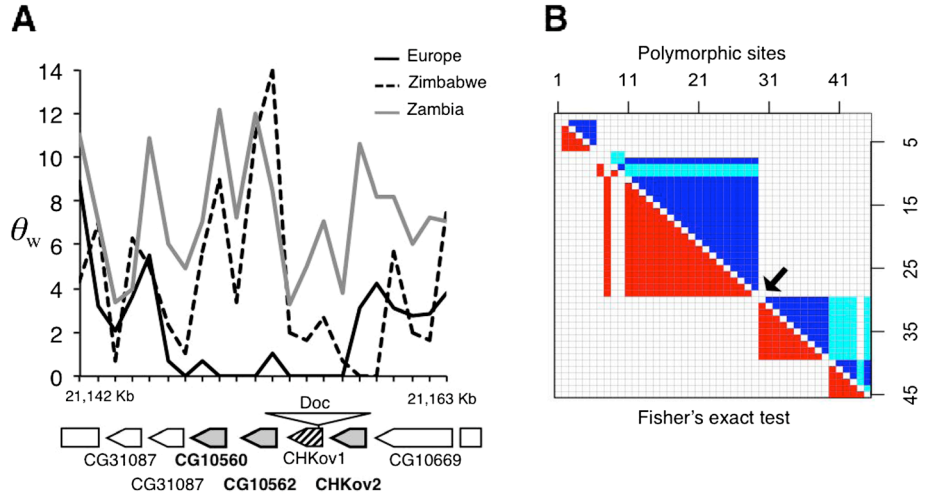


Figure 6 A cluster of differentially expressed genes in a region of low nucleotide polymorphism and strong linkage disequilibrium within Europe. (**A**) Nucleotide diversity (θ_W) in two African populations (Zimbabwe, dashed line; Zambia, gray line) and a European population (the Netherlands, solid line) along a 22 kb region of chromosome 3R. The genes located in this region are depicted below the plot, with arrowheads indicating the direction of transcription. Gray boxes represent genes that are over-expressed in the African population. The hatched box represents *CHKov1*, which is disrupted by a *Doc* element insertion in all of the European lines. (**B**) Linkage disequilibrium between all pairs of polymorphisms (excluding singletons) within the same 22 kb region in the European population. Values of r^2 are indicated in the upper matrix, with $r^2 \leq 0.4$ (white), $0.4 < r^2 < 0.8$ (light blue), and $r^2 \geq 0.8$ (dark blue). The lower matrix indicates the results of Fisher's exact test, with $P \leq 0.05$ (red) and $P > 0.05$ (white). The arrow indicates the location of the *CG10560*, *CG10562*, *CHKov1*, and *CHKov2* gene cluster.

case of a selective sweep [46]. To further test this, we computed the statistic ω, which quantifies LD on either side of a selected site relative to LD spanning the selected site [46,47]. High values of ω are expected following a selective sweep. The maximum value of ω in the *CHKov1* region was 33.95. A value this high occurred in only 7.8% of 10,000 neutral simulations that took into account the demographic history of the European population [48], indicating that the LD pattern is unlikely to be caused by demography alone.

Discussion

RNA-seq analysis of dissected brains revealed over 300 genes that differ in expression between population samples from Africa and Europe. Importantly, the analysis of a single tissue uncovered many differentially expressed genes that were not found in previous studies that compared gene expression in whole flies from the same populations [18,19]. In total, the previous studies identified 708 genes that differed in expression between the African and European populations in either males [18] or females [19]. Of these, only 15 genes also differed significantly in brain expression between the two populations, with 11 showing the same direction of difference (i.e., African or European over-expression). The only gene common to all three data sets was the insecticide resistance gene, *Cyp6g1*, which always showed high over-expression (>3-fold) in European flies. Increased expression of *Cyp6g1* is associated with an *Accord* element insertion in its upstream region and DDT resistance [45]. This insertion is present in all of our European lines, but only in half of the African lines [19]. Although the *Accord* insertion upstream of *Cyp6g1* mainly affects its expression in midgut, Malpighian tubule, and fat body [49], our results suggest that the *Accord* insertion has a similar effect on *Cyp6g1* expression in the brain, where it is expressed at low levels. In contrast, the gene *CG9509*, which shows 2–3 times greater expression in whole flies from Europe than in those from Africa due to variation in a *cis*-regulatory sequence [50], did not differ significantly in its brain expression between the two populations. In this case, the between-population expression difference appears to be specific to the Malpighian tubules, where *CG9509* shows highly enriched expression [51].

Four choline kinase genes that differ in expression between the African and European populations are located in a 22-kb region of chromosome arm 3R that shows evidence for a recent selective sweep in non-African populations (Figure 6) [34,52]. Three of these genes (*CG10560, CG10562* and *CHKov2*) had significantly higher expression in Africa than in Europe. The fourth gene, *CHKov1*, did not differ in expression between populations in our gene-based analysis. However, it did differ in the transcript-based analysis. Two truncated *CHKov1* transcripts that do not contain the choline kinase domain and are associated with a *Doc* element insertion showed significant over-expression in the European population, while the full-length transcript showed strong (1.7-fold), but not significant, over-expression in Africa. This is consistent with our finding that the *Doc* element insertion is in high frequency in Europe (12 out of 12 lines), but low frequency in Africa (1 out of 11 lines). Taken together, these results suggest that selection has favored a reduction in the choline kinase activity of all four genes in the European population. Previous work has shown that the *Doc* element insertion in *CHKov1* is associated with increased resistance to an organophosphate pesticide [34] and to sigma virus [52]. However, it is not known if the *Doc* insertion itself was the target of selection, or if it has hitchhiked to high frequency due to linkage with another selected variant. It is also not known if the *Doc* insertion is responsible for the expression difference of all choline kinase genes in this region, or if other regulatory changes are involved. Although the *Doc* insertion in *CHKov1* and the *Accord* insertion upstream of *Cyp6g1* are both associated with insecticide resistance, a key difference is that the former is associated with reduced expression, while the latter is associated with increased expression. Thus, resistance to organophosphate and organochlorine insecticides appears to occur through different mechanisms with respect to gene regulation.

The proportion of genes that were found to be differentially expressed in brain between the African and European populations of *D. melanogaster* (~2%) is similar to what has been reported for comparisons of wild and domesticated populations of fish and mammals (~1%) [53,54]. In contrast, comparisons of brain transcriptomes between nursing and foraging bees [55] and male Atlantic salmon with different mating strategies [56] revealed much higher proportions of differentially expressed genes (15% and 35%, respectively). Although the number of differently expressed genes that is detected in an experiment is highly sensitive to the experimental design, replication scheme, pooling of samples, and method of statistical analysis, these findings suggest that divergence in brain gene expression may be greater when individuals are separated into two very distinct behavioral classes.

In contrast to studies using whole flies or heads [22,30,32,57,58], we detected relatively little sexual dimorphism in brain gene expression. The difference between brains and whole flies is expected, as the vast majority of genes showing sex-biased expression in whole flies are expressed in reproductive tissues [22]. In head, it appears that most sex-biased gene expression occurs outside of the brain. A previous microarray study

found 754 sex-biased genes in the head, but only four in the central nervous system (brain plus ventral nerve cord) [58]. Similarly, an RNA-seq study identified 1,381 sex-biased genes in the head [32], while our RNA-seq analysis found only 91 sex-biased genes in the brain. These results suggest that nearly all of the sex-biased expression in the head occurs in non-nervous tissues such as the fat body, which is thought to play an important role in regulating sex-specific reproductive behavior and physiology [59].

Despite the relatively low level of sexual dimorphism in brain gene expression, we detected a significant over-representation of sex-biased genes (both male- and female-biased) on the X chromosome. Previous studies of whole flies observed a paucity of male-biased genes and an excess of female-biased genes on the X chromosome [22,57]. For male-biased genes, this pattern appears to be driven by gene expression in reproductive tissues [60] and an excess of X-linked male-biased genes in somatic tissue (head) has been reported [32,60]. It has been suggested that the over-representation of somatic male-biased genes on the X chromosome is related to the mechanism of dosage compensation [32]. Our data are consistent with this interpretation, as we find that X-linked, male-biased genes are significantly closer to MSL binding sites than female-biased or unbiased genes (Table 3). This suggests that their proximity to the MSL binding site leads to an "over-compensation" of expression in male somatic tissue. Our results contrast with those of a previous study of sex-biased gene expression in gonadectomized flies [61], which suggested that MSL binding might interfere with sex-specific regulation, leading to a reduction of male-biased expression. A possible explanation for this discrepancy is that the identification of sex-biased genes in gonadectomized flies might be confounded by variation in sex-bias among tissues. In comparison to male-biased genes, female-biased genes show a weaker enrichment on the X chromosome. Although there is some indication that X-linked female-biased genes tend to be located farther from MSL binding sites than male-biased or unbiased genes (Table 3), the sample size is too small to draw a firm conclusion. Thus, the excess of X-linked, female-biased genes may not be related to dosage compensation, but instead may reflect an overall feminization of the X chromosome [60], possibly caused by sexually antagonistic selection [21,62,63].

Conclusions

Organisms may adapt to new or changing environmental conditions by altering levels of gene expression. Since expression profiles can vary greatly among tissues, it is likely that some adaptive regulatory changes are tissue- or organ-specific. For example, gene expression changes occurring specifically in the brain may underlie adaptive behavioral or physiological responses to the environment. However, there are few cases where gene expression and behavioral polymorphisms have been linked in an evolutionary framework [64] and more work is needed in order to understand the relationships among selection, gene expression, and behavior/physiology. To this end, we used RNA-seq to compare brain gene expression between two *D. melanogaster* populations from different habitats. In total, we identified 343 genes that differed in brain expression between the populations, the vast majority of which were not detected in analogous studies that used whole flies as the source of RNA.

Among the differentially expressed genes were those with functions in stress response, olfaction, and detoxification, including two genes previously implicated in insecticide resistance. Brain tissue is thought to be especially sensitive to heat, cold, and oxidative stress, and also can be affected by chemical stressors, such as insecticides. For example, some insects are known to metabolize insecticides specifically in nerve ganglia [65,66]. Thus, the alteration of gene expression in the brain may be particularly important for environmental adaptation. Further studies are needed to elucidate the contribution of specific gene expression changes to behavioral and physiological differences between populations and to determine the selective agents and regulatory mechanisms responsible for them.

Methods

Fly strains and brain dissection

The population samples consisted of 11 isofemale lines (*A84*, *A95*, *A131*, *A145*, *A157*, *A186*, *A191*, *A229*, *A377*, *A384*, and *A398*) collected from Lake Kariba, Zimbabwe and 12 isofemale lines (*E01*, *E02*, and *E11–E20*) collected from Leiden, the Netherlands [8,18,19]. An estimate of cosmopolitan admixture for our African population is not available. However, admixture estimates are available for two nearby populations, Siavonga, Zambia (9 km away) and Sengwa, Zimbabwe (33 km away) [67]. Both of these populations show admixture proportions below 3%, suggesting that admixture in our population should be negligible. All flies were maintained on standard cornmeal-molasses medium at 22° with a 14 h light:10 h dark cycle. Adult flies aged 2–4 days were anesthetized on ice and brains were dissected in 1xPBS (phosphate buffered saline) and stored in RNAlater (Qiagen) to prevent RNA degradation. Five or six brains from each of the African and European lines were dissected and pooled following the scheme shown in Figure 1. Two biological replicates were performed for each population and sex.

RNA extraction and high-throughput sequencing

Total RNA extraction and DNase I digestion were performed using the MasterPure RNA Purification Kit (Epicentre). cDNA library construction and high-throughput sequencing were performed by GATC Biotech (Konstanz, Germany). Briefly, poly-A mRNA was purified and fragmented by sonication. First-strand, single-end cDNA was synthesized using random primers. Eight tagged libraries were generated, pooled and run on two lanes of a HiSeq 2000 sequencer (Illumina) to generate single reads of 50 bases. All sequences have been submitted to the GEO database under the series GSE40907.

Read mapping

Illumina sequence reads were mapped to the reference *D. melanogaster* transcriptome (FlyBase release 5.43) [29] using Stampy (version 1.0.17) [68] with default parameters, except that expected divergence to the reference sequence was set to 1%. This divergence corresponds roughly to the upper limit of what is observed when comparing exonic regions of African sequences from the *Drosophila* Population Genomics Project (DPGP) [67] to the reference genome (0.5% – 0.7% divergence, depending on strain). For comparison, we also mapped all reads using Bowtie (version 0.12.7) [69] and allowing a maximum of three mismatches over the length of the sequence read (option –v 3). The two methods gave highly concordant results, with a nearly perfect correlation in the number of reads mapped per gene over all libraries (Pearson's $R > 0.99$ in all cases). Overall, a higher proportion of reads were mapped to the transcriptome with Stampy (71%) than with Bowtie (68%). For this reason, we used the Stampy results for all subsequent analyses.

Two approaches were used to estimate expression levels. First, expression was quantified on a "per gene" basis. For this, if a sequence read mapped to one or more transcripts of the same gene, it was counted as one "hit" for that gene and all subsequent analyses were performed at the gene level. For calculations of RPKM, the length of the longest transcript of each gene was used. The use of the longest transcript systematically overestimates the true transcript length for genes with multiple transcripts, but is a reasonable compromise in situations where the relative abundance of the different transcript isoforms is unknown. In the second approach expression was quantified on a "per transcript" basis. For this, only reads that mapped uniquely to a single transcript were considered and all subsequent analyses were performed at the transcript level. Because there is high overlap among alternate transcript isoforms of the same gene, many reads could not be mapped to a specific transcript and were discarded. Thus, the "per transcript" approach results in a considerable loss of information. For this reason, all results are presented on a "per gene" basis unless otherwise noted. Reads that did not map to any transcript were mapped to other features of the *D. melanogaster* genome using annotation 5.43 and the procedure described above. For this, the single best match was used or, if there were multiple matches of equal quality, one was chosen arbitrarily. Reads that did not match any sequence in the genome were considered unmapped.

Differential expression analysis

To test for differential expression between sexes or populations, we used the DESeq package (version 2.10) [70] implemented in R (version 2.14.1) [71]. This approach is based on the negative binomial distribution and accounts for the dispersion in read counts per gene across replicate samples. We analyzed data from all eight of our samples (Table 1) together using a two-factor design that included both population (Africa or Europe) and sex (male or female). Significant effects of population or sex were determined by comparing the fit of the two-factor model to that of a one-factor model that excluded the factor of interest using the function *fitNbinomGLMs*. The FDR was determined using Benjamini-Hochberg adjusted *P*-values [72]. A total of 2,390 genes with fewer than 16 mapped reads summed over all libraries, which was the minimum needed to detect significant differential expression under our experimental design, were removed prior to statistical analysis.

Simulation analysis of mapping efficiency of African and European RNA-seq reads

To test for a potential population bias during the mapping process, we simulated RNA-seq samples using the whole genome sequences for three African (Zimbabwe) and eleven European (the Netherlands) strains made available by the DPGP as part of the DPGP2 data set (SRP005599) in the NCBI short read archive [67]. The African strains used were *A84*, *A131*, and *A186* (denoted by *ZK* in the DPGP data). The European strains used were *E01*, *E02*, and *E11–E19*. For each transcript in the genome we randomly sampled 100 fragments of length 50 bp covering the exonic content for each available strain. Fragments that contained stretches of four or more consecutive uncalled or masked bases were removed from the sample, assuming the lack of data was the result of missing coverage during the sequencing and/or assembly process rather than indel polymorphism within coding regions. For some transcripts this quality control step resulted in the removal of all simulated reads for one or more strains. These transcripts were removed from the data set and a total of 13,520 genes located on the five major chromosomal arms remained for subsequent analysis. The simulated reads

were then mapped to the reference transcriptome using Stampy [68] as described above and the proportion of fragments that could be mapped to the correct gene (referred to as mapping efficiency) was recorded for the African and European strains for each gene. For genes with multiple transcripts, mapping efficiencies were recorded for each transcript individually and then averaged to obtain per gene mapping efficiencies. The above analysis was repeated for the subset of genes detected as differentially expressed between populations in our RNA-seq experiment, but the number of sampled fragments was increased to 10,000 in order to improve the sensitivity of our method. Of the 343 genes in this data set, six were removed because they were not located on one of the five major chromosomal arms and another nine genes, all being part of the European over-expressed subset, were removed by the quality control step that excluded reads containing too many consecutive uncalled bases. All of these latter nine genes were located either close to the centromere or in regions containing DNA duplications, which can explain the lack of properly assembled genomic sequence. This lack of data did not show a population bias, as it was found in both European and African strains for all nine genes.

qRT-PCR

On the basis of the RNA-seq data, 12 genes were chosen for qRT-PCR analysis. Brain dissection, sample pooling, and total RNA extraction were performed as described above for RNA-seq, with the exception that four biological replicates were performed for each population and sex. cDNA was synthesized using random primers and SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer's instructions. TaqMan Gene Expression Assays (Applied Biosystems) were used for the following genes (TaqMan IDs are given in parentheses): *Ace* (Dm02134758_g1), *robo3* (Dm01800570_g1), *Hsc70-2* (Dm02330923_gH), *dsf* (Dm01842631_g1), *TotA* (Dm02363547_s1), *Cyp6a23* (Dm01824231_g1), *CG31157* (Dm02148979_g), *GstD3* (Dm02153755_s1), *Est-Q* (Dm01792292_g1), *mtg* (Dm02137882_g1), *CG13331* (Dm01797351_g1), and *CG16772* (Dm02365807_s1). The ribosomal protein gene *RpL32* (Dm02151827_g1) was used as a reference gene for expression normalization. qRT-PCR was performed using a Real-Time thermal cycler CFX96 (Bio-Rad). The ΔΔCt method was used to compute the normalized expression of the genes of interest [73].

PCR assay for the *Doc* element insertion in *CHKov1*

Specific forward (5'-GAACTCCGTGGGATCGACTA-3') and reverse (5'-GCGGAGCTTTTGAGAGAAGA-3') primers were designed to detect the presence or absence of the *Doc{}CHKov1*^{Doc1420} transposable element insertion in *CHKov1*. The 23 isofemale lines from the African and European populations (described above) were tested for the presence of the *Doc* element. DNA was extracted from a single fly of each line using the MasterPure DNA Purification Kit (Epicentre). PCR was conducted under standard conditions and the presence or absence of the *Doc* element was determined by the size of the amplified fragment (5.5 kb with the insertion, 1 kb without) as determined by 1% agarose gel electrophoresis.

Population genetic analysis

To analyze DNA sequence polymorphism in the *CHKov1* region, we used whole genome sequences generated by the DPGP [67]. The European population sample consisted of 11 of the Netherlands lines used in our RNA-seq analysis. One African population sample consisted of three of the Zimbabwe lines used in our RNA-seq analysis, while a second African population sample consisted of four lines from Siavonga, Zambia. Nucleotide diversity was calculated using Watterson's estimator, θ_W [74]. Linkage disequilibrium (LD) between any two polymorphic sites was calculated using Lewontin's $r^2 = D^2/p_1q_1p_2q_2$ where D is the frequency of the haplotypes and p and q represent the allele frequencies [75]. We calculated r^2 between all pairs of polymorphic sites, excluding singletons. Significance of the r^2 values was assessed with Fisher's exact test. The ω statistic was calculated using the sliding window approach implemented in the software OmegaPlus [47], with a minimum of four polymorphisms per window. To determine if the observed ω value could be expected under a neutral non-equilibrium demographic model, we conducted a parametric bootstrap analysis [76]. For this, we generated 10,000 simulated data sets using a coalescent demographic model that takes into account our current knowledge of the demographic history of the African and European populations of *D. melanogaster* [48]. This model assumes that the African and European populations diverged 128,430 generations ago from an ancestral population with an effective size (N_e) of 1,705,328. At the time of divergence, the European N_e was reduced to 32,128 before directly entering a phase of exponential growth until reaching a current N_e of 878,506. The simulated data sets were identical to the observed data set in terms of mutation rate, recombination rate, number of sites, and number of sampled individuals. For every simulated data set we performed a sliding-window analysis of the ω statistic and recorded the maximal value. The *P*-value was defined as the proportion of simulations with ω greater than or equal to the observed value.

Additional files

Additional file 1: Expression levels of all genes in all replicates. Table of read counts and relative expression levels of all genes in all replicates, including *P*-values for comparisons between populations and sexes.

Additional file 2: Sex-biased gene expression in the brain, Table of genes showing significant sex-biased gene expression in the brain and their expression bias in whole flies.

Additional file 3: Transcripts of multiple-transcript genes that differ in expression between the sexes. Table of individual transcripts that show significant sex-biased expression.

Additional file 4: Transcripts of multiple-transcript genes that differ in expression between populations. Table of individual transcripts that show a significant expression difference between the African and European populations.

Additional file 5: Results of qRT-PCR. Figure showing the relative expression of genes in the African and European populations as determined by qRT-PCR.

Additional file 6: PCR assay for the *Doc* element insertion in *CHKov1*. Figure showing the results of the PCR assay to detect the presence of the *Doc* element insertion in *CHKov1* in all African and European lines.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
AC and JP conceived of the study and its design. AC performed the brain dissections, RNA extractions and qRT-PCR. AC, SH and JP analyzed the RNA-seq data. AC and SH performed the population genetic analyses. AC and JP wrote the manuscript with input from SH. All authors read and approved the final manuscript.

Acknowledgements
We would like to thank Charles Langley and John Pool for making the DPGP2 data available prior to publication and for providing assembled genomes for our simulation analysis. We thank Hedwig Gebhart for technical assistance in the lab, Pablo Duchen, Stefan Laurent, and Pavlos Pavlidis for help with the population genetic analysis, and all members of the Munich evolutionary biology 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* grant PA 903/5.

Received: 31 July 2012 Accepted: 16 November 2012
Published: 21 November 2012

References
1. David JR, Capy P: **Genetic variation of *Drosophila melanogaster* natural populations.** *Trends Genet* 1988, **4:**106–111.
2. Lachaise D, Silvain JF: **How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster–D. simulans* palaeogeographic riddle.** *Genetica* 2004, **120:**17–39.
3. 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.
4. 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.
5. 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.
6. Stephan W, Li H: **The recent demographic and adaptive history of *Drosophila melanogaster*.** *Heredity* 2007, **98:**65–68.
7. 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.
8. 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.
9. Li H, Stephan W: **Inferring the demographic history and rate of adaptive substitution in *Drosophila*.** *PLoS Genet* 2006, **2:**e166.
10. 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.
11. González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA: **High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*.** *PLoS Biol* 2008, **6:**e251.
12. González J, Karasov TL, Messer PW, Petrov DA: **Genome-wide patterns of adaptation to temperate environments associated with transposable elements in *Drosophila*.** *PLoS Genet* 2010, **6:**e1000905.
13. King MC, Wilson AC: **Evolution at two levels in humans and chimpanzees.** *Science* 1975, **188:**107–116.
14. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA: **The evolution of transcriptional regulation in eukaryotes.** *Mol Biol Evol* 2003, **20:**1377–1419.
15. Whitehead A, Crawford DL: **Variation within and among species in gene expression: raw material for evolution.** *Mol Ecol* 2006, **15:**1197–1211.
16. Wray GA: **The evolutionary significance of *cis*-regulatory mutations.** *Nat Rev Genet* 2007, **8:**206–216.
17. 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.
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. Müller L, Hutter S, Stamboliyska R, Saminadin-Peter SS, Stephan W, Parsch J: **Population transcriptomics of *Drosophila melanogaster* females.** *BMC Genomics* 2011, **12:**81.
20. Wurmser F, Ogereau D, Mary-Huard T, Loriod B, Joly D, Montchamp-Moreau C: **Population transcriptomics: insights from *Drosophila simulans*, *Drosophila sechellia* and their hybrids.** *Genetica* 2011, **139:**465–477.
21. Ellegren H, Parsch J: **The evolution of sex-biased genes and sex-biased gene expression.** *Nat Rev Genet* 2007, **8:**689–698.
22. 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.
23. Mery F, Belay AT, So AK-C, Sokolowski MB, Kawecki TJ: **Natural polymorphism affecting learning and memory in *Drosophila*.** *Proc Natl Acad Sci USA* 2007, **104:**13051–13055.
24. Melcher C, Pankratz MJ: **Candidate gustatory interneurons modulating feeding behavior in the *Drosophila* brain.** *PLoS Biol* 2005, **3:**e305.
25. Broughton SJ, Piper MDW, Ikeya T, Bass TM, Jacobson J, Driege Y, Martinez P, Hafen E, Withers DJ, Leevers SJ, Partridge L: **Longer lifespan, altered metabolism, and stress resistance in *Drosophila* from ablation of cells making insulin-like ligands.** *Proc Natl Acad Sci USA* 2005, **102:**3105–3110.
26. Stehlík J, Závodská R, Shimada K, Sauman I, Kostál V: **Photoperiodic induction of diapause requires regulated transcription of timeless in the larval brain of *Chymomyza costata*.** *J Biol Rhythms* 2008, **23:**129–139.
27. Anholt RR, Mackay TF: **Quantitative genetic analyses of complex behaviours in *Drosophila*.** *Nat Rev Genet* 2004, **5:**838–849.
28. Anholt RR, Mackay TF: **Genetics of aggression.** *Annu Rev Genet* 2012, **46:**145–164.
29. McQuilton P, St Pierre SE, Thurmond J, FlyBase Consortium: **FlyBase 101–the basics of navigating FlyBase.** *Nucleic Acids Res* 2012, **40:**706–714.
30. Gnad F, Parsch J: **Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression.** *Bioinformatics* 2006, **22:**2577–2579.
31. Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, Brown JB, Cherbas L, Davis CA, Dobin A, Li R, Lin W, Malone JH, Mattiuzzo NR, Miller D, Sturgill D, Tuch BB, Zaleski C, Zhang D, Blanchette M, Dudoit S, Eads B, Green RE, Hammonds A, Jiang L, Kapranov P, *et al*: **The developmental transcriptome of *Drosophila melanogaster*.** *Nature* 2011, **471:**473–479.

32. Chang PL, Dunham JP, Nuzhdin SV, Arbeitman MN: **Somatic sex-specific transcriptome differences in *Drosophila* revealed by whole transcriptome sequencing.** *BMC Genomics* 2011, **12:**364.
33. Alekseyenko AA, Larschan E, Lai WR, Park PJ, Kuroda MI: **High-resolution ChIP-chip analysis reveals that the *Drosophila* MSL complex selectively identifies active genes on the male X chromosome.** *Genes Dev* 2006, **20:**848–857.
34. Aminetzach YT, Macpherson JM, Petrov DA: **Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*.** *Science* 2005, **309:**764–767.
35. Turner TL, Levine MT, Eckert ML, Begun DJ: **Genomic analysis of adaptive differentiation in *Drosophila melanogaster*.** *Genetics* 2008, **179:**455–473.
36. Zhang J, Marshall KE, Westwood JT, Clark MS, Sinclair BJ: **Divergent transcriptomic responses to repeated and single cold exposures in *Drosophila melanogaster*.** *J Exp Biol* 2011, **214:**4021–4029.
37. Carmel J, Rashkovetsky E, Nevo E, Korol A: **Differential expression of small heat shock protein genes *Hsp23* and *Hsp40*, and heat shock gene *Hsr-omega* in fruit flies (*Drosophila melanogaster*) along a microclimatic gradient.** *J Hered* 2011, **102:**593–603.
38. Colinet H, Lee SF, Hoffmann A: **Temporal expression of heat shock genes during cold stress and recovery from chill coma in adult *Drosophila melanogaster*.** *FEBS J* 2010, **277:**174–185.
39. Svetec N, Werzner A, Wilches R, Pavlidis P, Alvarez-Castro JM, Broman KW, Metzler D, Stephan W: **Identification of X-linked quantitative trait loci affecting cold tolerance in *Drosophila melanogaster* and fine mapping by selective sweep analysis.** *Mol Ecol* 2010, **20:**530–544.
40. Ekengren S, Hultmark D: **A family of *turandot*-related genes in the humoral stress response of *Drosophila*.** *Biochem Biophys Res Commun* 2001, **284:**998–1003.
41. Xu P, Atkinson R, Jones DNM, Smith DP: ***Drosophila* OBP LUSH is required for activity of pheromone-sensitive neurons.** *Neuron* 2005, **45:**193–200.
42. Ha TS, Smith DP: **A pheromone receptor mediates 11-*cis*-vaccenyl acetate-induced responses in *Drosophila*.** *J Neurosci* 2006, **26:**8727–8733.
43. Wang L, Anderson DJ: **Identification of an aggression-promoting pheromone and its receptor neurons in *Drosophila*.** *Nature* 2010, **463:**227–231.
44. van der Goes van Naters W, Carlson JR: **Receptors and neurons for fly odors in *Drosophila*.** *Curr Biol* 2007, **17:**606–612.
45. 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.
46. Kim Y, Nielsen R: **Linkage disequilibrium as a signature of selective sweeps.** *Genetics* 2004, **167:**1513–1524.
47. Alachiotis N, Stamatakis A, Pavlidis P: **OmegaPlus: A scalable tool for rapid detection of selective sweeps in whole-genome datasets.** *Bioinformatics* 2012, **28:**2274–2275.
48. Laurent SJ, Werzner A, Excoffier L, Stephan W: **Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia.** *Mol Biol Evol* 2011, **28:**2041–2051.
49. Chung H, Bogwitz MR, McCart C, Andrianopoulos A, ffrench-Constant RH, Batterham P, Daborn PJ: ***Cis*-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*.** *Genetics* 2007, **175:**1071–1077.
50. Saminadin-Peter SS, Kemkemer C, Pavlidis P, Parsch J: **Selective sweep of a *cis*-regulatory sequence in a non-African population of *Drosophila melanogaster*.** *Mol Biol Evol* 2012, **29:**1167–1174.
51. Chintapalli VR, Wang J, Dow JA: **Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease.** *Nat Genet* 2007, **39:**715–720.
52. Magwire MM, Bayer F, Webster CL, Cao C, Jiggins FM: **Successive increases in the resistance of *Drosophila* to viral infection through a transposon insertion followed by a duplication.** *PLoS Genet* 2011, **7:**e1002337.
53. Drew RE, Settles ML, Churchill EJ, Williams SM, Balli S, Robison BD: **Brain transcriptome variation among behaviorally distinct strains of zebrafish (*Danio rerio*).** *BMC Genomics* 2012, **13:**323.
54. Albert FW, Somel M, Carneiro M, Aximu-Petri A, Halbwax M, Thalmann O, Blanco-Aguiar JA, Plyusnina IZ, Trut L, Villafuerte R, Ferrand N, Kaiser S, Jensen P, Pääbo S: **A comparison of brain gene expression levels in domesticated and wild animals.** *PLoS Genet* 2012, **8:**e1002962.
55. Whitfield CW, Cziko A-M, Robinson GE: **Gene expression profiles in the brain predict behavior in individual honey bees.** *Science* 2003, **302:**296–299.
56. Aubin-Horth N, Landry CR, Letcher BH, Hofmann HA: **Alternative life histories shape brain gene expression profiles in males of the same population.** *Proc Biol Sci* 2005, **272:**1655–1662.
57. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL: **Sex-dependent gene expression and evolution of the *Drosophila* transcriptome.** *Science* 2003, **300:**1742–1745.
58. Goldman TD, Arbeitman MN: **Genomic and functional studies of *Drosophila* sex hierarchy regulated gene expression in adult head and nervous system tissues.** *PLoS Genet* 2007, **3:**e216.
59. Dalton JE, Kacheria TS, Knott SR, Lebo MS, Nishitani A, Sanders LE, Stirling EJ, Winbush A, Arbeitman MN: **Dynamic, mating-induced gene expression changes in female head and brain tissues of *Drosophila melanogaster*.** *BMC Genomics* 2010, **11:**541.
60. Meisel RP, Malone JH, Clark AG: **Disentangling the relationship between sex-biased gene expression and X-linkage.** *Genome Res* 2012, **22:**1255–1265.
61. Bachtrog D, Toda NR, Lockton S: **Dosage compensation and demasculinization of X chromosomes in *Drosophila*.** *Curr Biol* 2010, **20:**1476–1481.
62. Rice WR: **Sex chromosomes and the evolution of sexual dimorphism.** *Evolution* 1984, **38:**735–742.
63. Charlesworth B, Coyne JA, Barton NH: **The relative rates of evolution of sex chromosomes and autosomes.** *Am Nat* 1987, **130:**113–146.
64. Fitzpatrick MJ, Feder E, Rowe L, Sokolowski MB: **Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene.** *Nature* 2007, **447:**210–212.
65. Zhu F, Parthasarathy R, Bai H, Woithe K, Kaussmann M, Nauen R, Harrison DA, Palli SR: **A brain-specific cytochrome P450 responsible for the majority of deltamethrin resistance in the QTC279 strain of *Tribolium castaneum*.** *Proc Natl Acad Sci USA* 2010, **107:**8557–8562.
66. Korytko PJ, Scott JG: **CYP6D1 protects thoracic ganglia of houseflies from the neurotoxic insecticide cypermethrin.** *Arch Insect Biochem Physiol.* 1998, **37:**57–63.
67. Drosophila Population Genomics Project. http://www.dpgp.org.
68. Lunter G, Goodson M: **Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads.** *Genome Res* 2011, **21:**936–939.
69. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10:**R25.
70. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11:**R106.
71. Development Core Team R: *R: a language and environment for statistical computing, reference index version 2.14.1*. Vienna, Austria: R Foundation for Statistical Computing; 2011.
72. 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.
73. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic Acids Res* 2001, **29:**e45.
74. Watterson GA: **On the number of segregating sites in genetical models without recombination.** *Theor Popul Biol* 1975, **276:**256–276.
75. Lewontin RC: **The interaction of selection and linkage. I. General considerations; heterotic models.** *Genetics* 1964, **49:**49–67.
76. Pavlidis P, Laurent S, Stephan W: **msABC: a modification of Hudson's ms to facilitate multi-locus ABC analysis.** *Mol Ecol Resour* 2010, **10:**723–727.

doi:10.1186/1471-2164-13-654
Cite this article as: Catalán *et al.*: Population and sex differences in *Drosophila melanogaster* brain gene expression. *BMC Genomics* 2012 **13**:654.

Additional file 2: Sex-biased gene expression in the brain

		Sex bias		
Gene	Symbol	Brain	Whole fly[a]	Whole fly[b]
CG7024	*CG7024*	Male	Male	Male
CG9806	*CG9806*	Male	Male	Male
CG7592	*Obp99b*	Male	Male	Male
CG4979	*sxe2*	Male	Male	Male
CG6999	*CG6999*	Male	Male	Male
CG3360	*Cyp313a1*	Male	Male	Male
CG6730	*Cyp4d21*	Male	Male	Male
CG11315	*Npc2h*	Male	Male	Male
CG17052	*obst-A*	Male	Male	Male
CG13057	*retinin*	Male	Male	Male
CG3699	*CG3699*	Male	Male	Male
CG3466	*Cyp4d2*	Male	Male	Male
CG32706	*CG32706*	Male	Male	NA
CG2681	*CG2681*	Male	Male	Unbiased
CG14787	*CG14787*	Male	Male	Unbiased
CG13762	*CG13762*	Male	Male	Unbiased
CG6788	*CG6788*	Male	Male	Unbiased
CG32673	*Rab9E*	Male	Male	Unbiased
CG42398	*SteXh:CG42398*	Male	Male	Unbiased
CG11648	*Abd-B*	Male	Male	Unbiased
CG14195	*CG14195*	Male	Male	Unbiased
CG18467	*CG18467*	Male	Male	Unbiased
CG3540	*Cyp4d14*	Male	Male	Unbiased
CG40305	*FucTC*	Male	Male	Unbiased
CG3630	*CG3630*	Male	Male	Unbiased
CG2709	*CG2709*	Male	Unbiased	Unbiased
CG2706	*fs(1)Yb*	Male	Unbiased	Unbiased
CG4872	*CG4872*	Male	Unbiased	Unbiased
CG2904	*ec*	Male	Unbiased	Unbiased
CG3706	*CG3706*	Male	Unbiased	Unbiased
CG9307	*Cht5*	Male	Unbiased	Unbiased
CG4542	*CG4542*	Male	Unbiased	Female
CG32763	*l(1)G0045*	Male	Female	Unbiased
CG4293	*CG4293*	Male	Female	Female
CG9904	*CG9904*	Male	Female	Female
CG6461	*CG6461*	Male	Female	Female
CG15914	*CG15914*	Male	Female	Female
CG10803	*CG10803*	Male	Female	Female
CG9203	*CG9203*	Male	Female	Female
CG8939	*CG8939*	Male	Female	Female
CG9938	*Ndc80*	Male	Female	Female
CG2079	*Dok*	Male	Female	Female
CG1994	*l(1)G0020*	Male	Female	Female
CG14814	*CG14814*	Male	Female	Female
CG11164	*CG11164*	Male	Female	Female
CG4206	*Mcm3*	Male	Female	Female
CG4586	*CG4586*	Male	Female	Female
CG11420	*png*	Male	Female	Female
CG3692	*CalpC*	Male	Female	Female
CG2979	*Yp2*	Female	Female	Female
CG11129	*Yp3*	Female	Female	Female
CG4790	*fs(1)M3*	Female	Female	Female

CG2985	*Yp1*	Female	Female	Female
CG17820	*fit*	Female	Female	Female
CG9772	*CG9772*	Female	Female	Female
CG2054	*Cht2*	Female	Female	Female
CG2003	*CG2003*	Female	Female	Female
CG6806	*Lsp2*	Female	Female	Unbiased
CG14645	*CG14645*	Female	Female	Unbiased
CG4020	*CG4020*	Female	Unbiased	Female
CG31077	*CG31077*	Female	Unbiased	Unbiased
CG9743	*CG9743*	Female	Unbiased	Unbiased
CG3027	*pyd3*	Female	Unbiased	Unbiased
CG2962	*CG2962*	Female	Unbiased	Unbiased
CG14167	*Ilp3*	Female	Unbiased	Unbiased
CG9280	*Glt*	Female	Unbiased	Unbiased
CG15308	*CG15308*	Female	Unbiased	Unbiased
CG6667	*dl*	Female	Unbiased	Unbiased
CG4950	*CG4950*	Female	Unbiased	Unbiased
CG3757	*y*	Female	Unbiased	Unbiased
CG8942	*nimC1*	Female	Unbiased	Unbiased
CG4786	*Rcd2*	Female	Unbiased	Unbiased
CG3823	*CG3823*	Female	Unbiased	Unbiased
CG7002	*Hml*	Female	Male	Unbiased
CG4797	*CG4797*	Female	Male	Unbiased
CG42639	*proPO-A1*	Female	Male	Unbiased
CG33103	*Ppn*	Female	Male	Unbiased
CG6698	*NtR*	Female	Male	Unbiased
CG4099	*Sr-CI*	Female	Male	Unbiased
CG33273	*Ilp5*	Female	Male	Unbiased
CG34324	*CG34324*	Female	Male	Unbiased
CG15279	*CG15279*	Female	Male	Male
CG3906	*CG3906*	Female	Male	Male
CG14125	*CG14125*	Female	Male	Male
CG14277	*CG14277*	Female	Male	Male
CG13360	*CG13360*	Female	Male	Male
CG14456	*CG14456*	Female	Male	Male
CG15459	*CG15459*	Female	Male	Male
CG7106	*lectin-28C*	Female	Male	Male
CG34220	*CG34220*	Female	Male	NA
CG43673	*CG43673*	Female	Male	NA

[a]Data from modENCODE [31].

[b]Data from Sebida meta-analysis (v. 3) [30].

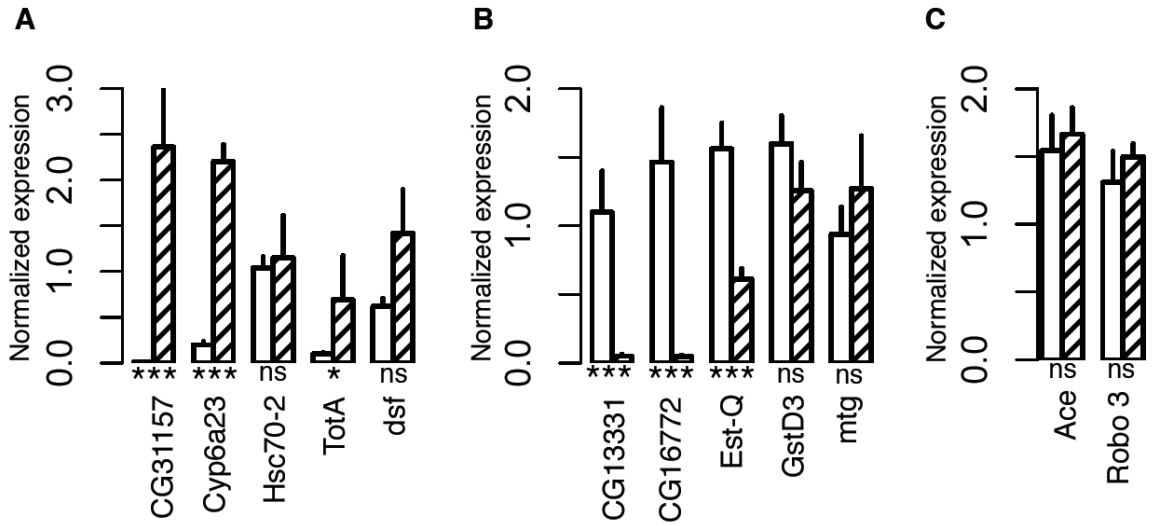
Additional file 3: Transcripts of multiple-transcript genes that differ in expression between the sexes

Transcript	Gene	Bias	Male/Female	Adj. *P*
FBtr0075364	*tra*	Female	0	1.97E-08
FBtr0081759	*dsx*	Male	4.09628443	8.14E-08
FBtr0083384	*Abd-B*	Male	7.73982604	8.14E-08
FBtr0305966	*Cht2*	Female	0.29785654	9.99E-05
FBtr0100476	*CG4293*	Male	9.84391044	0.00036357
FBtr0070982	*RpL17*	Female	0.12628794	0.00153393
FBtr0071136	*RpS6*	Female	0.24282851	0.00183333
FBtr0307287	*lawc*	Female	0.3621625	0.00537838
FBtr0071035	*inx7*	Female	0.28913877	0.0112404
FBtr0073576	*regucalcin*	Female	0.41537879	0.04654299

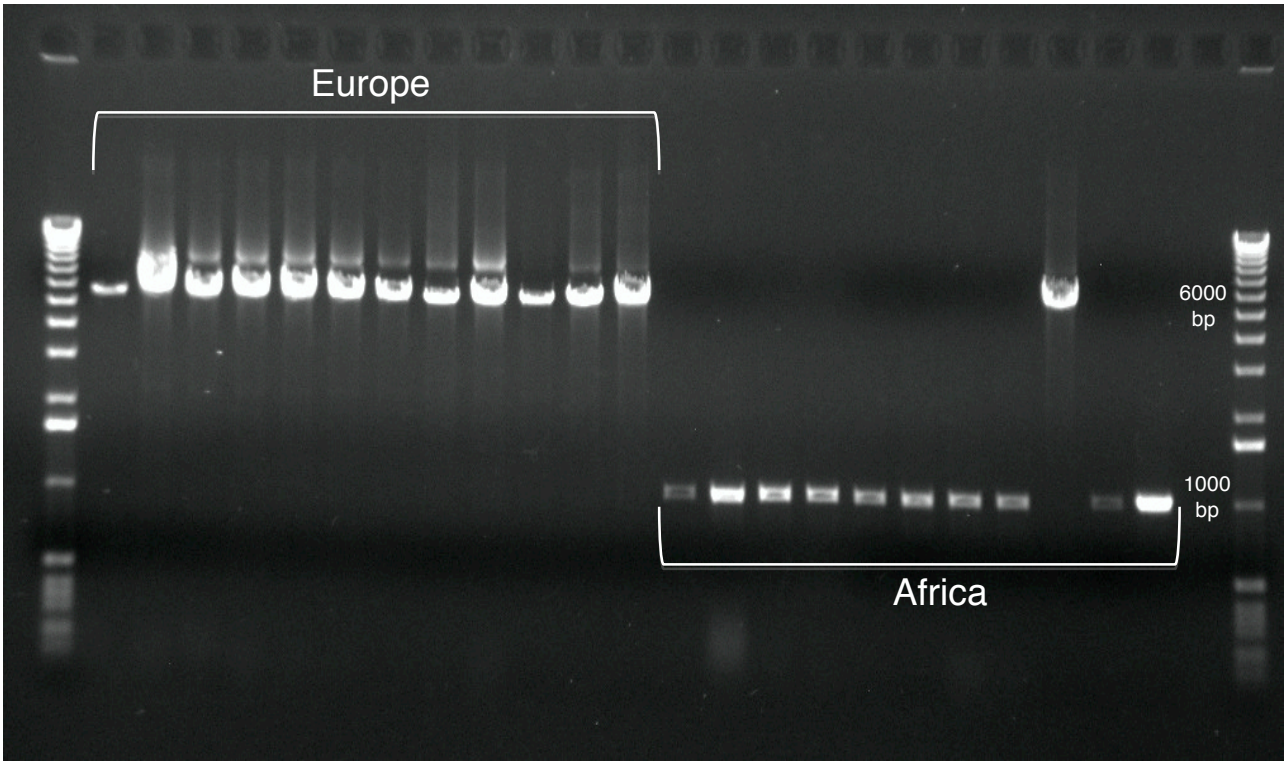
Additional file 4: Transcripts of multiple-transcript genes that differ in expression between populations

Transcript	Gene	Bias	Eur/Afr	Adj. *P*
FBtr0113212	*CG12947*	Europe	11.5055802	0
FBtr0089939	*Sod3*	Europe	39.5166675	0
FBtr0085427	*CG2010*	Africa	0.05405252	2.36E-10
FBtr0113213	*CG12947*	Europe	8.77504172	2.36E-10
FBtr0100496	*CG13430*	Europe	99.99999999	6.84E-09
FBtr0073556	*Karl*	Africa	0.04130866	3.42E-08
FBtr0303709	*CG42797*	Europe	12.46487	1.33E-07
FBtr0113353	*CHKov1*	Europe	4.30716314	2.48E-07
FBtr0073557	*Karl*	Africa	0.16611369	9.25E-07
FBtr0301854	*CG34423*	Europe	6.27349023	9.51E-07
FBtr0299720	*Ir93a*	Europe	13.2747227	1.11E-06
FBtr0305582	*Ag5r*	Europe	5.14790009	1.30E-06
FBtr0086900	*RhoGAP54D*	Europe	3.88335571	2.49E-06
FBtr0078118	*CG11455*	Europe	3.16993035	3.10E-06
FBtr0074365	*Sep4*	Europe	6.96126718	3.76E-06
FBtr0077540	*CG31955*	Europe	4.84346749	4.07E-06
FBtr0074186	*CG8974*	Africa	0.23442985	4.63E-06
FBtr0076527	*Tequila*	Europe	2.92458226	0.00010193
FBtr0299561	*Cpr62Ba*	Africa	0.14923376	0.00012116
FBtr0099994	*CG33958*	Africa	0.25373311	0.0001363
FBtr0308624	*scrib*	Europe	5.84416268	0.00029838
FBtr0306906	*Klp54D*	Europe	3.50297321	0.00031137
FBtr0306234	*lok*	Europe	99.99999999	0.00037285
FBtr0307170	*CG14985*	Europe	40.9845856	0.00118455
FBtr0084139	*CG6656*	Europe	2.88882775	0.0016121
FBtr0077538	*CG2818*	Europe	3.86592944	0.0026213
FBtr0100658	*Sod3*	Europe	2.37640694	0.00279063
FBtr0308224	*mRpL12*	Africa	0.34018592	0.00450618
FBtr0299586	*fz2*	Africa	0.40735149	0.00623311
FBtr0302847	*Gfat1*	Africa	0.30290828	0.00623311
FBtr0113351	*CHKov1*	Europe	2.51571059	0.006266
FBtr0290031	*Vha44*	Africa	0.07125104	0.00639859
FBtr0084101	*CG3301*	Europe	26.1226508	0.00893565
FBtr0307389	*fau*	Europe	3.96744765	0.00900808
FBtr0301887	*CG11455*	Europe	2.39466239	0.01082625

FBtr0077243	*Mgstl*	Europe	2.10821709	0.01093632
FBtr0112717	*Snoo*	Europe	3.47495816	0.01270455
FBtr0304846	*Octbeta2R*	Europe	2.67054076	0.01287655
FBtr0084858	*CG10550*	Europe	3.7322172	0.01406873
FBtr0100343	*sv*	Africa	0.0642603	0.01668524
FBtr0078084	*cbt*	Europe	3.51599084	0.01670706
FBtr0112927	*Trf2*	Europe	3.45847601	0.01670706
FBtr0302130	*Pde8*	Africa	0.37089452	0.01695904
FBtr0309042	*sls*	Europe	2.48208772	0.0171988
FBtr0076114	*Bmcp*	Europe	3.29602191	0.02127831
FBtr0301433	*CG31279*	Africa	0	0.02133633
FBtr0082557	*KLHL18*	Africa	0.32758103	0.02228499
FBtr0308340	*CG33521*	Europe	3.09209872	0.02529823
FBtr0074659	*bnb*	Africa	0.1721355	0.02686506
FBtr0113101	*CG10320*	Europe	2.26550063	0.03103432
FBtr0301823	*Mnt*	Europe	2.57736925	0.03247037
FBtr0088792	*pnut*	Europe	2.92656826	0.03324336
FBtr0305994	*l(2)06225*	Europe	3.16660201	0.03437594
FBtr0087106	*RpLP2*	Europe	5.5665086	0.03991185
FBtr0085847	*CG1971*	Africa	0.08854299	0.03991185
FBtr0076511	*mRpL12*	Europe	2.83526372	0.03991185
FBtr0077342	*fog*	Africa	0.3211506	0.03991185
FBtr0073947	*CG14411*	Europe	12.1139135	0.04119429
FBtr0073690	*CG3775*	Europe	11.5409369	0.04119429
FBtr0070362	*trr*	Europe	2.72418524	0.04236765
FBtr0299929	*sqa*	Africa	0.23500644	0.04525254
FBtr0079777	*CG31708*	Africa	0.21417141	0.04664426
FBtr0089712	*CG1677*	Europe	2.08470231	0.04664426



Additional file 5 - Results of qRT-PCR
The relative expression in Africa (open bars) and Europe (hatched bars) for genes tested by qRT-PCR. (A) Genes showing European over-expression in the RNA-seq analysis. (B) Genes showing African over-expression in the RNA-seq analysis. (C) Control genes showing equal expression in Europe and Africa in the RNA-seq analysis. Differences between population were tested with a Wilcoxon test. *P < 0.05, **P < 0.001, ***P < 0.0001, ns, not significant.



Additional file 6 - PCR assay for the *Doc* element insertion in *CHKov1*
The 12 European and 11 African lines used in the expression analysis were tested for the *Doc* element in *CHKov1* by PCR using primers that span the insertion site. The presence of the *Doc* element results in a PCR product of 5.5 kb, while its absence results in a product of 1 kb.

Chapter 2

An indel polymorphism in the *MtnA* 3' untranslated region is associated with gene expression variation and local adaptation in *Drosophila melanogaster*

Ana Catalán, Pablo Duchen and John Parsch

(Unpublished manuscript)

An indel polymorphism in the *MtnA* 3' untranslated region is associated with gene expression variation and local adaptation in *Drosophila melanogaster*

Ana Catalán, Pablo Duchen and John Parsch

(Unpublished manuscript)

Insertions and deletions (indels) are a major source of genetic variation within species and may result in functional changes to coding or regulatory sequences. In this study we report that an indel polymorphism in the 3' untranslated region (UTR) of the metallothionein gene *MtnA* is associated with gene expression variation in natural populations of *Drosophila melanogaster.* A derived allele of *MtnA* with a 49-bp deletion in the 3' UTR segregates at high frequency in populations outside of sub-Saharan Africa. The deletion shows a clinal distribution, with frequencies approaching 100% in northern Europe. Flies with the deletion have about 4-fold greater *MtnA* expression than flies with the ancestral sequence. Using reporter gene constructs in transgenic flies, we show that the 3' UTR deletion contributes to the observed expression difference. Population genetic analyses uncovered signatures of a selective sweep in the *MtnA* region within populations from northern Europe. These results suggest that the 3' UTR deletion has been a target of local adaptation for its ability to confer increased levels of *MtnA* expression in northern populations.

Introduction

Natural populations adapt constantly to their changing environments, with alterations in protein sequences and gene expression providing the main sources of variation upon which natural selection can act. At present, understanding how changes in gene expression contribute to adaptation is one of the major challenges in evolutionary genetics. The fruit fly *Drosophila melanogaster* has populations distributed throughout the world, with environments ranging from tropical to temperate. On the basis of biogeographical, anatomical and population genetic studies, the center of origin of *D. melanogaster* has been inferred to be in sub-Saharan Africa (David et al. 1988, Lachaise et al. 2001, Li and Stephan 2007). Several genomic studies concluded that *D. melanogaster* underwent a population expansion around 60,000 years ago that set the ground for a much broader species expansion leading to the colonization of Europe and Asia 13,000–19,000 years ago (Laurent et al. 2011, Duchén et al. 2012, Werzner et al. 2013). Because the colonization of new habitats requires that species adapt to new environmental conditions, there has been considerable interest in identifying the genetic and phenotypic changes that occurred during the out-of-Africa expansion of *D. melanogaster* (Saminadin-Peter et al. 2012, Werzner et al. 2012, Pool et al. 2012, Glaser-Schmitt et al. 2013).

In order to identify genes that differed in expression between a *D. melanogaster* population from Europe (the Netherlands) and one from sub-Saharan Africa (Zimbabwe), whole-transcriptome comparisons were carried out using adult males and females (Hutter et al. 2008, Müller et al. 2011), as well as the dissected brains of each sex (Catalán et al. 2012). These studies identified several hundred genes that were differentially expressed between the two populations and which represent candidates for adaptive regulatory evolution. One of the candidate genes that showed a large difference in expression between populations in the brains of both sexes was the metallothionein (MT) gene *Metallothionein A* (*MtnA*). *MtnA* lies on chromosome arm 3R (Figure 1) and belongs to a gene family of five members that also includes *MtnB*, *MtnC*, *MtnD* and *MtnE* (Egli et al. 2003, Pérez-Rafael et al. 2012). Metallothioneins are present in all eukaryotes and also have been identified in some prokaryotes (Guirola et al. 2011, Capdevila et al. 2012). In general, MTs are cysteine-rich proteins, a feature that makes them thermostable, and have a strong affinity to metal ions, especially Zn^{++} and Cu^{+} (Capdevila et

al. 2012). Some of the biological functions that have been described for MTs include: sequestration and dispersion of metal ions; zinc and copper homeostasis; regulation of the biosynthesis of zinc metalloproteins, enzymes and zinc dependent transcription factors; and protection against reactive oxygen species, ionizing radiation and metals (Nath 2000). In natural isolates of *D. melanogaster*, increased *MtnA* expression has been linked to copy number and indel variation and is associated with increased tolerance to heavy metals (Maroni et al. 1986, Lange et al. 1990).

In this paper we show that the expression difference of *MtnA* between a European and a sub-Saharan African population is not associated with copy number variation, but is associated with a derived 49-bp deletion in the *MtnA* 3' untranslated region (UTR). This deletion shows a latitudinal cline that spans from tropical sub-Saharan Africa to temperate northern Europe. Using transgenic reporter genes, we show that the indel polymorphism in the 3' UTR contributes to the expression difference observed between populations. Population genetic analyses indicate that *MtnA* has been the target of positive selection in non-African populations. Taken together, these results suggest that a *cis*-regulatoy polymorphism in the *MtnA* 3' UTR has undergone recent positive selection to increase *MtnA* expression in derived northern populations of *D. melanogaster*.

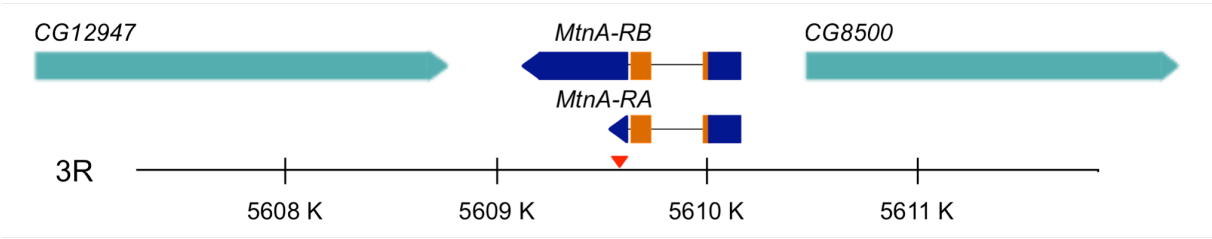


Figure 1. Structure of the *MtnA* locus.
Two transcripts that differ only in their 3' UTRs have been annotated for *MtnA* (*MtnA-RA* and *MtnA-RB*). Dark blue boxes represent the UTRs with the arrowheads indicating the direction of transcription. Orange boxes represent the coding exons. The thin lines joining the coding exons represent introns. The location of the polymorphic indel, which is shared by both transcripts, is indicated by the red triangle. For the genes flanking *MtnA* only the whole gene model is shown.

Materials and Methods

Fly strains

This study used isofemale lines from 11 populations of *D. melanogaster*, including: Zimbabwe (Lake Kariba), Zambia (Lake Kariba), Rwanda (Gikongoro), Cameroon (Oku), Egypt (Cairo), Malaysia (Kuala Lumpur), France (Lyon), Germany (Munich), the Netherlands (Leiden), Denmark (Aarhus) and Sweden (Umeå). The lines from Zimbabwe and the Netherlands were the same as those used in previous expression studies (Hutter et al. 2008, Müller et al. 2011, Catalan et al. 2012). Flies from Germany were collected locally from different locations in the greater Munich area. Flies from Denmark were kindly provided by Volker Loeschcke (Aarhus University). Flies from Sweden and Malaysia were kindly provided by Ricardo Wilches and Wolfgang Stephan (University of Munich). The remaining fly lines were collected as part of the *Drosophila* Population Genetics Project (DPGP: http://www.dpgp.org) and were kindly provided by John Pool and Charles Langley (University of California, Davis). Flies were maintained on standard cornmeal-molasses medium at a constant temperature of 22° with a 14 hour light/10 hour dark cycle.

Quantitative reverse transcription PCR (qRT-PCR)

Validation of the *MtnA* expression results obtained from brain RNA-seq data (Catalán et al. 2012) was performed by qRT-PCR using TaqMan probes (Applied Biosystems, Foster City, California, USA). For population-level comparisons, six brains were dissected from males and females of each of the 11 lines from Zimbabwe (*A84*, *A95*, *A131*, *A145*, *A157*, *A186*, *A191*, *A229*, *A377*, *A384*, *A398*) and five brains were dissected from males and females of each of the 12 lines from the Netherlands (*E01*, *E02*, *E11*, *E12*, *E13*, *E14*, *E15*, *E16*, *E17*, *E18*, *E19*, *E20*). The dissected brains of each population and sex were pooled in accordance with the RNA-seq strategy employed by Catalán et al. (2012). To compare the *MtnA* expression of individual lines within populations, subsets of eight lines were chosen from Zimbabwe (*A84*, *A95*, *A131*, *A145*, *A157*, *A186*, *A377*, *A384*) and the Netherlands (*E01*, *E02*, *E11*, *E12*, *E15*, *E16*, *E17*, *E18*). Thirty whole brains and digestive tracts (from foregut to hindgut) were dissected per line. Two biological replicates were performed for both the pooled and unpooled procedures. Tissue was dissected from flies 4–6 days old in 1X PBS (phosphate buffered saline). The tissue was stored in RNAlater (Life Technologies, Carlsbad, CA, USA) at -80° until RNA

extraction. Total RNA extraction and DNase I digestion was performed using the MasterPure RNA Purification Kit (Epicentre, Madison, WI, USA). One microgram of total RNA was reverse transcribed using random primers and SuperScript II reverse transcriptase (Life Technologies) following the manufacturer's instructions. TaqMan gene expression assays (Applied Biosystems) were used for *MtnA* (Dm12362764_s1) and *RpL32* (Dm02151827_g1). qRT-PCR was performed using a Real-Time thermal cycler CFX96 (Bio-Rad, Hercules, CA, USA). Two biological replicates, each with two technical replicates, were performed for each sample. The ΔΔCt method was used to compute the normalized expression of *MtnA* using the ribosomal protein gene *RpL32* as the reference (Pfaffl 2001).

Copy number variation assays

The paralogous genes *AttacinA* (*AttA*) and *AttacinB* (*AttB*) were used as positive controls for copy number variation (CNV) assays, because they share 97% nucleotide identity (Lazzaro and Andrew 2001) and can be co-amplified with the same primer set. The sequences for *AttA* and *AttB* were downloaded from FlyBase (Marygold et al. 2013) and aligned using the ClustalW2 algorithm implemented in SeaView (version 4) (Gouy et al. 2010). Primers were designed to the second coding exon, where the nucleotide identity of *AttA* and *AttB* is 100%. The primer sequences were as follows: forward (5'-GGTGCCTCTTTGACCAAAAC- 3') and reverse (5'-CCAGATTGTGTCTGCCATTG- 3'). The ribosomal protein gene *RpL32*, which has no report of CNV, was used as a negative control. The *RpL32* specific primers were: forward (5'-GACAATCTCCTTGCGCTTCT-3') and reverse (5'-AGCTGGAGGTCCTGCTCAT-3'). The specific primers for *MtnA* were: forward (5'-CACTTGACCATCCCATTTCC-3') and reverse (5'-GGTCTGCGGCATTCTAGGT-3'). CNV was assessed among 12 lines from the Netherlands and 11 lines from Zimbabwe. Individual DNA extractions were performed separately for three flies of each line and copy number was assessed individually for each fly. Genomic DNA was extracted using the MasterPure DNA Purification Kit (Epicentre). The assessment of CNV from genomic DNA was done with iQ SYBR Green Supermix (Bio-Rad) following the manufacturer's instructions. CNV assays were performed using a Real-Time thermal cycler CFX96 (Bio-Rad). The relative copy numbers of *MtnA* and *AttA/AttB* were obtained by the ΔCt method using *RpL32* as the reference gene.

Sequencing of the *MtnA* locus

Approximately 6 kb of the *MtnA* genomic region, spanning from the second intron of *CG12947* to the 3' UTR of *CG8500* (genome coordinates 3R: 5,606,733–5,612,630), were sequenced in 12 Dutch, 11 Zimbabwean and 12 Swedish lines (Figure 1). The following primer pairs were used (all 5' to 3'): GATGGTGGAATACCCTTTGC and AAAGCGGGTTTACCAGTGTG, GTTGGCCTGGCTTAATAACG and ACTGGCACTGGAGCTGTTTC, GCTCTTGCTAGCCATTCTGG and AGAACCCGGCATATAAACGA, GATATGCCCACACCCATACC and GTAGAGGCGCTGCATCTTGT, CACTTGACCATCCCATTTCC and CAAGTCCCCAAAGTGGAGAA, CTTGATTTTGCTGCTGACCA and ATCGCCACGATTATGATTGC, CAGGACAATCAAGCGGAAGT and TTATGAAGCGCAGCACCAGT, GACCCACTCGAATCCGTATC and TGCTTCTTGGTGTCCAGTTG. PCR products were purified with ExoSAP-IT (Affymetrix, Santa Clara, CA, USA) and sequenced using BigDye chemistry on a 3730 automated sequencer (Applied Biosystems). Trace files were edited using Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, MI, USA) and a multiple sequence alignment was generated using the ClustalW2 algorithm in SeaView (version 4) (Gouy et al. 2010).

***MtnA* indel polymorphism screening and environmental correlation study**

For individual flies of the isofemale lines described above, the presence or absence of the *MtnA* 3' UTR deletion was assessed by performing a two-step PCR reaction, which did not include an extension step, using the following primers: forward (5'-GCCGCAGACCAATTGATTA-3') and reverse (5'-CATTTGCATCCTGGAAAGAA-3'). The frequency of the deletion was estimated on an allelic basis, as heterozygous individuals were detected in some populations. Binomial 95% confidence intervals were calculated for the frequency of the deletion using the probit method implemented in R (Hornik 2013). The correlation between the frequency of the deletion and abiotic factors (latitude, minimum yearly temperature, maximum yearly temperature, minimum yearly rainfall and maximum yearly rainfall) was calculated using temperature and precipitation data from the World Meteorological Organization (www.wmo.in). A Spearman's rank test was performed to assess the strength and significance of the correlation. A least-squares regression line was fitted to the data for visualization.

Cloning and transgenesis

To test whether the indel polymorphism found in *MtnA* 3' UTR can account for the difference in expression observed between the European and the African populations, we constructed transgenic flies using the phiC31 transgenesis system (Bischof et al. 2007). Two expression vectors containing a green fluorescent protein (GFP) reporter gene were constructed using standard cloning techniques (Sambrook et al. 2001). *MtnA* 3' UTR sequences from the Netherlands (line *E20*) and Zimbabwe (line *A84*), corresponding to chromosome arm 3R coordinates 5,607,448–5,611,691, were PCR-amplified with forward (5'-TTTCCTCGAAC TTGTTCACTTG -3') and reverse (5'- GCCCGATGTGACTAGCTCTT -3') primers and cloned into the p*CR2.1-TOPO* vector (Invitrogen). The promoter region of *MtnA* (corresponding to genome coordinates 3R: 5,607,983–5,612,438), which is identical in the European and the African populations, was also PCR amplified and cloned separately into the p*CR2.1-TOPO* vector using forward (5'-GCCGCAGACCAATTGATTA-3') and reverse (5'-CATTTGCATCCTGGAAAGAA-3') primers. To generate the GFP expression construct, the *MtnA* promoter was excised with *Eco*RI and ligated into the *Eco*RI site at the 5' end of GFP in the plasmid p*RSET/EmGPP* (Invitrogen). Using *Ava*I and *Xba*I, the fragment containing the *MtnA* promoter and GFP was excised from the p*RSET/EmGPP* plasmid and ligated into the *Ava*I–*Xba*I sites proximal to the *MtnA* 3' UTR in the p*CR2.1-TOPO* vector. The whole construct (promoter + GFP + 3' UTR) was then excised with *Xba*I and *Kpn*I and ligated into the *Xba*I–*Kpn*I sites of the *pattB* integration vector (Bischof et al. 2007). PhiC31 site-specific transgenesis was used to generate flies that differed only in the presence or the absence of the 49–bp sequence in the 3' UTR of the reporter gene. Two different landing sites were chosen for transgenesis, with the lines *M{vas-int.Dm}ZH-2A, M{3xP3-RFP.attP}ZH-51D* and *M{vas-int.Dm}ZH-2A, M{3xP3-RFP.attP'}ZH-68E* being used for embryo microinjections (Bischof et al. 2007). Microinjection and screening for transformants were carried out by Fly Facility (Clermont-Ferrand Cedex, France) and Rainbow Transgenic Flies (Camarillo, CA, USA). The successfully transformed flies were crossed to a *yellow*, *white (yw)* strain for two generations to eliminate the integrase.

Reporter gene assays

The expression of the reporter gene GFP was measured in heterozygous flies generated by crossing transformant males to *yw* females. We tested for differences in the expression of GFP

in bodies and heads separately. Differences in GFP expression between lines were tested by qRT-PCR. For this, total RNA was extracted from five bodies and 10 heads of each transformant line using the RNA extraction and reverse transcription protocols described above. Thirteen biological replicates were performed for each line, each with two technical replicates. The primers used for the detection of GFP and *RpL32* were*:* forward (5'-GCTGACCCTGAAGTTCATCT -3') and reverse (5'- ATGCCCAACATCGGTTAC -3'), and forward (5'- GCTGACCCTGAAGTTCATCT -3') and reverse (5'-GCACCAGGAACTTCTTGAAT-3'), respectively. The qRT-PCR was performed with iQ SYBR Green Supermix (Bio-Rad) following the manufacturer's instructions and the samples were run using a Real-Time thermal cycler CFX96 (Bio-Rad). The data analysis was performed as described above for *MtnA* gene expression. A *t*-test was performed to assess significance.

Brain confocal imaging

Brain tissue was dissected in ice-cold 1X PBS and fixed with PLP (8% paraformaldehyde in NaOH and PBS with lysine (l)-HCl) for 1 h at room temperature as described in Cayirlioglu et al. (2008). After fixation the tissue was washed twice for 15 min with PBS-0.5% Triton X and then incubated for 1 hour in blocking solution (20% donkey serum, 0.5% Triton X in PBS) at room temperature. The primary antibody, mouse anti-disclarge (Developmental Studies Hybridoma Bank, University of Iowa, USA) was used at a 1:200 dilution and incubated overnight at 4° in blocking solution. After washing twice with PBS-0.5% Triton X, the tissue was incubated with the secondary antibodiy, 1:200 anti-rat-CY3 (Dianova, Hamburg, Germany). The brains were mounted in Vectashield mounting medium (Vector Laboratories, Burlingame, CA, USA) and scanned using confocal microscopy with a Leica SP5-2. The images were analyzed using the StackGroom plugin in ImageJ (Schneider et al. 2012).

Population genetic analysis and tests for selection

Summary statistics, including the number of segregating sites (*S*), number of haplotypes and Tajima's *D* (Tajima 1989) were calculated using DnaSP v.5.10.1 (Rozas 2009). The mean pairwise nucleotide diversity (π) (Tajima 1983), Watterson's estimate of nucleotide diversity (θ) and F_{st} (Hudson et al. 1992) were calculated as described in Duchen et al. (2012). To test for a

selective sweep, a *SweepFinder* analysis was performed using the *SweeD* software (Pavlidis et al. 2013). The background site frequency spectrum (SFS) was calculated for the entire 3R chromosome arm using 11 whole genome sequences from the Netherlands population and one whole genome sequence from the French (Lyon) population (http://www.dpgp.org). The French sequence was included in order to have a constant sample size of 12 sequences for the calculation of the SFS. This approach did not bias the background, as the French sequence did not differ more from the Netherlands sequences than the Netherlands sequences did from each other (Table S1, Figure S1a). Furthermore, the inclusion of a French line did not lead to a skew in the background SFS (Figure S1b). For the Swedish population, the background SFS of chromosome arm 3R was determined from 12 whole genome sequences from the Umeå population (PD, unpublished data). In order to increase the power of the test, the invariant sites in the alignment were also included (Nielsen 2005). To assess the significance of the composite likelihood ratio (CLR) statistic, neutral simulations were performed using *ms* (Hudson 2002). In the neutral simulations two demographic models were taken into account (Duchen et al. 2012, Werzner et al. 2013). These models differ in several parameters, including: the timing of the out-of-Africa bottleneck, the current effective population sizes of the European and African populations, and the ancient demographic history of the African population (expansion *versus* bottleneck). For our analyses, it is the estimated time of the out-of-Africa bottleneck that has the largest impact on the results. Duchen et al. (2012) infer this bottleneck to have occurred around 19,000 years ago, while Werzner et al. (2013) place it at around 13,000 years ago. However, the 95% confidence intervals of both estimates are very wide, ranging from 7,359–24,953 years ago. Thus, the two estimates are not incompatible with each other. The recombination rate of the *MtnA* genomic region was obtained from the *D. melanogaster* recombination rate calculator (Fiston-Lavier et al. 2010). A total of 10,000 simulations were performed. For each simulation, the maximum value of the CLR statistic was extracted and used to determine the 5% significance threshold.

Results

Differential expression of *MtnA* between an African and a European population of *D. melanogaster*.

A previous RNA-seq study of gene expression in the brain found *MtnA* to have four times higher expression in a European population (the Netherlands) than in a sub-Saharan African population (Zimbabwe) (Catalán et al. 2012). Of the members of the *Mtn* gene family, only *MtnA* showed high levels of expression and a significant difference in expression between populations (Figure 2A). To confirm this expression difference, we performed qRT-PCR on RNA extracted from dissected brains of flies from each population following the same pooling strategy used by Catalán et al. (2012). With this approach, we found *MtnA* to have 5-fold higher expression in the European population than in the African population (Figure 2B).

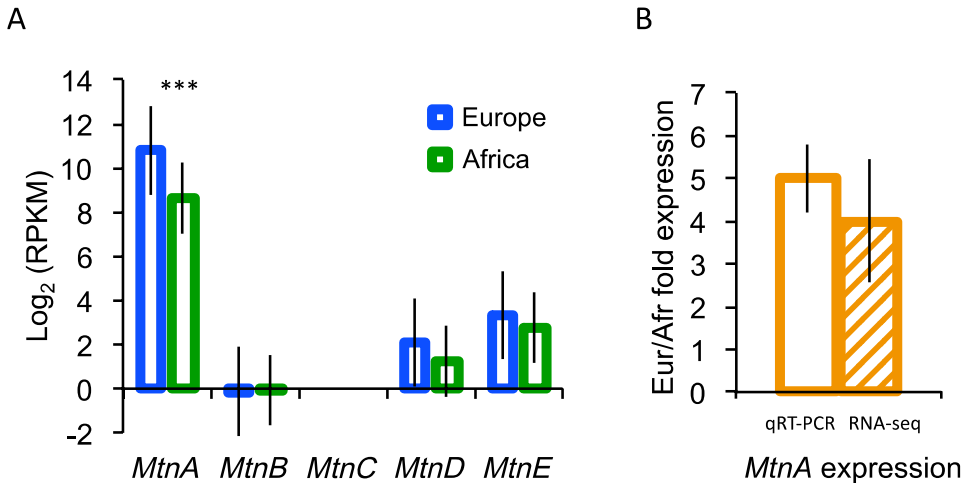


Figure 2. Expression of metallothionein genes in the brain in two populations of *D. melanogaster*. (A) Expression level of *Mtn* paralogs in the brain from RNA-seq data. Expression is reported in terms of reads per kilobase per million mapped reads (RPKM). Only *MtnA* showed a significant difference in expression between populations (adjusted $P < 10^{-7}$ in the RNA-seq analysis of Catalán et al. 2012). Expression of *MtnC* was not detected. (B) *MtnA* expression in the brains of flies from a European (the Netherlands) and an African (Zimbabwe) population, as determined by qRT-PCR. The expression difference between populations is highly significant (*t*-test, $P = 5 \times 10^{-5}$). In both panels, the error bars indicate the standard deviation among replicates.

The above RNA-seq and qRT-PCR analyses were performed on a "per gene" basis and did not discriminate between the two annotated transcripts of *MtnA*, which differ only in the length of their 3' UTR (Figure 1). The *MtnA-RA* transcript completely overlaps with that of *MtnA-RB* and contains no unique sequence. The *MtnA-RB* transcript, however, contains and extra 371 bp at the 3' end that can be used to assess isoform-specific expression. Using the RNA-seq data of Catalan et al. (2012), we found that the *MtnA-RB* isoform represents only a small proportion of total *MtnA* expression (1.50% in the European population and 0.13% in the African population). Thus, almost all of the observed expression difference in *MtnA* can be attributed to the *MtnA-RA* isoform. Although present at very low levels, the *MtnA-RB* transcript showed much higher expression (50-fold) in Europe than in Africa (Table 1).

Table 1. Isoform-specific expression of *MtnA* in the brain.

	Expression (RPKM)		
Population	*MtnA* shared (329 bp)	*MtnA-RB* specific (371 bp)	*% MtnA-RB*
Europe	3867.74	57.94	1.50
Africa	859.74	1.12	0.13

Data from Catalan et al. (2012)

Absence of *MtnA* copy number variation

Previous studies found copy number variation (CNV) for *MtnA* in natural isolates of *D. melanogaster* and showed that an increase in copy number was associated with higher *MtnA* expression (Maroni et al. 1987, Lange et al. 1990). To determine if CNV could explain the observed expression difference between the European and the African populations, assayed *MtnA* copy number in flies of both populations by quantitative PCR (qPCR). We found no evidence for CNV within or between the populations (Figure 3). In both populations, *MtnA* copy number was equal to that of the control single-copy gene *RpL32* and was about half that of the nearly-identical paralogs *AttA* and *AttB* (Lazzaro and Andrew 2001), which can be co-amplified by the same PCR primers and serve as a positive control for CNV. These results indicate that CNV cannot account for the observed variation in *MtnA* gene expression.

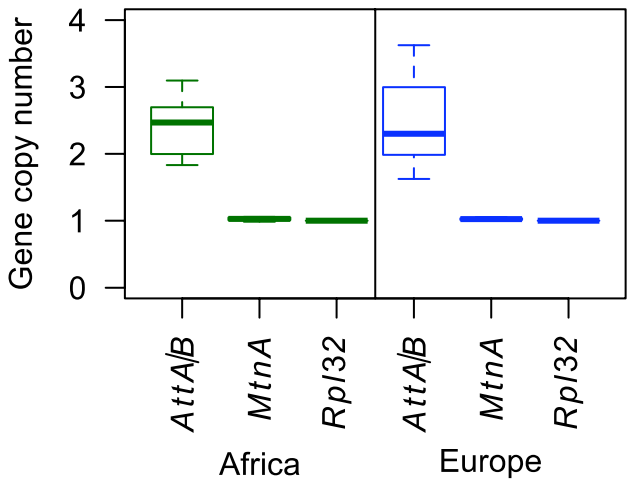


Figure 3. Results of copy number variation (CNV) assays.
Flies from Africa (Zimbabwe) and Europe (the Netherlands) were tested for *MtnA* CNV. The close paralogs *AttA* and *AttB* were used as a positive control for multiple gene copies, while *RpL32* was used as a single-copy reference.

An indel polymorphism in the *MtnA* 3' UTR is associated with expression variation

To identify *cis*-regulatory variants that might be responsible for the difference in *MtnA* expression between European and African flies, we sequenced a 6-kb region encompassing the *MtnA* transcriptional unit (figure 1) in 12 lines from the Netherlands and 11 lines from Zimbabwe. In addition, we quantified *MtnA* expression in a subset of eight lines from each population in both the brain and the gut by qRT-PCR. Across the 6-kb region, only a polymorphic 49-bp indel and a linked single nucleotide polymorphism (SNP) in the *MtnA* 3' UTR showed a large difference in frequency between the populations (Figure 4A). This indel was previously observed to segregate in natural populations from North America (Lange et al. 1990). A comparison with three outgroup species (*D. sechellia*, *D. simulans*, and *D. yakuba*) indicated that the deletion was the derived variant. This deletion was present in 10 of the 12 European lines, but was absent in Africa (Figure 4A). The qRT-PCR data revealed that the two European lines that lacked the deletion had *MtnA* expression that was similar to that of the African lines, but much lower than the other European lines. This result held for both brain and gut (Figure 4B) expression. Taken together, these results suggest that the 3' UTR polymorphism contributes to *MtnA* expression variation in natural populations. Furthermore, the expression variation is not limited to the brain, but shows a correlated response in at least one other tissue.

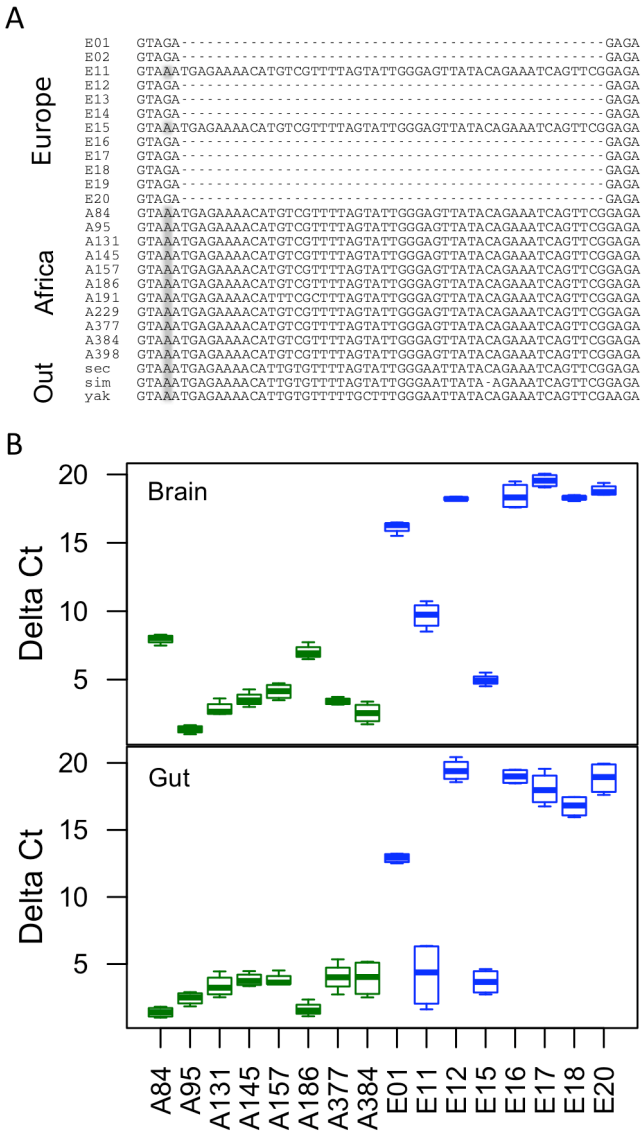


Figure 4. Association between an indel polymorphism in the *MtnA* 3' UTR and gene expression variation.

(A) An indel (and a linked SNP marked in grey) in the *MtnA* 3' UTR are the only polymorphisms within the 6-kb *MtnA* region that show a large difference in frequency between an African and a European population of *D. melanogaster*. A comparisons with three outgroup species, *D. sechellia* (sec), *D. simulans* (sim) and *D. yakuba* (yak), indicated that the deletion is the derived variant. (B) *MtnA* expression in the brain and the gut of eight European and eight African lines. The two European lines lacking the deletion, *E11* and *E15*, show lower *MtnA* expression than those with the deletion.

Functional test of the effect of the *MtnA* 3' UTR polymorphism on gene expression

To test if the 49-bp deletion in the *MtnA* 3' UTR has an effect on gene expression, we designed expression constructs in which the *MtnA* promoter was placed upstream of a GFP reporter gene. Two versions of the reporter gene were made, one with the ancestral *MtnA* 3' UTR sequence and one with the derived *MtnA* 3' UTR sequence, which has the 49-bp deletion (Figure 5A). The reporter genes were then introduced into the *D. melanogaster* genome by PhiC31 site-specific integration (Groth et al. 2004, Bischof et al. 2007).

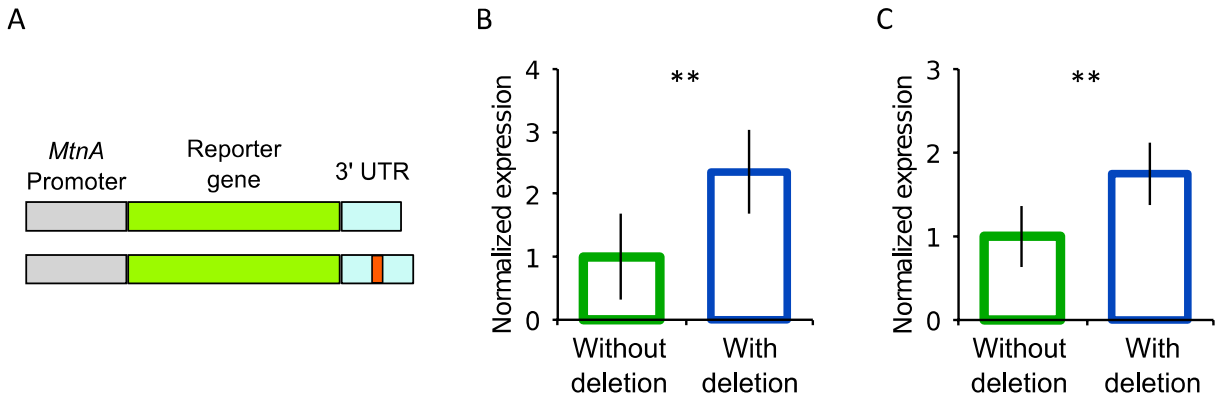


Figure 5. Schematic diagram of the reporter gene constructs and expression of GFP reporter genes fused to *MtnA* 3' UTR variants.

(A) The grey boxes represent the *MtnA* promoter. The green boxes represent the GFP reporter gene. The light blue boxes represent the *MtnA* 3' UTR, with the portion marked in red indicating the 49-bp sequence present at high frequency in the African population, but low frequency in the European population. (B) The two versions of the reporter gene differ in head (*t*-test, $P = 0.0019$) and (C) bodies (*t*-test, $P = 0.0046$). Expression was assayed by qRT-PCR.

Our analysis of *MtnA* expression in the brain and gut indicated that the difference in expression observed between African and European populations is not brain specific (Figure 4B). For this reason, GFP expression was measured in both heads and bodies of transformed flies. In both head and body, expression was significantly higher when the reporter gene had the 49-bp deletion in the 3' UTR (Figure 5B), with the difference in expression being about 2.3-fold in heads and 1.75 in bodies.

MtnA expression in the brain

MtnA shows a high expression in most *D. melanogaster* organs, including the fat body, digestive system, Malpighian tubules and brain (Chintapalli et al. 2010). Although it has been documented that *MtnA* and its paralogs are involved in heavy metal metabolism and tolerance, it is poorly understood which other functions *MtnA* might have and in which cells it is expressed. To get a more detailed picture of *MtnA* expression in the brain, we examined GFP expression in our transfromant flies (Figure 6).

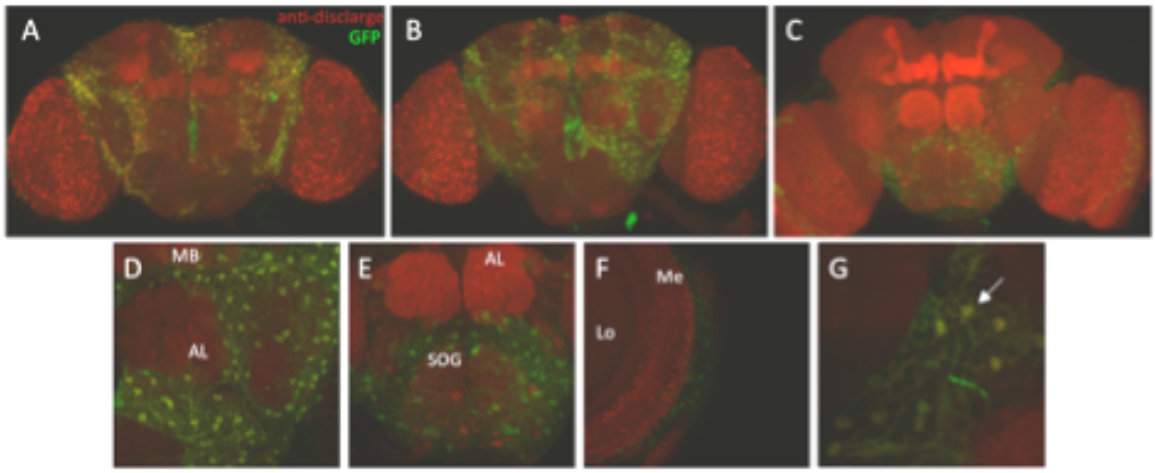


Figure 6. Expression of a *MtnA*-GFP reporter gene in the brain.
(A-C) GFP expression driven by the reporter gene construct with the ancestral *MtnA* 3' UTR variant. (D-G) Higher magnification of the brain regions where GFP is expressed. AL: antennal lobe, MB: mushroom bodies, SOG: subesophageal ganglion, Lo: lobula, Me: medulla. In (G) the arrow indicates cells expressing GFP. Green: GFP, red: anti-disclarge.

GFP expression driven by the *MtnA* promoter is evident in cells that form a mesh-like structure surrounding the brain and in between the neuropiles (Figure 6). *MtnA* does not appear to be expressed in neurons, as the cells expressing GFP do not have dendrites or axonal processes. The shape and localization of the cells expressing GFP in the brain suggest that they might be glia cells. Glia cells provide neurons with developmental, structural and trophic support as well as with protection against toxic elements (Hartenstein et al. 2008, Edwards et al. 2010, Hartenstein 2011). The expression of *MtnA* in glia cells serve as neuronal protection against oxidative stress that can be produced by contact with xenobiotics (Egli et al. 2006, Gruenewald et al. 2009). To determine if *MtnA* expression is specific to glial cells of the brain, it will be necessary to perform additional experiments with antibodies that target glia

cells, such as anti *reversed polarity* (*repo*) or anti *glia cell missing* (*gcm*). Without such data it cannot be ruled out that other types of brain cells express *MntA.* Nonetheless, our results provide direct evidence that *MtnA* is expressed in cell types other than the copper cells of the midgut and Malpighian tubules, as previously reported by Egli et al (2006).

Frequency of the *MtnA* 3' UTR deletion in additional populations

To better characterize the geographical distribution of the indel polymorphism in the *MtnA* 3' UTR, we used a PCR-based assay to screen six additional *D. melanogaster* populations across a latitudinal range spanning from tropical sub-Saharan Africa to northern Europe (Table 2). We found that the deletion was at very low frequency in sub-Saharan Africa, but nearly fixed in population from northern Europe. There was a highly significant correlation between the frequency of the *MtnA* deletion and latitude (Spearman's rank correlation, $P = 2 \times 10^{-16}$) (Figure 7A).

Table 2. Frequency of the *MtnA* 3' UTR deletion in different populations of *D melanogaster*.

Population	*N*	Latitude	Frequency of deletion [95% CI]
Sweden	12	63.83	1.00 [0.857–1.000]
Denmark	12	55.68	0.96 [0.797–0.995]
The Netherlands	12	52.16	0.83 [0.644–0.941]
Germany	11	48.13	0.91 [0.726–0.980]
France	12	45.77	0.92 [0.745–0.982]
Egypt	14	30.05	0.60 [0.421–0.771]
Cameroon	6	6.25	0.00 [0.000–0.264]
Malaysia	12	3.14	0.45 [0.271–0.654]
Rwanda	12	-2.49	0.08 [0.017–0.254]
Zambia	10	-16.54	0.05 [0.005–0.236]
Zimbabwe	11	-17.28	0.00 [0.000–0.154]

N, number of lines. Because the deletion was polymorphic in some lines, its frequency was calculated on the basis of two alleles per line.

The frequency of the *MtnA* deletion was also significantly correlated with yearly mean minimum and maximum temperatures ($P = 0.0051$ and $P = 0.0047$, respectively) (Figure 7B). There was also a weaker, but still significant, correlation between the frequency of the deletion

and the yearly maximum rainfall ($P = 0.0076$). The correlation between deletion frequency and yearly minimum rainfall was not significant ($P = 0.094$).

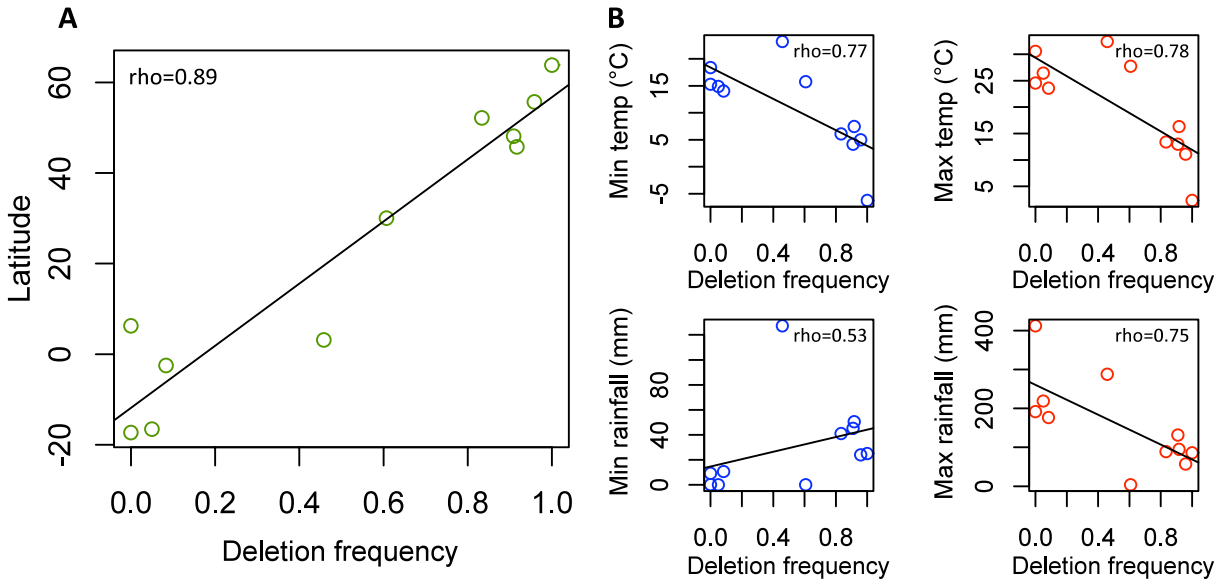


Figure 7. Correlation between the frequency of the *MtnA* 3' UTR deletion and abiotic factors. (A) Correlation between the deletion frequency and latitude. (B) Correlation between the deletion frequency and mean minimum yearly temperature, mean maximum yearly temperature, mean minimum yearly rainfall and mean maximum yearly rainfall. A least-squares linear regression line is shown for each graph. The rho value indicates the Spearman correlation coefficient. All correlations were significant ($P < 0.01$), except for minimum yearly rainfall ($P = 0.09$)

Evidence for positive selection at the *MtnA* locus

To test for a history of positive selection at the *MtnA* locus, we performed a population genetic analysis of the 6-kb *MtnA* region in the original European (the Netherlands) and African (Zimbabwe) populations. In addition, we sequenced this region in 12 lines of a Swedish population, in which the 49-bp 3' UTR deletion was at a frequency of 100% (Table 2). Across the entire region, the Zimbabwean population showed the highest nucleotide diversity, having 1.43- and 2.50-fold higher values of π than the Dutch and Swedish populations, respectively (Table 3). Tajima's *D* was negative in all three populations, and was significantly negative in both Zimbabwe and the Netherlands (Table 3). This could reflect a history of past positive or negative selection at this locus, but could also be caused by demographic factors, such as population expansion.

Table 3. Summary statistics for the *MtnA* locus.

Population	***n***	***S***	***θ***	***π***	***TajD***	***nHap***
Zimbabwe	11	54	0.312	0.194	-1.89*	11
The Netherlands	12	41	0.231	0.138	-1.85*	11
Sweden	12	17	0.096	0.078	-0.83	9

n, number of sequences; *S*, number of segregating sites; θ, Watterson's (Watterson 1975) estimate of nucleotide diversity (per 100 sites); π, mean pairwise nucleotide diversity (per 100 sites) (Tajima 1983); *TajD*, Tajima's *D* (Tajima 1989); *nHap*, number of haplotypes.

* $P < 0.05$

A sliding window analysis was performed to determine the distribution of nucleotide diversity (θ) (Figure 8A) and population differentiation (F_{st}) (Figure 8B) across the *MtnA* region. The region flanking the 3' UTR indel polymorphism showed very low sequence variation in Zimbabwe and Sweden, but higher variation in the Netherlands. This pattern is due to the fact that the ancestral state of the indel polymorphism is fixed in the Zimbabwean population and the derived state is fixed in the Swedish population. In the Dutch population, the *MtnA* 3' UTR is polymorphic (two of the 12 lines have the ancestral state), leading to an increase in nucleotide diversity. On average, Sweden and Zimbabwe showed the greatest population differentiation, with F_{st} reaching a peak in the 3' UTR of *MtnA* (Figure 8B). The Dutch population showed greater differentiation with the Zimbabwean population than with the Swedish population (Figure 8B). Across the 6-kb region, values of F_{st} were low for the comparison of the Dutch and Swedish populations, indicating that there is very little differentiation between them.

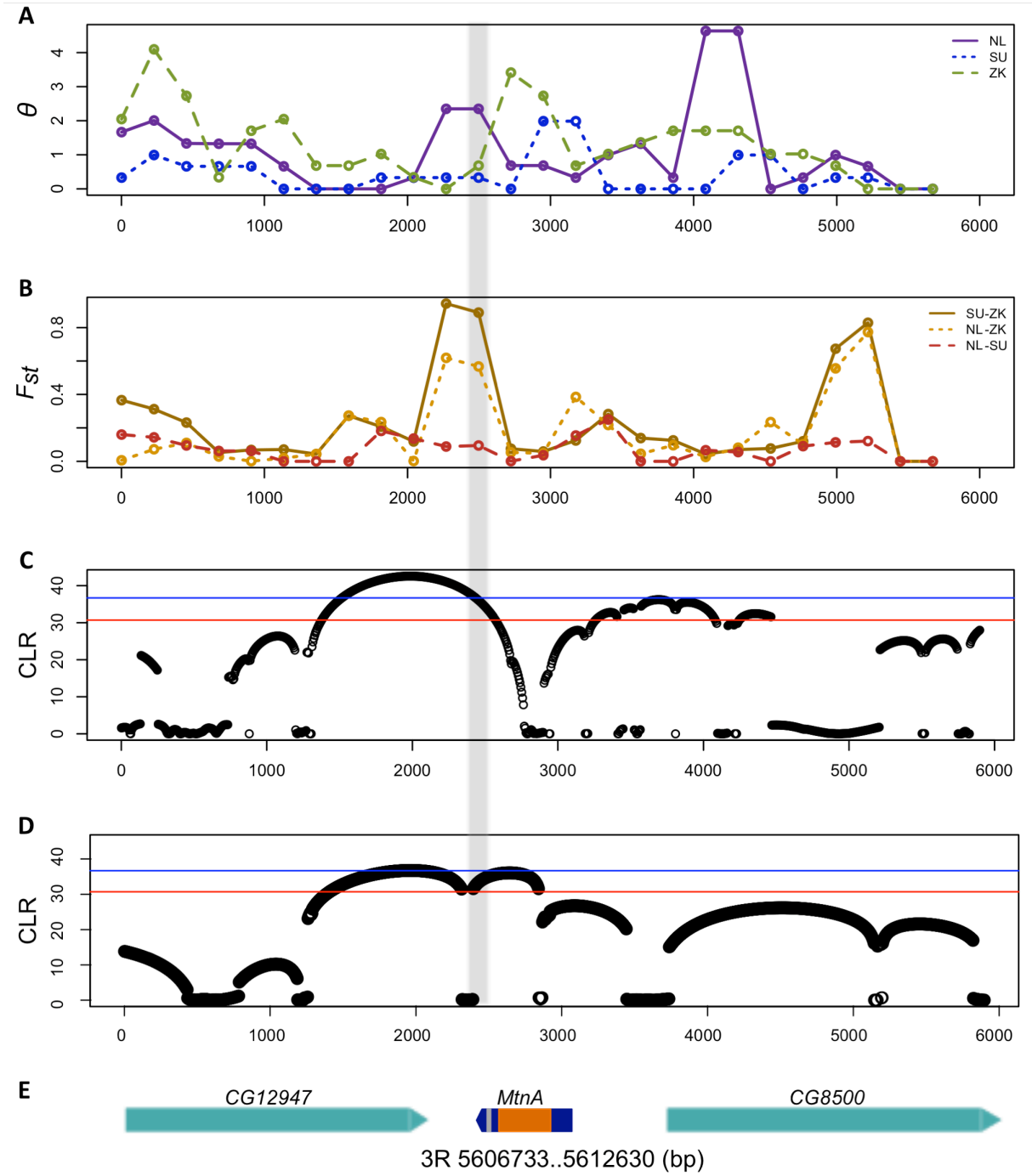


Figure 8. Evidence for positive selection at the *MtnA* locus.

(A) Watterson's θ of *D. melanogaster* populations from Zimbabwe (ZK), the Netherlands (NL) and Sweden (SU) calculated in sliding windows of 500 bp with a step size of 250 bp. (B) F_{st} values for pairwise comparisons of ZK, NL and SU calculated in sliding windows of 500 bp with a step size of 250 bp. (C) Selective sweep *(SweepFinder)* analysis of the Netherlands population showing the CLR statistic in sliding windows of 1000 bp. (D) Selective sweep (*SweepFinder*) analysis of the Swedish population showing the CLR statistic in sliding windows of 1000 bp. The blue line indicates the 5%

significance threshold calculated using the demographic model of Werzner et al. (2012) for neutral simulations. The red line indicates the 5% cutoff calculated using the demographic model of Duchen et al. (2012) for neutral simulations. (E) Gene models for the 6-kb region analyzed. The grey highlighted region indicates the position of the 49-bp indel polymorphism in the *MtnA* 3' UTR.

If positive selection has favored the derived *MtnA* allele (with the 49-bp 3' UTR deletion) in northern populations, then in this region of the genome one would expect there to be less variation among chromosomes containing the deletion than among those with the ancestral form of the allele. Indeed, this is what we observe in the Netherlands, where both alleles are segregating. Across the 6-kb region, there are 41 segregating sites within the Dutch population (Table 3). Among the 10 chromosomes with the deletion, there are 18 segregating sites, while between the two chromosomes lacking the deletion there are 23 segregating sites. This indicates that chromosomes with the deletion, which are in high frequency, shared a much more recent common ancestor. To test if this pattern differs from that expected under neutral evolution, we performed Hudson's haplotype test (HHT) (Hudson et al. 1993) using two different demographic models of the *D. melanogaster* out-of-Africa bottleneck for neutral simulations (Duchen et al. 2012; Werzner et al. 2013). Under the model of Werzner et al. (2013), HHT was significant ($P = 0.031$). Under the model of Duchen et al. (2012), HHT was marginally significant ($P = 0.094$). These results suggest that neutral evolution and demography are unlikely to explain the observed patterns of DNA sequence variation.

To further test if the *MtnA* locus has experienced recent positive selection in northern Europe, we used the composite likelihood ratio (CLR) test to calculate the likelihood of a selective sweep at a given position in the genome, taking into account the recombination rate, the effective population size, and the selection coefficient of the selected mutation (Nielsen et al. 2005, Pavlidis et al. 2012). Within the Dutch population, the CLR statistic shows a peak in the region just adjacent to the *MtnA* 3' UTR deletion (Figure 8C). This peak is significant when the demographic models of both Duchen et al. (2012) and Werzner et al. (2013) are used for neutral simulations, which provides compelling evidence for a recent selective sweep at the *MtnA* locus in the Netherlands population. A similar result was obtained for the Swedish population (Figure 8D), where the CLR statistic was above the 5% significance threshold determined from both the Duchen et al. (2012) and the Werzner et al. (2013) bottleneck

models, suggesting that the selective sweep was not limited to a single population, but instead affected multiple European populations.

Discussion

Differential expression of *MtnA* between a European and an African population of *D. melanogaster* was first detected in a brain-specific RNA-seq analysis (Catalán et al. 2012). In the present study, we confirm this inter-population expression difference by qRT-PCR and show that it is associated with an indel polymorphism in the *MtnA* 3' UTR. We also perform reporter gene experiments to demonstrate that a large proportion of the expression difference can be attributed to variation in the 3' UTR. The ancestral state of the 3' UTR contains a 49-bp sequence that is deleted in a derived allele that segregates in worldwide populations. The deletion is nearly absent from sub-Saharan Africa, but present in frequencies >80% in northern Europe (Table 2). The deletion is present at intermediate frequency in Egypt (60%) and Malaysia (45%). These findings suggest that positive selection has favored the 3' UTR deletion, at least within northern European populations. This interpretation in supported by population genetic analyses that indicate a recent selective sweep at the *MtnA* locus in populations from the Netherlands and Sweden (Figure 8).

The maintenance of the *MtnA* 3' UTR deletion at intermediate frequency in Egypt and Malaysia could be explained by four possible scenarios: relaxation of selective constraint, migration between derived and ancestral populations, population sub-structure, or balancing selection. In the case of the Malaysian population, an analysis of multiple loci from the X and third chromosome found that it has lower nucleotide diversity than both the Dutch and the Zimbabwean populations and shows no evidence of population sub-structure (Laurent et al. 2011). This contrasts with what is observed in the 3' UTR of *MtnA*, indicating that the pattern observed at *MtnA* is not shared by the rest of the genome and, thus, may be better explained by balancing selection.

The significant correlation between the frequency of the 3' UTR deletion and abiotic factors such as latitude, temperature and rainfall (Figure 7) suggests that the *MtnA* expression polymorphism could play a role in adaptation to the abiotic environment. However, all of

these factors are correlated with each other and, presumably, with many other unobserved factors, making it difficult to draw a conclusion about causation. Furthermore, abiotic factors determine how ecosystems evolve, including the evolution of biotic interactions. Thus, the correlations with the abiotic factors mentioned above could be a more general indication of adaptation to novel environments. It has previously been shown that increased expression of *MtnA* is associated with increased tolerance to heavy metals (Maroni et al. 1986). Since environmental contamination with heavy metals (e.g. copper, cadmium, zinc, or lead) is highest in industrialized regions, it may be that the high-expression allele of *MtnA* is favored in northern Europe, where there is a history of industrial pollution. However, a direct association between *MtnA* genetic variation and local heavy metal contamination has not been found (Lange et al. 1991).

At present, the mechanism by which the 3' UTR deletion affects *MtnA* gene expression is unknown. Although the deletion appears to have an effect on the usage of the *MtnA-RB* transcript isoform (Table 1), this isoform is too rare (<2% of all *MtnA* transcripts) to account for the observed 4-fold difference in *MtnA* expression. Another possibility is that the deleted 3' UTR region contains one or more binding sites for a microRNA (miRNA). miRNAs are short, non-coding RNAs that modulate the expression of genes by inhibiting transcription or inducing mRNA degradation (Chen et al. 2007). They are known to bind to a seed region that consists of 6–8 nucleotides in the 3' UTR of their target mRNA. Post-transcriptional gene expression regulation by miRNAs can result in the fine-tuned regulation of a specific transcript or can cause the complete silencing of a gene in a particular tissue or developmental stage (Berezikov 2011, Chen et al. 2007 and Flynt et al. 2008). To identify miRNAs that might bind specifically to the 49-bp sequence present in the ancestral form of the *MtnA* 3' UTR, we used the UTR predictor developed by Kertesz et al. (2007). The UTR predictor takes into account the three-dimensional structure of the miRNA and the 3' UTR, as well as the energetic stability of the miRNA-3' UTR base-pair binding. The score given by the UTR predictor is an energetic score, with the most negative scores indicating the most probable interactions. Our analysis of the *MtnA* 3' UTR identified five candidate miRNAs with scores below -6 that had predicted binding sites overlapping the 49-bp indel region (Table 4). These candidates should serve as a good starting point for future functional tests of putative miRNA-3' UTR interactions.

Table 4. Top scoring microRNAs predicted to bind within the polymorphic 49-bp sequence in the *MtnA* 3' UTR.

microRNA	Binding position	binding sites	Seed	ddG
dme-miR-284	52	1	8:0:1	-12.68
dme-miR-954	102	1	8:1:0	-10.61
dme-miR-956	43	1	8:1:1	-6.39
dme-miR-9c	74	1	8:1:1	-6.13
bantam	52	1	8:1:1	-6.13

The binding position coordinate indicates the distance between the start of the 3' UTR and first miRNA binding site. The notation describing the seed (X:Y:Z) represents the size of the seed (X), the number of mismatches (Y) and the number of G:U wobble pairs (Z). ddG is the energetic score for the probability and stability of the binding. The more negative the score is, the more probable is the interaction between the 3' UTR and the microRNA.

Genetic variation provides the substrate upon which natural selection acts, resulting in an increase in the frequency of alleles that are beneficial in a given environment. Because changes in gene expression, especially those caused by variation in *cis*-regulatory elements, are predicted to have fewer pleiotropic effects than changes occurring within coding regions, it has been proposed that they are the most frequent targets of positive selection (Carroll 2000, Wray 2007, Carroll 2008). In contrast to structural changes in protein sequences, changes in gene expression can be specific to a particular a tissue or developmental stage. Our results indicate that the observed variation in *MtnA* expression is not specific to the brain, as a correlated expression pattern is also seen in the gut (Figure 4). This suggests that the 3' UTR deletion has a general effect on *MtnA* expression, which is present at high levels in almost all organs of *D. melanogaster* (Chintapalli et al. 2007). However, tissue-specific effects of the difference in *MtnA* expression cannot be ruled out. For example, it is possible that *MtnA* has a particularly important role in the response to reactive oxygen species, which might be more important in the brain than in other organs.

Supplementary materials

Table S1. Average pairwise differences per kb between French (FR) and Dutch (NL) lines.

	FR151	FR180	FR207	FR217	FR229	FR310	FR361	NL11	NL12	NL13	NL14	NL15	NL16	NL17	NL18	NL19	NL1	NL2
FR14	4.1	5.6	4.1	6.1	1.8	4.3	5.9	5.2	4.4	6.1	4.2	4.9	3.3	4.7	4.8	5.3	4.7	5.1
FR151		5.7	4.1	6.3	2.1	4.4	5.9	5.4	4.5	6.1	4.2	5.1	3.4	4.8	4.8	5.3	4.8	5.1
FR180			5.6	6.2	2	5.7	4.7	6.5	5.8	5.1	5.7	6.3	4.7	6	6.1	6.6	6	6.3
FR207				6.2	1.7	4.4	6	5.3	4.5	6	4.3	5.1	3.4	4.7	4.8	5.2	4.7	5.1
FR217					2.2	6.3	6.2	7	6.4	6.2	6.1	6.7	5.2	6.5	6.7	7.1	6.6	6.9
FR229						2.2	2.5	2.4	2.1	2.1	1.2	2	1.1	2.1	2.1	2.2	2.3	2.1
FR310							6	5.3	4.7	6.2	4.5	5.3	3.6	4.8	3.1	5.4	4.9	5.3
FR361								6.8	6.1	5.1	5.9	6.6	5	6.3	6.4	6.9	6.3	6.6
NL11									5.4	7	5.2	6	4.4	5.6	5.8	6.2	5.7	6.1
NL12										6.3	4.3	5.2	3.5	4.8	5.1	5.4	4.9	5.1
NL13											6	6.8	5.2	6.5	6.5	7.1	6.5	6.8
NL14												4.6	3.2	4.5	4.9	4.9	4.6	4.9
NL15													2.1	5.3	5.7	5.8	5.1	5.7
NL16														3.5	4.1	4.4	3.7	4.1
NL17															5.3	5.7	5.3	5.2
NL18																5.8	5.4	5.7
NL19																	5.7	6
NL1																		3.7

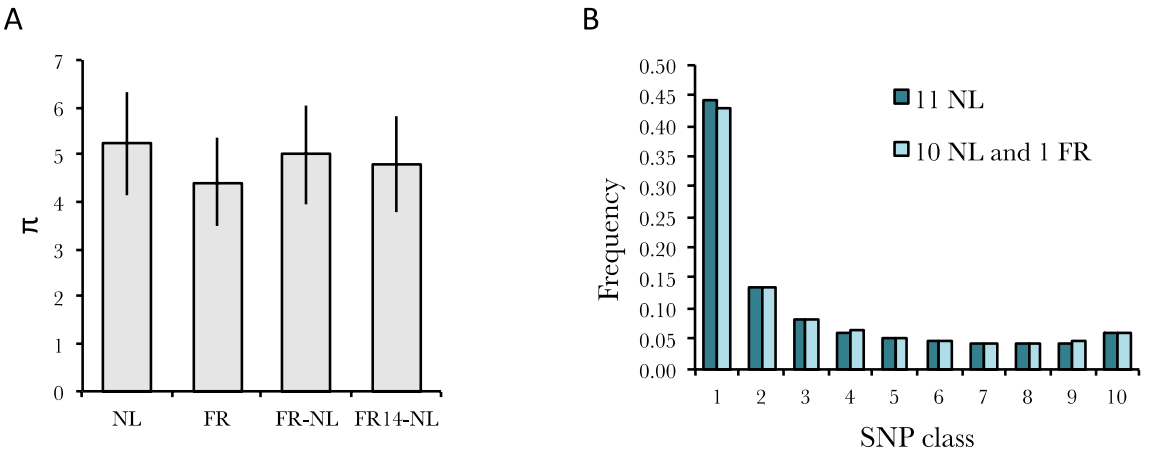


Figure S1. Nucleotide diversity (π) and site frequency spectrum (SFS) of chromosome arm 3R. (A) Nucleotide diversity (π) for 11 lines from the Netherlands (NL), eight lines from France (FR), all the Dutch and French lines combined (FR-NL), and the French line FR14 combined with 11 lines from the Netherlands (FR14-NL). (B) Dark blue bars indicate the SFS for the 11 Dutch lines for which complete genome sequences are available. Light blue bars indicate the SFS of 10 of these Netherlands lines plus one French line. In order to have a constant sample size of 12 for the SweepFinder analysis, one French line (FR14) was include with the NL lines to calculate the background site frequency spectrum.

Chapter 3

Adaptive divergence of a transcriptional enhancer between populations of *Drosophila melanogaster*

Amanda Glaser-Schmitt*, Ana Catalán*, and John Parsch

Phil. Trans. R. Soc. B 2013, **368:** 20130024

* Equal contributions

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

rstb.royalsocietypublishing.org

Research

Cite this article: Glaser-Schmitt A, Catalán A, Parsch J. 2013 Adaptive divergence of a transcriptional enhancer between populations of *Drosophila melanogaster*. Phil Trans R Soc B 368: 20130024.
http://dx.doi.org/10.1098/rstb.2013.0024

One contribution of 12 to a Theme Issue 'Molecular and functional evolution of transcriptional enhancers in animals'.

Subject Areas:
evolution, genetics

Keywords:
evolution, gene expression, population genetics

Author for correspondence:
John Parsch
e-mail: parsch@bio.lmu.de

†These authors contributed equally to this study.

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2013.0024 or via http://rstb.royalsocietypublishing.org.

Royal Society Publishing
Informing the science of the future

Adaptive divergence of a transcriptional enhancer between populations of *Drosophila melanogaster*

Amanda Glaser-Schmitt†, Ana Catalán† and John Parsch

Department of Biology II, University of Munich (LMU), Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany

As species colonize new habitats they must adapt to the local environment. Much of this adaptation is thought to occur at the regulatory level; however, the relationships among genetic polymorphism, expression variation and adaptation are poorly understood. *Drosophila melanogaster*, which expanded from an ancestral range in sub-Saharan Africa around 15 000 years ago, represents an excellent model system for studying regulatory evolution. Here, we focus on the gene *CG9509*, which differs in expression between an African and a European population of *D. melanogaster*. The expression difference is caused by variation within a transcriptional enhancer adjacent to the *CG9509* coding sequence. Patterns of sequence variation indicate that this enhancer was the target of recent positive selection, suggesting that the expression difference is adaptive. Analysis of the *CG9509* enhancer in new population samples from Europe, Asia, northern Africa and sub-Saharan Africa revealed that sequence polymorphism is greatly reduced outside the ancestral range. A derived haplotype absent in sub-Saharan Africa is at high frequency in all other populations. These observations are consistent with a selective sweep accompanying the range expansion of the species. The new data help identify the sequence changes responsible for the difference in enhancer activity.

1. Introduction

(a) The importance of gene regulation in adaptation

Differences in gene expression are thought to underlie many of the phenotypic differences between species and populations [1–3]. With the advent of transcriptomic technologies, such as microarrays and high-throughput RNA sequencing (RNA-seq), it has become possible to identify the genes that differ in expression between species or vary in expression among individuals of the same species. Such studies have revealed that there is considerable expression divergence between closely related species (e.g. human and chimpanzee [4] or *Drosophila melanogaster* and *Drosophila simulans* [5]) as well as abundant expression variation within species (e.g. human [4,6,7], mouse [8], *Drosophila* [9,10], yeast [11–13] and fish [14–16]). A current challenge in evolutionary genetics is to identify the specific genetic changes responsible for differences in gene expression and to determine how these changes impact an organism's fitness. In this context, much attention has been paid to *cis*-regulatory elements, such as transcriptional enhancers, as they are known to play a key role in regulatory evolution [17]. It has been argued that *cis*-regulatory evolution is the major driver of adaptive divergence between species, especially at the level of morphology [17–19]. However, the importance of *cis*-regulatory divergence in relation to other types of genetic changes (e.g. amino acid replacements within proteins) in adaptation is still a topic of debate [20].

A well-known example of adaptive *cis*-regulatory evolution in humans involves the lactase gene (*LCT*), where single-nucleotide polymorphisms (SNPs) in an upstream regulatory element are associated with persistent expression of *LCT* in adults and enable them to digest the milk sugar lactose [21]. Patterns

of DNA sequence polymorphism in the *LCT* region suggest that it has been the target of recent positive selection within northern European populations [22]. Furthermore, the discovery of different, independently derived SNPs in this region of the genome that are associated with lactase persistence in African pastoralist populations is indicative of convergent adaptive evolution [23]. In *D. melanogaster*, polymorphism in the expression of the cytochrome P450 gene *Cyp6g1* is associated with the insertion of an *Accord* transposable element into its upstream regulatory region [24]. Overexpression of *Cyp6g1* owing to the *Accord* insertion confers resistance to the insecticide DDT [25], a trait that is in high frequency in non-African populations [26]. Patterns of DNA sequence polymorphism are consistent with recent positive selection favouring the high-expression allele [26]. The *Cyp6g1* example illustrates how the powerful genetic resources available for *D. melanogaster* can be used to identify adaptive changes in gene expression.

(b) The demographic history of *Drosophila melanogaster*

Drosophila melanogaster is currently a cosmopolitan species with a worldwide distribution [27]. However, the global spread of the species from its ancestral range in sub-Saharan Africa is thought to have occurred relatively recently [27,28]. Genome-scale analyses of DNA sequence variation in multiple African and non-African populations have resulted in our current understanding of the species' biogeographic and demographic history [29–33]. A general pattern that has been observed is that DNA sequence polymorphism is greater among individuals from sub-Saharan Africa than among individuals from other worldwide locations [29,34–36], which is consistent with an Afrotropical origin of the species. Populations from southern-central Africa (e.g. Zambia and Zimbabwe) show the highest genetic diversity, suggesting that they best represent the centre of origin [32]. It is hypothesized that the initial expansion of *D. melanogaster* from its ancestral range occurred around 15 000 years ago with the colonization of human settlements in the Middle East [31]. The colonization of Europe and Asia from this original non-African source population is thought to have occurred more recently, within the past 2500–5000 years and been concomitant with the spread of human populations and agriculture [31]. Finally, the colonization of North America is documented to have occurred within the past 200 years [37] and appears to have involved the admixture of European and African *D. melanogaster* [33]. There is also evidence for recent non-African gene flow into sub-Saharan Africa, with the extent of admixture varying among African populations [32].

(c) Expression differences between *Drosophila melanogaster* populations

Its successful colonization of non-African territories suggests that *D. melanogaster* has undergone adaptation to new environmental conditions. Given our extensive knowledge of the *D. melanogaster* genome and its tractability as a model organism, there has been considerable interest in finding the genes and genetic changes that underlie this adaptation. One approach has been to look for regions of the genome that show patterns of sequence polymorphism indicative of recent positive selection [38,39]. These studies have identified genes or regions of the genome that are candidates for adaptive evolution [29,30,32,40], but in most cases it has been difficult to link genetic variants with functional or phenotypic differences between populations. Another approach has been to look for genes that differ in expression between African and non-African flies. This approach focuses on regulatory divergence. To date, such expression studies have been carried out using whole adult males [9,41], whole adult females [42] and the dissected brains of both sexes [43]. In all of these cases, hundreds of genes differing in expression between populations were identified. However, the overlap among the differentially expressed genes identified by each study was small, suggesting that regulatory evolution often occurs in a sex- and tissue-dependent fashion [42,43].

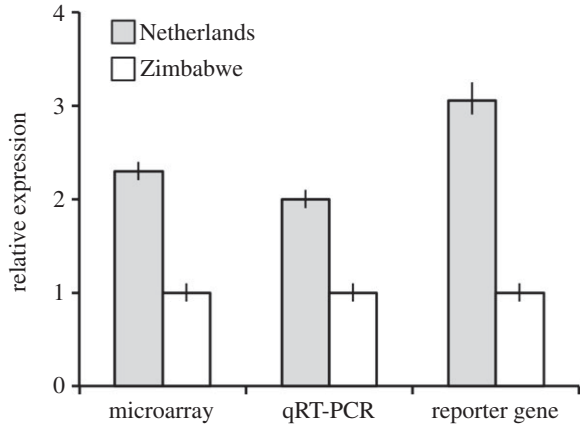


Figure 1. Expression of *CG9509* in a European (The Netherlands) and a sub-Saharan African (Zimbabwe) population. Shown are the relative expression levels in adult males as determined by microarrays or qRT-PCR. The 'reporter gene' comparison is for *lacZ* transgene expression driven by either the European or the African version of the *CG9509* enhancer. Error bars indicate ± 1 s.e. of the mean.

(d) Population genetics and expression of *CG9509*

One gene that shows a large and consistent expression difference between African and non-African flies of both sexes is *CG9509* [9,41,44]. The specific function of this gene in *D. melanogaster* is unknown, although sequence homology has led to it being annotated as a choline dehydrogenase [45]. In addition, its highly enriched expression in the Malpighian tubules [46] suggests that it may play a metabolic role in detoxification. The sequence and expression of *CG9509* have been studied in detail in population samples from Europe (The Netherlands) and Africa (Zimbabwe), revealing three major features [44]. First, *CG9509* shows two to three times higher expression in the European population than in the African population (figure 1). Second, sequence polymorphism in the *CG9509* region is greatly reduced in the European population, especially in the intergenic region just upstream of the *CG9509* coding sequence, which is consistent with a recent selective sweep. Third, this intergenic region (here denoted as the *CG9509* enhancer) is sufficient to drive differences in reporter gene expression equal to those observed for the *CG9509* gene in natural populations (figure 1). Taken together, these results provide strong evidence that positive selection has acted on the *CG9509* enhancer to increase expression in the European population. To better understand the timing and geographical scale of this positive selection, we extended the analysis of the *CG9509* enhancer to new population samples from Europe, Asia, northern Africa and sub-Saharan Africa. We find that sequence polymorphism is very low in all

populations outside the ancestral range, but much higher within sub-Saharan Africa. Furthermore, a derived haplotype associated with elevated *CG9509* expression is at high frequency in all populations outside sub-Saharan Africa but was not detected within the ancestral range. These results suggest that selection for increased expression of *CG9509* occurred during or soon after the out-of-Africa expansion of the species, before its spread into Europe and Asia.

2. Material and methods

(a) Population samples

Sequence polymorphism was surveyed in the following six *D. melanogaster* populations samples: 12 isofemale lines from The Netherlands (Leiden), 11 isofemale lines from Germany (Munich), 11 isofemale lines from Malaysia (Kuala Lumpur), 12 isofemale lines from Egypt (Cairo), 10 isofemale lines from Zambia (Siavonga) and 12 isofemale lines from Zimbabwe (Lake Kariba). The Zimbabwe and The Netherlands populations were used in a previous study of sequence and expression variation associated with the *CG9509* enhancer region [44], as well as in previous genome-wide studies [29,35,36,47]. The Malaysian population also was used in previous genome-wide demographic studies [31,48]. At least six strains from each population were used for quantitative reverse-transcription PCR (qRT-PCR) analysis. Flies from all populations were maintained as inbred, isofemale lines under standard conditions (22°C, 14 L : 10 D cycle, cornmeal-molasses medium) for at least 10 generations prior to expression analyses.

(b) DNA sequencing

New sequences of the *CG9509* intergenic region were obtained from isofemale lines of the German, Malaysian, Egyptian and Zambian populations. For each line, DNA was extracted from a single male fly using the MasterPure DNA Purification Kit (Epicentre). PCR was performed under standard conditions using four primer pairs published in Saminadin-Peter *et al*. [44] and one additional reverse primer (5′-AGCTGCAAGCAGA ACCGTAT-3′). The amplified region consisted of 1.2 kb of intergenic sequence, ranging from the stop codon of *CG14406* to the start codon of *CG9509*. PCR products were purified with ExoSAP-IT (USB) and sequenced using BigDye chemistry on a 3730 automated sequencer (Applied Biosystems). Both strands of DNA were sequenced using the PCR primers as sequencing primers. Trace files were edited using SeqTrace [49] and a multiple sequence alignment was generated with SeaView (v. 4) [50] using the ClustalW2 algorithm. All sequences have been submitted to the GenBank/EMBL database under the accession numbers HF913659–HF913726.

(c) Population genetic analyses

The following summary statistics were calculated using DnaSP v. 5.10.1 [51]: mean pairwise nucleotide diversity (π), Watterson's estimate of nucleotide diversity (θ) [52], number of segregating sites, haplotype number, haplotype diversity, F_{st} and D_{xy} (average pairwise differences between populations). Within each population, the 95% CIs of π and θ were estimated from 10 000 coalescent simulations. A neighbour-joining tree of all sequences was constructed using MEGA v. 5.05 [53]. For this, the evolutionary distances were calculated using the maximum composite likelihood method. Clade support was assessed from 1000 bootstrap replicates.

To determine whether the observed features (number of segregating sites, number of haplotypes and number of fixed, derived variants) in the populations outside sub-Saharan Africa could be explained solely by an out-of-Africa bottleneck, we performed coalescent simulations with *ms* [54], using bottleneck parameters inferred previously for the X chromosome [31,55]. To match the structure of our observed data, we simulated samples from two present-day populations of sizes N and $0.34N$, with sample sizes of 22 and 46 sequences, respectively. The larger sample was drawn from a population that experienced a bottleneck approximately 15 000 years ago, which reduced the population to 0.5% of its ancestral size. The smaller sample was drawn from a population that maintained a constant population size. Prior to the bottleneck, the two populations were assumed to be part of a single panmictic population of size N. Simulations were conditioned on the observed number of segregating sites in the total sample with a local recombination rate of 3.47 cM/Mb [56]. A total of 100 000 simulations were performed and the p-value was determined as the proportion of simulated datasets in which one of the above features in the bottlenecked population (46 sequences) was equal to (or more extreme than) the observed value in the combined non-sub-Saharan African populations.

(d) Expression analysis

Total RNA was extracted from 10 to 15 adult males (aged 4–6 days) and DNAse I digestion was performed using the MasterPure RNA Purification Kit (Epicentre). For each strain, at least two biological replicates were performed. For each replicate, 3 µg total RNA was reverse-transcribed using random hexamer primers and Superscript II reverse transcriptase (Invitrogen) following the manufacturer's protocol. A TaqMan Gene Expression Assay (Invitrogen) was then performed on the resulting cDNA using a probe specific to *CG9509* (Dm01838873_g1) as well as a probe specific to the ribosomal protein gene *RpL32* (Dm02151827_g1), which was used as an endogenous control. Since the amplification efficiencies of the two probes were nearly identical (within the range 96–99%), the ΔΔCt method was used to calculate normalized gene expression [57]. Briefly, the average threshold cycle (Ct) was determined for two technical replicates per biological replicate and ΔCt was calculated as the mean Ct difference between the *CG9509* and *RpL32* probes. The fold-change difference in expression for each biological replicate relative to the Zimbabwe population was then calculated as $2^{-(\Delta CtB - \Delta CtZK)}$, where ΔCtB is the mean ΔCt value for each biological replicate and ΔCtZK is the mean ΔCt value of the Zimbabwe strains. In order to ensure a balanced design, a total of six strains per population, each with two biological replicates, was used. For strains where more than two biological replicates were performed, the two replicates with ΔCt closest to the median were used.

3. Results

(a) Sequence polymorphism in the *CG9509* enhancer

A previous population genetic analysis of the *CG9509* enhancer examined only one population from Europe (The Netherlands) and one population from sub-Saharan Africa (Zimbabwe) [44]. To obtain a broader view of genetic variation, we sequenced the 1.2 kb intergenic region between *CG9509* and *CG14406* (figure 2) in new populations samples from Europe (Germany), Asia (Malaysia), northern Africa (Egypt) and sub-Saharan Africa (Zambia). In the following, we refer to the populations from outside sub-Saharan Africa as 'cosmopolitan'. Overall, we find that nucleotide diversity is very low in all the cosmopolitan populations (mean θ of 0.07%), with many individuals sharing the same haplotype (table 1). By contrast, nucleotide diversity is at least 12-fold higher in the Zambia and Zimbabwe

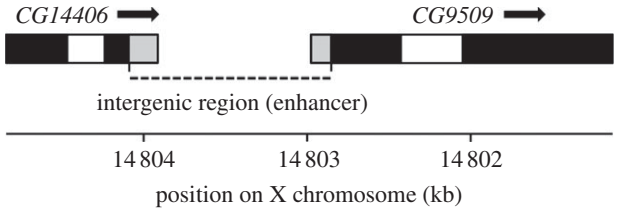


Figure 2. Map of the *CG9509* region of *D. melanogaster*. Transcriptional units are indicated by boxes, with coding regions in black, introns in white and untranslated regions in grey. The arrows indicate the direction of transcription. The intergenic region between the stop codon of *CG14406* and the start codon of *CG9509* was used for the population genetic analysis. This region has been shown to contain the transcriptional enhancer responsible for the expression difference between European and African alleles.

populations (θ of 1.3% and 1.1%, respectively), where each individual has a unique haplotype (table 1).

To determine whether the reduction in polymorphism observed in the cosmopolitan populations could be explained solely by an out-of-Africa bottleneck, we performed coalescent simulations using a demographic model inferred from X chromosome-wide polymorphism data [31,55]. Of 100 000 simulated datasets, none showed a reduction in θ as great as that observed in the real data, indicating that the probability of it being caused by a bottleneck alone is less than 0.00001. Two other features of the observed data, the number of haplotypes and the number of derived variants fixed in the cosmopolitan populations, were also highly unlikely to have been caused by a bottleneck alone ($p < 0.00001$).

(b) Sequence divergence between populations

For the cosmopolitan populations, there is not only low sequence diversity within each population, but also very little sequence divergence between populations. On average, F_{st} is 0.09 among these populations, while the average pairwise nucleotide divergence between populations (D_{xy}) is 0.08% (see electronic supplementary material, table S1). By contrast, these populations show much greater sequence divergence than the sub-Saharan African populations, with F_{st} averaging 0.46 and D_{xy} averaging 1.12%. There is little sign of population structure between the Zambia and Zimbabwe populations, where F_{st} is 0.001. The above features are also evident in a neighbour-joining tree, where the cosmopolitan sequences form an exclusive clade with very short branch lengths (figure 3), suggesting that they descend from a very recent common ancestor. By contrast, the Zambian and Zimbabwean sequences are separated by longer branches, which is consistent with an older age of these alleles (figure 3).

(c) Sequence differences between cosmopolitan and sub-Saharan African populations

Experiments using a transgenic reporter gene have shown that the twofold to threefold *CG9509* expression difference observed between flies from The Netherlands and Zimbabwe is caused by sequence variation in a 1.2-kb enhancer located just upstream of the *CG9509* coding region (figure 1) [44]. Within this region, there are nine sites that show a fixed or nearly fixed difference between the cosmopolitan and the sub-Saharan African populations (figure 4). These include eight SNPs and one insertion/deletion (indel) polymorphism. Using *D. simulans*, *Drosophila sechellia* and *Drosophila yakuba* as outgroup species, the ancestral state could be inferred for all eight SNPs (figure 4). In all cases, the sub-Saharan African variant was the ancestral form, indicating that new mutations have risen to high frequency in the other populations. For the indel polymorphism, it was not possible to determine the ancestral state, as multiple, large indels have occurred across this region in the outgroup species. However, the tight linkage of this indel polymorphism with the surrounding SNPs suggests that it represents a deletion mutation and that a common derived haplotype is present in all cosmopolitan populations. One strain from Zambia has a deletion similar to the one observed outside sub-Saharan Africa (figure 4). However, this may represent an independent mutational event, as there is also a unique SNP directly adjacent to the deletion in this strain (figure 4). Consistent with this interpretation, the deletion in the Zambia strain is not linked to any of the derived SNPs found at high frequency in the cosmopolitan populations (figure 4).

(d) Expression differences between populations

It was shown previously that *CG9509* has higher expression in a cosmopolitan population (The Netherlands) than in a sub-Saharan African population (Zimbabwe; figure 1) [41,44]. Using qRT-PCR, we were able to confirm this result and extend it to three new cosmopolitan populations (Germany, Malaysia and Egypt) and a new sub-Saharan African population (Zambia). On average, the cosmopolitan strains showed nearly threefold higher expression than the sub-Saharan African strains, which was highly significant (figure 5). We also compared *CG9509* expression in each cosmopolitan population to that in sub-Saharan Africa. Since the Zambian and Zimbabwean populations showed no evidence of population structure (see electronic supplementary material, table S1) and had very similar *CG9509* expression (figure 5), they were pooled for comparison with the cosmopolitan populations. Individually, the populations from The Netherlands, Malaysia and Egypt each had significantly higher *CG9509* expression than the pooled sub-Saharan African populations (figure 5). The German population showed, on average, 1.6-fold higher *CG9509* expression than the pooled sub-Saharan African populations, but this difference was not significant (figure 5).

(e) Association between sequence variants and expression

To determine whether particular sites within the *CG9509* enhancer that show a fixed or nearly fixed difference between cosmopolitan and sub-Saharan African populations (figure 4) were associated with the observed difference in expression, we examined the expression of *CG9509* in additional strains from Zambia. However, we could not establish a clear link between any individual sequence variant and the expression difference. For example, Zambia strain *ZI273*, which is the only sub-Saharan African strain with the 5-bp deletion at positions 821–817 before the *CG9509* start codon (figure 4), did not show higher expression than the other sub-Saharan strains (see electronic supplementary material, figure S1). Similarly, strain *ZI112*, which has cosmopolitan variants at positions 1180, 1174 and 1155, and strain *ZI254*, which has cosmopolitan variants at positions 748 and 718 (figure 4), did

Table 1. DNA sequence polymorphism within populations. *n*, number of sequences; *S*, number of segregating sites; θ, Watterson's [52] estimate of nucleotide diversity (per 100 sites); π, mean pairwise nucleotide diversity (per 100 sites); *nHap*, number of haplotypes; *HapDiv*, haplotype diversity. The 95% CIs for θ and π are shown in brackets.

population	*n*	*S*	θ	π	*nHap*	*HapDiv*
The Netherlands	12	1	0.03 (0.00–0.13)	0.05 (0.00–0.16)	2	0.55
Germany	11	5	0.14 (0.00–0.27)	0.11 (0.00–0.31)	5	0.71
Malaysia	11	1	0.03 (0.00–0.14)	0.05 (0.00–0.15)	2	0.55
Egypt	12	3	0.08 (0.00–0.24)	0.09 (0.00–0.27)	4	0.76
Zambia	10	42	1.33 (0.31–2.06)	1.03 (0.27–2.32)	10	1.00
Zimbabwe	12	39	1.12 (0.29–1.67)	0.87 (0.23–1.95)	12	1.00

not show unusually high expression relative to other Zambian strains (see electronic supplementary material, figure S1).

Although the German population showed lower average *CG9509* expression than the other cosmopolitan populations (figure 5), this difference was not caused solely by strains *MU10* and *MU11*, which were the only ones with the sub-Saharan variant (G) at position 167 (figure 4 and electronic supplementary material, figure S1). Within the cosmopolitan populations, there is a SNP (a G/C polymorphism 67 bp before the *CG9509* start codon) segregating at intermediate frequency (32%; see electronic supplementary material, figure S2). The derived variant (G) is associated with a 1.5-fold increase in *CG9509* expression within cosmopolitan populations (*t*-test; $p = 0.016$; see electronic supplementary material, figure S3). While this variant can account for some of the *CG9509* expression variation among cosmopolitan strains, it cannot account for the large expression difference between cosmopolitan and sub-Saharan African strains, as cosmopolitan strains with the sub-Saharan African variant (C) still have over twofold higher expression than sub-Saharan African strains (*t*-test; $p < 10^{-3}$; see electronic supplementary material, figure S3).

4. Discussion

(a) Evidence for adaptive evolution of *CG9509* at the level of expression

Several lines of evidence suggest that *CG9509* has undergone adaptive regulatory evolution within the past 5000–15 000 years. First, this gene shows a large and consistent expression difference between cosmopolitan and sub-Saharan African populations (figure 5) [9,41,44]. Second, within cosmopolitan populations, DNA sequence polymorphism is greatly reduced in the intergenic region immediately upstream of the *CG9509* coding sequence (table 1), which is consistent with a selective sweep in this region of the genome [44]. Third, sequence variation within this intergenic region (designated as the *CG9509* enhancer) has been shown to account for the difference in expression between cosmopolitan and sub-Saharan African strains [44]. Finally, within the *CG9509* enhancer, there is a derived haplotype that is in high frequency in cosmopolitan populations, but is absent in sub-Saharan Africa (figure 4).

The *CG9509* enhancer also shows evidence for long-term adaptive evolution over the past 2–3 Myr (since the divergence of *D. melanogaster* and species of the *D. simulans* clade). Application of the McDonald-Kreitman (MK) test [58] to data on polymorphism within *D. melanogaster* and divergence between *D. melanogaster* and *D. sechellia* found a significant excess of between-species divergence in the enhancer compared to synonymous sites in the *CG9509* coding region [44]. Although the previous analysis did not polarize divergence to the *D. melanogaster* lineage, a re-analysis of the data using *D. yakuba* as an outgroup to polarize changes indicated that a significant excess of substitutions in the enhancer occurred on the *D. melanogaster* lineage (see electronic supplementary material, table S2). This suggests that there have been recurrent selective sweeps within the *D. melanogaster CG9509* enhancer since its divergence from *D. sechellia*.

(b) Evidence for adaptive evolution of *CG9509* at the level of protein sequence

In addition to showing evidence for adaptive regulatory evolution, *CG9509* also shows evidence for having undergone adaptive protein evolution within the past 2–3 Myr. A comparison of polymorphism and divergence within the *CG9509* coding region using the MK test revealed a significant excess of non-synonymous divergence between species [44], which is indicative of recurrent selection for amino acid replacements. A recent genome-wide study of polymorphism also identified *CG9509* as a target of positive selection using MK tests polarized to the *D. melanogaster* lineage [59]. Indeed, *CG9509* was ranked among the top 10 genes in the genome that showed evidence for adaptive protein evolution on the *D. melanogaster* lineage [59].

(c) *CG9509* sequence and expression variation within North America

Drosophila melanogaster is believed to have colonized North America within the past 200 years [37]. This colonization appears to be the result of admixture between European and African source populations, with the estimated proportion of European and African ancestry being 85% and 15%, respectively [33]. The *Drosophila* Genetic Reference Panel (DGRP) [60], consisting of 192 inbred, isofemale lines derived from a single outbred population from Raleigh, North Carolina, is an excellent resource for examining naturally occurring variation within a North American *D. melanogaster* population. Consistent with the inferred proportion of admixture in North America [33], the cosmopolitan variants at the sites showing fixed or nearly fixed differences between cosmopolitan and sub-Saharan African populations in the *CG9509*

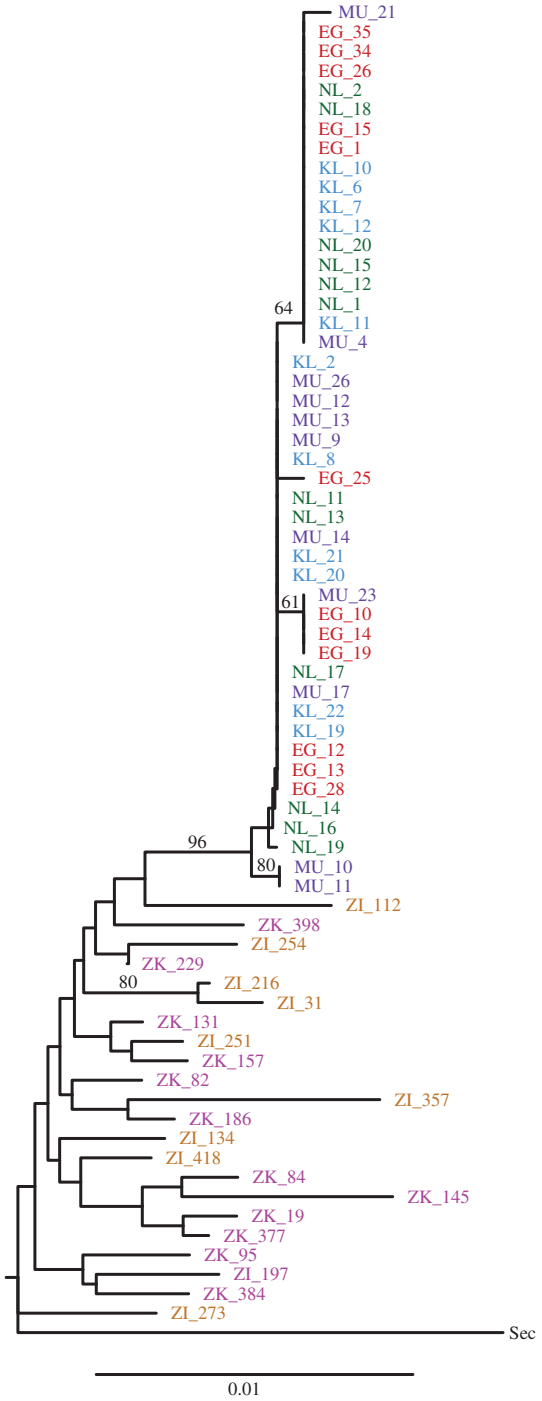


Figure 3. Neighbour-joining tree of all intergenic region sequences. The population abbreviations are as follows: The Netherlands (NL), Germany (MU), Malaysia (KL), Egypt (EG), Zambia (ZI) and Zimbabwe (ZK). *Drosophila sechellia* (Sec) was used as an outgroup. The branch lengths are proportional to the sequence distances, with the exception of the *D. sechellia* branch, which is shown at 20% of its actual length. Bootstrap values are shown for nodes with greater than 60% support. (Online version in colour.)

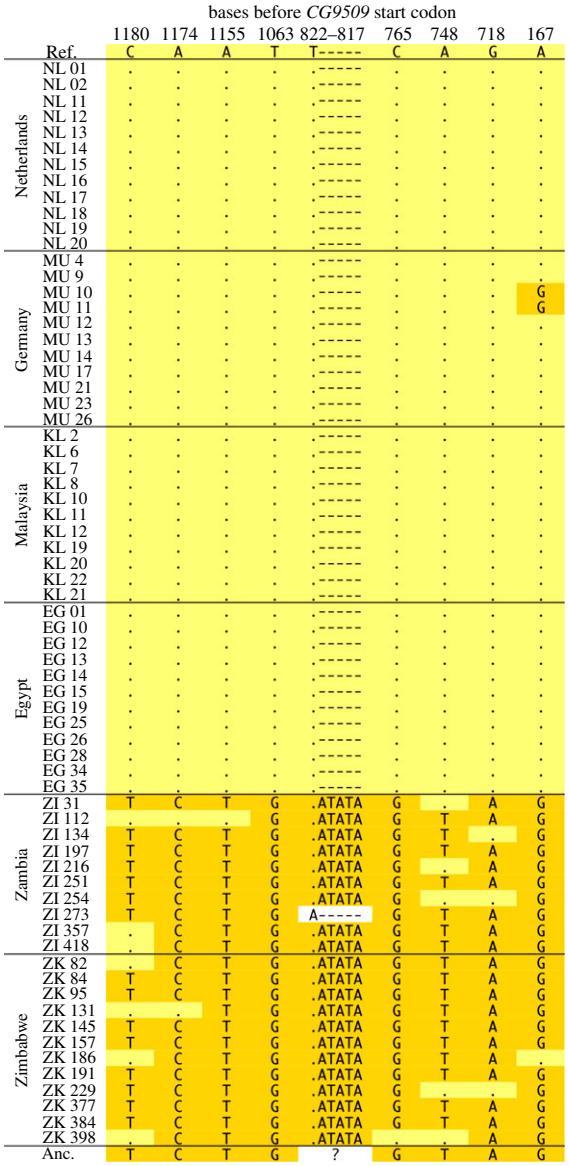


Figure 4. Fixed and nearly fixed differences in the *CG9509* enhancer region between cosmopolitan and sub-Saharan African populations. Cosmopolitan variants are indicated by light shading and sub-Saharan African variants by dark shading. Ambiguous variants are shown in white. The reference sequence (Ref.) was obtained from FlyBase release 5.48 [45] and the ancestral (Anc.) state was inferred from alignments with *D. simulans*, *D. sechellia* and *D. yakuba*. (Online version in colour.)

enhancer (figure 4) are present in approximately 75–85% of the DGRP lines [60], while the private cosmopolitan variant (G 67 bp before the start codon; see electronic supplementary material, figure S2) is present in 31%.

The results of an association study of sequence and expression variation in a subset of 39 DGRP lines [61] are consistent with some of the major features of *CG9509* sequence and expression variation identified in our study. First, in some DGRP lines the *CG9509* enhancer region shows greatly reduced variant density in comparison to the surrounding regions [61], which is similar to the greatly reduced sequence polymorphism observed in our cosmopolitan strains (table 1). Second, DGRP lines showing this low variant density correspond to cosmopolitan haplotypes of the *CG9509* enhancer that are associated with increased expression [44,61]. Third, the presence of cosmopolitan variants within the *CG9509* enhancer region in particular DGRP lines appears to be associated with a general increase of *CG9509* expression in these lines [61]. Analysis of the DGRP lines revealed an expression quantitative trait locus (eQTL) associated with *CG9509* expression within the *CG9509* enhancer region [61]. This eQTL corresponds to the segregating site 67 bp before the start codon (see electronic supplementary material, figure S2) that we found to be associated with *CG9509* expression variation within cosmopolitan populations (see electronic supplementary material, figure S3). The direction and magnitude of the expression change [61] agree well with our finding that

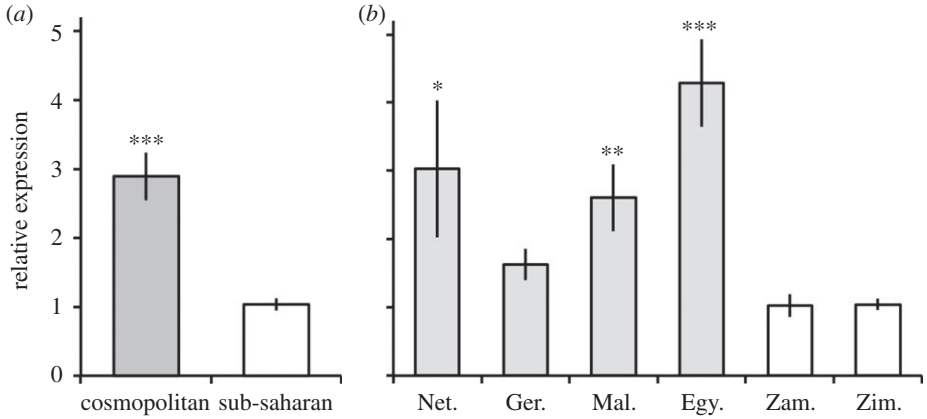


Figure 5. Expression of *CG9509* in cosmopolitan and sub-Saharan African populations. (*a*) Expression in pooled cosmopolitan and sub-Saharan African populations. (*b*) Expression in the Netherlands (Net.), Germany (Ger.), Malaysia (Mal.), Egypt (Egy.), Zambia (Zam.) and Zimbabwe (Zim.). Cosmopolitan populations are denoted by grey bars and sub-Saharan African populations by white bars. Error bars indicate ± 1 s.e. of the mean. Statistical significance was assessed using a *t*-test. In (*b*), each cosmopolitan population was compared to the pooled sub-Saharan African populations using a Bonferroni multiple test correction ($^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 10^{-6}$).

the G variant at this site is associated with a 1.5-fold increase in expression within cosmopolitan populations (see electronic supplementary material, figure S3). However, none of the fixed or nearly fixed differences between cosmopolitan and sub-Saharan African populations (figure 4) showed a significant association with *CG9509* expression within the DGRP lines [61]. This may be due to the fact that the analysis was performed on a single North American population in which sub-Saharan African variants were present only at low frequency, which reduces the statistical power to detect associations in genome-wide analyses.

(d) Possible functions of *CG9509*

At present, the specific function of *CG9509* in *D. melanogaster* and the effect that variation in its expression has on phenotypic differences between individuals are unknown. *CG9509* is predicted to encode a choline dehydrogenase with highly enriched expression in the Malpighian tubules [45,46], which is functionally analogous to the kidney of mammals. This suggests that *CG9509* may play a role in detoxification. Variation in other genes involved in choline metabolism, namely choline kinases, has been implicated in insecticide resistance, with resistant alleles being present at high frequency in cosmopolitan *D. melanogaster* populations [43,62]. Unlike *CG9509*, these choline kinases show reduced expression (or loss of function) outside sub-Saharan Africa [43,59]. By contrast, resistance to DDT is conferred by overexpression of the cytochrome P450 gene *Cyp6g1* [24], which also shows highest expression in the Malpighian tubules [46]. *CG9509*'s similarity in function and expression to these other insecticide resistance genes, as well as the strong signal for adaptive evolution outside sub-Saharan Africa, suggest that it may also play a role in the detoxification of insecticides or other chemicals present outside *D. melanogaster*'s ancestral home range.

It is also possible that *CG9509* plays a role in adaptation to temperature or humidity. For example, it has been shown in *Drosophila* that the ratio of phosphatidylcholine to phosphatidyethanolamine decreases during cold acclimation [63], suggesting that choline metabolism might be linked to cold tolerance. Additionally, choline dehydrogenases are known to catalyse the conversion of choline into betaine [64], which has been reported to play an osmoprotectant role in mammals [65] and has also been found in insects [66]. *CG9509*'s very high expression in the Malpighian tubules (and lower expression in the gut) is consistent with a role in osmoregulation, which is a critical process for environmental adaptation. A QTL study of *D. melanogaster* did not find *CG9509* to be among the major QTLs affecting desiccation resistance [67]. However, this study was carried out using recombinant inbred lines derived from two isofemale lines of a single North American (California) population and, thus, did not include genetic variation from sub-Saharan Africa.

Finally, knockout of the choline dehydrogenase gene (*Chdh*) in mice has been shown to decrease sperm motility [68]. Similarly, polymorphism in the human *Chdh* gene also is associated with variation in sperm motility [69]. Furthermore, dietary choline is required for proper sperm motility and reproductive behaviour in *Drosophila* [70]. Thus, it is possible that expression variation in the *Drosophila CG9509* gene affects male fertility and/or sperm competition. Genes expressed in the testes, especially those that are X-linked, tend to show the greatest signal of adaptive evolution in *Drosophila* [71]. However, *CG9509* shows only very low levels of expression in the testes that are several hundred-fold lower than those in the Malpighian tubules [46], making a role in male fertility unlikely.

5. Conclusion

Our finding that the selective sweep encompassing the *CG9509* enhancer extends to populations from Asia and northern Africa has three important implications. First, it indicates that the sweep is not restricted to a local population or region. Second, it helps to establish the timing of the sweep, which must have occurred after the out-of-Africa migration of the species, but before the divergence of the European and Asian populations (i.e. 5000–15 000 years ago). Third, it suggests that the sweep was not caused by adaptation to a temperate environment *per se*, as it spans populations from tropical and temperate latitudes. In this respect, the *CG9509* example differs from other well-studied polymorphisms in *D. melanogaster* that show latitudinal clines in frequency and are thought to reflect climatic adaptation [72–74]. Instead, the *CG9509* sweep may

be the result of adaptation to human commensalism or agriculture, which is consistent with the inferred role of *CG9509* in detoxification. The sequence variants differing in frequency between the cosmopolitan and sub-Saharan African populations represent candidates for the specific target(s) of selection and future studies that examine their functional effect on *CG9509* expression will help elucidate the molecular mechanism of gene regulatory evolution.

Acknowledgements. We thank John Baines, Sonja Grath, Francesco Paparazzo, Aparup Das, Korbinian von Heckel and John Pool for providing *Drosophila* stocks. We also thank Andreas Massouras and Bart Deplancke for access to polymorphism and eQTL association data for the DGRP lines. Hedwig Gebhart and Hilde Lainer provided excellent technical assistance in the laboratory.

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

References

1. King MC, Wilson AC. 1975 Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116. (doi:10.1126/science.1090005)
2. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. 2003 The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**, 1377–1419. (doi:10.1093/molbev/msg140)
3. Whitehead A, Crawford DL. 2006 Variation within and among species in gene expression: raw material for evolution. *Mol. Ecol.* **15**, 1197–1211. (doi:10.1111/j.1365-294X.2006.02868.x)
4. Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S. 2005 Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* **309**, 1850–1854. (doi:10.1126/science.1108296)
5. Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003 Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science* **300**, 1742–1745. (doi:10.1126/science.1085881)
6. Stranger BE *et al*. 2005 Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1**, e78. (doi:10.1371/journal.pgen.0010078)
7. Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM. 2007 Gene-expression variation within and among human populations. *Am. J. Hum. Genet.* **80**, 502–509. (doi:10.1086/512017)
8. Voolstra C, Tautz D, Farbrother P, Eichinger L, Harr B. 2007 Contrasting evolution of expression differences in the testis between species and subspecies of the house mouse. *Genome Res.* **17**, 42–49. (doi:10.1101/gr.5683806)
9. Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003 Rapid evolution of male-biased gene expression in *Drosophila*. *Proc. Natl Acad. Sci. USA* **100**, 9894–9899. (doi:10.1073/pnas.1630690100)
10. Ayroles JF *et al*. 2009 Systems genetics of complex traits in *Drosophila melanogaster*. *Nat. Genet.* **41**, 299–307. (doi:10.1038/ng.332)
11. Cavalieri D, Townsend JP, Hartl DL. 2000 Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc. Natl Acad. Sci. USA* **97**, 12 369–12 374. (doi:10.1073/pnas.210395297)
12. Townsend JP, Cavalieri D, Hartl DL. 2003 Population genetic variation in genome-wide gene expression. *Mol. Biol. Evol.* **20**, 955–963. (doi:10.1093/molbev/msg106)
13. Fay JC, McCullough HL, Sniegowski PD, Eisen MB. 2004 Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol.* **5**, R26. (doi:10.1186/gb-2004-5-4-r26)
14. Oleksiak MF, Churchill GA, Crawford DL. 2002 Variation in gene expression within and among natural populations. *Nat. Genet.* **32**, 261–266. (doi:10.1038/ng983)
15. Aubin-Horth N, Landry CR, Letcher BH, Hofmann HA. 2005 Alternative life histories shape brain gene expression profiles in males of the same population. *Proc. R. Soc. B* **272**, 1655–1662. (doi:10.1098/rspb.2005.3125)
16. Whitehead A, Crawford DL. 2006 Neutral and adaptive variation in gene expression. *Proc. Natl Acad. Sci. USA* **103**, 5425–5430. (doi:10.1073/pnas.0507648103)
17. Wray GA. 2007 The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216. (doi:10.1038/nrg2063)
18. Carroll SB. 2000 Endless forms: the evolution of gene regulation and morphological diversity. *Cell* **101**, 577–580. (doi:10.1016/S0092-8674(00)80868-5)
19. Carroll SB. 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36. (doi:10.1016/j.cell.2008.06.030)
20. Hoekstra HE, Coyne JA. 2007 The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016. (doi:10.1111/j.1558-5646.2007.00105.x)
21. Ingram CJ, Mulcare CA, Itan Y, Thomas MG, Swallow DM. 2009 Lactose digestion and the evolutionary genetics of lactase persistence. *Hum. Genet.* **124**, 579–591. (doi:10.1007/s00439-008-0593-6)
22. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. 2004 Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74**, 1111–1120. (doi:10.1086/421051)
23. Tishkoff SA *et al*. 2007 Convergent adaptation of human lactase persistence in Africa and Europe. *Nat. Genet.* **39**, 31–40. (doi:10.1038/ng1946)
24. Daborn PJ *et al*. 2002 A single p450 allele associated with insecticide resistance in *Drosophila*. *Science* **297**, 2253–2256. (doi:10.1126/science.1074170)
25. Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, Daborn PJ. 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. (doi:10.1534/genetics.106.066597)
26. Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH, Schlotterer C. 2004 World-wide survey of an *Accord* insertion and its association with DDT resistance in *Drosophila melanogaster*. *Mol. Ecol.* **13**, 2491–2504. (doi:10.1111/j.1365-294X.2004.02263.x)
27. Lachaise D, Silvain JF. 2004 How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster–D. simulans* palaeogeographic riddle. *Genetica* **120**, 17–39. (doi:10.1023/B:GENE.0000017627.27537.ef)
28. Stephan W, Li H. 2007 The recent demographic and adaptive history of *Drosophila melanogaster*. *Heredity* **98**, 65–68. (doi:10.1038/sj.hdy.6800901)
29. Ometto L, Glinka S, De Lorenzo D, Stephan W. 2005 Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. *Mol. Biol. Evol.* **22**, 2119–2130. (doi:10.1093/molbev/msi207)
30. Li H, Stephan W. 2006 Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet.* **2**, e166. (doi:10.1371/journal.pgen.0020166)
31. Laurent SJ, Werzner A, Excoffier L, Stephan W. 2011 Approximate Bayesian analysis of *Drosophila melanogaster* polymorphism data reveals a recent colonization of Southeast Asia. *Mol. Biol. Evol.* **28**, 2041–2051. (doi:10.1093/molbev/msr031)
32. Pool JE *et al*. 2012 Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genet.* **8**, e1003080. (doi:10.1371/journal.pgen.1003080)
33. Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S. 2013 Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics* **193**, 291–301. (doi:10.1534/genetics.112.145912)
34. Begun DJ, Aquadro CF. 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550. (doi:10.1038/365548a0)
35. Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D. 2003 Demography and natural

selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics* **165**, 1269–1278.

36. Hutter S, Li H, Beisswanger S, De Lorenzo D, 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. (doi:10.1534/genetics.107.074922)
37. Keller A. 2007 *Drosophila melanogaster*'s history as a human commensal. *Curr. Biol.* **17**, R77–R81. (doi:10.1016/j.cub.2006.12.031)
38. Jensen JD, Wong A, Aquadro CF. 2007 Approaches for identifying targets of positive selection. *Trends Genet.* **23**, 568–577. (doi:10.1016/j.tig.2007.08.009)
39. Stephan W. 2010 Detecting strong positive selection in the genome. *Mol. Ecol. Resour.* **10**, 863–872. (doi:10.1111/j.1755-0998.2010.02869.x)
40. Harr B, Kauer M, Schlötterer C. 2002 Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **99**, 12 949–12 954. (doi:10.1073/pnas.202336899)
41. Hutter S, Saminadin-Peter SS, Stephan W, Parsch J. 2008 Gene expression variation in African and European populations of *Drosophila melanogaster*. *Genome Biol.* **9**, R12. (doi:10.1186/gb-2008-9-1-r12)
42. Müller L, Hutter S, Stamboliyska R, Saminadin-Peter SS, Stephan W, Parsch J. 2011 Population transcriptomics of *Drosophila melanogaster* females. *BMC Genomics* **12**, 81. (doi:10.1186/1471-2164-12-81)
43. Catalán A, Hutter S, Parsch J. 2012 Population and sex differences in *Drosophila melanogaster* brain gene expression. *BMC Genomics* **13**, 654. (doi:10.1186/1471-2164-13-654)
44. Saminadin-Peter SS, Kemkemer C, Pavlidis P, Parsch J. 2012 Selective sweep of a *cis*-regulatory sequence in a non-African population of *Drosophila melanogaster*. *Mol. Biol. Evol.* **29**, 1167–1174. (doi:10.1093/molbev/msr284)
45. McQuilton P, St Pierre SE, Thurmond J, FlyBase Consortium. 2012 FlyBase 101–the basics of navigating FlyBase. *Nucleic Acids Res.* **40**, D706–D714. (doi:10.1093/nar/gkr1030)
46. Chintapalli VR, Wang J, Dow JA. 2007 Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat. Genet.* **39**, 715–720. (doi:10.1038/ng2049)
47. Parsch J, Zhang Z, Baines JF. 2009 The influence of demography and weak selection on the McDonald-Kreitman test: an empirical study in *Drosophila*. *Mol. Biol. Evol.* **26**, 691–698. (doi:10.1093/molbev/msn297)
48. Glinka S, Stephan W, Das A. 2005 Homogeneity of common cosmopolitan inversion frequencies in Southeast Asian *Drosophila melanogaster*. *J. Genet.* **84**, 173–178. (doi:10.1007/BF02715842)
49. Stucky BJ. 2012 SeqTrace: a graphical tool for rapidly processing DNA sequencing chromatograms. *J. Biomol. Tech.* **23**, 90–93. (doi:10.7171/jbt.12-2303-004)
50. Gouy M, Guindon S, Gascuel O. 2010 SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224. (doi:10.1093/molbev/msp259)
51. Rozas J. 2009 DNA sequence polymorphism analysis using DnaSP. *Methods Mol. Biol.* **537**, 337–350. (doi:10.1007/978-1-59745-251-9_17)
52. Watterson GA. 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **276**, 256–276. (doi:10.1016/0040-5809(75)90020-9)
53. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739. (doi:10.1093/molbev/msr121)
54. Hudson R. 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337–338. (doi:10.1093/bioinformatics/18.2.337)
55. Werzner A, Pavlidis P, Ometto L, Stephan W, Laurent S. 2013 Selective sweep in the *Flotillin-2* region of European *Drosophila melanogaster*. *PLoS ONE* **8**, e56629. (doi:10.1371/journal.pone.0056629)
56. Comeron JM, Ratnappan R, Bailin S. 2012 The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet.* **8**, e1002905. (doi:10.1371/journal.pgen.1002905)
57. Pfaffl MW. 2001 A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45. (doi:10.1093/nar/29.9.e45)
58. McDonald JH, Kreitman M. 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654. (doi:10.1038/351652a0)
59. Langley CH *et al*. 2012 Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* **192**, 533–598. (doi:10.1534/genetics.112.142018)
60. Mackay TFC *et al*. 2012 The *Drosophila melanogaster* genetic reference panel. *Nature* **482**, 173–178. (doi:10.1038/nature10811)
61. Massouras A *et al*. 2012 Genomic variation and its impact on gene expression in *Drosophila melanogaster*. *PLoS Genet.* **8**, e1003055. (doi:10.1371/journal.pgen.1003055)
62. Aminetzach YT, Macpherson JM, Petrov DA. 2005 Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science* **309**, 764–767. (doi:10.1126/science.1112699)
63. Kostal V, Korbelova J, Rozsypal J, Zahradnickova H, Cimlova J, Tomcala A, Simek P. 2011 Long-term cold acclimation extends survival time at 0°C and modifies the metabolomic profiles of the larvae of the fruit fly *Drosophila melanogaster*. *PLoS ONE* **6**, e25025. (doi:10.1371/journal.pone.0025025)
64. Lin CS, Wu RD. 1986 Choline oxidation and choline dehydrogenase. *J. Protein Chem.* **5**, 193–200. (doi:10.1007/BF01025488)
65. Petronini PG, De Angelis EM, Borghetti P, Borghetti AF, Wheeler KP. 1992 Modulation by betaine of cellular responses to osmotic stress. *Biochem. J.* **282**, 69–73.
66. Phalaraksh C, Reynolds SE, Wilson ID, Lenz EM, Nicholson KJ, Lindon JC. 2008 A metabonomic analysis of insect development: ^{1}H-NMR spectroscopic characterization of changes in the composition of the haemolymph of larvae and pupae of the tobacco hornworm, *Manduca sexta*. *ScienceAsia* **34**, 279–286. (doi:10.2306/scienceasia1513-1874.2008.34.279)
67. Foley BR, Telonis-Scott M. 2011 Quantitative genetic analysis suggests causal association between cuticular hydrocarbon composition and desiccation survival in *Drosophila melanogaster*. *Heredity* **106**, 68–77. (doi:10.1038/hdy.2010.40)
68. Johnson AR, Craciunescu CN, Guo Z, Teng YW, Thresher RJ, Blusztajn JK, Zeisel SH. 2010 Deletion of murine choline dehydrogenase results in diminished sperm motility. *FASEB J.* **24**, 2752–2761. (doi:10.1096/fj.09-153718)
69. Johnson AR, Lao S, Wang T, Galanko JA, Zeisel SH. 2012 Choline dehydrogenase polymorphism *rs12676* is a functional variation and is associated with changes in human sperm cell function. *PLoS ONE* **7**, e36047. (doi:10.1371/journal.pone.0036047)
70. Geer BW. 1967 Dietary choline requirements for sperm motility and normal mating activity in *Drosophila melanogaster*. *Biol. Bull.* **133**, 548–566. (doi:10.2307/1539917)
71. Baines JF, Sawyer SA, Hartl DL, Parsch J. 2008 Effects of X-linkage and sex-biased gene expression on the rate of adaptive protein evolution in *Drosophila*. *Mol. Biol. Evol.* **25**, 1639–1650. (doi:10.1093/molbev/msn111)
72. De Jong G, Bochdanovits Z. 2003 Latitudinal clines in *Drosophila melanogaster*: body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. *J. Genet.* **82**, 207–223. (doi:10.1007/BF02715819)
73. Sezgin E, Duvernell DD, Matzkin LM, Duan Y, Zhu CT, Verrelli BC, Eanes WF. 2004 Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in *Drosophila melanogaster*. *Genetics* **168**, 923–931. (doi:10.1534/genetics.104.027649)
74. Umina PA, Weeks AR, Kearney MR, McKechnie SW, Hoffmann AA. 2005 A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. *Science* **308**, 691–693. (doi:10.1126/science.1109523)

Supplementary Material

Adaptive divergence of a transcriptional enhancer between populations of *Drosophila melanogaster*

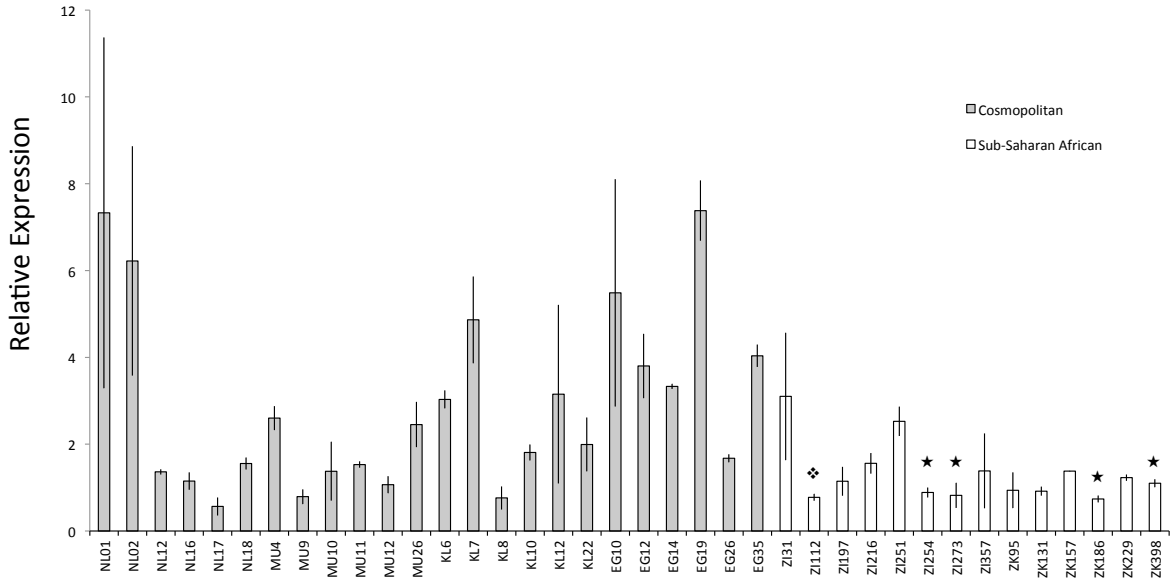
Amanda Glaser-Schmitt, Ana Catalán and John Parsch

Supplementary Table S1. Sequence divergence between populations. F_{st} values are shown above the diagonal. D_{xy} values (in percent) are shown below the diagonal.

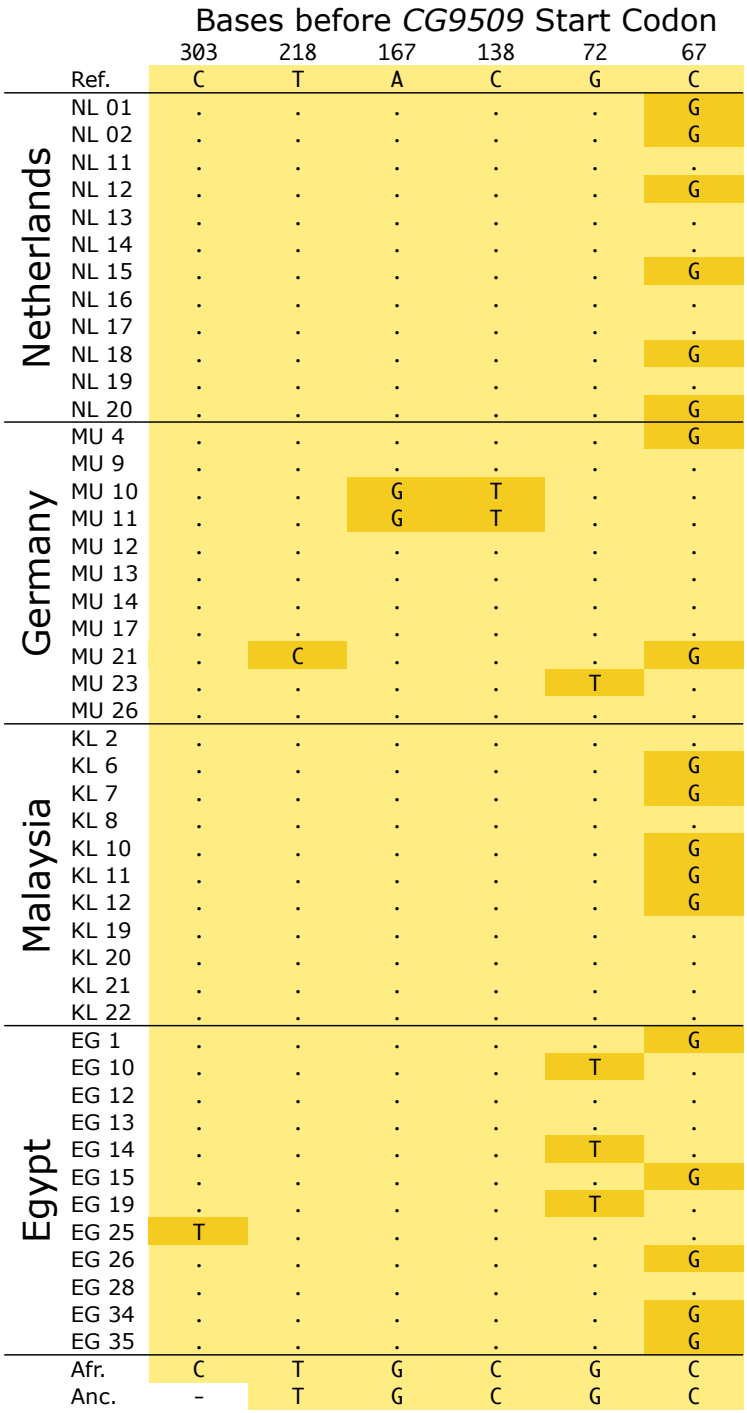
	Netherlands	Germany	Malaysia	Egypt	Zambia	Zimbabwe
Netherlands	–	0.156	0.153	0.153	0.455	0.414
Germany	0.095	–	0.042	0.026	0.495	0.443
Malaysia	0.045	0.092	–	0.011	0.511	0.459
Egypt	0.068	0.108	0.067	–	0.408	0.457
Zambia	1.159	1.144	1.155	1.175	–	0.001
Zimbabwe	1.079	1.070	1.075	1.095	0.882	–

Supplementary Table S2. Results of McDonald-Kreitman (MK) tests of the *CG9509* enhancer polarised to the *D. melanogaster* lineage. The number of fixed differences (*D*) occurring on the *D. melanogaster* lineage and the number of polymorphic sites (*P*) within each *D. melanogaster* population are shown. The subscripts indicate synonymous (s) or enhancer (enh) sites.

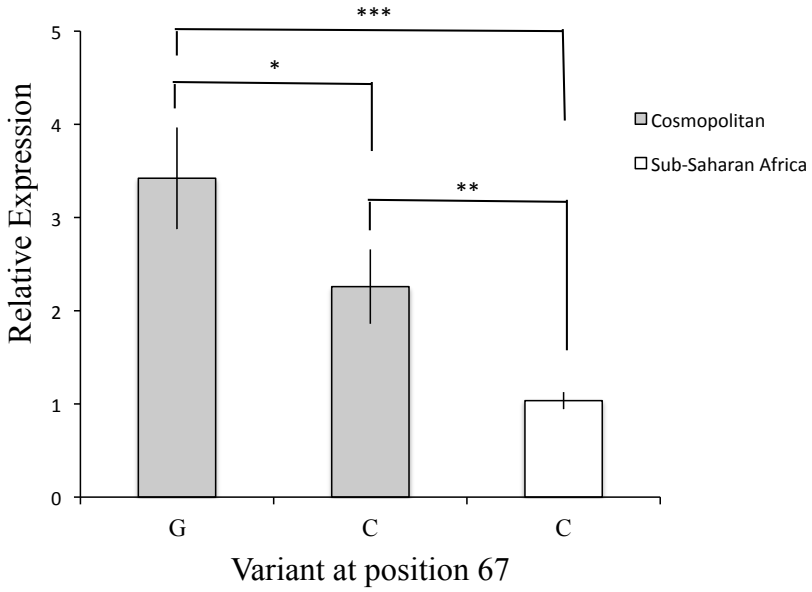
Population	D_s	P_s	D_{enh}	P_{enh}	*P*-value
Cosmopolitan (Netherlands)	17	5	38	1	0.020
Sub-Saharan African (Zimbabwe)	12	29	30	22	0.007



Supplementary Figure S1: Mean *CG9509* expression in adult males of all lines surveyed as determined by qRT-PCR. The population abbreviations are as follows: Netherlands (NL), Germany (MU), Malaysia (KL), Egypt (EG), Zambia (ZI) and Zimbabwe (ZK). Lines for which more than two biological replicates were performed are indicated by either a star (★, six replicates) or a diamond (❖, four replicates). Error bars indicate ±1 standard error of the mean.



Supplementary Figure S2: Polymorphic sites within the *CG9509* intergenic region of cosmopolitan populations. The reference sequence (Ref) was obtained from FlyBase and the ancestral (Anc) state was inferred from alignments with *D. simulans*, *D. sechellia* and *D. yakuba*. The SNP variant present in sub-Saharan Africa (Zambia and Zimbabwe) is given in the 'Afr' row. Derived variants are indicated by dark shading and inferred ancestral variants by light shading. For site 303, a gap was present in the outgroup species and, thus, the ancestral state could not be determined unambiguously.



Supplementary Figure S3: *CG9509* expression in cosmopolitan strains containing either a C or a G 67 bp before the start codon. Error bars indicate ±1 standard error of the mean. Statistical significance was assessed using a *t*-test (*$P < 0.05$, **$P < 0.01$, ***$P < 10^{-6}$).

Chapter 4

'Escaping' the X chromosome leads to increased gene expression in the male germline of *Drosophila melanogaster*

Claus Kemkemer, Ana Catalán, and John Parsch

Heredity (in press), doi: 10.1038/hdy.2013.86.

Heredity (2013), 1–7

www.nature.com/hdy

ORIGINAL ARTICLE

'Escaping' the X chromosome leads to increased gene expression in the male germline of *Drosophila melanogaster*

C Kemkemer[1], A Catalán[2] and J Parsch[2]

Genomic analyses of *Drosophila* species suggest that the X chromosome presents an unfavourable environment for the expression of genes in the male germline. A previous study in *D. melanogaster* used a reporter gene driven by a testis-specific promoter to show that expression was greatly reduced when the gene was inserted onto the X chromosome as compared with the autosomes. However, a limitation of this study was that only the expression regulated by a single, autosomal-derived promoter was investigated. To test for an increase in expression associated with 'escaping' the X chromosome, we analysed reporter gene expression driven by the promoters of three X-linked, testis-expressed genes (*CG10920*, *CG12681* and *CG1314*) that were inserted randomly throughout the *D. melanogaster* genome. In all cases, insertions on the autosomes showed significantly higher expression than those on the X chromosome. Thus, even genes whose regulation has adapted to the X-chromosomal environment show increased male germline expression when relocated to an autosome. Our results provide direct experimental evidence for the suppression of X-linked gene expression in the *Drosophila* male germline that is independent of gene dose.
Heredity advance online publication, 11 September 2013; doi:10.1038/hdy.2013.86

Keywords: gene expression; spermatogenesis; MSCI; gene movement; genome evolution

INTRODUCTION

Although the X chromosome is nearly identical to the autosomes in its gene density and organisation, genes residing on the X chromosome experience a very different environment than autosomal genes in terms of natural selection and gene expression (Vicoso and Charlesworth, 2006). The ploidy of the X chromosome differs between the sexes, with females having two copies and males having only one, and this has several important consequences. First, over the course of its evolution, the X chromosome is present twice as often in females as in males. This may lead to 'feminization' of the X chromosome and the accumulation of sexually antagonistic mutations with dominant female-beneficial effects (Rice, 1984; Charlesworth *et al.*, 1987; Sturgill *et al.*, 2007). Second, the hemizygosity of the X chromosome in males allows selection to be more effective on X-linked than autosomal recessive mutations. This may result in faster adaptive evolution at X-linked loci and the accumulation of sexually antagonistic mutations with recessive male-beneficial effects on the X chromosome (Rice, 1984; Charlesworth *et al.*, 1987; Vicoso and Charlesworth, 2006; Baines *et al.*, 2008). Third, the difference in copy number between the X chromosome and the autosomes can create an imbalance in expression, which is often overcome by mechanisms of dosage compensation (Mank, 2009; Vicoso and Bachtrog, 2009).

A fourth difference between the X chromosome and the autosomes is that the X chromosome appears to be transcriptionally silenced in the male germline, a phenomenon also known as meiotic sex chromosome inactivation (MSCI). MSCI was proposed on the basis of cytological and genetic observations (Lifschytz and Lindsley, 1972). For example, precocious condensation of the X chromosome in spermatocytes has been reported in various species, including *Drosophila melanogaster* (for example, Henking, 1891; Cooper, 1951). In *D. pseudoobscura*, this condensation has been observed for the ancestral X chromosome but not for the neo-X, which was derived recently from an autosome (Lifschytz and Lindsley, 1972). However, cytological studies have produced conflicting results (McKee and Handel, 1993) and their support for MSCI should be considered tentative (Cooper, 1951).

More recent studies have provided empirical support for MSCI in a variety of species, including mammals (Richler *et al.*, 1992; Handel *et al.*, 1994; Turner, 2007), *Caenorhabditis elegans* (Fong *et al.*, 2002; Kelly *et al.*, 2002) and *D. melanogaster* (Hense *et al.*, 2007; Vibranovski *et al.*, 2009a). However, there is currently debate regarding the extent of X-linked germline expression silencing in *Drosophila* and whether it occurs through the same mechanism described as MSCI in other taxa. Vibranovski *et al.* (2009a) performed a microarray analysis of gene expression in dissected regions of testes that were enriched for mitotic and meiotic cells and found a significant excess of genes whose expression was downregulated in the meiotic region, which is consistent with MSCI. In contrast, subsequent studies by Meiklejohn *et al.* (2011) and Mikhaylova and Nurminsky (2011) failed to find evidence for MSCI. This discrepancy has been attributed to the statistical methods that were employed, and a re-analysis of the data under the statistical framework of Vibranovski *et al.* (2009a) revealed a significant excess of meiotically downregulated genes in both data sets (Vibranovski *et al.*, 2012). However, it has been pointed out that, even if the effect is significant, the X-chromosomal downregulation seen in *Drosophila* is much weaker than the well-known MSCI that occurs in mammals (Mikhaylova and Nurminsky, 2012).

[1]Department of Ecology and Evolution, University of Chicago, Chicago, IL, USA and [2]Department of Biology II, University of Munich (LMU), Planegg-Martinsried, Germany
Correspondence: Professor J Parsch, Department of Biology II, University of Munich (LMU), Grosshaderner Strasse 2, 82152 Planegg-Martinsried, Germany.
E-mail: parsch@bio.lmu.de
Received 29 April 2013; revised 27 June 2013; accepted 29 July 2013

Meiklejohn *et al.* (2011) also reported an absence of dosage compensation in the male germline, which results in an average 1.5-fold reduction in the expression of X-linked genes relative to autosomal genes. However, this result has been called into question, as both Meiklejohn *et al.* (2011) and Deng *et al.* (2011) found that the difference between X-linked and autosomal gene expression was greater in the testes of wild-type males than those of *bag of marbles* mutants, in which germ cell differentiation does not progress beyond mitosis. This suggests that at least some level of dosage compensation occurs in mitotic cells of the male germline, and suppression of X-chromosomal gene expression increases at meiosis (Deng *et al.*, 2011, but see Meiklejohn and Presgraves, 2012).

Hense *et al.* (2007) showed that autosomal insertions of a transgenic construct containing the promoter of the testis-specific *ocnus* (*ocn*) gene fused to a *lacZ* reporter gene had significantly higher expression than X-linked insertions of the same construct. As the copy number of the reporter gene was the same for both the autosomal and X-linked insertions, these results could not be explained by a lack of dosage compensation and, thus, suggested that another mechanism functions to suppress X-linked gene expression in the male germline. However, a limitation of the Hense *et al.* (2007) study was that it used only a single promoter sequence that came from an autosomal gene. Thus, it is not known whether the results are relevant to other promoters and, in particular, to promoters of X-linked testis-expressed genes, which presumably have evolved to provide high expression in the male germline. In other words, the previous experiment showed that relocating an autosomal gene to the X chromosome decreased its expression but not that the relocation of an X-linked gene to an autosome increased its expression. In the present study, we demonstrate the latter using transgenic reporter genes driven by promoter sequences of three different X-linked testis-expressed genes (*CG10920*, *CG12681* and *CG1314*). In all cases, we find significantly higher expression of transgenes inserted on the autosomes relative to those inserted on the X chromosome. Our results provide direct experimental evidence for the general, dosage-independent transcriptional suppression of X-linked genes during spermatogenesis.

MATERIALS AND METHODS

Transformation vector construction

Putative promoter sequences of three X-linked genes (*CG10920*, *CG12681* and *CG1314*) were PCR-amplified from genomic DNA of the *Canton S* strain of *D. melanogaster*. The *CG10920* promoter corresponds to bases 7 748 179–7 748 758 of the X chromosome (FlyBase release 5.50). The *CG12681* promoter corresponds to bases 4 769 051–4 769 815, and the *CG1314* promoter corresponds to bases 20 740 370–20 740 877. All of the amplified sequences lie just upstream of their respective coding sequences and end at base –28 (*CG10920*), –10 (*CG12681*) and –4 relative to the start codon.

The amplified PCR products were cloned directly into the pCR2.1-TOPO vector (Invitrogen, Carlsbad, CA, USA). The identity and orientation of the PCR fragments were confirmed by restriction analysis. A 3.6-kb *Not*I fragment of the pCMV-SPORT-βgal plasmid (Invitrogen) containing the *Escherichia coli lacZ*-coding region was cloned into the *Not*I site of the promoter-containing plasmid. Afterward, we performed restriction analysis to ensure that both the promoter and the *lacZ*-coding sequence were in the same transcriptional orientation. In a final step, an *Spe*I/*Xba*I fragment containing both the promoter and the *lacZ*-coding sequence was ligated into the p*P[wFl]* transformation vector (Siegal and Hartl, 1996). This vector is derived from the *P* transposable element and contains the *D. melanogaster white* (*w*; here in the form of *mini-white*) gene as a selectable marker.

Germline transformation

All transformation vectors were purified with the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) and were eluted from the column with injection buffer (0.1 mM sodium phosphate pH 6.8, 5 mM KCl). Vector DNA at a concentration of 200 ng μl^{-1} was used for microinjection of early-stage embryos of the strain *yw*; *Δ2-3, Sb/TM6*. The stable genomic *P*-element transposase *Δ2-3* on the third chromosome served as the source of transposase. After microinjection, all surviving flies were crossed to a *yw* strain to remove the transposase source and establish stable lines. The offsprings of this cross were screened for red eye colour (imparted by the wild-type w^+ gene of the vector), which was the diagnostic for stable germline transformants. Additional mobilisations of transgenes to and from the X chromosome were carried out through genetic crosses with the *Δ2-3* transposing-containing stock as described previously (Hense *et al.*, 2007).

The chromosomal location of each transgene (X or autosome) was mapped initially by genetic crosses. Transformed males were mated to *yw* females, and inheritance of the w^+ marker was observed in the next generation. Transformed lines with X-linked insertions were identified as those producing only daughters that carry the w^+ allele. Subsequently, the exact chromosomal position of each transgene insertion was determined using inverse PCR (Bellen *et al.*, 2004). Briefly, genomic DNA was digested with *Hpa*II or *Hinp*1I, and the resulting fragments were self-ligated with T4 DNA-Ligase (New England Biolabs, Ipswich, MA, USA). The target sequence, the inserted expression construct, was amplified with two primer pairs either Pry1 (5′-CCTTAGCATG TCCGTGGGGTTTGAAT-3′) and Pry2 (5′-CTTGCCGACGGGACCACCTTAT GTTATT-3′) or Plac1 (5′-CACCCAAGGCTCTGCTCCCACAAT-3′) and Plac4 (5′-ACTGTGCGTTAGGTCCTGTTCATTGTT-3′). The resulting PCR products were sequenced using the above primers and BigDye v1.1 chemistry on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA, USA). DNA sequences were used for a BLAST search of the *D. melanogaster* genome (FlyBase release 5.50) to determine the exact position of transgene insertion.

β-galactosidase assays

To avoid any confounding effects of transgene dosage on comparisons when comparing transformant flies with X-linked and autosomal insertions, all β-galactosidase assays were performed on flies heterozygous (autosomal) or hemizygous (X-linked) for the transgene insertion. These flies were generated by mating transformants to a *yw* stock. Offsprings were collected separated by sex shortly after eclosion and then maintained in standard food vials for 4–6 days prior to protein extraction.

For each enzymatic assay, six flies were homogenised in 150 μl of a buffer containing 0.1 M Tris-HCl, 1 mM EDTA and 7 mM 2-mercaptoethanol at pH 7.5. The homogenate was kept on ice for 15 min, then centrifuged at 12 000 *g* for 15 min at 4 °C. Enzymatic assays were performed using 50 μl of supernatant and 50 μl of assay buffer (200 mM sodium phosphate pH 7.4, 2 mM $MgCl_2$, 100 mM 2-mercaptoethanol) containing 1.33 mg ml^{-1} o-nitro-phenyl-β-D-galactopyranoside. β-galactosidase activity was measured spectrophotometrically at a wavelength of 420 nm over a period of 45 min at 25 °C. The slope of the absorbance in relation to the incubation time was used to determine the amount of β-galactosidase and the relative expression between the autosomal and X-linked insertions. For each transformed line, β-galactosidase activity was measured for three biological replicates, each with two technical replicates.

Testis *in situ* hybridisations

Testes were dissected from males that were either heterozygous (autosomal) or hemizygous (X-linked) for the reporter gene insertion and were used for whole-tissue *in situ* hybridisations following the procedure described by Morris *et al.* (2009). The specific lines used for *in situ* hybridisation are indicated in Supplementary Table S1. The probe was prepared using specific forward (5′-CAAAACTCTCAAGCAGCA-3′) and reverse (5′-GATGTGGATTGGCGA TAA-3′) primers to amplify ~1 kb of the pCMV-SPORT-βgal plasmid, which included a portion of the *lacZ*-coding region as well as the T7 promoter of the vector. The PCR products were purified using the QIAquick PCR Purification Kit (Qiagen), and an antisense RNA DIG-labelled probe was synthetised using T7 RNA polymerase (Roche, Basel, Switzerland) and DIG RNA Labeling Mix (Roche) as described by the manufacturer. Testes from autosomal and X-linked transformants were processed in parallel and a constant staining time of 1.5 h was used for all samples.

Quantitative reverse-transcription PCR (qRT-PCR)

Total RNA was extracted from flies heterozygous (or hemizygous) for the transgene insertion using Trizol (Invitrogen) and following the manufacturer's protocol. Beginning with 5 µg of total RNA, DNaseI treatment was carried out for 1 h at room temperature. Afterwards, the RNA was reverse-transcribed using the Superscript II reverse transcriptase and random hexamer primers (Invitrogen). A custom-designed TaqMan probe (Applied Biosystems) was used to quantify relative *lacZ* mRNA abundance using a Bio-Rad CFX 96 real-time PCR machine (Bio-Rad, Hercules, CA, USA). As an internal reference, a probe to the ribosomal protein gene *RpL32* (probe number Dm 02151827_g1) was used. Relative transcript abundance was measured as the difference in threshold cycle (ΔC_t) between the target and the reference gene. The difference in transcript abundance between lines with X-linked and autosomal transgene insertions was measured as the average difference in ΔC_t among lines ($\Delta\Delta C_t$).

Quantiative analysis of *mini-white* expression

As a proxy for *mini-white* expression, we measured eye pigmentation in 4- to 6-day-old flies of both sexes using the approach of Majumder *et al.* (2009). Briefly, 20 heads were homogenised in 50 µl AEA (30% EtOH, 0.1% concentrated HCl) buffer and incubated at 22 °C for 30 min while shaking at 800 r.p.m. Afterwards, 1 µl of 0.5% H_2O_2 was added and the solution was centrifuged for 10 min at 10 000 *g*. The supernatant was used for spectrophotometrical measurement of the eye pigmentation at 480 nm. In total, we performed four replicate measurements (two biological replicates, each with two technical replicates) for each genotype and sex. In all cases, we used flies carrying only a single copy of the transgene (that is, males were either hemizygous or heterozygous and females were heterozygous for the insertion).

RESULTS

Functional analysis of three X-linked, testis-specific promoters

To functionally test for an increase in male germline gene expression associated with escaping the X chromosome, we performed experiments using the upstream regulatory sequences of three X-linked, testis-specific genes: *CG10920*, *CG12681* and *CG1314*. These genes are located in different regions of the X chromosome and were chosen because they show significantly male- and testis-biased expression (Table 1). In addition, for all three genes the McDonald–Kreitman test (McDonald and Kreitman, 1991) indicates a significant excess of amino-acid replacements between *D. melanogaster* and its sister species *D. simulans* (Baines *et al.*, 2008), which is a hallmark of adaptive evolution.

As functional information about the regulatory sequences of *CG10920*, *CG12681* or *CG1314* was not available, we identified putative promoter sequences responsible for the testis expression of the three genes by comparative sequence analysis. Using aligned upstream sequences from *D. melanogaster*, *D. simulans*, *D. yakuba*, *D. erecta* and *D. sechellia*, we chose conserved regions of 580 bp (*CG10920*), 765 bp (*CG12681*) and 508 bp (*CG1314*) for further functional analysis (see Materials and methods).

Table 1 Summary of genes used in promoter analysis

Gene	*Cytogenetic map position*	*Male/female expression*[a]	*Testis/carcass expression*[b]	α[c]	*MK-test P-value*[d]
CG10920	7C	4.75	76.7	0.65	0.010
CG12681	4D	12.52	96.3	0.77	0.049
CG1314	19E	7.60	112.3	0.86	0.001

[a]Ratio of male-to-female expression from Sebida database (release 3.0; Gnad and Parsch, 2006).
[b]Ratio of testis-to-carcass expression from FlyAtlas database (Chintapalli *et al.*, 2007).
[c]Estimated proportion of positively selected amino-acid replacements (Smith and Eyre-Walker, 2002).
[d]*P*-value of McDonald and Kreitman (1991) test.

Putative promoter sequences were fused to the *E. coli lacZ* gene (encoding β-galactosidase) and cloned into the p*P[wFl]* transformation vector (Siegal and Hartl, 1996) (Figure 1). Stably transformed *D. melanogaster* strains were generated by embryo microinjection and subsequent genetic crosses. We recovered eight independent autosomal insertions each of the *CG10920*, *CG12681* and *CG1314* reporter gene constructs. β-galactosidase enzymatic assays indicated that all three reporter gene constructs showed highly male-biased expression (Table 2). In all cases, the difference in expression between males and females was highly significant (Mann–Whitney test, $P<10^{-4}$). Additionally, we compared the β-galactosidase activity in dissected testes to that in the remaining carcass of male flies transformed with each reporter gene construct. In all cases, expression was at least 140-fold higher in the testes than in the carcass. Furthermore, *in situ* hybridisations indicated that there was a high level of reporter gene expression in the testis (Figure 2a). All three constructs showed lower expression in the apical tip of the testis, which is enriched for mitotic cells, than in the mid- and posterior testes, which are enriched for meiotic and post-meiotic cells, respectively (Figure 2b). This pattern was especially pronounced for the *CG10920* and *CG12681* constructs (Figure 2b). The *CG1314* construct consistently displayed lower reporter gene expression than the other two constructs (Table 2), presumably because the *CG1314* promoter fragment was a relatively weak driver of gene expression.

Comparison of X-linked and autosomal reporter gene insertions

In addition to the autosomal insertions described above, we also recovered seven, eight and nine X-linked insertions of the *CG10920*, *CG12681* and *CG1314* reporter constructs, respectively. As expected, all three constructs showed male- and testis-biased expression (Table 2). In all cases, the difference in expression between males and females was significant (Mann–Whitney test, $P<10^{-4}$). Additionally, we compared the expression in the dissected testis with that in the remaining carcass of male flies transformed with each reporter

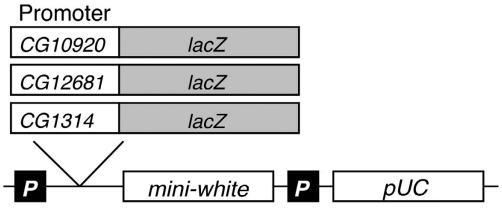


Figure 1 Reporter gene constructs. Promoter sequences of three X-linked, testis-expressed genes were fused to the *E. coli lacZ* reporter gene and independently inserted into the p*P[wFl]* transformation vector (Siegal and Hartl, 1996). This vector contains terminal repeat sequences of a *Drosophila* transposable element (*P*) and the *mini-white* gene as a selectable marker (eye colour). The portion of the plasmid required for replication in *E. coli* is labelled 'pUC'.

Table 2 Mean β-galactosidase activity of transformants

Promoter	*Autosomal*			*X-linked*		
	n	*Male*	*Female*	n	*Male*	*Female*
CG10920	8	6.83 (2.42)	0.08 (0.08)	7	2.44 (0.32)	0.01 (0.10)
CG12681	8	5.20 (1.34)	0.14 (0.10)	8	1.35 (0.19)	0.11 (0.06)
CG1314	8	2.08 (0.29)	0.14 (0.09)	9	0.72 (0.22)	0.05 (0.07)

n, number of independent transgene insertions. s.d.'s are given in parentheses.

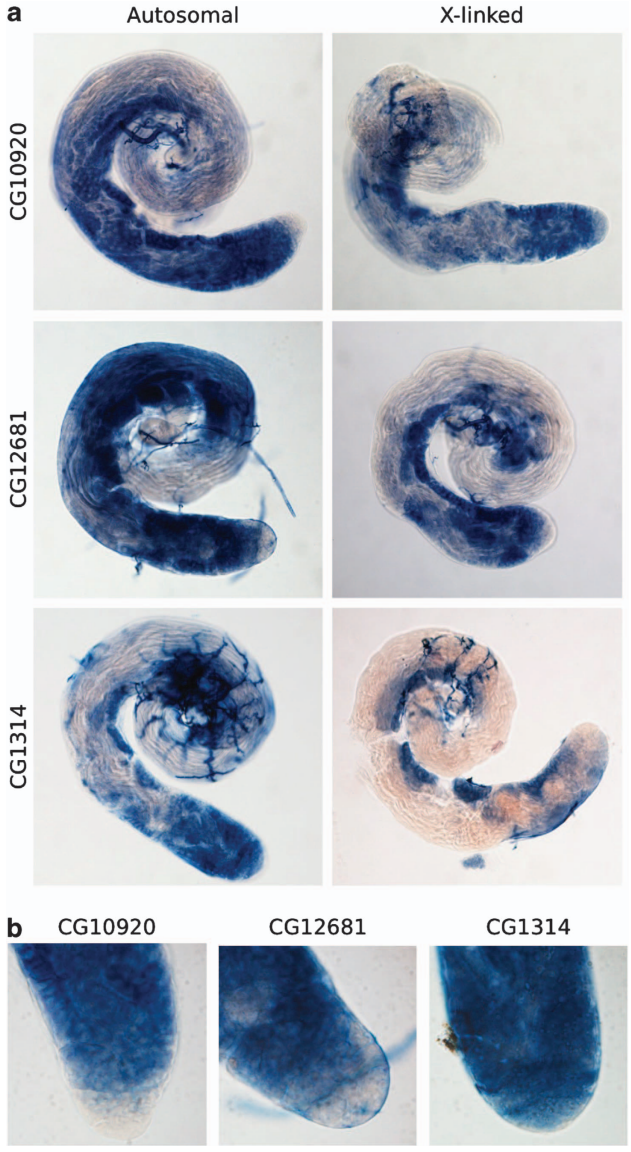


Figure 2 *In situ* hybridisation of a *lacZ* probe in the testes. (**a**) Testes were dissected from males containing a single autosomal or X-linked insertion of each reporter gene construct and hybridised with a probe specific to the *lacZ* reporter gene. Dark blue areas indicate the presence of reporter gene mRNA. The magnification is ×200. (**b**) Enlargement of the testis apex from males with autosomal insertions of each reporter gene construct. The *CG10920* and *CG12681* show very low expression in the apical tip, which is enriched for mitotic cells. The magnification is ×600.

gene construct. In all cases, expression was at least 12-fold higher in the testes than in the carcass.

Although the X-linked insertions of all three promoter constructs showed expression in the testis (Figure 2a), their level of expression was significantly lower than that of autosomal insertions (Figure 3). The average differences in β-galactosidase activity between autosomal and X-linked insertions were 2.8-, 3.9- and 2.9-fold for the *CG10920*, *CG12681* and *CG1314* reporter constructs, respectively.

To confirm the above results at the level of transcript abundance, we performed qRT-PCR to estimate relative levels of *lacZ* mRNA. For all three reporter gene constructs, the *lacZ* transcript abundance was significantly higher for autosomal insertions than for X-linked insertions (Figure 3). The average differences in *lacZ* mRNA concentration between autosomal and X-linked insertions were 2.33-, 3.01- and 3.32-fold for the *CG10920*, *CG12681* and *CG1314* reporter constructs, respectively. Thus, the estimates of transcript abundance agree well with the estimates of protein abundance. Furthermore, there was a strong correlation between expression levels measured using qRT-PCR and β-galactosidase activity (*CG10920*: Spearman's $\rho = 0.78$; $P < 10^{-5}$; *CG12681*: $\rho = 0.82$, $P < 10^{-7}$; *CG1314*: $\rho = 0.66$, $P < 0.0025$).

Fine-scale mapping of transgene insertions

In order to determine the local context of the transgene insertions, we performed inverse PCR to map their precise position in the genome (Bellen *et al.*, 2004). With this method, we were able to map eight autosomal and seven X-linked insertions for the *CG10920* construct, eight autosomal and eight X-linked insertions for the *CG12681* construct, and eight autosomal and nine X-linked insertions for the *CG1314* construct (Figure 4). Overall, we were able to precisely map 88% of the autosomal insertions and 92% of the X-linked insertions. For all constructs, the insertions were distributed throughout the euchromatin and most (63%) were associated with genes (within a 5′ untranslated region, coding region or intron; Supplementary Table S1). The remaining insertions were in intergenic regions; however, all were within 10 kb of an annotated gene. There were no significant differences in expression among transgenes inserted into different gene regions and, within gene regions, autosomal transgene expression was always greater than X-linked transgene expression. In addition, the genomic regions surrounding autosomal transgenes did not show a significant excess of genes with testis-enriched expression in comparison to the regions surrounding X-linked transgenes (Supplementary Table S2) Thus, the observed differences in expression between autosomal and X-linked transgenes cannot be explained by differences in the local context into which they are inserted.

Analysis of somatic *mini-white* expression

To determine the effect of X linkage on gene expression in somatic tissues, we took advantage of the fact that all of our transformation vectors contained the *mini-white* gene as a selectable marker (Figure 1). This gene is derived from the X-linked *white* gene and shows enriched expression in the eye, where the degree of pigmentation (ranging from pale yellow to dark red) serves as an indicator of *mini-white* expression. We performed a spectrophotometric assay to quantify the amount of red pigment in the eyes of males and females of all of our transformed lines, using flies that were either heterozygous or hemizygous for the transgene insertion (that is, the gene dose of *mini-white* was always one). Comparison of the expression of individual inserts between males and females revealed a general pattern of higher expression in males (Table 3). Of the 48 independent insertions, 39 showed higher expression in males (sign test, $P < 0.0001$). This difference is mainly attributable to X-linked insertions: 23 out of 24 X-linked insertions showed higher expression in males (sign test, $P < 0.0001$), whereas 16 out of 24 autosomal insertions showed higher expression in males (sign test, $P = 0.08$). These results are consistent with there being dosage compensation of X-linked *mini-white* gene expression in somatic tissues of males.

Unlike the testis-promoter constructs, which showed significantly greater expression when inserted on autosomes, the *mini-white* gene showed a trend towards greater expression when inserted on the X chromosome (Table 3). The difference between X-chromosomal and autosomal expression was marginally significant in males (Mann–Whitney test, $P = 0.060$) but not significant in females

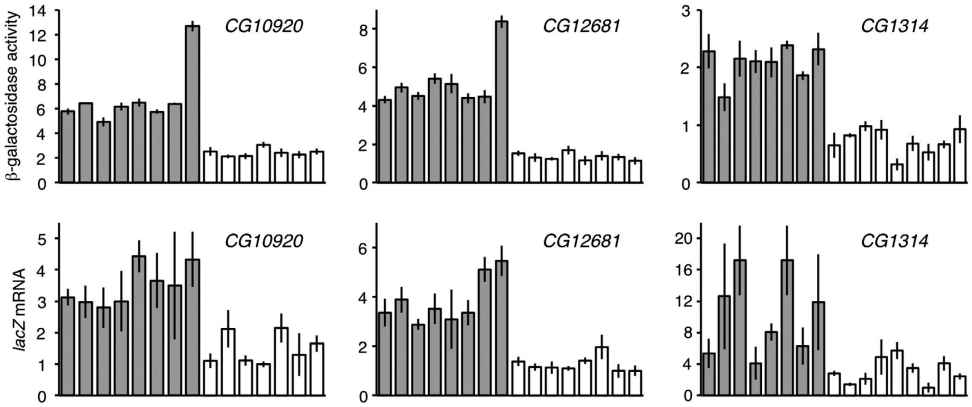


Figure 3 Expression of autosomal and X-linked reporter gene insertions in adult males. The upper row shows the mean β-galactosidase activity of transformants with autosomal (grey bars) and X-linked (open bars) insertions of each reporter gene construct. Each bar represents an independent insertion at a different genomic location. The lower row shows the relative expression of the *lacZ* gene as determined using qRT-PCR. For each construct, the expression of the lowest line is set to 1 and all other expression values are scaled accordingly. The order of the bars corresponds to that in the upper row. In all cases, autosomal expression was significantly greater than X-linked expression (Mann–Whitney test, $P<0.001$). Error bars indicate the s.d.

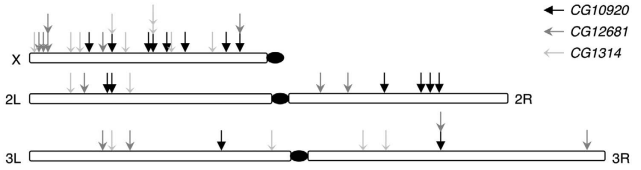


Figure 4 Map of transgene insertion locations. The precise chromosomal location of each insertion was determined using inverse PCR. Each arrow indicates an insertion at a unique site. Multiple arrows at the same position do not indicate insertions at the same site but indicate insertions that are too close to each other (within 400 kb) to be distinguished on the scale of the figure.

Table 3 ***mini-white*** **expression in the eye**

Location	n[a]	*Male OD_{480} (s.e.m.)*	*Female OD_{480} (s.e.m.)*	*M>F*[b]	P-*value*[c]
Autosome	24	86.2 (12.5)	53.4 (8.5)	16	0.0758
X chromosome	24	158.4 (28.6)	58.6 (5.9)	23	0.0001

Abbreviation: OD, optical density.
[a]Number of independent transgene insertions.
[b]Number of insertions showing greater expression in males than females.
[c]*P*-value of sign test comparing male and female expression.

($P=0.120$). These results are consistent with dosage compensation of X-linked genes in somatic tissues and indicate that reduced X-linked expression is not a general property of our transformation vector or its preferred integration sites but instead is a feature of the male germline.

DISCUSSION

We find that the level of testis expression driven by three different X-linked promoters is significantly increased when reporter genes are relocated to the autosomes. In combination with previous experiments that showed a reduction in testis expression when an autosomal promoter was moved to the X chromosome (Hense *et al.*, 2007; Kemkemer *et al.*, 2011), our results demonstrate that the X chromosome presents an unfavourable environment with respect to expression in the male germline. The three X-linked promoters used in the current study do not share sequence homology with each other or with other known testis-specific regulatory elements, which suggests that either they do not have a simple, shared regulatory mechanism or that any common regulatory sequences have diverged so extensively that they cannot be detected by a homology search. The *CG12681* promoter contains a 20-bp sequence that is identical to a sequence found upstream of the male- and testis-biased gene *CG5732* on chromosome arm 3R (Gnad and Parsch, 2006; Chintapalli *et al.*, 2007). This region is predicted to contain binding sites for the Even-skipped and Zerknuellt transcription factors (Messeguer *et al.*, 2002). However, both of these transcription factors are known to function during early embryogenesis and have no known function in spermatogenesis nor do they show enriched expression in males or testis (Gnad and Parsch, 2006; Chintapalli *et al.*, 2007).

The exact mechanism by which X-chromosomal gene expression is suppressed in the *Drosophila* male germline is unknown. One possibility is that a lack of dosage compensation in the male germline leads to a general reduction in the expression of X-linked genes. Whether or not dosage compensation occurs in the *Drosophila* male germline is currently a subject of debate (see Introduction). However, even a complete absence of dosage compensation cannot explain our observations. This is because all of the transformed flies used in the expression assays carried only a single copy of the reporter gene. Thus, the gene dose was equal in X-linked and autosomal transformants. For this reason, our experiments are conservative, as any amount of dosage compensation would be expected to increase the level of X-chromosomal gene expression relative to that of the autosomes. Indeed, we find that the *mini-white* gene, which is present in all of our transformation vectors, shows expression patterns consistent with dosage compensation in the somatic (eye) tissue (Table 3). This is in agreement with previous studies reporting that, in males, X-linked alcohol dehydrogenase (*Adh*) transgenes show higher expression than those inserted on autosomes (Laurie-Ahlberg and Stam, 1987; Parsch *et al.*, 1997). These findings indicate that the reduced X-linked expression seen for our testis-expressed transgenes is not an artifact of the *P*-element vector used for transformation, as this pattern is not observed for somatically expressed transgenes.

Another possible mechanism is MSCI, the transcriptional inactivation of the X chromosome during meiosis. A microarray analysis of gene expression during different stages of spermatogenesis indicated

that there is a significant excess of X-linked genes that are down-regulated during the transition from mitosis to meiosis (Vibranovski *et al.*, 2009a), which is consistent with MSCI. However, the average decline in expression between the two stages was relatively small, suggesting that a wholesale inactivation of the X chromosome does not occur. In addition, microarray and qRT-PCR data suggest that the expression of some spermatocyte-specific genes (including *CG10920*, *CG12681*, *CG1314* and *ocn*) increases during the mitosis–meiosis transition (Vibranovski *et al.*, 2009a; Meiklejohn *et al.*, 2011; Mikhaylova and Nurminsky, 2011). Our reporter gene experiments also revealed that the mRNA abundance of all three promoter constructs was relatively high in the regions of the testis enriched with meiotic and post-meiotic cells. However, there was very little expression of the *CG10920* and *CG12681* constructs in the the apical tip of the testis, which is enriched with mitotic cells (Figure 2b). This observation has two important implications. First, it indicates that the X chromosome is not completely inactivated at meiosis. Thus, the suppression of X-linked germline expression appears to be mechanistically different from the MSCI known to occur in mammals. Second, it suggests that contamination between stages may be an important confounding factor in studies that compare expression between dissected regions of the testes (for example, Vibranovski *et al.*, 2009a; Meiklejohn *et al.*, 2011). This is because the signal of expression observed in dissected 'mitotic' samples may come primarily from contamination with meiotic cells (Vibranovski *et al.* 2012). On the other hand, mRNA that is transcribed in mitotic cells will persist in meiotic and post-meiotic cells and will be detected by transcriptomic and reporter gene studies. This could explain why the observed expression difference between X-linked and autosomal insertions of our reporter gene constructs (approximately threefold) is greater than the expression difference between endogenous autosomal and X-linked genes detected using high-throughput RNAsequencing ($\sim$1.5-fold; Meiklejohn *et al.*, 2011). As the genes used in our study show very low expression in mitotic cells, there should be less residual signal of mitotic transcription for these genes than for many endogenous genes.

An excess of gene duplication from the X chromosome to the autosomes has been observed across the *Drosophila* genus (Betrán *et al.*, 2002; Meisel *et al.*, 2009; Vibranovski *et al.*, 2009b). This is mainly attributable to there being a significant over-representation of retroduplicate pairs in which the parental gene is X-linked and the retrogene is autosomal (Meisel *et al.*, 2009). Furthermore, X-to-autosome retroduplicates tend to show expression in the testis (Meisel *et al.*, 2009). It has been proposed that selection favours retroduplicate gene copies with beneficial functions in the testis that escape the X chromosome, as they can achieve higher levels of testis expression when they are autosomal (Betrán *et al.,* 2002). Our results are consistent with this interpretation, as all of our reporter gene constructs showed higher expression in the testis when they were relocated from the X chromosome to an autosome. In the case of retrotransposition, it is typically assumed that flanking regulatory elements are not duplicated and that new regulatory sequences are acquired from the insertion site, either by recruiting pre-existing elements or by evolving them *de novo* (Bai *et al.*, 2008). Thus, it is likely that the increased expression of the autosomes in the male germline makes it easier to recruit or evolve regulatory sequences that drive high expression in the testis. Although it is difficult to establish a direct link between an increase in a gene expression in the testis and an increase in male reproductive fitness, previous findings that testis-expressed genes show high rates of adaptive evolution at the protein level (Pröschel *et al.*, 2006; Baines *et al.*, 2008) suggest that positive selection has an important role in the evolution of genes expressed in the male germline. Similarly, positive selection has been shown to act on testis-expressed retrogenes that have relocated from the X chromosome to an autosome (Betrán and Long, 2003; Quezada-Diaz *et al.*, 2010; Tracy *et al.*, 2010).

CONCLUSION

Previous work in *D. melanogaster* found that a transgenic reporter gene had significantly lower expression in the testis when inserted onto the X chromosome than the autosomes (Hense *et al.*, 2007). This result is consistent with the suppression of X-chromosomal gene expression in the male germline. However, a caveat to the previous study was that only a single promoter derived from an autosomal gene was used (Vibranovski *et al.*, 2012). The present study shows that the suppression of X-chromosomal gene expression extends to three additional promoters derived from X-linked genes. Thus, the observed X suppression in the male germline is independent of the promoter or its chromosome of origin. The use of transgenes allows us to examine the expression of identical genes within different chromosomal contexts and to control for gene dose, both of which are not possible in genome-wide studies of endogenous gene expression. Importantly, it allows us to rule out an absence of dosage compensation in the male germline as a cause of the reduced X-linked expression. The expression patterns of the reporter genes (Figure 2) suggest that the difference in expression between the X chromosome and the autosomes is most pronounced in meiotic and post-meiotic cells. This could be caused by a mechanism similar to the MSCI that occurs in mammals. However, the high expression of all three reporter genes meiotic and post-meiotic cells suggests that if MSCI occurs in *Drosophila*, it is to a much lesser extent that the MSCI that occurs in mammals. Regardless of the specific molecular mechanism, our results demonstrate that X linkage limits the expression of genes in the male germline and provide experimental support for a selective process driving the excess of X-to-autosome retroduplication that has been observed across the *Drosophila* genus.

DATA ARCHIVING

Data have been deposited at Dryad: doi:10.5061/dryad.qt652.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Hedwig Gebhart for technical assistance and the Munich evolutionary biology group for helpful comments and discussions. We also thank four anonymous reviewers for their constructive comments on the manuscript. This work was supported by the *Deutsche Forschungsgemeinschaft* grants PA 903/5 and PA 903/6.

Bai Y, Casola C, Betrán E (2008). Evolutionary origin of regulatory regions of retrogenes in *Drosophila*. *BMC Genomics* **9**: 241.

Baines JF, Sawyer SA, Hartl DL, Parsch J (2008). Effects of X-linkage and sex-biased gene expression on the rate of adaptive protein evolution in *Drosophila*. *Mol Biol Evol* **25**: 1639–1650.

Bellen HJ, Levis RW, Liao G, He Y, Carlson JW, Tsang G *et al.* (2004). The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* **167**: 761–781.

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

Betrán E, Thornton K, Long M (2002). Retroposed new genes out of the X in *Drosophila*. *Genome Res* **12**: 1854–1859.

Charlesworth B, Coyne JB, Barton NH (1987). The relative rates of evolution of sex-chromosomes and autosomes. *Am Nat* **130**: 113–146.

Chintapalli VR, Wang J, Dow JA (2007). Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet* **39**: 715–720.
Cooper KW (1951). Normal spermatogenesis in *Drosophila*. In: Demerec M (ed) *Biology of Drosophila*. Hafner: New York, NY, USA, pp 1–61.
Deng X, Hiatt JB, Nguyen DK, Ercan S, Sturgill D, Hillier LW *et al.* (2011). Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. *Nat Genet* **43**: 1179–1185.
Fong Y, Bender L, Wang W, Strome S (2002). Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of *C. elegans*. *Science* **296**: 2235–2238.
Gnad F, Parsch J (2006). Sebida: a database for the functional and evolutionary analysis of genes with sex-biased expression. *Bioinformatics* **22**: 2577–2579.
Handel MA, Park C, Kot M (1994). Genetic control of sex-chromosome inactivation during male meiosis. *Cytogenet Cell Genet* **66**: 83–88.
Henking H (1891). Über Spermatogenese und deren spermatogenesis Beziehung zur Entwicklung bei *Pyrrhocoris apterus L. Z Wiss Zool* **51**: 685–736.
Hense W, Baines JF, Parsch J (2007). X chromosome inactivation during *Drosophila* spermatogenesis. *PLoS Biol* **5**: e273.
Kelly WG, Schaner CE, Dernburg AF, Lee MH, Kim SK, Villenueve LM *et al.* (2002). X-chromosome silencing in the germline of *C. elegans*. *Development* **129**: 479–492.
Kemkemer C, Hense W, Parsch J (2011). Fine-scale analysis of X chromosome inactivation in the male germline of *Drosophila melanogaster*. *Mol Biol Evol* **28**: 1561–1563.
Laurie-Ahlberg CC, Stam LF (1987). Use of *P*-element-mediated transformation to identify the molecular basis of naturally occurring variants affecting *Adh* expression in *Drosophila melanogaster*. *Genetics* **115**: 129–140.
Lifschytz E, Lindsley DL (1972). The role of X-chromosome inactivation during spermatogenesis. *Proc Natl Acad Sci USA* **69**: 182–186.
Majumder P, Roy S, Belozerov VE, Bosu D, Puppali M, Cai HN (2009). Diverse transcription influences can be insulated by the *Drosophila* SF1 chromatin boundary. *Nucleic Acids Res* **37**: 4227–4233.
Mank JE (2009). The W, X, Y and Z of sex-chromosome dosage compensation. *Trends Genet* **25**: 226–233.
McDonald JH, Kreitman M (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
McKee BD, Handel MA (1993). Sex chromosomes, recombination, and chromatin conformation. *Chromosoma* **102**: 71–80.
Meiklejohn CD, Landeen EL, Cook JM, Kingan SB, Presgraves DC (2011). Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. *PLoS Biol* **9**: e1001126.
Meiklejohn CD, Presgraves DC (2012). Little evidence for demasculinization of the *Drosophila* X chromosome among genes expressed in the male germline. *Genome Biol Evol* **4**: 1007–1016.
Meisel RP, Han MV, Hahn MW (2009). A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. *Genome Biol Evol* **1**: 176–188.
Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM (2002). PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics* **18**: 333–334.
Mikhaylova LM, Nurminsky DI (2011). Lack of global meiotic sex chromosome inactivation, and paucity of tissue-specific gene expression on the *Drosophila* X chromosome. *BMC Biol* **9**: 29.
Mikhaylova LM, Nurminsky DI (2012). No severe and global X chromosome inactivation in meiotic male germline of *Drosophila*. *BMC Biol* **10**: 50.
Morris CA, Benson E, White-Cooper H (2009). Determination of gene expression patterns using *in situ* hybridization to *Drosophila* testes. *Nat Protoc* **4**: 1807–1819.
Parsch J, Tanda S, Stephan W (1997). Site-directed mutations reveal long-range compensatory interactions in the *Adh* gene of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **94**: 928–933.
Pröschel M, Zhang Z, Parsch J (2006). Widespread adaptive evolution of *Drosophila* genes with sex-biased expression. *Genetics* **174**: 893–900.
Quezada-Diaz JE, Muliyil T, Rio J, Betrán E (2010). *Drcd-1 related*: a positively selected spermatogenesis retrogene in *Drosophila*. *Genetica* **138**: 925–937.
Rice WR (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution* **38**: 735–742.
Richler C, Soreq H, Wahrman J (1992). X inactivation in mammalian testis is correlated with inactive X-specific transcription. *Nat Genet* **2**: 192–195.
Siegal ML, Hartl DL (1996). Transgene Coplacement and high efficiency site-specific recombination with the *Cre/loxP* system in *Drosophila*. *Genetics* **144**: 715–726.
Smith NG, Eyre-Walker A (2002). Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022–1024.
Sturgill D, Zhang Y, Parisi M, Oliver B (2007). Demasculinization of X chromosomes in the *Drosophila* genus. *Nature* **450**: 238–241.
Tracy C, Rio J, Motiwale M, Christensen SM, Betrán E (2010). Convergently recruited nuclear transport retrogenes are male biased in expression and evolving under positive selection in *Drosophila*. *Genetics* **184**: 1067–1076.
Turner JM (2007). Meiotic sex chromosome inactivation. *Development* **134**: 1823–1831.
Vibranovski MD, Lopes HF, Karr TL, Long M (2009a). Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. *PLoS Genet* **5**: e1000731.
Vibranovski MD, Zhang Y, Long M (2009b). General gene movement off the X chromosome in the *Drosophila* genus. *Genome Res* **19**: 897–903.
Vibranovski MD, Zhang YE, Kemkemer C, Lopes HF, Karr TL, Long M (2012). Re-analysis of the larval testis data on meiotic sex chromosome inactivation revealed evidence for tissue-specific gene expression related to the *Drosophila* X chromosome. *BMC Biol* **10**: 49.
Vicoso B, Bachtrog D (2009). Progress and prospects toward our understanding of the evolution of dosage compensation. *Chromosome Res* **17**: 585–602.
Vicoso B, Charlesworth B (2006). Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet* **7**: 645–653.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)

Table S2 Analysis of transgene insertion neighborhoods

I. Number of transgene insertions in or near testis-expressed genes

		Insertions in genes		*Insertions near genes (within 10 kb)*	
Chromosome	*Mapped*[a]	*Testis-enriched*[b]	*Testis-biased*[c]	*Testis-enriched*[b]	*Testis-biased*[c]
Autosomal	21	1	1	8	11
X-linked	22	1	1	6	8

[a]Excluding insertions inside genes for which expression data were not available from FlyAtlas

[b]Defined as having higher expression in testis than any other tissue (Chintapalli *et al*., 2007)

[c]Defined as having 2-fold higher expression in testis than in ovary (Chintapalli *et al*., 2007)

II. *P*-values (Fisher's exact test) for comparisons of autosomes and X chromosomes

Comparison	*Expression*	*P-value*
In genes	testis-enriched	1.00
In genes	testis-biased	1.00
Near genes	testis-enriched	0.53
Near genes	testis-biased	0.36

Table S1 Location and expression of transgene insertions

Promoter	Chrom.	Cyt. Loc.	Coordinate	Context	Affected Gene	Proximal Gene (10 kb)	Distal Gene (10 kb)	Testis enriched[a]	Testis biased[b]	Bgal mean	Bgal s.d.	RT-PCR mean	RT-PCR s.d.
CG10920	X	10D1	11516084	5' UTR	CG1817				near	2.16	0.26	1.12	0.17
CG12681*	3L	65D5	6972569	5'UTR	CG10060			near	near	5.42	0.26	3.51	0.61
CG12681	2R	46B1	5599879	5'UTR	CG1772			near	near	4.51	0.20	2.88	0.22
CG12681	3R	94E5	19016930	5'UTR	CG17894					4.40	0.27	3.34	0.51
CG1314	3L	80A2	22781342	5'UTR	CG14448			near	near	2.10	0.20	4.12	2.10
CG10920	2L	28B1	7576521	5'UTR	CG34374					5.76	0.24	3.13	0.26
CG1314	3R	85D22	5358515	Exon	CG9379					2.09	0.26	8.10	1.01
CG12681	3R	99F2	26214768	Exon	CG1469					4.47	0.35	5.10	0.50
CG10920*	2R	53D8	12670334	Exon	CG15920			near	near	4.93	0.37	2.80	0.65
CG10920	2L	27F3	7421490	Exon	CG5229				near	6.45	0.06	2.98	0.52
CG10920	3L	75B1	17955937	Exon	CG8127					6.40	0.10	3.50	1.72
CG10920	2R	49F10	9107394	Exon	CG4646			near	near	6.16	0.31	3.00	0.97
CG10920	2R	55C4	14244239	Exon	CG5580				near	6.49	0.34	4.43	0.51
CG1314*	X	15F3	17106995	Exon	CG18258			near	near	0.67	0.07	4.11	0.93
CG1314	X	10D8	11623204	Exon	inaF cluster					0.53	0.15	1.00	0.62
CG12681	X	1D2	828749	Exon	CG32815			near	near	1.52	0.15	1.38	0.19
CG1314	X	10B5	11590075	Exon	CG1830					0.68	0.13	3.40	0.62
CG1314	X	4B1	4014702	Exon	CG4857					0.82	0.04	1.43	0.24
CG1314	X	4D6	4823106	Exon	CG4068			in	in	0.98	0.08	2.13	0.77
CG10920	X	11E1	13022777	Exon	CG32638			near	near	3.06	0.23	1.00	0.10
CG1314*	2L	30B1	9387298	Intergen			CG3752			1.48	0.24	12.62	6.72
CG1314	3R	86E10	7393239	Intergen		CG6783	CG14709	near	near	2.39	0.07	17.08	4.38
CG12681	2R	43A2	3136383	Intergen		CG1851	CG11086		near	4.95	0.25	3.88	0.53
CG1314	2L	24C4	3788360	Intergen		CG31958		near		2.28	0.30	5.38	1.87

CG1314	3L	66A17	7860777	Intergen		CG12151	CG32364			2.15	0.31	17.19	4.43
CG12681	2L	25C1	5027473	Intergen		CG16858	CG4145			4.31	0.19	3.35	0.55
CG12681	3L	67B10	9498960	Intergen		CG3424	CG3408			5.15	0.49	3.09	1.20
CG10920	3R	94E1	18968035	Intergen		CG4637			near	12.68	0.42	4.34	0.88
CG1314	X	7C2	7802374	Intergen		CG10946	CG1444		near	0.92	0.10	4.90	2.26
CG12681	X	6E4	6879859	Intergen			CG14430			1.16	0.26	1.42	0.13
CG10920	X	17C2	18428513	Intergen			CG6500	near	near	2.28	0.29	1.30	0.68
CG10920	X	12F5	14720137	Intergen		CG9533	CG14408			2.42	0.34	2.15	0.46
CG10920	X	5C6	5780651	Intergen		CG16721		near	near	2.53	0.37	1.11	0.23
CG10920	2R	54B16	13347396	Intron	CG14478			near	near	5.75	0.17	3.66	0.88
CG12681*	X	2B17	1842812	Intron	CG3600					1.69	0.22	1.10	0.09
CG1314	X	12A9	13536139	Intron	CG11172					0.93	0.24	2.45	0.40
CG12681	X	18F4	19780935	Intron	CG11937					1.39	0.24	1.96	0.49
CG1314	X	8C4	8936538	Intron	CG42388					0.32	0.11	5.73	1.09
CG12681	X	1E5	1130460	Intron	CG3638					1.31	0.23	1.16	0.15
CG1314	X	1B2	323934	Intron	CG32816					0.65	0.22	2.82	0.31
CG12681	X	2B4	1513944	Intron	CG11491					1.24	0.09	1.14	0.25
CG10920*	X	7B6	7586656	Intron	CG12690					2.13	0.15	2.12	0.59
CG10920	X	18F3	19743488	Intron	CG11940					2.52	0.23	1.66	0.27
CG12681	Auto	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	8.38	0.33	5.46	0.62
CG1314	Auto	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.86	0.08	6.29	2.39
CG1314	Auto	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	2.32	0.28	11.86	6.10
CG12681	X	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.34	0.17	1.01	0.28
CG12681	X	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	1.14	0.20	1.00	0.21

*lines used for *in situ* hybridization (see Figure 2).

n.d., not determined.

β-galactosidase (Bgal) activity and mRNA abundance (RT-PCR) values are for adult males. RT-PCR expression values are relative within each promoter type.

[a]defined as having higher expression in testis than any other tissue (Chintapalli *et al*., 2007); "in" indicates within a gene, "near" indicates within 10 kb of a gene.

[b]defined as having 2-fold higher expression in testis than in ovary (Chintapalli *et al*., 2007); "in" indicates within a gene, "near" indicates within 10 kb of a gene.

General discussion

COMPARING the complete transcriptomes of two populations that have diverged is one of the approaches used to detect candidate genes for adaptive evolution. To narrow down the candidate genes that might be involved in a particular adaptive trait, assessing transcriptome differences at the tissue-specific level might facilitate linking gene expression differences and DNA polymorphism patterns with a beneficial phenotype. The different organs found in an organism have different and unique physiological roles that enable the individual to survive and reproduce in a particular ecosystem. In the case of the brain, it is the organ where all the information, either biotic or abiotic, collected with the sensorial organs is analyzed and processed, with the final result being an appropriate physiological or behavioral response. In our model system for the study of adaptive evolution, we use one *D. melanogaster* population from the ancestral species range (sub-Saharan Africa) and one from the derived species range (Europe), which allows us to study adaptation to novel environments. Clearly, the environment that northern European populations of *D. melanogaster* encounter differs greatly from the tropical environment of the ancestral species range. Factors like photoperiod, seasonality and exposure to freezing temperatures, different feeding and oviposition sites, and cues from predators and/or parasitoids might differ between a tropical and a temperate habitat and thus be factors that might induce adaptive evolution. From the RNA-seq data obtained from brains of an African and a European population of *D. melanogaster* we detected over three hundred candidate genes for adaptive evolution. We found many genes involved in olfactory and gustatory reception, genes involved in stress response, genes that are involved in behavioral responses and genes that might be involved in reproductive isolation (Table 4).

From behavioral assays, assortative mating has been found between African and cosmopolitan populations of *D. melanogaster*. African females prefer to mate with African males instead of cosmopolitan males, while cosmopolitan females do not exhibit any mating preference (Hollocher et al. 1997). From our RNA-seq brain data set, some of the genes we found to be differentially expressed could have a role in reproductive isolation. Genes like *beethoven* (*btv*), *retained* (*retn*) and *dissatisfaction* (*dsf*) have already been implicated in having a role in reproductive behavior. For example, *dsf* and *retn* are directly involved in female receptivity and the latter is also involved in male and female courtship behavior (Finley et al. 1998, Shirangi et al. 2006). Null mutants of *btv* display defective wing rowing behavior and sound perception, suggesting that *btv* might have an influence on the differential courtship behavior observed between African and cosmopolitan *D. melanogaster* (Eberl et al. 2000, Gleason 2005, Colegrave et al. 2000).

In insect courtship behavior, pheromone communication is an essential part of species-specific recognition, as well as in mate quality assessment and female receptivity (Smadja et al. 2009, Wicker-Thomas 2007). In *D. melanogaster,* 11-*cis*-vaccenyl acetate (cVA) is the only volatile pheromone that has been identified (Xu et al. 2005). This male specific pheromone induces aggregation behavior in both female and male flies and it also influences male courtship behavior (Xu et al. 2005, Ejima et al. 2007). *D. melanogaster* cVA acts through binding to the odorant receptor 67d (*Or67d*) and female and male null mutants for *Or67d* fail to show wild-type courtship behavior (Kurtovic et al. 2007). In our data set, *Or67d* is over-expressed in the brain of European flies in comparison to African flies. The difference in expression of *Or67d* could cause a stronger or a weaker response to cVA and thus cause changes in the behaviors controlled by cVA.

Another set of genes that are differentially expressed in the brain between African and European flies is genes coding for cuticular proteins (*Cpr65Ec*, *Cpr49Ae*, *desat2*, *Cpr72Ea*, *Cpr100A*). These genes are usually expressed in the cuticle and are important in species and gender recognition and in assessing fertility and mating status (Ferveur 2005). Changes in cuticular proteins and hydrocarbon profiles have been associated with inter- and intraspecific mating preferences (Ferveur 2005, Gleason et al. 2005, Legendre et al. 2007). We found cuticular proteins to be differentially expressed in the brains of African and European flies. At

present there is no literature available about the possible functions of cuticular proteins in the brain. Some cuticular proteins, like *Cpr49Ae*, had a medium-to-high expression level (up to 300 mapped reads per million), so contamination with cuticle coming from trachea can be neglected. In our fly populations, it remains to be tested how the differential expression of the genes mentioned above (*btv*, *retn*, *dsf, Or67d*, *desat2* and the cuticular proteins) could be involved in incipient speciation and /or assortative mating, as well as in other phenotypes that could have an effect on fitness.

Table 4. Candidate genes for local adaptation. Genes differentially expressed in the brains of European and African *D. melanogaster*.

Name	Symbol	Chromosome	Differential fold expression E/A
Sensorial detection			
Odorant receptor 88A	*Or88a*	3R	3.90
Odorant receptor 45b	*Or45b*	2R	3.5
Odorant receptor 45a	*Or45a*	2R	2.5
Glutamate receptor IIA	*GluRIIA*	2L	2.67
Odorant-binding protein 18a	*Obp18a*	X	2.29
Odorant receptor 67d	*Or67d*	3L	2.71
Odorant receptor 63a	*Or63a*	3L	1.64
Ionotropic receptor 93a	*Ir93a*	3R	2.81
Gustatory receptor 61a	*Gr61a*	3L	0.47
Odorant-binding protein 49a	*Obp49a*	2R	0.43
Stress response			
Heat shock protein cognate 2	*Hsc70-2*	3R	69.61
Turandot C	*TotC*	3R	8.18
Cyp6a23	*Cyp6a23*	2R	6.57
Turandot A	*TotA*	3R	5.67
Cyp6w1	*Cyp6w1*	2R	3.64
Cyp6g1	*Cyp6g1*	2R	4.23
Cyp313a4	*Cyp313a4*	3R	4.14
Esterase P	*Est-P*	3L	2.61
Cytochrome P450-4d2	*Cyp4d2*	X	2.54
Heat shock protein 23	*Hsp23*	3L	1.72

Cyp4d14	*Cyp4d14*	X	1.93
Glutathione S transferase O1	*GstO1*	3L	1.72
Heat shock protein 68	*Hsp68*	3R	0.66
Glutathione S transferase D3	*GstD3*	3R	0.38
Esterase Q	*Est-Q*	3L	0.36
Glutathione S transferase E10	*GstE10*	2R	0.30
Glutathione S transferase D4	*GstD4*	3R	0.25
Glutathione S transferase D 5	*GstD5*	3R	0.27
Behavioral response			
Cuticular protein 65Ec	*Cpr65Ec*	2R	2.85
Beethoven	*btv*	2L	3.66
Cuticular protein 49Ae	*Cpr49Ae*	2R	1.73
Fatty acid binding protein	*fabp*	3R	1.75
Desat2	*Deat2*	3R	0.33
Cuticular protein 72Ea	*Cpr72Ea*	3L	0.27
Retained	*retn*	2R	2.02
RhoGAP18B	*RhoGAP18B*	X	1.52
Cuticular protein 100A	*Cpr100A*	3R	0.36
Cuticular protein 62Ba	*Cpr62Ba*	3L	0.19

One important aspect of adaptation to novel habitats, especially in the case of fruit flies, is the exposure to novel food and oviposition sources. The fly's organs for olfactory perception are the antennae and the maxillary palps, where housed in sensilla (hair like structures) are odor receptor neurons. These odor receptor neurons (ORNs) transmit olfactory information to the antennal lobes in the fly's brain. The antennal lobes are the first brain centers where olfactory information is processed and forwarded to further brain centers like the mushroom body and the lateral horns. From our brain expression data, we found at least ten genes differentially expressed between African and European flies that are involved in olfactory response, including odorant receptors, ionotropic receptors and odorant binding proteins (Table 4). Studies documenting adaptive behavioral shifts within populations are scarce, but there are some examples of behavioral changes between closely related species. *Drosophila sechellia*, which shared a common ancestor with *D. melangoaster* about 5 million years ago (Kliman et al. 2000), has undergone ecological specialization to feed and oviposit on the fruit *Morinda citrifolia*

(Farine et al. 1996, Jones 2005). Olfactory adaptation and specialization has occurred in the *D. sechellia* lineage, as *D. sechellia* is specifically attracted to *M. citrifolia* and is resistant to the fruit's toxic compounds, whereas other *Drosophila* species like *D. melanogaster* and *D. simulans* are repelled by the odors produced by *M. citrifolia* (Dekker et al. 2006). *D. sechellia* has a higher number of sensilla in the antenna housing ORNs that bind to specific *M. citrifolia* compounds in comparison to *D. melanogaster*. Within the brain, the glomerulus in the antennal lobe, where these specific ORNs project, is twice the size in *D. sechellia* as it is in closely related species (Dekker et al. 2006). Besides the changes in glomerulus size and in the number of sensilla in the antenna, DNA polymorphism in the odorant binding proteins, OBP57d and OBP57e, has been associated with the behavioral and ecological shift that occurred in *D. sechellia* (Matsuo et al. 2007). In *D. melanogaster*, SNPs and indel variation in *Obp57e* and the three *Obp99* paralogs has been found to be segregating in some populations. In the case of the *Obp99* group, SNP variation has been correlated with response strength to benzaldehyde, an odor that occurs in the fruits of plants from which these flies were collected (Takahashi et al. 2005, Wang et al. 2009). From our brain data set, the odorant binding protein *Obp18a* shows about two-fold higher expression in European flies than in African flies. *Obp18a* showed an overall high expression in the brain and could be a candidate gene for olfactory adaptation. Changes in gene expression of ORs, GRs or Obps might relax or fine-tune olfactory behavior toward specific food sources, which could be selected to optimize feeding and oviposition sites.

An important environmental factor driving adaptation to novel environments is toxic agents coming from natural sources, like plant toxins, or toxic agents coming from human sources like waste, insecticides and herbicides. Our study found at least 30 genes involved in stress response, including some genes having a role in xenobiotic tolerance and resistance, resistance to cold or heat shock, oxidative stress, and response to infection (Table 4). Some of these genes, like *Cyp6g1* and genes of the *CHKov* family, have transposable element insertions that have been linked to differences in gene expression and to a higher resistance to insecticides (Schmidt et al. 2010, Amenetzach et al. 2005, Catalán et al. 2012). A recent paper has shown that in the genome of the moth *Plutella xylostella* there is a high density of transposable element insertions near genes involved in xenobiotic metabolism (You et al. 2013). In the case of *Cyp6g1*, it has been found to be differentially expressed between European and African populations of *D. melanogaster*, when whole flies and when only brains were compared (Hutter

et al. 2008, Müller et al. 2011, Catalán et al. 2012). Some genes coding for glutathione S-transferases and cytochrome P450 monooxygenases already have a basal constant expression level that is triggered when flies are exposed to xenobiotics, with the transcriptional response becoming stronger at high concentration levels of the toxic substance (Misra et al. 2011). In this sense, flies having a constant high expression of a gene involved in stress response can react in a faster manner when it comes to stress exposure. Most of the stress response genes that are differentially expressed have a high read count, suggesting that they might be expressed at high levels throughout the brain. A ubiquitous and high expression of these genes in the brain might be of adaptive importance in certain environments, especially because many insecticides and plant toxins target the nervous system (Zlotkin 1999, Wittstock and Gershenzon 2003).

It is now widely accepted that changes in *cis*-regulatory elements constitute a considerable proportion of the genetic variation upon which natural selection acts to drive adaptation (Saminadin-Peter et al. 2012, Catalán et al. 2012, Linnen et al. 2013). Much of the variation taking place at *cis*-regulatory elements has been reported to occur at promoter or enhancer regions, suggesting that most of the regulatory variation happens at the transcriptional level. It has been reported that untranslated regions (UTRs) harbor a considerable amount of polymorphism, including SNPs, transposable elements (TEs) and indels. In *D. melanogaster,* 3' UTR sequences have been suggested to be under strong selective constraint, but are also frequent targets of positive selection (Andolfatto 2005, Lipatov et al. 2005). From our RNA-seq brain data, we found an indel polymorphism on the 3' UTR of *MtnA* that appears to be a target of positive selection, according to population genetic data and a selective sweep analysis (Chapter 2). The ancestral state of the 3' UTR has a 49-bp fragment in the 3' UTR. This sequence is completely absent in a derived allele that is in high frequency in northern European populations. The presence of this derived deletion is perfectly correlated with a higher expression of *MtnA* in the flies from the Netherlands population. We hypothesize that the up-regulation of *MtnA* observed in the European population might be due to the loss of microRNA (miRNA) binding sites in the derived state of the 3' UTR. miRNAs bind to the 3' UTRs of mature mRNAs and either sequester them or target them for degradation, thus regulating gene expression at the post-transcriptional level (Flynt and Lai 2008, Berezikov 2001). Gene regulation by miRNAs can make a large contribution to morphological and

behavioral variation. For example, within *D. melanogaster* populations there is variation for the density of trichomes on the femur of the second leg known as the "naked valley" (Sucena et al. 2003). This morphological polymorphism has been mapped to the micro-RNA *mir92a* (Arif et al. 2013). Differential expression of *mir92a* directly affects the trichome density in the femur of *D. melanogaster* (Arif et al. 2013). Another example of how a miRNA can have a strong impact on the phenotype comes from null knockouts of *mir-279* in *D. melanogaster*. Null mutant flies of *mir-279* develop odor receptor neurons responding to CO_2 in the maxillary palps, instead of only having CO_2 specific ORs in the antenna like wild type flies (Cayirlioglu et al. 2008). In another study, Li et al. (2012) found three different binding sites of miRNAs in three different genes to be evolving under positive selection in human populations. Two of the polymorphic SNPs that they found are located in the seeding region of *mir-155* in the 3' UTR of TYRP1, an important melanosomal enzyme associated with human pigmentation differences. These segregating SNPs have different frequencies in ancestral and derived human populations and are positively correlated with latitude (Li et al. 2012). These studies illustrate how variation in the expression of miRNAs or variation at the binding sites of miRNAs can have strong effects at the phenotypic level and thus might be targets for adaptation.

Structural variation accounts for a large proportion of the divergence and polymorphism in the genomes of flies, humans, dogs and many other organisms (Väli et al. 2008, Mills et al. 2011, Leushkin et al. 2012). Insertion and deletion (indel) variation has the potential to have strong effects at the phenotypic level and thus might trigger adaptive walks at the protein and regulatory level (Leushkin et al. 2012, Massouras et al. 2012). It has been suggested that indels themselves are mutagenic agents, as a higher mutation rate has been detected in the flanking sequences of indels (Tian et al. 2008). In *D. melanogaster*, deletions affect 4.2% of the genome and insertions can add up to 2.1 Mb to the genome. Additionally, data from a *cis*-expression quantitative trait locus (*cis*-eQTL) study in *D. melanogaster* show that about 10% of the *cis*-eQTLs are explained by indel variation (Massouras et al. 2012). In accordance with these data, indel variation can strongly be affected by natural selection, as we show in the case of the indel polymorphism found in the 3' UTR of *MtnA.* Often when using Illumina sequence data, genome or cDNA reads obtained from different populations are mapped to a reference genome, which in the case of *D. melanogaster* is derived from a European fly. Most of the genome-wide population genetic studies take into account only SNP data for the calculation of

summary statistics, as well as for the inference of positive selection and for the assessment of genome variation linked to differences in gene expression. They do not take into account other types of genetic variation like inversions, transposable elements and indels as potential targets for adaptive evolution. One way of circumventing the former issue is to do whole genome *de novo* assembly from different ancestral and derived populations. In this manner comparative studies of genetic polymorphism across different populations could be done in order to find variants other than SNPs that may be targets of positive selection in natural populations.

The study of *D. melanogaster* in the areas of medicine, neuroscience, genetics, molecular biology and evolution has produced a huge amount of important and interesting data. Despite all the research done using *D. melanogaster*, very little is known about its biology and ecological niches, factors that are of special interest in the study of adaptation by natural selection. The spread of *D. melanogaster* throughout the world has been associated with its commensal relationship with humans (Keller 2007). Nevertheless, there are very few studies about the ecological characterization of the new habitats that were colonized by derived populations of *D. melanogaster*. Very little information is available about local ecological factors like food and oviposition sources, natural enemies, toxin exposure, parasites and parasitoids. Additionally, there is a lack of information about life history traits in nature and how these might change over time and under different circumstances. In the case of the study of local adaptation to abiotic factors, there is very little data about how *D. melanogaster* copes with cold, heat or drought in the wild. For example, there is no information about how northern European populations of *D. melanogaster* survive during the long and cold winter months. For other species of *Drosophila* living in temperate habitats, like *D. littoralis* and *D. obscura*, there is detailed information about how these species cope with winter and in which stage of their life cycle these species hibernate (Lumme 1978). Surprisingly, this kind of information is absent for *D. melanogaster*. Research efforts to characterize the ecological niches of *D. melanogaster* populations across the world might generate new and interesting hypotheses about adaptation, which might not be obvious from observations made under laboratory conditions. Combining ecological information with the genetic and genomic resources that are available for *D. melanogaster* might lead to a better understanding of how natural populations adapt to novel and changing ecosystems.

Bibliography

ALACHIOTIS N., STAMATAKIS A., PAVLIDIS P., 2012 OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. Bioinformatics **28**: 2274–2275.

AMINETZACH Y. T., MACPHERSON J. M., PETROV D. A., 2005 Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila. Science **309**: 764–767.

ANASTASI J., 1984 Hemoglobin S-mediated membrane oxidant injury: protection from malaria and pathology in sickle cell disease. Med. Hypotheses **14**: 311–320.

ANDERSON A. R., HOFFMANN A. A., MCKECHNIE S. W., UMINA P. A., WEEKS A. R., 2005 The latitudinal cline in the In(3R)Payne inversion polymorphism has shifted in the last 20 years in Australian Drosophila melanogaster populations. Mol. Ecol. **14**: 851–858.

ANDERSON P. R., OAKESHOTT J. G., 1986 Ethanol tolerance and alcohol dehydrogenase activity in Australian populations of Drosophila simulans. Heredity. **56**: 185–190.

ANDOLFATTO P., 2005 Adaptive evolution of non-coding DNA in Drosophila. Nature **437**: 1149–1152.

ARIF S., MURAT S., ALMUDI I., NUNES M. D. S., BORTOLAMIOL-BECET D., MCGREGOR N. S., CURRIE J. M. S., HUGHES H., RONSHAUGEN M., SUCENA É., LAI E. C., SCHLÖTTERER C., MCGREGOR A. P., 2013 Evolution of mir-92a underlies natural morphological variation in Drosophila melanogaster. Curr. Biol. **23**: 523–528.

BARTEL D. P., 2004 MicroRNAs. Cell **116**: 281–297.

BEGUN D. J., AQUADRO C. F., 1993 African and North American populations of Drosophila melanogaster are very different at the DNA level. Nature **365**: 548–550.

BEREZIKOV E., 2011 Evolution of microRNA diversity and regulation in animals. Nat. Rev. Genet. **12**: 846–860.

BISCHOF J., MAEDA R. K., HEDIGER M., KARCH F., BASLER K., 2007 An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc. Natl. Acad. Sci. U. S. A. **104**: 3312–3317.

BRETZ F., HOTHORN T., WESTFALL P., 2010 Multiple Comparisons Using R (Google eBook). CRC Press.

CARROLL S. B., 2005 Evolution at two levels: on genes and form. PLoS Biol. **3**: e245.

CARROLL S. B., 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell **134**: 25–36.

CATALÁN A., HUTTER S., PARSCH J., 2012 Population and sex differences in Drosophila melanogaster brain gene expression. BMC Genomics **13**: 654.

CAYIRLIOGLU P., KADOW I. G., ZHAN X., OKAMURA K., SUH G. S. B., GUNNING D., LAI E. C., ZIPURSKY S. L., 2008 Hybrid neurons in a microRNA mutant are putative evolutionary intermediates in insect CO_2 sensory systems. Science **319**: 1256–1260.

CHEN K., RAJEWSKY N., 2007 The evolution of gene regulation by transcription factors and microRNAs. Nat. Rev. Genet. **8**: 93–103.

CHINTAPALLI V. R., WANG J., DOW J. A. T., 2007 Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat. Genet. **39**: 715–720.

COLEGRAVE N., HOLLOCHER H., HINTON K., RITCHIE M. ., 2000 The courtship song of African Drosophila melanogaster. J. Evol. Biol. **13**: 143–150.

COLLINGE J. E., ANDERSON A. R., WEEKS A. R., JOHNSON T. K., MCKECHNIE S. W., 2008 Latitudinal and cold-tolerance variation associate with DNA repeat-number variation in the hsr-omega RNA gene of Drosophila melanogaster. Heredity (Edinb). **101**: 260–270.

CRISPONI G., CAPDEVILA M., BOFILL R., PALACIOS Ò., ATRIAN S., 2012 State-of-the-art of metallothioneins at the beginning of the 21st century. Coord. Chem. Rev. **256**: 46–62.

DINGLE H. (Ed.), 1978 Evolution of insect migration and diapause. Springer US, New York, NY.

DITCH L. M., SHIRANGI T., PITMAN J. L., LATHAM K. L., FINLEY K. D., EDEEN P. T., TAYLOR B. J., MCKEOWN M., 2005 Drosophila retained/dead ringer is necessary for neuronal pathfinding, female receptivity and repression of fruitless independent male courtship behaviors. Development **132**: 155–164.

DUCHEN P., ZIVKOVIC D., HUTTER S., STEPHAN W., LAURENT S., 2013 Demographic inference reveals African and European admixture in the North American Drosophila melanogaster population. Genetics **193**: 291–301.

EBERL D. F., HARDY R. W., KERNAN M. J., 2000 Genetically similar transduction mechanisms for touch and hearing in Drosophila. J. Neurosci. **20**: 5981–5988.

EGLI D., SELVARAJ A., YEPISKOPOSYAN H., ZHANG B., HAFEN E., GEORGIEV O., SCHAFFNER W., 2003 Knockout of "metal-responsive transcription factor" MTF-1 in Drosophila by homologous recombination reveals its central role in heavy metal homeostasis. EMBO J. **22**: 100–108.

EJIMA A., SMITH B. P. C., LUCAS C., GOES VAN NATERS W. VAN DER, MILLER C. J., CARLSON J. R., LEVINE J. D., GRIFFITH L. C., 2007 Generalization of courtship learning in Drosophila is mediated by cis-Vaccenyl Acetate. Curr. Biol. **17**: 599–605.

FARINE J.-P., LEGAL L., MORETEAU B., QUERE J.-L. LE, 1996 Volatile components of ripe fruits of Morinda citrifolia and their effects on Drosophila. Phytochemistry **41**: 433–438.

FERVEUR J.-F., 2005 Cuticular hydrocarbons: their evolution and roles in Drosophila pheromonal communication. Behav. Genet. **35**: 279–295.

FINLEY K. D., EDEEN P. T., FOSS M., GROSS E., GHBEISH N., PALMER R. H., TAYLOR B. J., MCKEOWN M., 1998 Dissatisfaction encodes a tailless-like nuclear receptor expressed in a subset of CNS neurons controlling Drosophila sexual behavior. Neuron **21**: 1363–1374.

FISH M. P., GROTH A. C., CALOS M. P., NUSSE R., 2007 Creating transgenic Drosophila by microinjecting the site-specific phiC31 integrase mRNA and a transgene-containing donor plasmid. Nat. Protoc. **2**: 2325–2331.

FISTON-LAVIER A.-S., SINGH N. D., LIPATOV M., PETROV D. A., 2010 Drosophila melanogaster recombination rate calculator. Gene **463**: 18–20.

FLYNT A. S., LAI E. C., 2008 Biological principles of microRNA-mediated regulation: shared themes amid diversity. Nat. Rev. Genet. **9**: 831–842.

FREEMAN M. R., DOHERTY J., 2006 Glial cell biology in Drosophila and vertebrates. Trends Neurosci. **29**: 82–90.

FRANKS S. J., HOFFMANN A. A., 2012 Genetics of climate change adaptation. Annu. Rev. Genet. **46**: 185–208.

GLEASON J. M., 2005 Mutations and natural genetic variation in the courtship song of Drosophila. Behav. Genet. **35**: 265–277.

GONZÁLEZ J., KARASOV T. L., MESSER P. W., PETROV D. A., 2010 Genome-wide patterns of adaptation to temperate environments associated with transposable elements in Drosophila. PLoS Genet. **6**: e1000905.

GOUY M., GUINDON S., GASCUEL O., 2010 SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. **27**: 221–224.

GUIROLA M., NARANJO Y., CAPDEVILA M., ATRIAN S., 2011 Comparative genomics analysis of metallothioneins in twelve Drosophila species. J. Inorg. Biochem. **105**: 1050–1059.

HAMILTON M., 2009 *Population Genetics*. Wiley.

HARTENSTEIN V., 2011 Morphological diversity and development of glia in Drosophila. Glia **59**: 1237–1252.

HARTENSTEIN V., SPINDLER S., PEREANU W., FUNG S., 2008 The development of the Drosophila larval brain. Adv. Exp. Med. Biol. **628**: 1–31.

HENSE W., BAINES J. F., PARSCH J., 2007 X chromosome inactivation during Drosophila spermatogenesis. PLoS Biol. **5**: e273.

HUDSON R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics **18**: 337–338.

HUDSON R. R., BAILEY K., SKARECKY D., KWIATOWSKI J., AYALA F. J., 1994 Evidence for positive selection in the superoxide dismutase (Sod) region of Drosophila melanogaster. Genetics **136**: 1329–1340.

HUTTER S., SAMINADIN-PETER S. S., STEPHAN W., PARSCH J., 2008 Gene expression variation in African and European populations of Drosophila melanogaster. Genome Biol. **9**: R12.

JENSEN J. D., KIM Y., DUMONT V. B., AQUADRO C. F., BUSTAMANTE C. D., 2005 Distinguishing between selective sweeps and demography using DNA polymorphism data. Genetics **170**: 1401–1410.

JOHNSON C. W., 1913 The Distribution of some species of Drosophila. Psyche (Stuttg). **20**: 202–205.

JONES C. D., 2005 The genetics of adaptation in Drosophila sechellia. Genetica **123**: 137–145.

JONES W. D., CAYIRLIOGLU P., KADOW I. G., VOSSHALL L. B., 2007 Two chemosensory receptors together mediate carbon dioxide detection in Drosophila. Nature **445**: 86–90.

KELLER A., 2007 Drosophila melanogaster's history as a human commensal. Curr. Biol. **17**: R77–81.

KERTESZ M., IOVINO N., UNNERSTALL U., GAUL U., SEGAL E., 2007 The role of site accessibility in microRNA target recognition. Nat. Genet. **39**: 1278–1284.

KIM Y., STEPHAN W., 2002 Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics **160**: 765–777.

KIM Y., STEPHAN W., 2003 Selective sweeps in the presence of interference among partially linked loci. Genetics **164**: 389–398.

KIMURA M., 1984 The neutral theory of molecular evolution. Cambridge University Press.

KLIMAN R. M., ANDOLFATTO P., COYNE J. A., DEPAULIS F., KREITMAN M., BERRY A. J., MCCARTER J., WAKELEY J., HEY J., 2000 The population genetics of the origin and divergence of the Drosophila simulans complex species. Genetics **156**: 1913–1931.

LACHAISE D., CARIOU M-L., DAVID J.R, LEMEUNIER F., TSACHAS L., ASHBURNER M., Historical biogeography of the Drosophila melanogaster species subgroup. Evol. Biol. **22**: 159–225.

LANGE B. W., LANGLEY C. H., STEPHAN W., 1990 Molecular evolution of Drosophila metallothionein genes. Genetics **126**: 921–932.

LAURENT S. J. Y., WERZNER A., EXCOFFIER L., STEPHAN W., 2011 Approximate Bayesian analysis of Drosophila melanogaster polymorphism data reveals a recent colonization of Southeast Asia. Mol. Biol. Evol. **28**: 2041–2051.

LAZZARO B. P., CLARK A. G., 2001 Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the Attacin genes of Drosophila melanogaster. Genetics **159**: 659–671.

LEGENDRE A., MIAO X.-X., LAGE J.-L. DA, WICKER-THOMAS C., 2008 Evolution of a desaturase involved in female pheromonal cuticular hydrocarbon biosynthesis and courtship behavior in Drosophila. Insect Biochem. Mol. Biol. **38**: 244–255.

LEUSHKIN E. V, BAZYKIN G. A., KONDRASHOV A. S., 2012 Insertions and deletions trigger adaptive walks in Drosophila proteins. Proc. Biol. Sci. **279**: 3075–3082.

LEWIS B. P., BURGE C. B., BARTEL D. P., 2005 Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell **120**: 15–20.

LI J., LIU Y., XIN X., KIM T. S., CABEZA E. A., REN J., NIELSEN R., WRANA J. L., ZHANG Z., 2012 Evidence for positive selection on a number of MicroRNA regulatory interactions during recent human evolution. PLoS Genet. **8**: e1002578.

LINNEN C. R., POH Y.-P., PETERSON B. K., BARRETT R. D. H., LARSON J. G., JENSEN J. D., HOEKSTRA H. E., 2013 Adaptive evolution of multiple traits through multiple mutations at a single gene. Science **339**: 1312–6.

LIPATOV M., LENKOV K., PETROV D. A., BERGMAN C. M., 2005 Paucity of chimeric gene-transposable element transcripts in the Drosophila melanogaster genome. BMC Biol. **3**: 24.

LIU N., LANDREH M., CAO K., ABE M., HENDRIKS G.-J., KENNERDELL J. R., ZHU Y., WANG L.-S., BONINI N. M., 2012 The microRNA miR-34 modulates ageing and neurodegeneration in Drosophila. Nature **482**: 519–523.

MARONI G., OTTO E., LASTOWSKI-PERRY D., 1986 Molecular and cytogenetic characterization of a metallothionein gene of Drosophila. Genetics **112**: 493–504.

MARYGOLD S. J., LEYLAND P. C., SEAL R. L., GOODMAN J. L., THURMOND J., STRELETS V. B., WILSON R. J., 2013 FlyBase: improvements to the bibliography. Nucleic Acids Res. **41**: D751–7.

MBOUP M., FISCHER I., LAINER H., STEPHAN W., 2012 Trans-species polymorphism and allele-specific expression in the CBF gene family of wild tomatoes. Mol. Biol. Evol. **29**: 3641–3652.

MCKECHNIE S. W., BLACKET M. J., SONG S. V, RAKO L., CARROLL X., JOHNSON T. K., JENSEN L. T., LEE S. F., WEE C. W., HOFFMANN A. A., 2010 A clinally varying promoter polymorphism associated with adaptive variation in wing size in Drosophila. Mol. Ecol. **19**: 775–784.

MEIKLEJOHN C. D., PARSCH J., RANZ J. M., HARTL D. L., 2003 Rapid evolution of male-biased gene expression in Drosophila. Proc. Natl. Acad. Sci. U. S. A. **100**: 9894–9899.

MILLS R. E., PITTARD W. S., MULLANEY J. M., FAROOQ U., CREASY T. H., MAHURKAR A. A., KEMEZA D. M., STRASSLER D. S., PONTING C. P., WEBBER C., DEVINE S. E., 2011 Natural genetic variation caused by small insertions and deletions in the human genome. Genome Res. **21**: 830–839.

MISRA J. R., HORNER M. A., LAM G., THUMMEL C. S., 2011 Transcriptional regulation of xenobiotic detoxification in Drosophila. Genes Dev. **25**: 1796–1806.

MÜLLER L., HUTTER S., STAMBOLIYSKA R., SAMINADIN-PETER S. S., STEPHAN W., PARSCH J., 2011 Population transcriptomics of Drosophila melanogaster females. BMC Genomics **12**: 81.

NATH R., KUMAR D., LI T., SINGAL P. K., 2000 Metallothioneins, oxidative stress and the cardiovascular system. Toxicology **155**: 17–26.

NIELSEN R., WILLIAMSON S., KIM Y., HUBISZ M. J., CLARK A. G., BUSTAMANTE C., 2005 Genomic scans for selective sweeps using SNP data. Genome Res. **15**: 1566–1575.

OKAMURA K., HAGEN J. W., DUAN H., TYLER D. M., LAI E. C., 2007 The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell **130**: 89–100.

PAVLIDIS P., ŽIVKOVIC D., STAMATAKIS A., ALACHIOTIS N., 2013 SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Mol. Biol. Evol. **30**: 2224–2234.

PÉREZ-RAFAEL S., KURZ A., GUIROLA M., CAPDEVILA M., PALACIOS O., ATRIAN S., 2012 Is MtnE, the fifth Drosophila metallothionein, functionally distinct from the other members of this polymorphic protein family? Metallomics **4**: 342–349.

PFAFFL M. W., 2001 A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. **29**: e45.

PHIFER-RIXEY M., HECKMAN M., TRUSSELL G. C., SCHMIDT P. S., 2008 Maintenance of clinal variation for shell colour phenotype in the flat periwinkle Littorina obtusata. J. Evol. Biol. **21**: 966–978.

POOL J. E., CORBETT-DETIG R. B., SUGINO R. P., STEVENS K. A., CARDENO C. M., CREPEAU M. W., DUCHEN P., EMERSON J. J., SAELAO P., BEGUN D. J., LANGLEY C. H., 2012 Population Genomics of sub-saharan Drosophila melanogaster: African diversity and non-African admixture. (HS Malik, Ed.). PLoS Genet. **8**: e1003080.

RUBY J. G., STARK A., JOHNSTON W. K., KELLIS M., BARTEL D. P., LAI E. C., 2007 Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res. **17**: 1850–1864.

SAMBROOK J., FRITSCH EF., MANIATIS T., 1989 Molecular Cloning: A Laboratory Manual. Ed. 2. Cold Spring Harbor (NY): Cold Spring Harbor Laboratorz Press. 1659 p.

SAMINADIN-PETER S. S., KEMKEMER C., PAVLIDIS P., PARSCH J., 2012 Selective sweep of a cis-regulatory sequence in a non-African population of Drosophila melanogaster. Mol. Biol. Evol. **29**: 1167–74.

SCHMIDT J. M., GOOD R. T., APPLETON B., SHERRARD J., RAYMANT G. C., BOGWITZ M. R., MARTIN J., DABORN P. J., GODDARD M. E., BATTERHAM P., ROBIN C., 2010 Copy number variation and transposable elements feature in recent, ongoing adaptation at the Cyp6g1 locus. PLoS Genet. **6**: e1000998.

SCHNEIDER C. A., RASBAND W. S., ELICEIRI K. W., 2012 NIH Image to ImageJ: 25 years of image analysis. Nat. Methods **9**: 671–675.

SHIRANGI T. R., TAYLOR B. J., MCKEOWN M., 2006 A double-switch system regulates male courtship behavior in male and female Drosophila melanogaster. Nat. Genet. **38**: 1435–1439.

SMADJA C., BUTLIN R. K., 2009 On the scent of speciation: the chemosensory system and its role in premating isolation. Heredity. **102**: 77–97.

SMITH J. M., HAIGH J., 2009 The hitch-hiking effect of a favourable gene. Genet. Res. **23**: 23.

STEPHAN W., LI H., 2007 The recent demographic and adaptive history of Drosophila melanogaster. Heredity. **98**: 65–68.

STEPHAN W., RODRIGUEZ V. S., ZHOU B., PARSCH J., 1994 Molecular evolution of the metallothionein gene Mtn in the melanogaster species group: results from Drosophila ananassae. Genetics **138**: 135–143.

SUCENA E., DELON I., JONES I., PAYRE F., STERN D. L., 2003 Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. Nature **424**: 935–938.

SZUPLEWSKI S., KUGLER J.-M., LIM S. F., VERMA P., CHEN Y.-W., COHEN S. M., 2012 MicroRNA transgene overexpression complements deficiency-based modifier screens in Drosophila. Genetics **190**: 617–626.

TAJIMA F., 1983 Evolutionary relationship of DNA sequences in finite populations. Genetics **105**: 437–60.

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

THEODORE L., HO A. S., MARONI G., 1991 Recent evolutionary history of the metallothionein gene Mtn in Drosophila. Genet. Res. **58**: 203–210.

TIAN D., WANG Q., ZHANG P., ARAKI H., YANG S., KREITMAN M., NAGYLAKI T., HUDSON R., BERGELSON J., CHEN J.-Q., 2008 Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. Nature **455**: 105–108.

UMINA P. A., WEEKS A. R., KEARNEY M. R., MCKECHNIE S. W., HOFFMANN A. A., 2005 A rapid shift in a classic clinal pattern in Drosophila reflecting climate change. Science **308**: 691–693.

VÄLI U., BRANDSTRÖM M., JOHANSSON M., ELLEGREN H., 2008 Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. BMC Genet. **9**: 8.

VIGNIERI S. N., LARSON J. G., HOEKSTRA H. E., 2010 The selective advantage of crypsis in mice. Evolution **64**: 2153–2158.

WATTERSON G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **276**: 256–276.

WICKER-THOMAS C., 2007 Pheromonal communication involved in courtship behavior in Diptera. J. Insect Physiol. **53**: 1089–1100.

WITTSTOCK U., GERSHENZON J., 2002 Constitutive plant toxins and their role in defense against herbivores and pathogens. Curr. Opin. Plant Biol. **5**: 300–307.

YOU M., YUE Z., HE W., YANG X., YANG G., *et al.*, 2013 A heterozygous moth genome provides insights into herbivory and detoxification. Nat. Genet. **45**: 220–225.

ZHANG J., 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. **95**: 3708–3713.

ZHANG J., 2003 Evolution by gene duplication: an update. Trends Ecol. Evol. **18**: 292–298.

ZLOTKIN E., 1999 The insect voltage-gated sodium channel as target of insecticides. Annu. Rev. Entomol. **44**: 429–455.

Aknowledgments

I will first thank my family, for their support through all these years. Gracias, Mama, Papa, Isa y Oli, por estar siempre ahí para mí, brindándome su apoyo incondicional. Sin las largas conversaciones casi todos los domingos desde que me vine, seguramente hubiera colapsado mentalmente. ¡Los quiero mucho!

Secondly, I want to thank Sebastian. Du hast mein Leben schöner und besser gemacht. Die Konferenzen hätten nicht so viel Spass gemacht ohne dich. I hab dich lieb!

Thirdly, I want to thank my supervisor John Parsch. Thanks a lot for supporting me through my PhD. I have learned a lot and it has been a very interesting experience being able to do a PhD in your group.

Fourthly, I want to thank Hedwig Gebhart and Hilde Lainer for their amazing help that I have received from you in the lab. Vielen Dank!

I also want to thank my office mates, Amanda, Ina and Liza for making my PhD time more enjoyable; I have had fun with you girls!

Ricardo, Pablo y Meike (¡¿qué pumas?!), ¡gracias por su amistad! La vida en Alemania no hubiera sido tan agradable sin ustedes. Espero que nuestra amistad continue por mucho más tiempo.

And finally, thanks to all the people in the evolutionary biology department. It has been nice meeting you all!

ANA CATALÁN

LUDWIG MAXIMILIANS UNIVERSITÄT
MUNICH (LMU)
EVOLUTIONARY BIOLOGY, DEPARTMENT BIOLOGIE II
GROßHADERNER STR. 2, D-82152 PLANEGG-MARTINSRIED
CATALAN@BIO.LMU.DE

RESEARCH INTERESTS: adaptation, insect biology, molecular evolution, functional genomics, field work, population genetics, neuroethology.
LANGUAGES: English, German and Spanish (all fluent).

ACADEMIC QUALIFICATIONS

- 2010 - present
  Senior PhD student at the Evolutionary and Functional Genomics research group under the supervision of Prof. Dr. John Parsch.

- 2010
  Master Degree in Evolution, Ecology and Systematics. Ludwig Maximilians Universität, Munich.

- 2005
  Licentiate in Biology, Universidad del Valle de Guatemala.

- 2000
  High-school diploma, Austrian Institute of Guatemala

RESEARCH EXPERIENCE

- 2010 – present.
  PhD student. Evolutionary biology, LMU.
  Linking gene expression differences between populations with DNA polymorphism patterns, which may influence specific phenotypes relevant for adaptation to novel habitats. RNA-seq, qRT-PCR, molecular cloning, transgenesis, functional assays, population genetics analysis, *in-situ* hybridization.

- 2009 - 2010
  Master thesis. Sensory Neurogenetics, Max Planck Institute of Neurobiology.
  "Evolution of CO_2 behavioral response in the *Drosophila* genus". Supervisor: Dr. Ilona Kadow.
  Behavioral assays, brain immunohistochemestry, confocal imaging, character evolution study.

- 2008-2009
  Research assistant. LMU.
  Sponge molecular phylogenetics, DNA extraction and PCR under the supervision of Prof. Dr. Wörheide (2009). Fungal infection experiments on two different populations of *Drosophila melanogaster* under the supervision of Dr. Stephan Hutter (2008).

- 2007
  Junior research grant applicant. Center for Health Studies, Universidad del Valle de Guatemala.

"Gene expression in *Anopheles albimanus* resistant to deltamethrin". Project funded by the CONCYT (Guatemalan Science and Technology Council). Selection of an *Anopheles albimanus* deltamethrin resistant strain and insect histology.

- 2006 - 2007
  Research assistant. Center for Health Studies, Universidad del Valle de Guatemala.
  Established the base line for deltamethrin susceptible *Anopheles albimanus* and sentinel network in northern Guatemala, for monitoring insecticide resistance, as part of a technical assistance to the Ministry of Health.
- 2005 - 2006
  Entomology assistant. Center for Health Studies, Universidad del Valle de Guatemala and US Center for Disease Control and Prevention (CDC).
  West Nile Virus prevalence survey in Guatemala. Mosquito field collection and taxonomy. RNA mosquito extraction.

- 2005
  Licentiate thesis. Universidad del Valle de Guatemala.
  "Identification of potential vectors of potato zebra chip disease in Guatemala". Molecular identification of leafhoppers (Hemiptera:Cicadellidae) as vectors of phytoplasmas and *Xylella fastidiosa*. Supervisor: Dr. Jack Schuster. Phytoplasma diagnosis, field work and insect taxonomy.

- 2003 - 2004
  Entomology assistant. Universidad del Valle de Guatemala.
  Identification of alternative vectors of lethal yellowing of palm in northern Guatemala.

- 2003
  Entomology assistant. Universidad del Valle de Guatemala and Defensores de la Natureleza Fundation.
  Insect taxonomy as part of a conservation initiative of the pine forest *Dendroica chrysoparia*.

PUBLICATIONS

Catalán, A., S. Hutter, and J. Parsch (2012)
Population and sex differences in Drosophila melanogaster brain gene expression.
BMC Genomics 13: 654.

Glaser-Schmitt*, A., **A. Catalán***, and J. Parsch (2013)
Adaptive divergence of a transcriptomal enhancer between populations of *Drosophila melanogaster*.
Philosophical Transactions B (accepted for publication).
* Equal contributions.

Kemkemer, C., **A. Catalán**, and J. Parsch (2013)
"Escaping" the X chromosome leads to increased gene expression in the male germline of *Drosophila melanogaster*.
Heredity 11 September 2013 (online access).

Catalán, A. Pablo Duchén and J. Parsch (2013).
An indel polymorphism in the *MtnA* 3' untranslated region is associated with gene expression variation and local adaptation in *Drosophila melanogaster*.
Manuscript in preparation.

GRANTS AND AWARDS

- 2011-2013
  The Graduate School Life Science Munich (LSM) travel grant award (3x).

- 2011
  Evolution, Ecology and Systematics (EES) travel award, Munich.

- 2010
  EES master award, LMU Munich.

- 2009
  LSM fast-track award, LMU Munich.

- 2007
  CONCYT research grant award, Guatemala.

TEACHING EXPERIENCE

- 2010-2013
  Student supervisor at the Evolutionary Biology department at the University of Munich. Two bachelor students and five master students. Project design and instruction in laboratory work.

- 2005
  General entomology. Licentiate level. Universidad del Valle de Guatemala. Lectures on insect anatomy, physiology and ecology. Laboratory assistance.

- 2002 – 2003
  General biology. Licentiate level. Universidad del Valle de Guatemala. Laboratory tutorials on general biology.

CONFERENCE PRESENTATIONS

- 2010 – 2013
  Oral presentation. Evolutionary biology department, LMU.
  Seminar talks twice a year about current ongoing research.

- 2013
  Oral presentation. Symposium on Population Genomics, LMU.
  "A putative case for adaptive pos-transcriptional gene regulation in *Drosophila melanogaster*".

- 2012
  Poster presentation. SMBE, Dublin.
  "Population and sex differences in *Drosophila* brain gene expression".

- 2011
  Poster presentation. ESEB, Tübingen.
  "Sequence and expression divergence of the *foraging* gene between Africa and European populations of *Drosophila melanogaster*".

- 2011
  Oral presentation. Evolution, Oklahoma.

"Sequence and expression divergence of the *foraging* gene between Africa and European populations of *Drosophila melanogaster*".

- 2010
  Oral presentation. EES conference, Munich.
  "Evolution of CO_2 behavioral response in the *Drosophila* genus"

- 2010
  Poster presentation. SMBE, Lyon.
  "Differential CO_2 behavioral response in the *D. melanogaster* subgroup. A trait evolving neutrally?"

- 2007
  Seminar. National Institute of Health Laboratories (INLASA), Bolivia.
  "Implementation of the CDC bottle bioassay for insecticide resistance detection"

- 2006.
  Oral presentation. Research Seminar, Universidad del Valle de Guatemala.
  "Identification of potential vectors of potato zebra chip disease in Guatemala".

VOLUNTEER ACTIVITIES

- 2011
  LMU
  Conference organization. Evolution, Ecology and Systematics (EES) annual conference.

- 2003 - 2008
  Universidad San Carlos de Guatemala, volunteer program.
  Sea turtle conservation. Adult healing, biometry and monitoring. Sea turtle egg conservation.

- 2001 - 2003
  Universidad del Valle de Guatemala**.**
  Entomology collection. Insect taxonomy and curator.