Gene regulatory divergence between populations of *Drosophila melanogaster*

Dissertation

der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

vorgelegt von

Amanda Nicole Glaser-Schmitt

aus Upland, California, USA

München, 15 Juni 2016

Erstgutachter: Prof. Dr. John Parsch

Zweitgutachter: Prof. Dr. Wolfgang Enard

Tag der Mundliche Prüfung: 27. Oktober 2016

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.

_____

Amanda Glaser-Schmitt, 15. Juni 2016, München

For Markus and Lukas

Table of Contents

List of Figures

List of Tables

Declaration of Contribution

In this dissertation I present the results of my doctoral research, which was conducted from November 2011 to June 2016. All of the work is the result of collaborations with other scientists and is presented in three chapters. Chapters 1 and 3 have been published, while chapter 2 is presented as an unpublished manuscript.

Chapter 1:

Glaser-Schmitt A*, Catalán A*, Parsch J. Adaptive divergence of a transcriptional enhancer between populations of *Drosophila melanogaster*. *Phil. Trans. R. Soc. B.* 2013; 368 (1632): 20130024. *These authors contributed equally

For chapter 1, I maintained fly stocks, performed analyses of expression and its association with sequent variants, performed PCR and sequencing, and performed the polarized MK test. I wrote the parts of the materials and methods and results sections corresponding to the expression analysis and its association with sequence variants as well as the part of the discussion section concerning *CG9509* sequence and expression variation in North America. In addition, I helped with manuscript revision and prepared figures 4, 5, S1, S2, and S3.

Chapter 2:

Glaser-Schmitt A, Parsch J. Functional characterization of adaptive variation within a *cis*-regulatory element influencing *Drosophila melanogaster* growth. (Unpublished Manuscript)

In chapter 2, John Parsch and I conceived the study and its design. I maintained fly stocks and performed the experiments as well as performed the analyses. John Parsch and I wrote the manuscript. A paper containing parts of this chapter is in preparation.

Chapter 3:

Catalán A, **Glaser-Schmitt A**, Argyridou E, Duchen P, Parsch J. An Indel Polymorphism in the *MtnA* 3' Untranslated Region Is Associated with Gene Expression Variation and Local Adaptation in *Drosophila melanogaster*. *PLoS Genetics* 2016; 12 (4): e1005987.

For chapter 3, I maintained fly stocks, performed and analyzed the β–galactosidase reporter gene assays, and in collaboration with Eliza Argyridou performed the copper and oxidative stress tolerance assays. I also analyzed the tolerance assay data. In collaboration with John Parsch (and Ana Catalán for copper tolerance assays), I conceived and designed the copper and oxidative stress tolerance assays. I wrote the parts of the materials and methods and results sections corresponding to the β–galactosidase assays as well as the copper and oxidative stress tolerance assays. I also wrote parts of the discussion section and helped with manuscript revision. I prepared tables S4-S12 and, in collaboration with Ana Catalán, prepared figures 5C, 8, 9, and S3.

_____ _____

Amanda Glaser-Schmitt Prof. Dr. John Parsch

Abstract

The pervasiveness of gene expression variation at both the population and species level is well-documented. Underlying this expression variation are gene regulatory changes. It has been hypothesized that regulatory changes, especially *cis*-regulatory changes, are especially important in phenotypic evolution since they are more easily fine-tuned both temporally and tissue-specifically than protein-coding changes. This dissertation aims to examine the genetic basis of adaptive regulatory changes, including the effects of adaptive regulatory polymorphisms on both gene expression and organismal phenotype. This thesis centers around the analysis of adaptive *cis*-regulatory changes associated with two *Drosophila melanogaster* genes: *CG9509*, a gene of unknown function, and *Metallothionein A* (*MtnA*), a gene involved in metal homeostasis and the general stress response.

Chapters 1 and 2 provide an in-depth analysis of a case of previously identified adaptive regulatory divergence in a *cis*-regulatory element of *CG9509*. Adult *CG9509* expression is 2–3-fold higher in a European population in comparison to an ancestral, sub-Saharan African population and transgenic reporter gene assays have previously shown that this expression divergence is driven by a 1.2-kb *cis*-regulatory enhancer element upstream of the *CG9509* coding region, which shows signs of recent positive selection in the European population. In Chapter 1, I show that the observed *CG9509* expression divergence extends to adults in other cosmopolitan and sub-Saharan African populations, and in chapter 2 I show that it extends to larvae as well. This suggests that the previously identified positive selection on the cosmopolitan variant of the *CG9509* enhancer element occurred during or

shortly after *D. melanogaster*'s expansion out of Africa, before the separation of European and Asian populations. In chapter 2, I use site directed-mutagenesis and transgenic reporter gene constructs to identify the three single nucleotide polymorphisms (SNPs) within the *CG9509* enhancer responsible for the observed expression divergence. Interestingly, two of these SNPs have a relatively small effect on expression and appear to have been the targets of a selective sweep, while the third SNP has a much larger effect on expression and appears to have been a recent target of balancing selection. In chapter 2, I further use a series of functional and tolerance assays to show that *CG9509* expression affects *D. melanogaster* growth and propose that the organismal phenotype under selection is reduced wing loading, which likely improves flight ability at cooler temperatures.

Chapter 3 identifies a new case of adaptive *cis*-regulatory divergence in the 3' untranslated region (UTR) of *MtnA*. *MtnA* expression in the brain is 5-fold higher in a European in comparison to an African population and within the *MtnA* 3'UTR is a 49-basepair insertion/deletion (indel) polymorphism. I performed transgenic reporter gene assays to show that the deletion in the 3'UTR, which is the derived variant and is at high frequency in the European population, is associated with increased *MtnA* expression. In conjunction with population genetic evidence, this suggests that the deletion in the *MtnA* 3'UTR has been the target of selection for an increase in *MtnA* expression in the European population. Using hydrogen peroxide tolerance assays, I further show that *MtnA* expression is involved in oxidative stress tolerance and that the 3'UTR indel polymorphism is associated with oxidative stress tolerance variation in natural populations, suggesting that improved oxidative stress tolerance is the organismal phenotype under selection.

General Introduction

> "But Natural Selection, as we shall hereafter see, is a power incessantly ready for action, and is immeasurably superior to man's feeble efforts, as the works of Nature are to those of Art."
>
> Charles Darwin, *On the Origin of Species*

After Charles Darwin first proposed the theory of natural selection in 1859, it has remained in the zeitgeist of evolution ever since. Although some of what Darwin proposed in *On the Origin of Species* has proven untenable, the idea of natural selection has stood the test of time. While the theory of neutral (or nearly neutral) evolution, which states that the majority of observed molecular evolutionary changes are the result of the genetic drift of neutral or nearly mutations rather than natural selection (Kimura 1983, Ohta 1992), is an integral part of evolution and establishes the baseline evolutionary biologists use to determine if selection is occurring, adaptation is arguably the more interesting phenomenon. When a species encounters new habitats or a change in environmental conditions within the current habitat, it must adapt in order to survive and flourish in the newly encountered conditions and the new fitness challenges they may impose. This is especially important as species expand their range. As they colonize their new habitat, they must adapt to the new environmental factors, such as temperature, humidity, elevation, parasite pressure, predation, or resources, if they are to establish in the new area. Furthermore, the initial range expansion itself can sometimes be the product of natural

selection, such as adaptation to becoming a human commensal, a change in dispersal method, or a host or resource switch that has allowed the species to spread more quickly or more easily than previously. Thus, species that have been able to successfully undergo range expansions make excellent subjects for the study of adaptation, especially if much is known about their demographic history and the method of their range expansion.

Identifying adaptation in the genome

DNA sequences contain information about their evolutionary history. When the DNA sequences of individuals within a population are aligned, one can see standing variation along a locus in the form of mutations at various nucleotide positions, known as single nucleotide polymorphisms (SNPs). Although genetic variability is also increased by insertion/deletion (indel) polymorphisms, translocations, transposable elements, and inversions, for convenience, SNPs are usually used to assess the genetic variation of a given locus within a population. Patterns in this genetic variation can reveal much about the history of a sequence, since evolutionary forces such as genetic drift and natural selection in the form of balancing, positive, and negative, or purifying, selection leave footprints in the genome (Ohta 1992, Stephan et al 1992, Fu and Li 1993, Charlesworth 2006). However, demographic events, such as bottlenecks or expansions, can also leave footprints in the genome that can be very similar to those left by selection (Jensen et al 2005, Hahn et al 2002). Therefore, knowledge of a species' demographic history is key in identifying instances of selection.

Drosophila melanogaster

Drosophila melanogaster is an excellent model organism with which to study adaptation. It has a long history as a model organism and is employed extensively in many fields of research. As a result, it has a plethora of genetic tools available as well as a well-annotated genome and an increasing availability of information on its genomic functional elements. Its short generation time combined with the *Drosophila* community's extensive and continually expanding knowledge of its regulatory elements, chromatin states, proteins, signaling pathways, development, etc. make it a powerful model system. However, what suits *D. melanogaster* most to the study of adaptation is our knowledge of its biogeographic and demographic history.

D. melanogaster is a cosmopolitan, domestic species that can be found as far north as Finland and as far south as Tasmania (Keller 2007). Based on genome-wide analyses of DNA sequence polymorphism, *D. melanogaster* is believed to have originated in sub-Saharan Africa (Ometto et al 2005, Glinka et al 2003, Li and Stephan 2006), specifically somewhere in southern-central Africa (Figure 1, Pool et al 2012). It is thought to have expanded from its ancestral range approximately 15,000–17,000 years ago with the establishment of human settlements in the Middle East (Laurent et al 2011). From this non-sub-Saharan African source population, *D. melanogaster* is believed to have colonized Asia and Europe approximately 2,500–5,000 years ago as humans and agriculture spread (Figure 1, Laurent et al 2011). *D. melanogaster* has only more recently colonized North America, within the last 200 years, through an admixture of European and African populations (Duchen et al 2013). Sub-Saharan African admixture with non-African populations is also thought to have occurred recently, with more extensive admixture in urban areas, potentially due to a bigger connection to international trade (Pool et al 2012).

Figure 1: Demography of *D. melanogaster* in Africa and Eurasia. The black circle shows *D. melanogaster*'s hypothesized point of origin in sub-Saharan Africa. Arrows represent direction of colonization, with the solid line representing *D. melanogaster*'s initial expansion out of Africa and hatched lines representing the subsequent colonization of Europe and Asia from a putative source population in the Middle East (white circle). Shown are estimated dates of range expansions in years ago (ya) based on genome-wide analyses of DNA sequence polymorphism (Laurent et al 2011).

D. melanogaster is a part of the *melanogaster* group of the subgenus *Drosophila*, which is composed of nine species, all of which are Afrotropical in origin. However, only *D. melanogaster* and *D. simulans*, which split approximately 2–3 million years ago, are human commensals and have expanded to a cosmopolitan range (Lachaise and Silvain 2004). Colonization capability often depends upon the breadth of a species' ecological niche, with species subsisting on a broader range of food or capable of withstanding wider temperature ranges being more capable of expanding their species range (Keller 2007). Of the two species, *D. melanogaster* is the more efficient colonizer, with *D. simulans* often lagging

several years behind *D. melanogaster* in the colonization of new areas (Keller 2007). This slower speed of dispersal could be in part due to *D. simulans* being less domestic than *D. melanogaster*. While *D. simulans* is only semi-domestic and less likely to enter homes or other human buildings (David et al 2004, Keller 2007), *D. melanogaster* is more closely associated with humans and is thought to be a strict human commensal, although the timing of the shift from wild to domestic habits is still under debate (Lachaise and Silvain 2004). *D. melanogaster*'s expansion of its species range and shift to domestication suggest that the species has likely undergone adaptation both to environmental factors as well as being a human commensal.

Gene expression variation and its effect on phenotype

Gene expression variation among both species and populations is a widespread phenomenon. As the use of transcriptomic technologies such as microarrays and RNA-seq has become more pervasive, the opportunity to study gene expression variation within and among species has arisen. Indeed, many such studies have been carried out, revealing the pervasiveness of gene expression variation within and among species. Expression divergence has been well-documented among closely related species, including human and chimpanzee (Khaitovich et al 2005) and *D. melanogaster* and *D. simulans* (Ranz et al 2003), as well as on evolutionary branches of closely related species, such as primates (Enard et al 2002, Khaitovich et al 2006) and within the *Drosophila* group (Kalinka et al 2010, Zhang et al 2007). There have even been studies documenting the gene expression divergence across large evolutionary distances such as across mammalian (Brawand et al 2011) or reptilian (Tzika et al 2011) lineages. Studies documenting the abundant gene expression variation within species have also been performed across many taxa, including humans (Khaitovich et al

2005, Stranger et al 2005, Storey et al 2007), fish (Oleksiak et al 2002, Aubin-Horth et al 2005, Whitehead et al 2006), *Drosophila* (Meiklejohn et al 2003, Ayroles et al 2009), yeast (Cavalieri et al 2000, Townsend et al 2003), and mice (Voolstra 2007). Studies of gene expression divergence among populations of *Drosophila melanogaster* are particularly abundant. Indeed, expression divergence has been measured between the same derived, European (the Netherlands) and ancestral, sub-Saharan African (Zimbabwe) populations of *D. melanogaster* in both sexes of whole flies (Hutter et al 2008, Müller et al 2011) as well as in the brain and Malpighian tubules (Catalán et al 2012, Huylmans and Parsch 2014). These studies have revealed extensive gene expression divergence between these European and sub-Saharan African populations (Table 1), some of which likely reflects adaptation to newly encountered biotic and abiotic conditions as *D. melanogaster* expanded out of Africa.

Table 1. Studies of expression divergence between Dutch and Zimbabwean *D. melanogaster*

Source	No. Genes[a]	Study
Whole males	1600	Hutter et al (2008)
Whole females	1364	Müller et al (2011)
Brain	328	Catálan et al (2012)
Malpighian tubules	2474	Huylmans and Parsch (2014)

[a]Number of genes showing differential expression between the populations at a false discovery rate of 0.05

Gene expression variation is thought to underlie much of the phenotypic divergence observed within and between species (King and Wilson 1975, Wray et al 2003). This variation is controlled by regulatory elements, which can either be located near a gene and directly affect expression (*cis*-regulatory elements, Figure 2) or more distant from a gene and

indirectly affect expression via interaction with a *cis*-regulatory element (*trans*-regulatory element or *trans*-acting factor, Figure 2). A well-known example of gene expression variation among closely related species is the extensive phenotypic and behavioral divergence observed between humans and chimpanzees, despite sharing approximately 98% sequence identity (Chimpanzee Sequencing and Analysis Consortium 2005). Thus, much of the observed phenotypic divergence is thought to be due to expression rather than protein divergence (King and Wilson 1975), especially in the brain (Preuss et al 2004). Another example from the *Drosophila* clade is male wing pigmentation pattern. In some species of *Drosophila*, spots of dark pigmentation form on the wingtips of males. These spots are involved in courtship display and have been gained and lost multiple times in the *melanogaster* and *obscura* groups (Prud'homme et al 2006). Although the formation of these spots is under the control of multiple genes (Gompel et al 2005), the addition or loss of spots in all of the species examined was due to changes in *cis*-regulatory elements of the pigmentation gene, *yellow* (Gompel et al 2005, Prud'homme et al 2006).

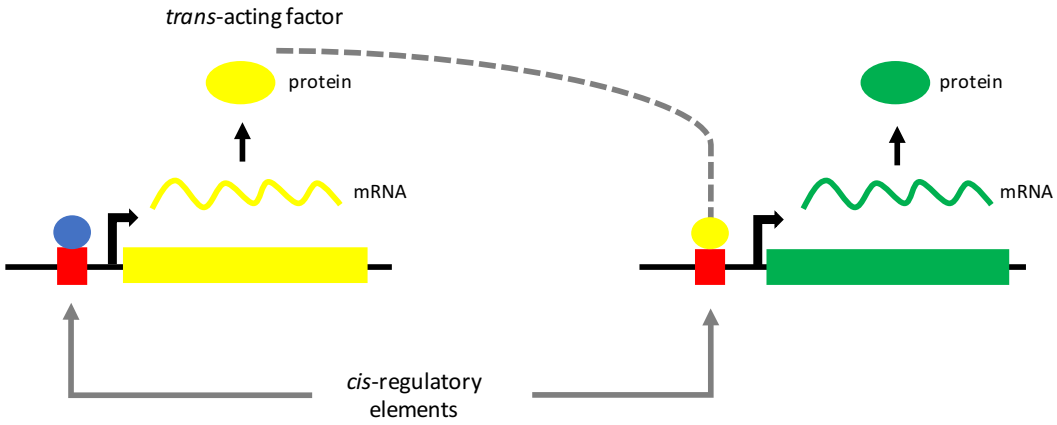


Figure 2: *cis*- versus *trans*-regulatory elements. Schematic of how *cis*-regulatory elements (shown in red) act to directly affect gene expression versus how *trans*-acting factors (shown in yellow or blue) act indirectly to affect gene expression. Hatched line indicates the yellow *trans*-acting factor's interaction with a *cis*-regulatory element.

A striking example within human populations is hair color, with darker hair more prevalent in Asian, African, and southern European populations and lighter colors, including blond, more common in Northern Europeans. Hair color determination is a complex trait and blond hair in Europeans has been associated with genetic variants linked to eight genes (Sulem et al 2007, Han et al 2008, Sulem et al 2008, Zhang et al 2013). While some variants associated with blond hair alter protein-coding regions of known pigmentation genes (Sturm 2009, Kenny et al 2012), many are located outside of protein-coding regions and are thought to be located in regulatory sequences (Sturm 2009, Hindorff 2009, ENCODE Project Consortium 2012). A SNP over 350 kb upstream of the *KITLG* gene, which encodes a ligand for the KIT receptor tyrosine kinase (Morrison-Graham et al 1993), is associated with blond hair in northern European populations and can, together with variants associated with other genes, explain 3–6% of blond hair color variation (Sulem et al 2007). Using mice, it was confirmed that this SNP's association with blond hair color is indeed a regulatory change,

with the blond-associated SNP altering a binding site for the lymphoid enhancer-binding factor 1 within a regulatory enhancer driving expression in developing hair follicles, leading to significant differences in hair pigmentation (Guenther et al 2014).

Adaptation via gene expression regulation

Gene expression divergence is thought to play a particularly important role in phenotypic evolution, since alterations in gene expression are temporally and tissue-specifically more easily adjusted and fine-tuned than changes in protein structure (Carroll 2000, Carroll 2008). The adaptive evolution of *cis*-regulatory elements, in particular, is thought to be especially abundant (Carroll 2000, Carroll 2008, Wray et al 2003, Prud'homme et al 2007, Wray 2007). This is because *cis*-regulatory changes generally affect only one gene or group of genes and alter allele-specific expression, while *trans*-regulatory changes tend to be more pleiotropic, affecting a larger number of genes across multiple tissues and/or developmental stages and will affect both alleles (Carroll 2000, Carroll 2008). While studies have demonstrated the abundance of *cis*-regulatory changes available for evolution between (Wittkopp et al 2008) and within several species, including humans (Rockman and Wray 2002) and *D. melanogaster* (Négre et al 2011, Massouras et al 2012), the relative importance of *cis*-regulatory changes versus *trans*-regulatory and other genetic changes, such as amino acid replacements within proteins, remains a topic of debate (Hoekstra and Coyne 2007, Holloway et al 2007, Wagner and Lynch 2008).

A well-known example of adaptive *cis*-regulatory divergence in *D. melanogaster* is the insertion of an *Accord* transposable element into the upstream regulatory sequence of the cytochrome P450 gene, *Cyp6g1* (Daborn et al 2002). The insertion results in increased *Cyp6g1* expression and subsequent increased tolerance of the insecticide DDT and shows

evidence of positive selection outside of Africa (Catania et al 2004, Chung et al 2007). *cis*-regulatory mutations affecting gene expression can also occur within the gene itself. A recent example of such adaptive *cis*-regulatory expression divergence in *D. melanogaster* is the insertion of the *pogo* transposable element in the *kinetochore Mis12-Ndc80 network component 1* (*Kmn1*) transcript (Mateo et al 2014). This insertion affects the choice of the polyadenylation signal for the *CG11699* gene, resulting in the production of only the *CG11699* transcript with a shortened 3' untranslated region (UTR) and higher *CG11699* expression, which was associated with increased xenobiotic resistance (Mateo et al 2014). The flanking regions of the *pogo* insertion show evidence of a recent selective sweep in non-African *D. melanogaster*, suggesting that it has undergone positive selection outside of Africa (Gonzàlez et at 2008).

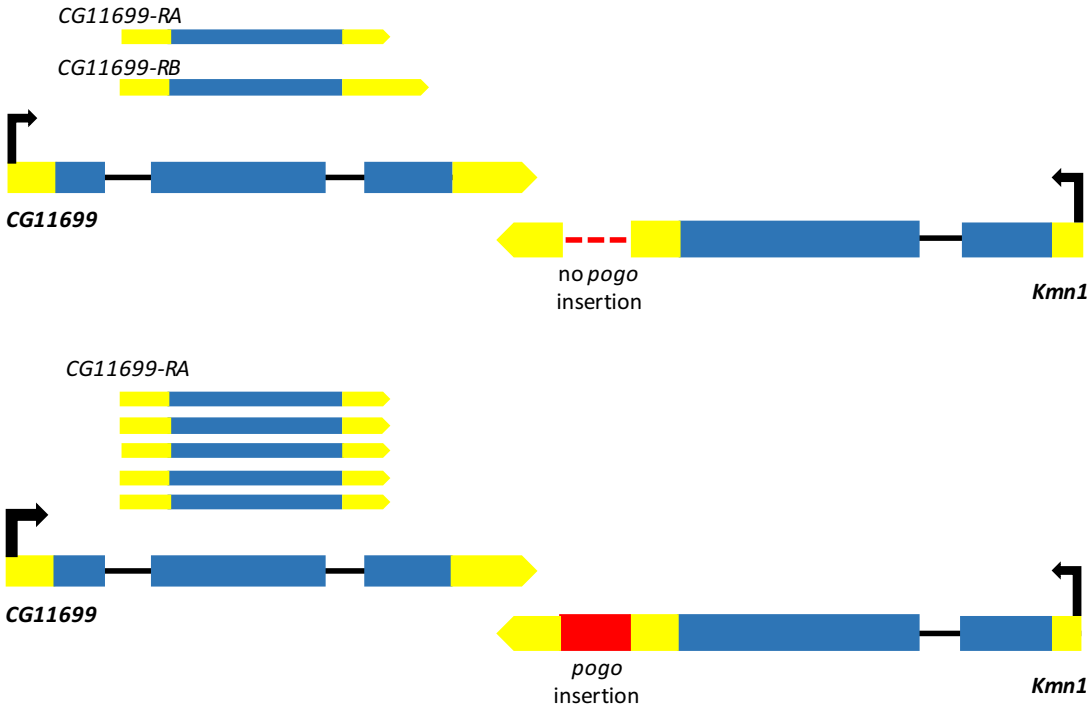


Figure 3: Effect of *pogo* insertion in *Kmn1* 3'UTR on *CG11699* expression. Exons are shown as blue boxes and UTRs as yellow boxes. The *pogo* transposon is represented as a red box, while its absence is indicated by a red hatched line. When the *pogo* transposon is inserted in the *Kmn1* 3'UTR, only the

CG11699-RA transcript, which has a shorter 3'UTR, is expressed and at higher expression levels than when the *pogo* insertion is absent.

Functional analyses of regulatory adaptations

Despite predictions of the pervasiveness of adaptive *cis*-regulatory divergence, comparatively few examples currently exist, although the number of examples continues to grow (Ingram et al 2009, González et al 2009, Daborn et al 2002, Fraser et al 2012, Sucena and Stern 2000, Saminadin-Peter et al 2012, Guio et al 2014, Mateo et al 2014, Koshikawa et al 2015). The discrepancy between predicted and verified instances of adaptive regulatory evolution is due in part to the difficulty in detecting the regulatory adaptation itself as well as determining the connection between genotype and organismal phenotype and, therefore, the trait under selection. Since phenotypes are often the result of the interaction of multiple genes, genes can have pleiotropic effects, and the effects of a large number of genes on phenotype remain unknown, it can be particularly difficult to correlate putatively adaptive *cis*-regulatory divergence with its effect on organismal phenotype. While linking an expression change to an adaptive organismal phenotype is challenging, it has proven even more difficult to elucidate the precise molecular and genetic mechanisms within adaptively evolving *cis*-regulatory elements underlying the observed change in expression. It is important to identify the genetic and molecular mechanisms that underlie adaptive *cis*–regulatory divergence, since knowledge of individual examples helps to further our understanding of the mechanisms of phenotypic evolution as a whole (Wittkopp and Kalay 2012).

Objectives

The overarching aim of this dissertation is to examine the genetic basis of adaptive regulatory changes, as well as the effects of adaptive regulatory polymorphisms on both gene expression and organismal phenotype. By analyzing specific cases of adaptive regulatory divergence, this dissertation further aims to shed light on the mechanisms through which *D. melanogaster* has adapted to new environmental conditions outside of sub-Saharan Africa. This thesis examines two genes initially identified in studies documenting gene expression divergence between European and sub-Saharan African *D. melanogaster* as candidates for adaptive evolution in the European population: *CG9509*, which was originally identified in males (Table 1, Hutter et al 2008), and *Metallothionein A* (*MtnA*), which was originally identified in the brain (Table 1, Catalán et al 2012). Chapter 3 identifies a new case of adaptive *cis*-regulatory divergence in the 3'UTR of *MtnA*. Chapters 1 and 2 provide an analysis of a case of previously identified regulatory divergence in a *cis*-regulatory element of *CG9509*. Expression of *CG9509* is 2–3-fold higher in the European population than in the sub-Saharan African population (Hutter et al 2008, Saminadin-Peter et al 2012). Transgenic reporter gene assays have shown that this expression divergence is driven by a 1.2-kb *cis*-regulatory enhancer element upstream of the *CG9509* coding region, which shows signs of recent positive selection in the European population (Saminadin-Peter et al 2012).

In Chapter 1, I show that the observed *CG9509* expression divergence between European and sub-Saharan African *D. melanogaster* extends to other cosmopolitan (here defined as outside of Africa) and sub-Saharan African populations. Together with population genetic analyses revealing low sequence variation and a high frequency, derived haplotype in the enhancer element in all cosmopolitan populations, this suggests that the

cosmopolitan variant of the *CG9509 cis*-regulatory element underwent a selective sweep for increased *CG9509* expression during or shortly after *D. melanogaster*'s expansion out of Africa. In chapter 2, I perform an in-depth functional analysis of *CG9509* and its associated *cis*-regulatory element. Up until this study, the function of *CG9509* has remained unknown. Using a series of functional and tolerance assays, I show that *CG9509* plays a role in *D. melanogaster* growth and determine how variation in *CG9509* expression affects organismal phenotype. Using site directed-mutagenesis and transgenic reporter gene constructs, I further determine the sequence variants within the *CG9509* enhancer element responsible for the observed expression divergence between cosmopolitan and sub-Saharan African *D. melanogaster*.

Chapter 3 identifies a new instance of adaptive *cis*-regulatory divergence via a modification of a 3'UTR, in this case, involving an insertion/deletion polymorphism in the *MtnA* 3'UTR. *MtnA* was originally identified as a candidate for adaptive evolution in an RNA-seq study comparing brain expression between a European and an African population and is expressed 5-fold higher in the European population (Table 1, Catalán et al 2012). I perform transgenic reporter gene assays to show that the deletion in the 3'UTR, which is at high frequency in the European population, is associated with increased expression. In concert with *MtnA*'s increased expression and signatures of a selective sweep around the *MtnA* 3'UTR in European populations, this suggests that positive selection has acted on the deletion in the *MtnA* 3'UTR for an increase in *MtnA* expression in the European population. I further use RNA interference and hydrogen peroxide tolerance assays to show that *MtnA* expression plays a role in oxidative stress tolerance and that the 3'UTR indel polymorphism is associated with oxidative stress tolerance variation in natural populations. I propose that

selection occurred on the deletion in the *MtnA* 3'UTR in order to increase oxidative stress tolerance in populations outside of the ancestral species range in sub-Saharan Africa.

Chapter 1

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 (1632): 20130024.

*These authors contributed equally

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].

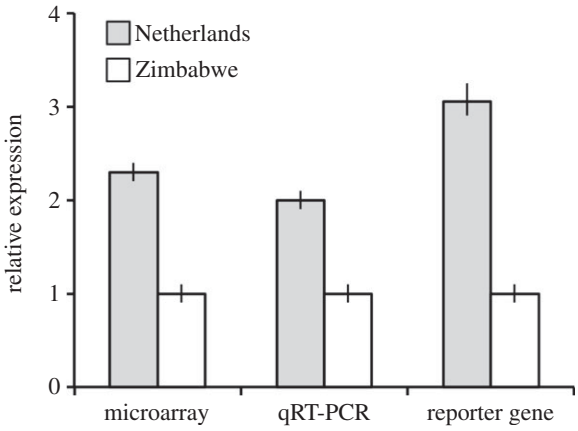


Figure 1. Expression of *CG9509* in a European (The Netherlands) and a sub-Saharan African (Zimbabwe) population. Shown are the relative expression levels in adult males as determined by microarrays or qRT-PCR. The 'reporter gene' comparison is for *lacZ* transgene expression driven by either the European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

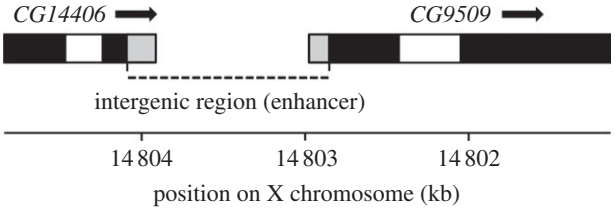


Figure 2. Map of the *CG9509* region of *D. melanogaster*. Transcriptional units are indicated by boxes, with coding regions in black, introns in white and untranslated regions in grey. The arrows indicate the direction of transcription. The intergenic region between the stop codon of *CG14406* and the 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*

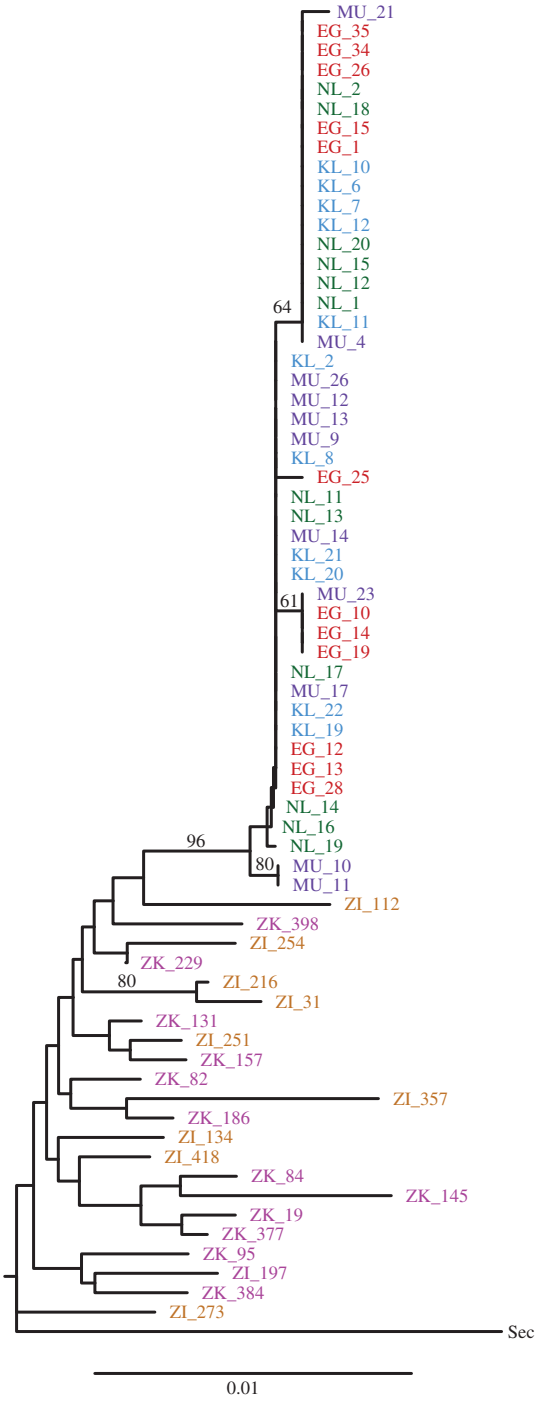


Figure 3. Neighbour-joining tree of all intergenic region sequences. The population abbreviations are as follows: The Netherlands (NL), Germany (MU), Malaysia (KL), Egypt (EG), Zambia (ZI) and Zimbabwe (ZK). *Drosophila sechellia* (Sec) was used as an outgroup. The branch lengths are proportional to 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.)

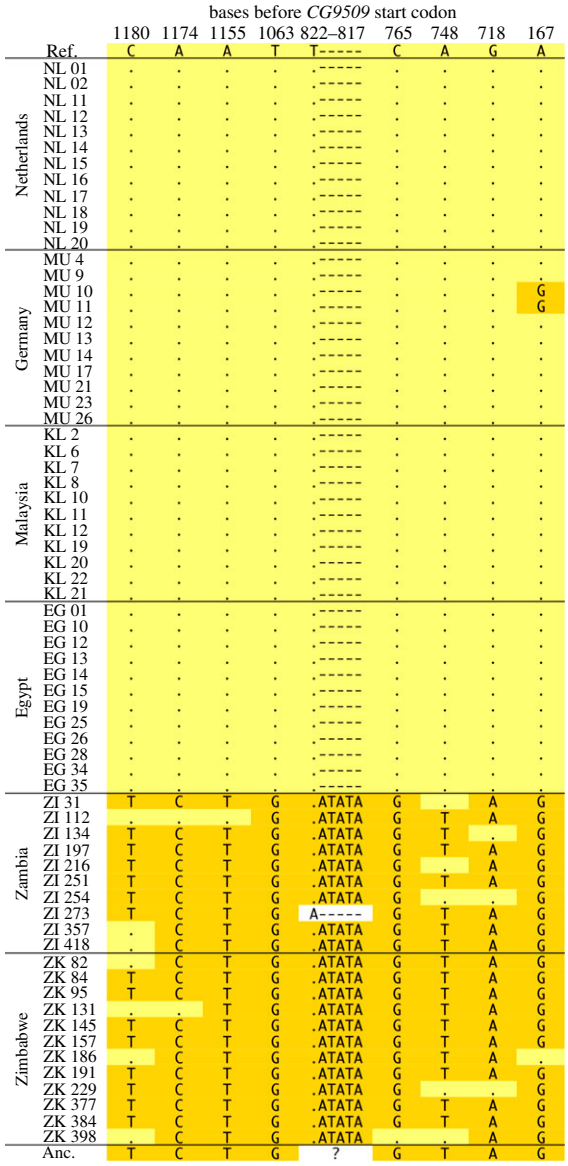


Figure 4. Fixed and nearly fixed differences in the *CG9509* enhancer region between cosmopolitan and sub-Saharan African populations. Cosmopolitan variants are indicated by light shading and sub-Saharan African variants by dark shading. Ambiguous variants are shown in white. The reference sequence (Ref.) 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

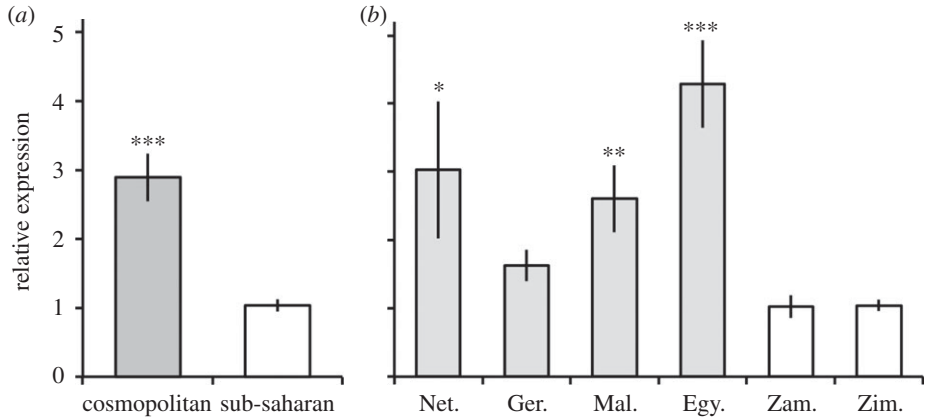


Figure 5. Expression of *CG9509* in cosmopolitan and sub-Saharan African populations. (*a*) Expression in pooled cosmopolitan and sub-Saharan African populations. (*b*) Expression in the Netherlands (Net.), Germany (Ger.), Malaysia (Mal.), Egypt (Egy.), Zambia (Zam.) and Zimbabwe (Zim.). Cosmopolitan 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*

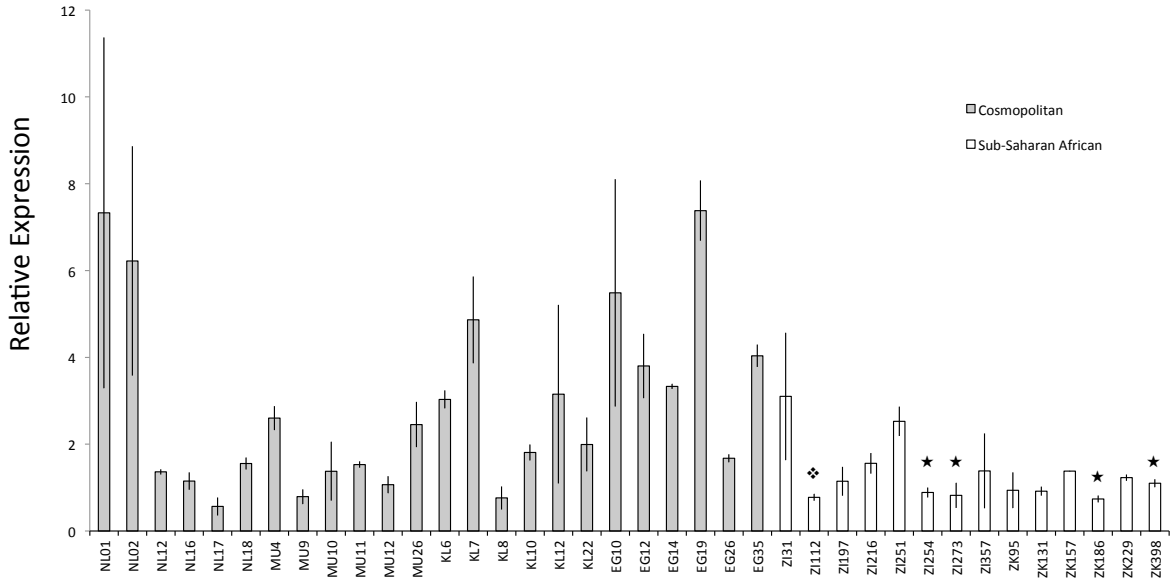
Amanda Glaser-Schmitt, Ana Catalán and John Parsch

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.

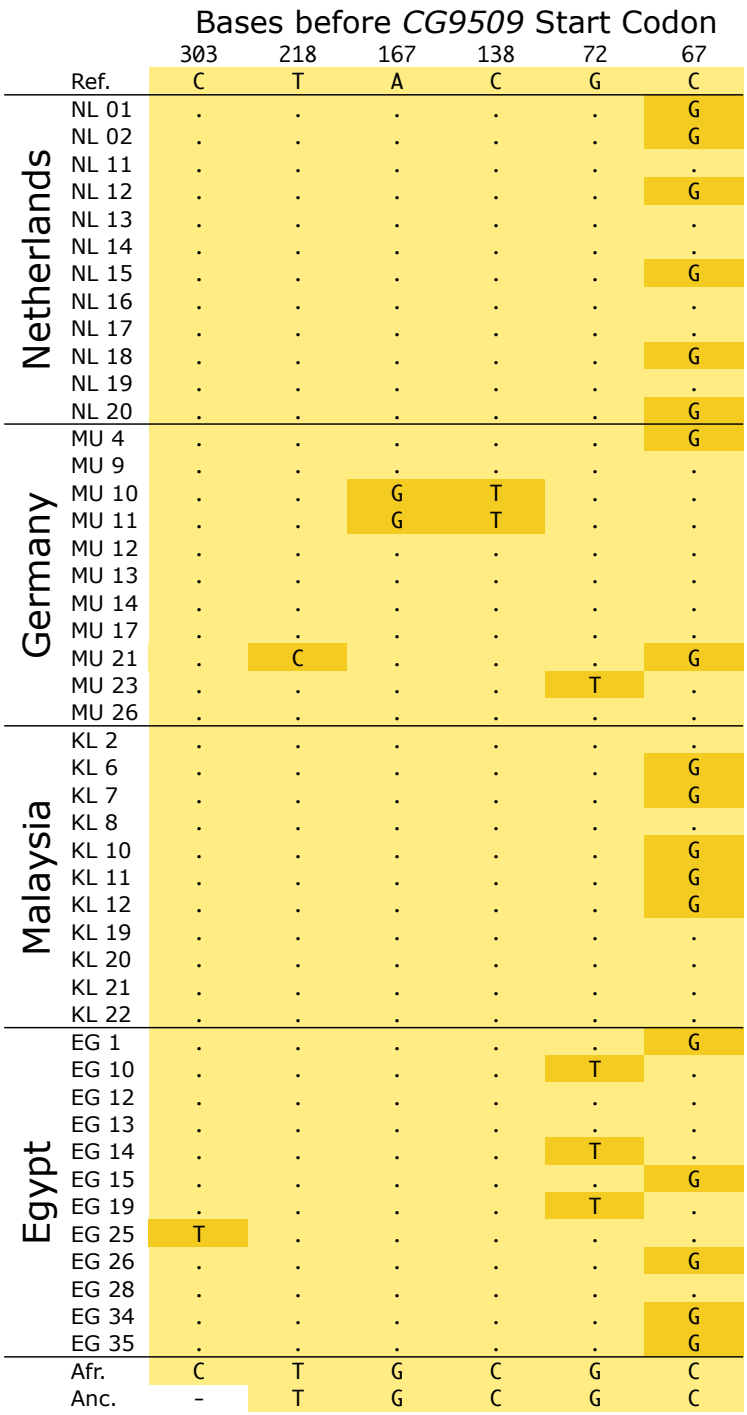
	Netherlands	Germany	Malaysia	Egypt	Zambia	Zimbabwe
Netherlands	–	0.156	0.153	0.153	0.455	0.414
Germany	0.095	–	0.042	0.026	0.495	0.443
Malaysia	0.045	0.092	–	0.011	0.511	0.459
Egypt	0.068	0.108	0.067	–	0.408	0.457
Zambia	1.159	1.144	1.155	1.175	–	0.001
Zimbabwe	1.079	1.070	1.075	1.095	0.882	–

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.

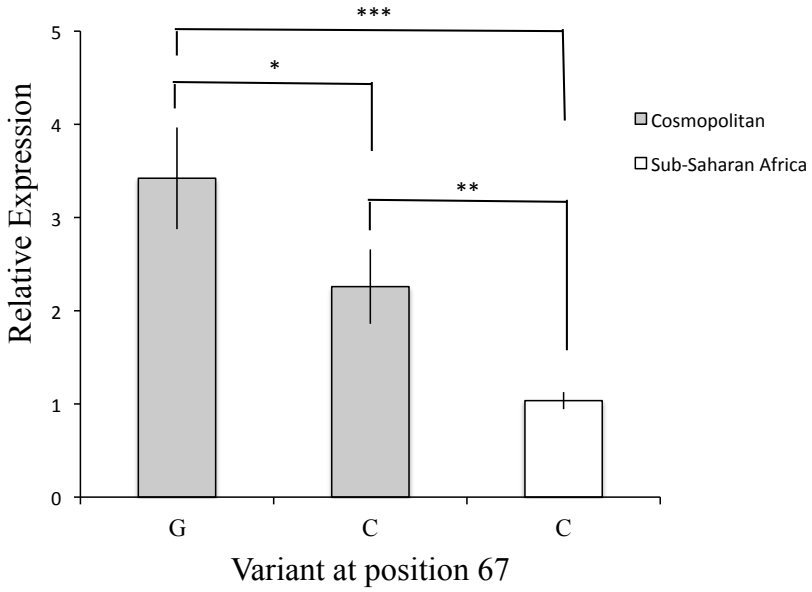
Population	D_s	P_s	D_{enh}	P_{enh}	*P*-value
Cosmopolitan (Netherlands)	17	5	38	1	0.020
Sub-Saharan African (Zimbabwe)	12	29	30	22	0.007



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 2

Functional characterization of adaptive variation within a *cis*-regulatory element influencing *Drosophila melanogaster* growth

Amanda Glaser-Schmitt and John Parsch

(Unpublished Manuscript)

Functional characterization of adaptive variation within a *cis*-regulatory element influencing *Drosophila melanogaster* growth

Amanda Glaser-Schmitt and John Parsch

(Unpublished Manuscript)

Abstract

Gene expression variation is a major contributor to phenotypic diversity within species and is thought to play an important role in adaptation. However, examples of adaptive regulatory polymorphism are rare, especially those that have been characterized at both the molecular genetic and the organismal level. In this study, we perform a functional analysis of the *Drosophila melanogaster CG9509* enhancer, a *cis*-regulatory sequence that shows evidence of adaptive evolution in populations outside the species' ancestral range in sub-Saharan Africa. Using site-directed mutagenesis and transgenic reporter gene assays, we determined that three single nucleotide polymorphisms are responsible for the difference in *CG9509* expression that is observed between sub-Saharan African and other cosmopolitan populations. To elucidate the function of *CG9509*, we performed a series of functional and tolerance assays on flies in which *CG9509* expression was manipulated by genetic mutations or RNA interference. We found that *CG9509* plays a role in larval growth and influences adult body and wing size, as well as wing loading. The effect on growth appears to result

from a modulation of active ecdysone levels and expression of growth factors. Taken together, our findings suggest that selection acted on three sites within the *CG9509* enhancer to increase *CG9509* expression and, as a result, reduce wing loading as *D. melanogaster* expanded out of sub-Saharan Africa. Interestingly, while two of the variants that provide an increase in *CG9509* expression appear to have been the target of a selective sweep outside of sub-Saharan Africa, the variant with the largest effect on expression remains polymorphic in cosmopolitan populations, suggesting that it has been subject to balancing selection.

Introduction

Gene expression variation is extensive both within and between species and is believed to underlie much of the phenotypic diversity observed among species, as well as among populations of the same species (King and Wilson 1975, Wray et al 2003). Furthermore, expression variation is thought to provide an abundant source of material for adaptation, as alterations in gene expression are more easily fine-tuned on a temporal and tissue-specific scale than changes in protein structure (Carroll 2000, Carroll 2008). In particular, *cis*-regulatory elements, which are adjacent to genes and directly affect their expression, are thought to be frequent targets of adaptive evolution (Carroll 2000, Carroll 2008, Wray et al 2003, Prud'homme et al 2007, Wray 2007). Despite this prediction, examples of adaptive *cis*-regulatory changes remain comparatively rare, although the number of such examples continues to grow (Ingram et al 2009, González et al 2009, Daborn et al 2002, Fraser et al 2012, Sucena and Stern 2000, Saminadin-Peter et al 2012, Guio et al 2014, Mateo et al 2014, Koshikawa et al 2015, Catalán et al 2016). The discrepancy between

the predicted abundance and actual instances of identified adaptive *cis*-regulatory divergence is likely in part due to the difficulty in detecting regulatory adaptation, as well as in determining its effect on an organismal phenotype under selection. Even in some of the best-studied species, the function of many genes remains unknown and alterations in those with known functions often have pleiotropic effects, making it difficult to determine the link between an expression change and an adaptive organismal phenotype. As more instances of adaptive *cis*-regulatory evolution are uncovered, it is important to identify the genetic and molecular mechanisms that underlie them, which can help to further our understanding of the mechanisms of phenotypic evolution (Wittkopp and Kalay 2012). However, studies performing in-depth functional analyses of individual adaptively-evolving *cis*-regulatory elements remain even more rare than those documenting adaptive *cis*-regulatory divergence (for examples see: Gompel et al 2005, Prud'homme et al 2006, Kwasnieski et al 2012, Chang et al 2013).

As RNA-seq has become more cost-effective and widespread, the identification of putatively adaptive alterations in gene expression within and among species has become more efficient and the utilization of it and similar methods has proven very effective in doing so (Cooper et al 2003, Holloway et al 2007, Hutter et al 2008, Müller et al 2011, Catalán et al 2012, Schoville et al 2012, Huylmans and Parsch 2014). *CG9509* is a gene initially identified as a candidate for adaptive *cis*-regulatory divergence through one such study that compared expression between a derived, European and an ancestral, sub-Saharan African (henceforth sub-Saharan) population of *D. melanogaster* (Hutter et al 2008). Until now, the function of *CG9509* has remained unknown, although it has been postulated to have choline dehydrogenase activity (McQuilton et al 2012) and/or play a role in the ecdysteroid metabolic process (Iida et al 2007). *CG9509* expression was found to be 2–3-fold higher in

adults of the European population than in the sub-Saharan population (Figure 4A, Hutter et al 2008, Saminadin-Peter et al 2012) and this expression difference has been shown to extend to additional cosmopolitan (here defined as populations outside of south and central Africa) and sub-Saharan populations (Glaser-Schmitt et al 2013). Transgenic reporter gene experiments revealed that the observed expression divergence can be fully accounted for by variation within a 1.2-kb *cis*-regulatory element (referred to here as the *CG9509* enhancer, Figure 5A) located just upstream of the gene (Saminadin-Peter et al 2012). Consistent with a selective sweep, the *CG9509* enhancer showed greatly reduced sequence variation and a high frequency, derived haplotype in cosmopolitan populations (Saminadin-Peter et al 2012, Glaser-Schmitt et al 2013). These findings suggest that the *CG9509* enhancer was the target of a selective sweep for increased *CG9509* expression after *D. melanogaster*'s expansion out of sub-Saharan Africa, which is estimated to have occurred around 15,000 years ago (Glinka et al 2003, Ometto et al 2005, Li and Stephan 2006), but before the separation of the European and Asian populations about 2,500–5,000 years ago (Laurent et al 2011, Glaser-Schmitt et al 2013). Within the *CG9509* enhancer, there are nine single nucleotide polymorphisms (SNPs) and one insertion/deletion (indel) polymorphism (Figure 5B) that show large differences in frequency between the populations and are candidates for the target(s) of selection responsible for the observed expression divergence. In all cases, the cosmopolitan variant was inferred to be the derived state, while the sub-Saharan variant was inferred to be ancestral (Glaser-Schmitt et al. 2013).

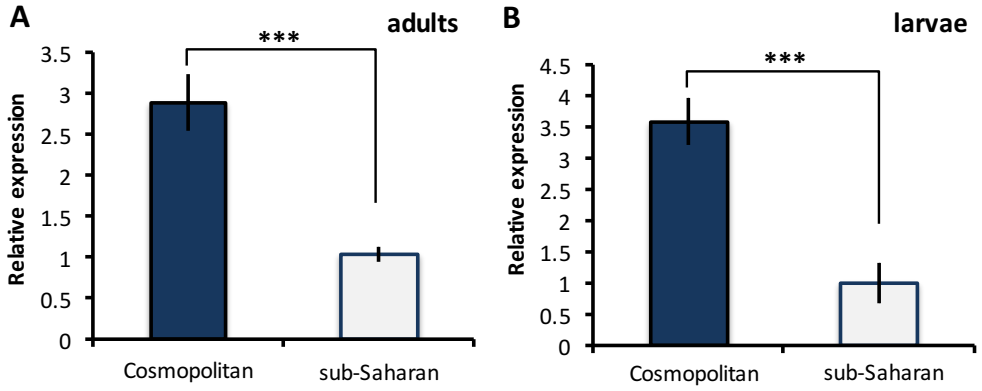
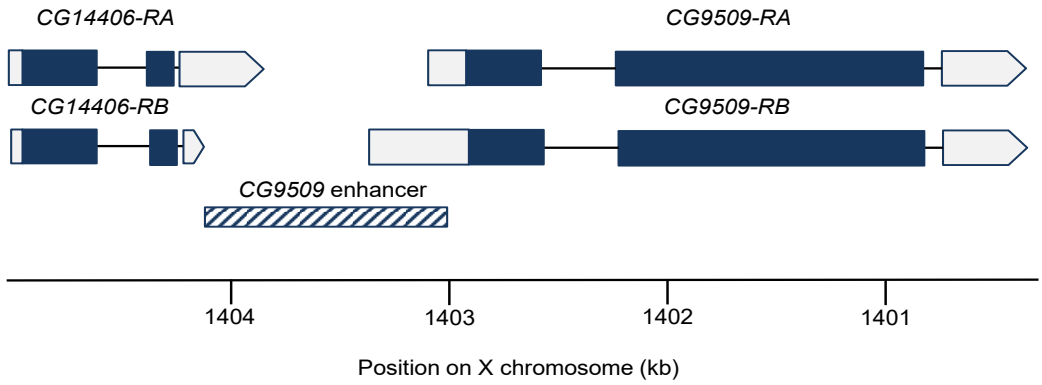


Figure 4: *CG9509* expression in cosmopolitan and sub-Saharan African *D. melanogaster.* Relative expression levels as determined by qRT-PCR in (A) adult males (Glaser-Schmitt et al 2013) and (B) late wandering 3rd instar larvae are shown. Expression is shown relative to the sub-Saharan populations. Blue bars represent cosmopolitan flies and white bars represent sub-Saharan flies. Error bars indicate standard error of the mean. Differences between populations were assessed with a *t*-test. **P*<0.05, ***P*<0.01, ****P*<0.005.

In this study, we use site-directed mutagenesis and transgenic reporter genes to determine the effect of individual SNP and indel variants within the *CG9509* enhancer on *CG9509* expression. In addition, we use RNA interference (RNAi) and a newly-discovered *CG9509* null allele to elucidate some of *CG9509*'s previously unknown biological functions. We find that three SNPs within the *CG9509* enhancer contribute to the observed expression divergence seen between cosmopolitan and sub-Saharan alleles. We also show that *CG9509* expression influences larval growth rate and, thus, plays a role in body size determination, including wing loading (i.e. the ratio of body mass to wing area). Because wing loading is thought to help facilitate flight at colder temperatures (Gilchrist and Huey 2004, Stalker 1980), we propose that selection on the three SNPs within the *CG9509* enhancer likely occurred in order to reduce wing loading outside of sub-Saharan Africa. Interestingly, two of

the SNPs with a small effect on *CG9509* expression are fixed in cosmopolitan populations and appear to have been the targets of a selective sweep, while the SNP with the largest effect on *CG9509* expression is at intermediate frequency in cosmopolitan populations, suggesting that it has been a more recent target of balancing selection.

A



B

		Position (bp before *CG9509* start codon)									
		1180	1174	1155	1063	821-817	765	748	718	167	67
Variant	sub-Saharan	T	C	T	G	ATATA	G	T	A	G	C
	Cosmopolitan	C	A	A	T	-----	C	A	G	A	G
Population	Netherlands	100	100	100	100	100	100	100	100	100	50
	Germany	100	100	100	100	100	100	100	100	81.8	18.1
	Malaysia	100	100	100	100	100	100	100	100	100	45.5
	Egypt	100	100	100	100	100	100	100	100	100	41.7
	Zambia	30	10	10	0	0	0	30	20	0	0
	Zimbabwe	30	8.3	0	0	0	8.3	16.7	8.3	8.3	0

Figure 5: The *CG9509* gene region. (A) Schematic of the *CG9509* gene and enhancer. Blue boxes represent exons and white boxes represent UTRs with the pointed ends indicating the direction of transcription. The *CG9509* enhancer is indicated by a hatched box. (B) SNPs and indels in the *CG9509* enhancer with >10% frequency in cosmopolitan populations. Cosmopolitan sequence variants are indicated in blue and sub-Saharan variants are indicated in white, with lighter shades of blue indicating a mixture of both variants in the population. For each population, the observed frequency of the cosmopolitan variant (in percent) is shown (Glaser-Schmitt et al 2013).

Results

CG9509 expression divergence in larvae

Previous studies of *CG9509* expression variation focused solely on adults (Saminadin-Peter et al 2012, Glaser-Schmitt et al 2013). In order to determine if the adult expression pattern is established earlier in development, we surveyed *CG9509* expression in late wandering 3rd instar larvae of three cosmopolitan populations (the Netherlands, Egypt, and Malaysia) and two sub-Saharan populations (Zimbabwe and Zambia). Since *CG9509* expression in the Zimbabwean and Zambian populations was very similar and did not differ significantly (*t*-test, $P = 0.749$), they were pooled for analysis. Similar to adults (Figures 4A and S1C), larval *CG9509* expression in cosmopolitan populations was significantly higher than in sub-Saharan populations by an average of 3.5-fold (*t*-test, $P < 10^{-10}$, Figure 4B). When examined individually, larval *CG9509* expression in each cosmopolitan population was significantly higher than in the sub-Saharan populations by 3–5.5-fold ($P < 5.10^{-4}$ for all populations, Figure S1A), which was also similar to the results seen for adults (Glaser-Schmitt et al 2013).

Functional analysis of sequence variants in the *CG9509* enhancer

In order to determine which variants in the *CG9509* enhancer region (Figure 5B) contribute to the observed expression divergence between cosmopolitan and sub-Saharan *D. melanogaster* (Figures 4 and S1), we performed site-directed mutagenesis to create a series of transgenic reporter gene constructs. Briefly, the 1.2-kb enhancer region of a cosmopolitan and a sub-Saharan *CG9509* allele was cloned in front of a *LacZ* reporter gene and six sites of interest at positions 1174, 1155, 1063, 821-817, 765, and 67 (Figure 5B) were

mutated individually and in various combinations. The tested sites were chosen as those at which the derived variant was fixed in all cosmopolitan populations, but at a frequency ≤10% in sub-Saharan populations. Position 67, at which the derived variant is at intermediate frequency in cosmopolitan populations, but absent from the sub-Saharan populations, was also tested because it had been identified previously as being associated with *CG9509* expression variation (Massouras et al 2012, Glaser-Schmitt et al 2013). The remaining three sites (1180, 748, 718), which had derived variants at frequencies ≥20% in at least one sub-Saharan population (Figure 5B) and showed no association with *CG9509* expression among sub-Saharan lines (Glaser-Schmitt et al 2013) were not tested.

Mutations were first introduced into the cosmopolitan version of the *CG9509* enhancer in order to change the nucleotide(s) to the ancestral (sub-Saharan) state. Sites found to have an effect on expression in the cosmopolitan background were subsequently mutated to the cosmopolitan state in the sub-Saharan background. The reporter gene constructs were then introduced into *D. melanogaster* via the PhiC31 integration system (Bischof et al 2007). In agreement with previous findings (Saminadin-Peter et al 2012), adult expression from the cosmopolitan version of the *CG9509* enhancer was 2–3-fold higher than that from the sub-Saharan African version and was highly significant for both males (*t*-test, $P < 5 \times 10^{-4}$, Figure 6A) and females ($P < 5 \times 10^{-6}$, Figure 6B). In larvae, expression driven by the cosmopolitan *CG9509* enhancer was about 4-fold higher than that driven by the sub-Saharan enhancer ($P < 10^{-6}$, Figure 6C), which is in line with the expression divergence found in natural populations (Figures 4B and S1A). Thus, similar to adults, the *CG9509* enhancer can account for the majority of the expression divergence between cosmopolitan and sub-Saharan larvae. Expression of all reporter gene constructs can be found in Figure S2.

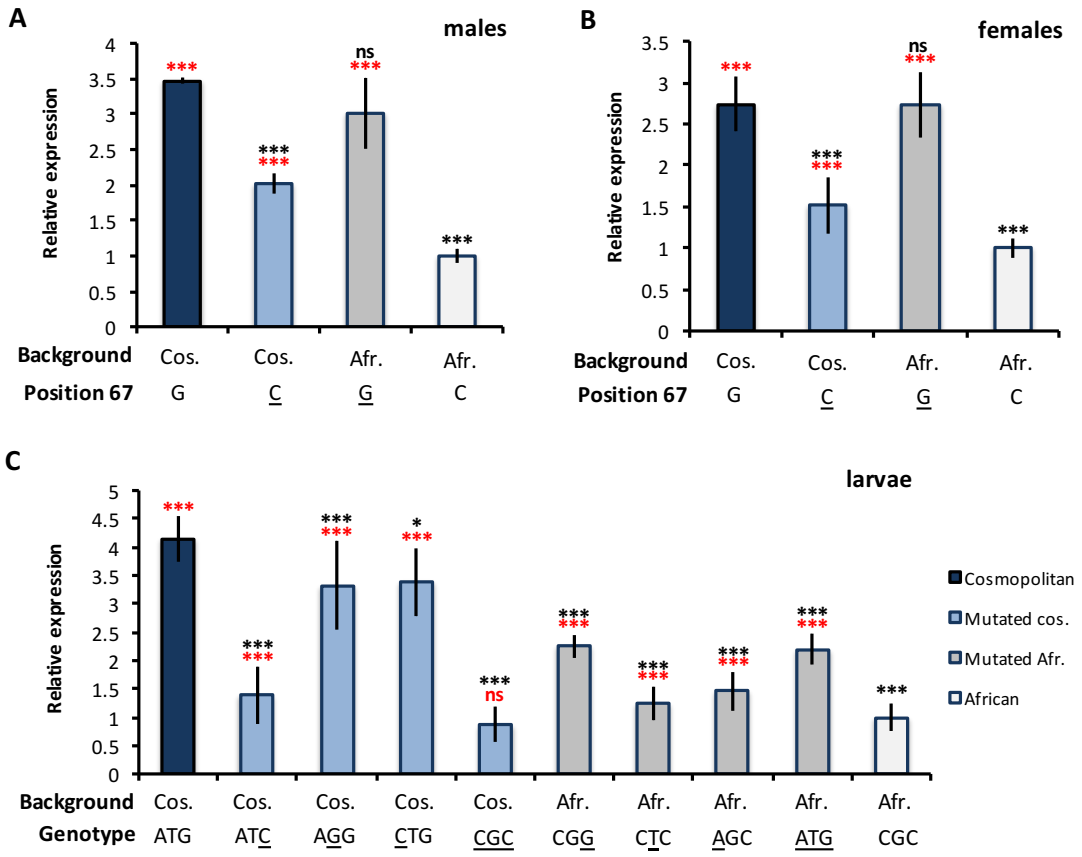


Figure 6: Transgenic reporter gene expression of variants contributing to *CG9509* expression divergence. *CG9509* reporter gene expression in adult (A) males, (B) females, and (C) 3rd instar wandering larvae. Expression of the *LacZ* reporter gene was measured as the change in absorbance per minute at 420 nm. Relative expression, which was calculated in comparison to expression of the sub-Saharan enhancer, is shown. Blue bars indicate expression driven by a wild-type cosmopolitan (Cos.) enhancer, while white bars show expression driven by a wild-type sub-Saharan African (Afr.) enhancer. Light blue bars indicate expression driven by the enhancer after mutations were introduced into a cosmopolitan background. Dark gray bars show expression driven by the enhancer after mutations were introduced in a sub-Saharan background. "Genotype" indicates the nucleotides at positions 1174, 1063, and 67 before the *CG9509* start codon, respectively. Mutated sites are underlined. Significance was assessed using a *t*-test with Bonferroni multiple test correction. Significance is represented in black for comparisons to the wild-type cosmopolitan enhancer and in

red for comparisons to the sub-Saharan enhancer. ns, not significant, *P<0.05, **P<0.01, ***P<0.005.

One SNP accounts for most of the observed adult expression divergence. We determined that the SNP at position 67 (Figure 5B) accounts for the majority of the expression divergence observed between adults of the cosmopolitan and sub-Saharan populations. The sub-Saharan "C" variant of this SNP segregates at intermediate frequency in cosmopolitan populations (Figure 5B) and, when introduced into the cosmopolitan *CG9509* enhancer background, leads to a ~45% reduction in expression in both males (*t*-test, $P < 5 \times 10^{-4}$, Figure 6A,) and females ($P < 10^{-6}$, Figure 6B). However, despite this reduction, expression was still significantly higher than that driven by the sub-Saharan version of the enhancer in both sexes (male $P < 5 \times 10^{-4}$, female $P = 0.002$, Figures 6A and 6B). When the cosmopolitan "G" variant was introduced into the sub-Saharan background, expression increased ~3-fold (male $P = 0.004$, female $P < 5 \times 10^{-4}$, Figures 6A and 6B), which brought expression up to the cosmopolitan level. There was no significant difference in expression driven by the mutated sub-Saharan enhancer and the normal cosmopolitan enhancer (male $P = 0.436$, female $P = 1$, Figures 6A and 6B,). Thus, while the cosmopolitan variant "G" is sufficient to produce cosmopolitan expression in a sub-Saharan background, there appear to be background-specific effects, presumably due to other site(s) within the cosmopolitan enhancer, that prevent a complete return to sub-Saharan expression when the "C" variant is introduced into a cosmopolitan background.

To examine the effect of the SNP at position 67 on expression in natural populations, we looked for an association between the sequence variant and adult male *CG9509*

expression in a population from the Netherlands (which was polymorphic at this site) and a population from Zimbabwe (which was monomorphic for the sub-Saharan "C" variant). We found that *CG9509* expression in Dutch flies with the "G" variant was 4-fold higher than that in Zimbabwean flies (*t*-test, $P < 5x10^{-4}$, Figure S1D), while expression in Dutch flies with the "C" variant was not significantly different from that of Zimbabwean flies ($P = 0.648$, Figure S1D). Thus, these results are generally consistent with those of our reporter gene experiments and suggest that the SNP at position 67 can explain nearly all of the expression difference observed between cosmopolitan and sub-Saharan African adults.

Three SNPs contribute to the observed larval expression divergence. We identified three SNPs in the *CG9509* enhancer at positions 67, 1063, and 1174 (Figure 5B) that can account for the majority of the expression divergence observed between cosmopolitan and sub-Saharan African larvae. Of the three, the SNP at position 67, which accounts for the majority of adult expression divergence (Figures 6A and 6B), has the largest individual effect on expression. Introduction of the cosmopolitan "G" variant into an otherwise sub-Saharan background increased expression by 2.25-fold (*t*-test, $P < 10^{-9}$, Figure 6C), while introduction of the sub-Saharan "C" variant into the cosmopolitan background reduced expression by 2-fold ($P < 5 \times 10^{-8}$, Figure 6C). At position 1063, introducing the cosmopolitan "T" variant into the sub-Saharan background increased expression by 25% ($P < 0.001$, Figure 6C), while introducing a sub-Saharan "G" variant into the cosmopolitan background reduced expression by 20% ($P = 0.001$, Figure 6C). Similarly, at position 1174, introducing a sub-Saharan "C" variant into the cosmopolitan background reduced expression by 20% in ($P = 0.012$, Figure

6C), while introducing a cosmopolitan "A" variant into the sub-Saharan African background increased expression by 50% (Figure 6C, $P < 5 \times 10^{-4}$).

When the sub-Saharan variants at all three positions (67, 1063, and 1174) were introduced together into the cosmopolitan background, expression was significantly reduced (*t*-test, $P < 5 \times 10^{-7}$) to levels equivalent to sub-Saharan expression and showed no significant difference from sub-Saharan expression ($P = 0.88$, Figure 6C). When the cosmopolitan variants at all three positions were introduced together into the sub-Saharan background, expression was increased by 2.2-fold ($P < 10^{-10}$, Figure 6C); however, it still remained 50% lower than cosmopolitan expression ($P < 5 \times 10^{-5}$, Figure 6C). Furthermore, when all three cosmopolitan variants were introduced into the sub-Saharan background expression was not significantly different from that when only the cosmopolitan "G" variant at position 67 was introduced ($P = 1$, Figures 6C and S2C). Thus, there appear to be background-specific effects, presumably caused by interactions with other sub-Saharan variants within the *CG9509* enhancer, which prevent expression from reaching the full cosmopolitan level.

To confirm the effect of the SNP at position 67 on expression in natural populations, we looked for an association between the sequence variant and larval *CG9509* expression in the cosmopolitan (all of which were polymorphic at this site) and sub-Saharan (which was monomorphic for the sub-Saharan "C" variant) populations surveyed in the preceding section. With the exception of the Egyptian population, where expression was 5–6-fold higher than sub-Saharan expression, regardless of the variant at position 67 (*t*-test, $P < 5 \times 10^{-4}$ for both, Figure S1B), expression in cosmopolitan larvae with the "G" variant was 2–4-fold higher than in larvae with a "C" variant ($P < 0.001$ for both, Figure S1B). However, only in the Dutch population was expression of larvae with the sub-Saharan "C" variant not significantly

different from sub-Saharan expression (P = 1, Figure S1B). These results are consistent with our reporter gene results, suggesting that while the sequence variant at position 67 makes a substantial contribution to the observed *CG9509* expression divergence between larval cosmopolitan and sub-Saharan African populations, it cannot fully account for it.

The prevalence of sex-, stage-, and background-specific effects. In addition to the background- and stage- specific effects mentioned in preceding sections, we identified several other stage-, background-, and sex-specific effects. We have already shown that introducing a cosmopolitan "G" at position 67 into a sub-Saharan background is enough to produce cosmopolitan expression in adults (Figures 6A and 6B). However, when we additionally introduced cosmopolitan variants "T" and "A" at positions 1063 and 1174, respectively, into a sub-Saharan background already containing a cosmopolitan "G" at positon 67, expression was reduced to 60–70% of cosmopolitan levels for both males (*t*-test, $P < 10^{-4}$, Figure S2A) and females ($P < 5 \times 10^{-5}$, Figure S2B). In males, the expression decrease associated with the additional introduction of cosmopolitan variants at positions 1063 and 1174 to a sub-Saharan African background already containing a cosmopolitan "G" at position 67 was significant ($P < 5 \times 10^{-4}$, Figure S2A), but not in females after multiple test correction (P = 0.182, Figure S2B). When the reverse was performed, and sub-Saharan African variants "G" and "C" at positions 1063 and 1174, respectively, were introduced into a cosmopolitan background already containing a sub-Saharan African "C" at position 67, there was also a 20% decrease in expression, which was significant for both sexes (male P = 0.008, female P = 0.007, Figure S2). Furthermore, expression in both sexes was also significantly reduced by 20–30% when cosmopolitan variants for both position 1063 (male P = 0.049, female P =

0.002, Figure S2) and position 1174 (*P* = 0.001 for both, Figure S2) were introduced individually into a sub-Saharan African background. These expression changes are in the opposite direction of those seen in larvae, where positions 1063 and 1174 contribute to the observed expression divergence between cosmopolitan and sub-Saharan African populations (Figure 6C). On the other hand, similar to its effect in larvae, the introduction of the sub-Saharan African variant "G" at position 1063 in a cosmopolitan background significantly reduced expression by 20%, but only in males (*P* = 0.012, Figure S2). Lastly, introducing a sub-Saharan African "T" at position 1155 in a cosmopolitan background significantly increased expression by 10–15% in both sexes (male *P* = 0.041, female *P* = 0.009, Figure S2).

Functional analysis of *CG9509*

In order to determine the function of *CG9509* and the effects of its expression on the organismal phenotype, we performed a series of functional and tolerance assays on flies in which *CG9509* expression was either knocked out by a frameshift mutation (*CG9509*$^{-}$) or knocked down by RNA interference (RNAi), as well as their respective controls. The *CG9509* knockout line (*CG9509*$^{-}$) was discovered as a spontaneous mutation in an isofemale line derived from Munich, Germany (Figure S6). As a control, we compared it to other isofemale lines derived from the same population at the same time. *CG9509* knockdown was achieved by crossing a ubiquitous *Act5C*-GAL4 driver line to a transgenic line expressing an RNAi hairpin construct specific to *CG9509* (RNAi-*CG9509*) flanked by a yeast upstream activating sequence (UAS). As a control, we used the host strain from which it was derived (UAS^{-}).

Effect of *CG9509* expression on cold, ethanol, and insecticide tolerance. *CG9509* knockdown flies were tested for cold, ethanol, and insecticide (DDT and malathion) tolerance. If *CG9509* expression plays a role in tolerance to cold temperatures or any of the tested compounds, we expect that a decrease in *CG9509* expression should lead to a decrease in tolerance. For cold tolerance in males and ethanol tolerance in both sexes, we found no effect when *CG9509* expression was knocked down (*t*-test and generalized linear model, respectively, Table 2, Figure S3). However, contrary to expectations, we found that knocking down *CG9509* expression led to an increase in DDT and malathion tolerance, as well as female cold tolerance (generalized linear models and *t*-test, respectively, Table 2, Figure S3).

Table 2. Effect of *CG9509* knockdown in adult tolerance assays

Tolerance Assay	Effect[a]	*P*-value[b]
DDT	+	0.013
Malathion	+	2.84×10^{-5}
Cold (females)	+	0.009
Cold (males)	0	0.248
Ethanol	0	0.615

[a]Directional effect of *CG9509* knockdown on tolerance. [b]*P*-values were obtained from generalized linear models with a quasi-binomial distribution using sex, fly strain, and compound concentration as factors or, for cold tolerance, a *t*-test.

Effect of *CG9509* expression on developmental timing and larval growth rate. *CG9509* knockdown and knockout larvae were measured for larval growth rate as well as time from 1st instar larvae to pupariation and duration of the wandering stage as measures of developmental timing. Knocking down or knocking out *CG9509* expression had no effect on

developmental timing, with no significant change in either time from 1st instar larvae to pupariation (*t*-test, Figures S4A and S4C) or duration of the wandering stage (Figures S4B and S4D). On the other hand, knocking down or knocking out *CG9509* expression was associated with a significant increase in the larval growth rate (Figure 7), with larvae in which *CG9509* expression had been knocked down (*t*-test, $P < 0.05$ for all stages, Figure 7A) or knocked out ($P \leq 0.001$ for all stages, Figure 7B) significantly larger than their respective control larvae for all stages (Figure 8A).

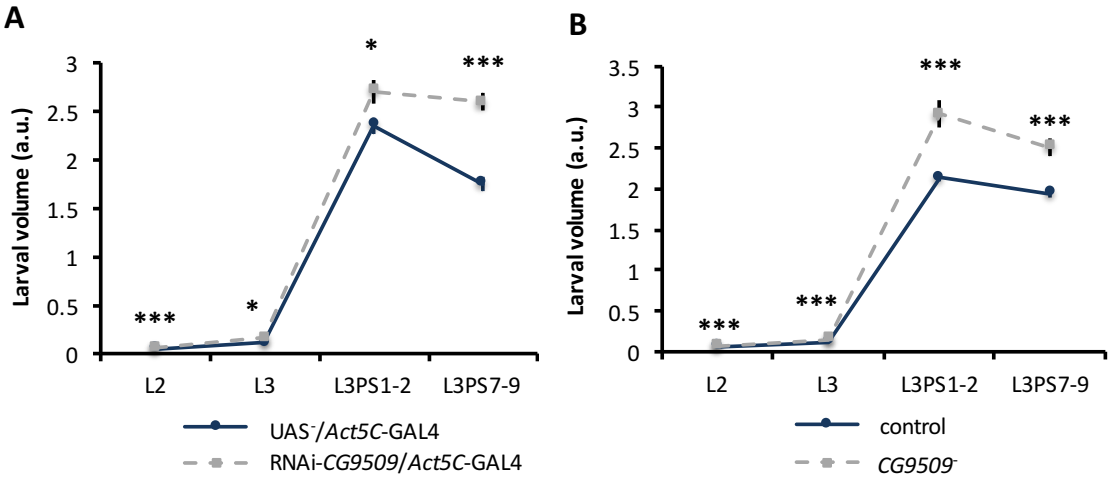


Figure 7: Effect of *CG9509* expression on larval growth rate. Larval volume in arbitrary units (a.u.) in (A) UAS-/*Act5C*-GAL4 (blue lines) and RNAi-*CG9509*/*Act5C*-GAL4 (gray, hatched lines) larvae and (B) *CG9509*- (gray, hatched lines) and control (blue lines) larvae. Four larval stages were examined: 2nd instar larvae (L2) after the 1st to 2nd instar larval molt approximately 48 hours after egg laying (AEL), 3rd instar larvae (L3) after the 2nd to 3rd instar molt (approximately 72 hours AEL), early wandering 3rd instar larvae (L3PS1-2, approximately 110 hours AEL), and late wandering 3rd instar larvae (L3PS7-9, approximately 116 hours AEL). Error bars represent standard error of the mean. Significance was assessed via *t*-test for each larval stage and a Bonferroni multiple test correction was applied. **P*<0.05, ***P*<0.01, ****P*<0.005.

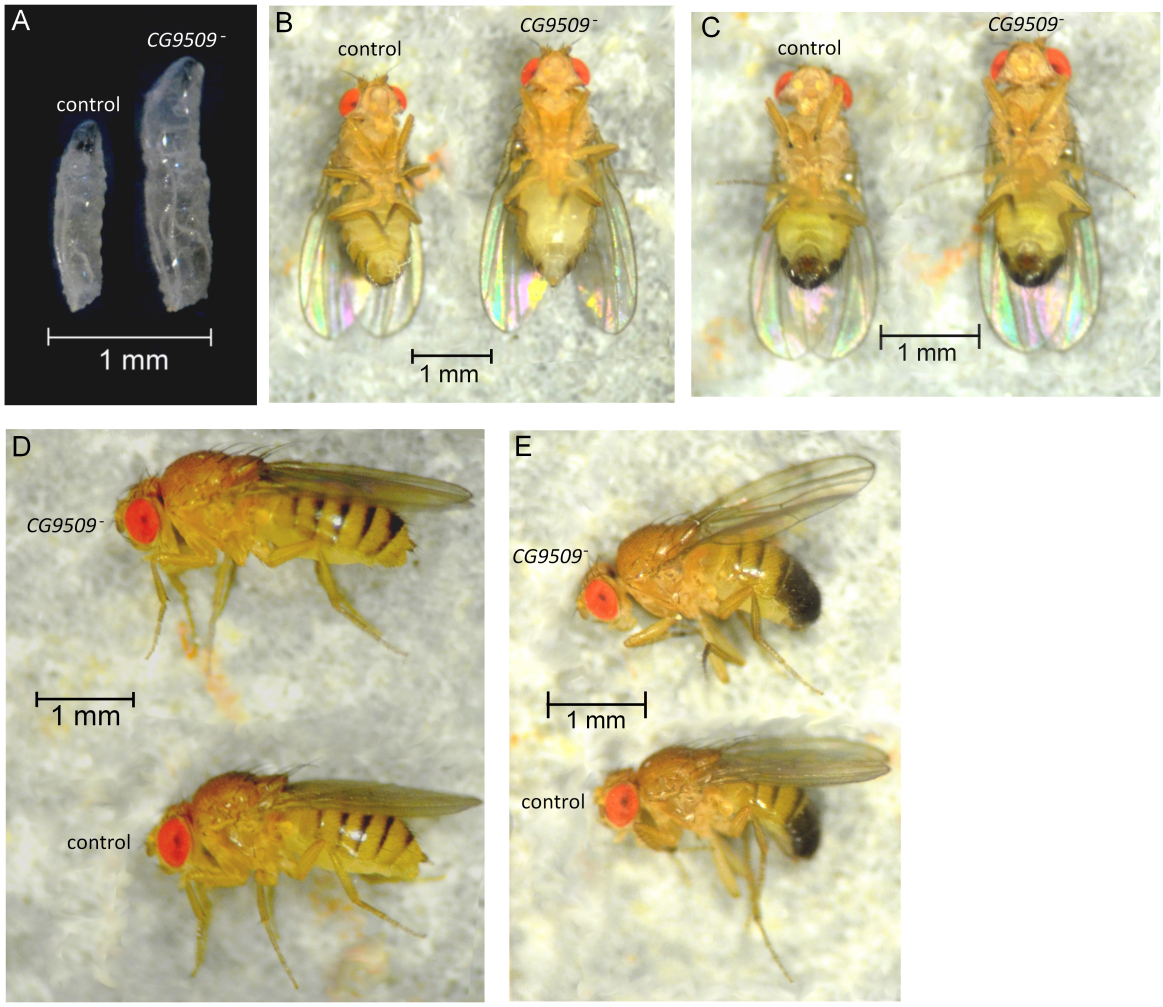


Figure 8: Body size in *CG9509*⁻ and control flies. Pictured are *CG9509*⁻ (A-C: right, D-E: top) and control (A-C: left, D-E: bottom) (A) 3rd instar larvae after the 2nd to 3rd instar molt approximately 72 hours after egg laying, adult female flies in (B) ventral view and (D) side view, and adult male flies in (C) ventral view and (E) side view.

Effect of *CG9509* expression on adult body size. We examined adult body size, as measured by body weight and wing size (length and area), in *CG9509* knockdown and knockout flies. Knocking out or knocking down *CG9509* expression increased body weight by approximately 8–15% in both males (*t*-test, $P < 10^{-4}$ for both, Figures 9A and 9B) and females ($P < 5 \times 10^{-6}$ and $P = 0.023$, respectively, Figures 9A and 9B). Similarly, knocking out or down *CG9509*

expression increased wing size by approximately 4–11% in both sexes, which was significant for all comparisons (*t*-test, wing length $P < 0.05$ for all, female wing area $P \leq 0.002$, Figures 9D and 9E) except male wing area in *CG9509*$^{-}$ flies ($P = 0.266$, Figure 9E).

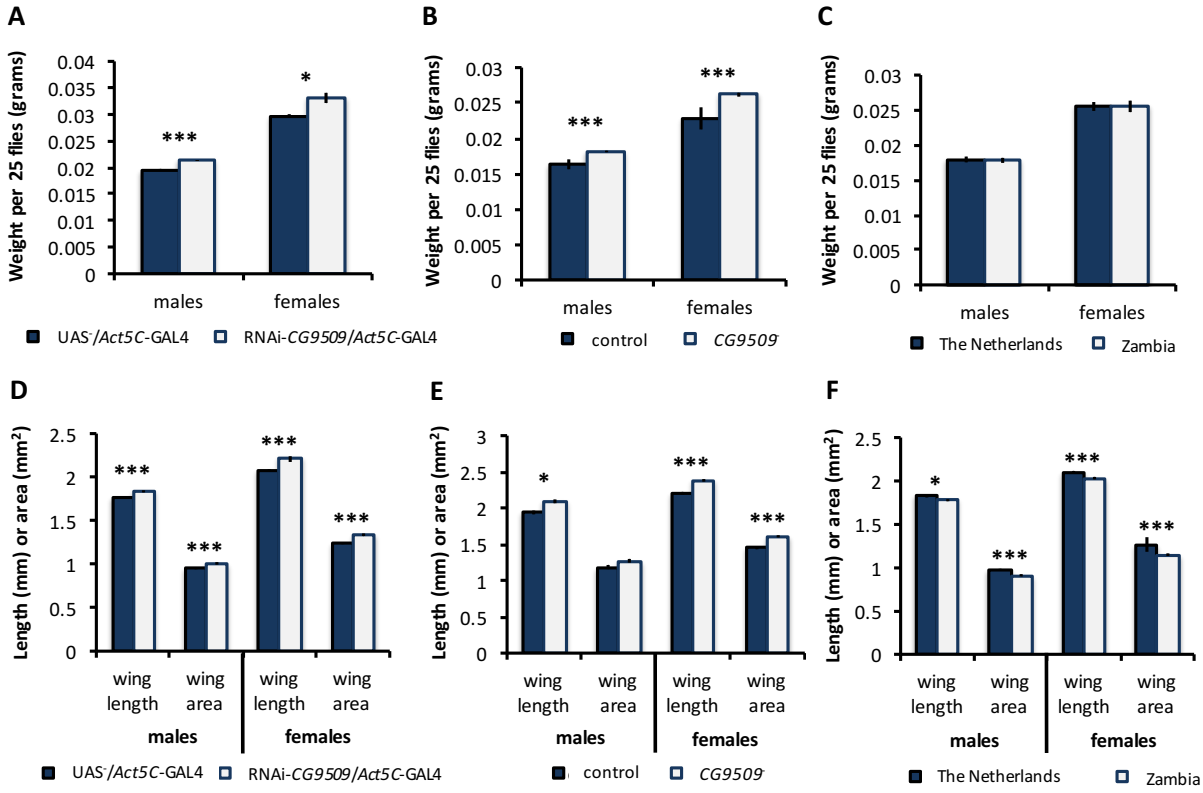


Figure 9: Effect of *CG9509* expression on body size. (A-C) Body weight per 25 flies in (A) UAS^{-}/*Act5C*-GAL4 (blue) and RNAi-*CG9509*/*Act5C*-GAL4 (white) flies, (B) *CG9509*$^{-}$ (white) and control (blue) flies, and (C) Zambian (white) and Dutch (blue) flies. (D-F) Wing length and wing area in (D) UAS^{-}/Act5C-GAL4 (blue) and RNAi-*CG9509*/Act5C-GAL4 (white) flies, (E) *CG9509*$^{-}$ (white) and control (blue) flies, and (F) Zambian (white) and Dutch (blue) flies. Error bars represent standard error of the mean. Significance was assessed via *t*-test. **P*<0.05, ***P*<0.01, ****P*<0.005.

Next, we examined body size in a Dutch and a Zambian population. For body weight, we found no significant difference between the two populations for both sexes (*t*-test,

Figure 9C). On the other hand, Dutch wing size was significantly larger for both sexes (Figure 9F), with an approximately 2–3% increase in wing length (male $P = 0.013$, female $P < 10^{-5}$) and a 7–10% increase in wing area (male $P < 5 \times 10^{-4}$, female $P < 10^{-9}$). This finding is the opposite of expectations if selection on the *CG9509* enhancer occurred to body size, but is in agreement with previous studies documenting clinal variation in body size (James and Partridge 1995, James et al 1995, Robinson and Partridge 2001, de Jong and Bochdanovits 2003). Interestingly, the sequence variant at position 67 (Figure 5B), which contributes to both larval and adult *CG9509* expression divergence (Figure 6), is associated with body and wing size variation within the Dutch population. Males and females with the high expression, cosmopolitan "G" variant had larger wings (*t*-test, wing length $P < 0.02$ for both, wing area $P < 0.03$ for both, Figure S5B) than flies with the sub-Saharan "C" variant. Flies with the high expression, cosmopolitan "G" variant were also heavier, although this was only significant for males ($P < 0.001$, Figure S5A).

Effect of *CG9509* expression on wing loading. Wing loading is the relationship between the mass of a flying object and the wing area, or more specifically, the pressure that the wings exert on the surrounding air (Gilchrist and Huey 2004) and is dependent upon relative body size. In order to determine if *CG9509* expression variation affects wing loading, we measured the wing load index (wet weight/wing area) in *CG9509* knockout and knockdown flies. Knocking down or knocking out *CG9509* expression significantly increased wing loading by approximately 10–35% in both sexes (*t*-test, male $P < 10^{-5}$ for both, female $P = 0.02$ for both, Figures 10A and 10B). Next, we examined wing loading in a Zambian and a Dutch population. Dutch wing loading was approximately 3–7% lower than Zambian wing loading (Figure 10C),

however, this difference was only significant in males (*t*-test, male $P = 0.04$, female $P = 0.136$). We also found an association of the sequence variant at position 67 (Figure 5B), which contributes to larval and adult *CG9509* expression divergence (Figure 6), with wing loading within the Dutch population, with the cosmopolitan, high expression "G" variant associated with decreased wing loading (Figure S5C); however, this was only significant for females (*t*-test, $P = 0.006$). It is likely that the lack of a significant difference in wing loading between Dutch and Zambian females is due to this observed variation associated with the sequence variation at position 67 within the Dutch population. Indeed, when flies are separated according to their variant at position 67, wing loading in females with the high expression, cosmopolitan "G" variant is significantly higher than in Zambian females ($P < 5 \times 10^{-4}$); while, wing loading in females with the low expression, sub-Saharan "C" variant is not significantly different from Zambian females ($P = 0.720$). Taken together, these results suggest that reduced wing loading is the likely organismal phenotype under selection.

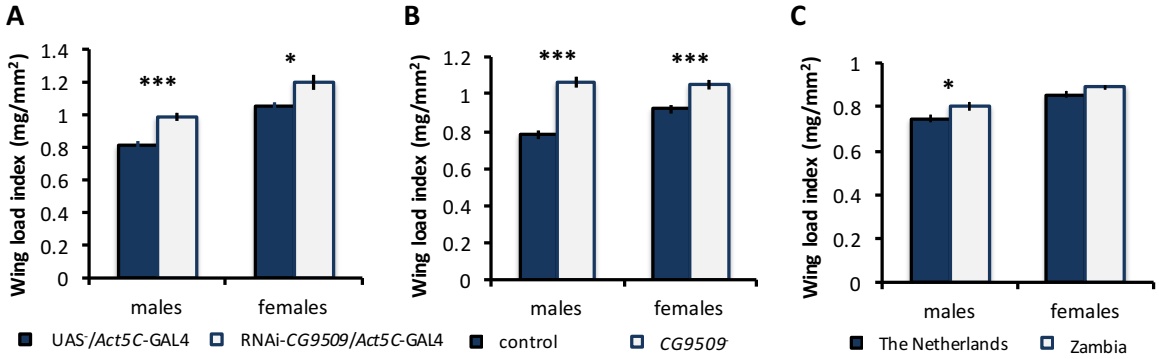


Figure 10: Effect of *CG9509* expression on wing loading. Wing load index in (A) UAS^{-}/*Act5C*-GAL4 (blue) and RNAi-*CG9509*/*Act5C*-GAL4 (white) flies, (B) *CG9509*$^{-}$ (white) and control (blue) flies, and (C) Zambian (white) and Dutch (blue) flies. Error bars represent standard error of the mean. Significance was assessed via *t*-test. **P*<0.05, ***P*<0.01, ****P*<0.005.

Effect of *CG9509* expression on active ecdysone levels. *CG9509* has been predicted to be involved in the metabolism of the steroid hormone ecdysone (Iida et al 2007), pulses of which act as temporal signals during *D. melanogaster* development to help trigger various developmental transitions (Quinn et al 2012, Mirth and Riddiford 2007). As an approximation of active ecdysone levels in early and late wandering 3rd instar larvae (puff stages 1–2 and 7–9, respectively), we measured expression levels of *E74B*, which is directly activated by ecdysone and often used as a readout of active ecdysone levels (Karim and Thummel 1991, Caldwell et al 2005, Layalle et al 2008), in *CG9509* knockdown flies. When *CG9509* expression was knocked down, *E74B* expression significantly decreased by approximately 50% in both early (*t*-test, $P = 0.038$, Figure 11A) and late ($P = 0.016$, Figure 11B) wandering 3rd instar larvae. Thus, *CG9509* expression is positively correlated with levels of active ecdysone.

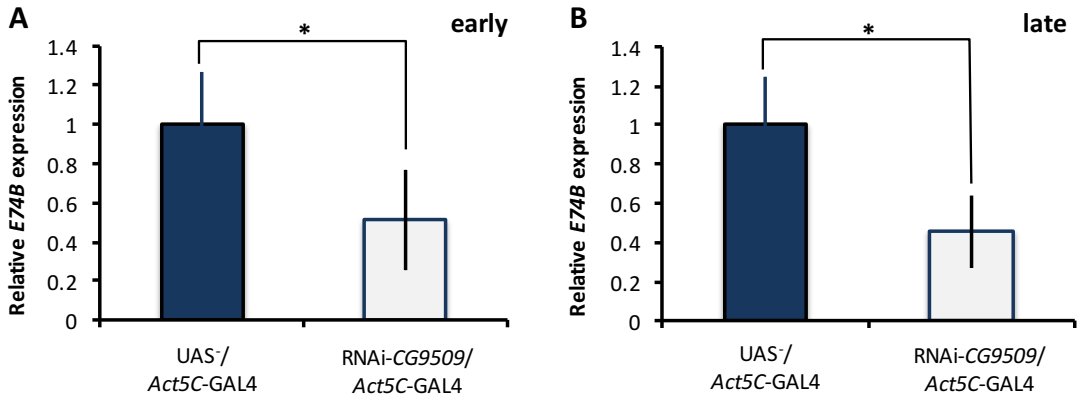


Figure 11: Effect of *CG9509* expression on active ecdysone levels. Levels of active ecdysone, approximated by relative *E74B* expression as measured by qRT-PCR in (A) early and (B) late wandering 3rd instar larvae in UAS-/*Act5C*-GAL4 (blue bars) and RNAi-*CG9509*/*Act5C*-GAL4 (white bars) flies. Expression is shown relative to UAS-/*Act5C*-GAL4. Error bars represent standard error of the mean. Significance was assessed via *t*-test. **P*<0.05, ***P*<0.01, ****P*<0.005.

Effect of *CG9509* expression on expression of two known growth regulators. During larval development, ecdysone and insulin signaling interact to influence growth (Tennessen and Thummel 2011, Yamanaka et al 2013), and integral to this interaction are the growth regulators forkhead box, sub-group O (dFOXO) and *Myc* (*dMyc*). We examined *dMyc* and *dFOXO* expression in both early (puff stages 1-2) and late (puff stages 7-9) wandering 3rd instar *CG9509* knockdown larvae, since these genes are known to have developmental stage-specific effects (Quinn et al 2012, Tennessen and Thummel 2011). Knocking down *CG9509* expression was significantly associated with an approximately 40–50% decrease in both *dMyc* and *dFOXO* expression in late wandering 3rd instar larvae (*t*-test, $P = 0.012$ and $P = 0.001$ respectively, Figure 12B). On the other hand, knocking down *CG9509* expression had no significant effect on *dMyc* or *dFOXO* expression in early wandering 3rd instar larvae (*t*-test, Figure 12A). Thus, *CG9509* expression is associated with the expression of these two important growth regulators and this association is stage-specific.

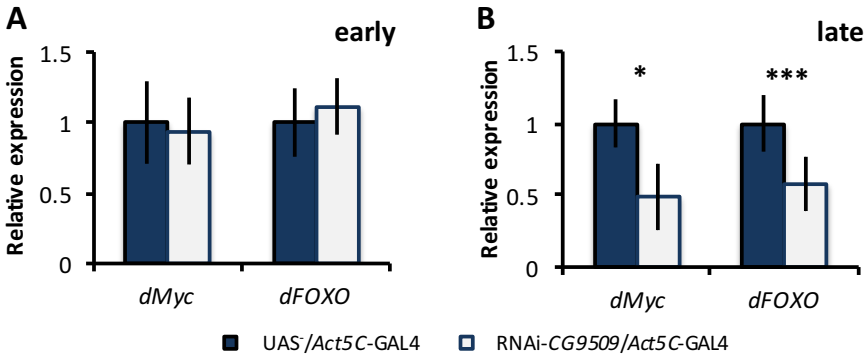


Figure 12: Effect of *CG9509* expression on *dMyc* and *dFOXO* expression. Relative gene expression of *dMyc* and *dFOXO* in UAS-/*Act5C*-GAL4 (blue) and RNAi-*CG9509*/*Act5C*-GAL4 (white) (A) early (puff stages 1-2) wandering 3rd instar larvae and (B) late (puff stages 7-9) wandering 3rd instar larvae. Expression is calculated relative to UAS-/*Act5C*-GAL4 for each stage. Error bars represent standard error of the mean. Significance was assessed with a *t*-test. **P*<0.05, ***P*<0.01, ****P*<0.005.

Discussion

We have shown that the previously identified *CG9509* expression divergence extends to larvae (Figure 4) and this divergence is driven by polymorphism in the *CG9509* enhancer (Figure 6). We identified three SNPs within the *CG9509* enhancer at positions 1174, 1063, and 67 (Figure 5B) that can account for the majority of the observed expression divergence (Figure 6). Two of the identified SNPs (positions 1174 and 1063, Figure 5B) only affect expression in larvae and have a relatively small effect on larval *CG9509* expression (Figure 6C). The cosmopolitan variants of these SNPs are fixed in European, Asian, and Middle Eastern populations (Figure 5B), but absent or at low frequency in the ancestral, sub-Saharan African populations and low-to-intermediate frequency in central African populations (Figure 5B, Table 3). The third SNP at positon 67 (Figure 5B) can account for the majority *of CG9509* expression divergence and affects expression in both adults and larvae (Figure 6). The cosmopolitan variant of this SNP segregates at moderate-to-high frequency in the European, Asian, and Middle Eastern populations (Figure 5B), but is absent from the ancestral, sub-Saharan and the central African populations (Figure 5B, Table 3).

Table 3. Percentage of cosmopolitan SNPs contributing to expression divergence in central African populations

		Frequency (%) cosmopolitan SNP variant		
Population	**N[§]**	**Pos.* 1174[L]**	**Pos. 1063[L] (N_A)**	**Pos. 67[L,A]**
Camaroon	10	30	0 (8)	0
Rwanda	25	16	9.5 (21)	0

[§]Number of lines surveyed (Pool et al 2012). *Position in bp before *CG9509* start codon. N_A, Number of lines available for comparison if different from N. [L]Site contributes to larval expression divergence. [A]Site contributes to adult expression divergence.

We propose that the *CG9509* enhancer was the target of a two-step selective sweep. First, the cosmopolitan variants at positions 1174 and 1063, each of which has a small effect on *CG9509* expression in larvae, were the target of a selective sweep. Based on the distribution of these SNPs in natural populations (Figure 5B Table 3), this sweep likely occurred during or shortly after *D. melanogaster*'s expansion out of sub-Saharan Africa and before the separation of European and Asian populations. The presence of these SNPs at low frequency in central African populations (Table 3) suggests that selection occurred on standing variation within a derived, non-sub-Saharan African progenitor population from which the cosmopolitan populations stem, potentially in the Middle East, which has been hypothesized to be the source population of European and Asian populations (Laurent et al 2011). Since a single haplotype spanning both position 1174 and 1063 is fixed in the cosmopolitan populations, it appears that the sweep targeted a chromosome on which both derived variants were present and that they were not fixed by two independent sweeps. In a second step, we propose that the cosmopolitan variant at position 67, which has a large effect on *CG9509* expression, arose as a new mutation on the selected haplotype and rose to intermediate frequency more recently in cosmopolitan populations.

The maintenance of this cosmopolitan SNP at intermediate frequency in cosmopolitan populations (Figure 5B) suggests that it may be a target of balancing selection. It has been proposed that advantageous mutations with large effects are likely to display heterozygotic advantage, and that this may be especially relevant to gene regulatory polymorphisms (Sellis et al 2011). Under such a scenario, individuals that are heterozygous at a regulatory locus may achieve an optimal level of expression, while those that are homozygous may overshoot the optimum (Sellis et al 2011). Thus, overdominant selection (heterozygote advantage) will maintain the polymorphism. The large effect of the variant at

position 67 on *CG9509* expression (Figure 6) and its intermediate frequency (~40–50%) in cosmopolitan populations (Figure 5B) are consistent with this model. However, more complex models involving interactions of alleles at multiple loci are also possible (Sellis et al 2011, Pavlidis et al 2012).

Until now, *CG9509*'s function and, therefore, the organismal phenotype(s) affected by variation in *CG9509* expression have remained unknown. Here, we used RNAi-mediated knockdown of *CG9509* expression and a newly identified *CG9509* null mutant to show that increased *CG9509* expression is associated with reduced larval growth (Figure 7) and final adult body size (Figures 8 and 9), as well as increased levels of the maturation hormone ecdysone (Figure 11). Most likely, the increased active ecdysone levels result in the reduced larval growth rate and a subsequently smaller body size, since adult body size is known to be influenced by larval growth, which is suppressed by ecdysone's antagonization of insulin signaling (Colombani et al 2005, Caldwell et al 2005). *CG9509*'s role in larval growth is further supported by the negative correlation of other growth associated genes with stress tolerance (Colombani et al 2005, Broughton et al 2005), which we also found for *CG9509* expression (Table 2). However, it must be noted that this correlation could simply be a by-product of body size, which has been shown to correlate with stress tolerance in *Drosophila* (Gibbs and Matzkin 2001).

In *Drosophila*, and indeed all flying animals, relative size is very important for the aerodynamics of flight. Cold has a particularly detrimental effect on *Drosophila* flight dynamics since, as small, flight-capable ectotherms, their wing beat frequency and power output decrease as temperature decreases (Curtsinger and Laurie-Ahlberg 1981, Unwin and Corbet 1984, Stevenson and Josephson 1990), resulting in reduced flight ability at cold

temperatures (Gilchrist and Huey 2004). Theory and evidence from previous studies in *Drosophila* suggest that reduced wing loading may facilitate better flight ability at cold temperatures (Gilchrist and Huey 2004, Stalker 1980). Here, we have shown that increased *CG9509* expression is associated with reduced wing loading (Figures 10 and S5C), wing loading is reduced in a European population in comparison to a sub-Saharan population (Figure 10C), and wing loading in the European population could be associated with polymorphism within the *CG9509* enhancer (Figure S5C). Coupled with *CG9509*'s increased larval expression in cosmopolitan populations (Figures 4B and S1), these findings suggest that the selection on the *CG9509* enhancer likely favored the reduction in wing loading conferred by increased larval *CG9509* expression as *D. melanogaster* expanded out of Africa. Thus, improved flight ability at cooler temperatures may have imposed an important selective constraint as *D. melanogaster* expanded its species range outside of sub-Saharan Africa.

The mechanism through which *CG9509* expression interacts with ecdysone and insulin signaling to fine-tune larval growth and reduce wing loading remains unknown. The effect of *CG9509* expression on wing loading (Figures 10 and S5C) is likely at least in part due to its effect on active ecdysone levels (Figure 11), since ecdysone is thought to play a key role in regulating proportional growth and integrating individual organ growth with each other as well as the entire body (Mirth and Shingleton 2012). *CG9509* is highly expressed in the larval fat body (Chintapelli et al 2007), which is functionally analogous to a mammalian liver and fat cells and acts as a coordinator of larval growth (Tennessen and Thummel 2011, Quinn et al 2012, Mirth and Riddiford 2007). Ecdysone signaling specifically in the fat body antagonizes insulin signaling in part via down-regulation of the positive growth regulator *dMyc* expression as well as translocation of the negative growth regulator dFOXO to the

nucleus where it can activate the expression of target genes (Dlanoue et al 2010, Tennessen and Thummel 2011, Yamanaka et al 2013). When we knocked down *CG9509* expression, we found a stage-specific decrease in both *dMyc* and *dFOXO* expression in late wandering third instar larvae (Figure 12), which is the stage during which the peak of the final and largest larval ecdysone titer occurs, signaling the onset of pupariation (Mirth and Riddiford 2007, Quinn et al 2012). It is interesting that we see this stage-specific reduction in *dFOXO* transcript expression when *CG9509* expression is decreased since it is the localization of the dFOXO protein that suppresses growth (Dlanoue et al 2010, Tennessen and Thummel 2011, Yamanaka et al 2013). A previous study on transcriptional variation within the insulin/TOR network found strong co-variance of gene expression for *dFOXO* and dFOXO-affected genes (Nuzhdin et 2009), suggesting that the expression of genes downstream of dFOXO may also be affected when *CG9509* expression is knocked down. Counterintuitively, we found a similar stage-specific decrease in expression of the positive growth regulator *dMyc* (Figure 12), which we would expect to be up-regulated in the fat body when *CG9509* expression is knocked down. However, the effect of ecdysone on *dMyc* expression is both stage- and tissue-specific (Quinn et al 2012), thus, the observed decrease in expression is likely to be in another tissue, and may represent a part of the mechanism through which *CG9509* expression is able to adjust proportional body size to affect wing loading.

It is important to elucidate both the mechanisms behind and the selective forces driving adaptive divergence of individual *cis*-regulatory sequences as these examples help us to understand the genetic mechanisms of evolution, especially phenotypic evolution. Our results provide evidence that in cosmopolitan populations of *D. melanogaster,* the selection identified in previous studies (Saminadin-Peter et al 2012, Glaser-Schmitt et al 2013) has acted on 3 SNPs within the *CG9509* enhancer for an increase in *CG9509* expression leading

to reduced wing loading. While two of these SNPs appear to be the targets of a selective sweep, the third, which has the largest effect on *CG9509* expression, shows signs of recent balancing selection. We propose that the reduced wing loading is an adaptation to improve flight ability at colder temperatures as *D. melanogaster* began its expansion out of Africa. Several studies have documented clinal variation in wing loading among *D. melanogaster* populations across multiple continents (Azevedo et al 1998, Gilchrist et al 2000, Gilchrist and Huey 2004, Bhan et al 2014), with wing loading being negatively correlated with latitude, and this cline is thought to be maintained by selection for improved flight at cooler temperatures.

Materials and Methods

Drosophila melanogaster lines

The *phiX-86Fb* stock (Bischof et al 2007) containing a mapped *attP* site on the third chromosome (cytological position: 3R 86F), was obtained from the Bloomington Stock Center (Indiana, USA) and used for phiC31 site-specific integration. All flies were maintained as inbred, isofemale lines under standard conditions (22°C, 14 hr light : 10 hr dark cycle, cornmeal-molasses medium).

Population samples. Expression of *CG9509* was surveyed in isofemale lines derived from the following locations: Leiden, The Netherlands (12 lines), Kuala Lumpur, Malaysia (11 lines), Cairo, Egypt (12 lines), Siavonga, Zambia (11 lines), and Lake Kariba, Zimbabwe (11 lines). Body size and wing loading assays were performed on the Dutch and Zambian populations.

All of the surveyed populations were used in a previous study of adult expression and sequence variation associated with the *CG9509* enhancer region (Glaser-Schmitt et al 2013). The Zimbabwean and Dutch populations were also used in another previous study of sequence and expression variation associated with the *CG9509* enhancer region (Saminadin-Peter et al 2012) as well as genome-wide expression studies (Hutter et al 2008, Müller et al 2011, Catalán et al 2012, Huylmans and Parsch 2014).

***CG9509* knockout and knockdown lines.** *CG9509*, *E74B*, *dFOXO*, and *dMyc* expression analysis as well as body size, larval growth rate, developmental timing, wing loading, and tolerance assays were performed on flies in which the open reading frame (ORF) of the *CG9509* gene was knocked out or in which *CG9509* expression was knocked down by RNAi. We discovered the *CG9509* knockout allele as a naturally-occurring variant in an isofemale line from a Munich population. In this line (*CG9509*$^{-}$), a deletion introduces a frameshift that leads to a premature stop codon 232 amino acids into the *CG9509* ORF (Figures S6B and C). Our qRT-PCR assay (see below) was able to detect the expression of *CG9509* mRNA in this line, but only at very low levels, suggesting that it is degraded by the nonsense-mediated decay pathway (Figure S6A). Given the disrupted ORF and the very low expression, we assume that *CG9509* is completely knocked out in this line. As a control, four lines from the same Munich population, showing representative *CG9509* expression for the population, were used. The knockdown of *CG9509* expression was achieved using an RNAi construct under the control of the yeast GAL4/upstream activating sequence (UAS) system. A *D. melanogaster* line producing a hairpin RNA complementary to *CG9509* mRNA under the control of a UAS (RNAi-*CG9509*, transformant ID: 107089) as well as a line containing an

empty vector at the same genomic location (UAS^-, transformant ID: 60100), which we used as a control, were obtained from the Vienna *Drosophila* Resource Center (Vienna, Austria) (Dietzl et al 2007). The RNAi-*CG9509* and UAS^- lines were crossed to an *Act5C*-GAL4/*Cyo* driver line and the progeny were used in subsequent body size, larval growth rate, developmental timing, wing loading, and tolerance assays, as well as in an expression analysis. Using qRT-PCR, *CG9509* expression knockdown efficiency was confirmed to be approximately 98.6% for adult females, 98.9% for adult males, 76% for early wandering third instar larvae, and 97.6% for late wandering third instar larvae.

Larval Staging

In order to control for size and age, all larval staging was performed at 25°C with a 14 hr light : 10 hr dark cycle. Adult females were placed in cages and allowed to lay eggs on molasses-agar plates supplemented with yeast for 48 hours for expression analysis and β-Galactosidase assays or for 4 hours for larval growth rate analysis. First instar larvae (L1) were then collected for staging. To stage second instar larvae (L2) or early third instar larvae (L3) , first instar larvae were transferred to small molasses-agar plates supplemented with cornmeal-molasses medium with a density of 50 larvae per plate. Second instar larvae were collected ~48 hours after egg laying (shortly after the L1-L2 transition) and identified by their mouth hooks. Early third instar larvae were collected ~72 hours after egg laying (shortly after the L2-L3 transition) and were identified by their spiracles. To stage wandering third instar larvae, first instar larvae were transferred to large vials containing cornmeal-molasses medium supplemented with 5% bromophenol blue (Sigma Aldrich, St.Louis, Missouri, USA) at a density of 150 larvae per vial. Early third instar wandering larvae (shortly after onset of

wandering, puff stages 1-2) were identified by the dark blue color of their guts, while, late third instar wandering larvae (shortly before pupariation, puff stages 7-9) were identified by their clear guts. Larvae were washed in PBS and gently dried then kept on ice until use. Larvae for RNA extraction were stored at -80°C before use.

Expression Analysis

Total RNA was extracted from 3–5 adult males (aged 4–6 days) or 1–3 early (L3 PS1-2) or late third instar wandering larvae (L3 PS7-9) and a DNAse I digestion was performed using the MasterPure RNA Purification Kit (Epicentre , Madison, Wisconsin, USA). Two biological replicates were performed for each line and/or stage. Using random hexamer primers and Superscript III reverse transcriptase (Invitrogen, Waltham, Massachusetts, USA), 3 mg total RNA for each replicate was reverse-transcribed following the manufacturer's protocol. TaqMan Gene Expression Assays (Invitrogen, Waltham, Massachusetts, USA) were then performed on the resulting cDNA using probes specific to *CG9509* (Dm01838873_g1), *dFOXO* (Dm02140207_g1), *dMyc* (Dm01843706_m1), and/or *E74B* (Dm01793592_m1) as well as a probe specific to the ribosomal protein gene *RpL32* (Dm02151827_g1), which was used as an endogenous control. The ΔΔCt method was used to calculate normalized gene expression (Pfaffl 2001). Briefly, for each biological replicate, the average threshold cycle (Ct) of two technical replicates was measured and ΔCt was calculated as the mean Ct difference between the probe of interest and the *RpL32* probe. The fold-change difference in expression relative to the Zimbabwe population for population comparisons, $CG9509^{-}$ for *CG0509* knockout comparisons, or the UAS^{-}/*Act5C*-GAL4 line for *CG9509* expression knockdown comparisons was then calculated as $2^{-(\Delta CtX-\Delta CtY)}$, where ΔCtX is the mean ΔCt

value for each biological replicate of the line of interest and ΔCtY is the mean ΔCt value of the either the Zimbabwe, *CG9509*$^{-}$, or UAS^{-}/*Act5C*-GAL4 lines. Significance was assessed with a t-test. When more than three comparisons were made using the same data, a Bonferroni multiple test correction was applied.

Transgenic reporter gene assays

The *CG9509* enhancer region, beginning 9 bp after the *CG14406* stop codon and ending 2 bp before the *CG9509* start codon, was PCR-amplified from one cosmopolitan and one sub-Saharan African strain as described in (Saminadin-Peter et al 2012) and cloned into the pCR2.1-TOPO vector (Invitrogen, Waltham, Massachusetts, USA). The effects of six sub-Saharan African sequence variants (Figure 5B, positions 67, 765, 821-817, 1063, 1155, and 1174 bp) in the cosmopolitan background were examined. The sub-Saharan African variant "C" at 67 bp is present in cosmopolitan populations and was cloned into the TOPO vector as described above. Due to the complexity of the region in which the 5 bp indel at 822-817 bp is located, a synthesized vector containing parts of the TOPO vector, the cosmopolitan upstream variant, and the desired 5 bp insertion was custom ordered from GeneArt (Invitrogen, Waltham, Massachusetts, USA). A 540 bp *SpeI*/*NdeI* fragment of this vector was then ligated into the TOPO vector containing the cosmopolitan enhancer sequence.

The four other sub-Saharan African variants were introduced into the cosmopolitan sequence via site directed mutagenesis, which was performed using thermal cycling with PfuUltra II Fusion HS DNA polymerase (Agilent Technologies, Santa Clara, California, USA) and complimentary, fully or partially overlapping primer pairs containing the desired base pair change (Zheng et al 2004, for primer list see Table S1). For sites that were shown to

affect reporter gene expression, the cosmopolitan variants were introduced into the sub-Saharan African enhancer and constructs with all contributing sites were generated in both a cosmopolitan and sub-Saharan African background using site directed mutagenesis as described above (for primer list see S1 Table) for a total of 13 reporter gene constructs. Mutants and parental enhancer sequences were confirmed via sequencing using the following primers (5'->3'): TTTGGTTTCCTTACCGTTTTG, GTGCAGTTTGGAACTCAG, CATTTATAGCACTTGGCTCG, GCTTCGCATTCTGGATGC, TGGCGCTAACCTGAATTCC, and GCGTTTTGCTTTTCCGTTAG. A 3.6-kb *NotI* fragment of the pCMV-SPORT-βgal plasmid (Invitrogen, Waltham, Massachusetts, USA) containing the *E. coli lacZ* coding region was then inserted into the *NotI* site of the each of the constructs and a *BamHI*/*XbaI* fragment containing the *CG9509* enhancer and *lacZ* sequences was ligated into the *pattB* integration vector (Bischof et al. 2007). The *pattB* vectors containing the *CG9509* enhancer and the *lacZ* reporter gene were microinjected into early-stage embryos of the *phiX-86Fb* (*attP* site at cytological band 86F) strain (Bischof et al. 2007), which contains a stable source of *phiC31* integrase on the X chromosome. After microinjection, surviving flies were crossed to a *white*$^{-}$ strain to remove the integrase source and stable lines homozygous for each of the constructs were established. A subset of the microinjections was performed by Rainbow Transgenic Flies (Camarillo, CA, USA). The majority of the constructs were also incorporated into the *phiX-68E* (*attP* site at cytological band 68E) strain, which yielded similar results (data not shown), therefore, only the strain *phiX-86Fb* was used.

For each reporter gene construct, β-galactosidase activity was measured in groups of 15 adult 4-6 day old males or females or 8 late wandering 3rd instar larvae. Soluble proteins were extracted and a β-galactosidase activity assay was performed as described in Hense et al (2007) with the following exceptions: Flies or larvae were frozen with liquid nitrogen and

homogenized before the addition of 200 µl of the 0.1 M Tris-HCl, 1 mM EDTA, and 7 mM 2-mercaptoethanol buffer (pH 7.5). β-galactosidase activity was measured spectrophotometrically by following the change in absorbance at 420 nm at 37 °C. 4–8 biological replicates were performed per stage or sex. Significance was assessed using a *t*-test and a Bonferroni multiple test correction was applied for each stage and sex.

Adult Staging

Since temperature and density during rearing as well as age can affect body size and therefore wing loading (James and Partridge 1995, Santos et al 1994, Angilletta et al 2004), *D. melanogaster* used in body size and wing loading assays were strictly controlled similarly to the larval staging. All staging took place at 25°C. Adults were placed in cages and allowed to lay eggs for 48 hours. First instar larvae were picked and placed in vials at a density of 50 larvae per vial. After eclosion, flies were collected and sorted into groups of males and females every 3-4 hours, so that only virgin flies were used. Flies were aged to one-day old before use in assays.

Body size assays

Weight. In order to keep variance down, wet weight of flies was measured in groups of 25 males or females. Groups of flies were lightly gassed with CO_2 and placed in pre-weighed 1.5 mL Eppendorf tubes on ice for 5 minutes before being weighed on a Mettler H51 scale (d = 0.01 mg, error = 0.05 mg). The weight of 25 flies was then calculated as the weight of 25 flies and tube – the weight of the tube. For each line and sex, 4 replicates were performed for

population comparisons, 5–7 replicates were performed for the *CG9509*$^{-}$ and RNAi-*CG9509*/*Act5C*-GAL4 lines, and 4–5 replicates were performed for all control lines. Significance was assessed using a t-test.

Wing size. Right wings were dissected in isopropanol, mounted in europal (Sigma Aldrich, St.Louis, Missouri, USA), and allowed to dry at least 1 week before being photographed. Wings were photographed using a Nikon D5100 camera and compound microscope. Images were analyzed in ImageJ (Schneider et al 2012). A piece of millimeter paper was included in all images for scale. Wing length was measured in a straight line from the humeral-costal break to the 3rd longitudinal vein and wing area was estimated as previously described (Gilchrist and Partridge 1999). For each line and sex, wing size was measured for 5 flies for population comparisons and *CG9509* knockdown control lines and 10–15 biological replicates per sex for RNAi-*CG9509*/*Act5C*-GAL4, UAS^{-}/*Act5C*-GAL4, and *CG9509*$^{-}$ lines. Significance was assessed using a *t*-test.

Wing loading assays

Wing load index was calculated as wet weight of a fly / right wing area. Flies were lightly gassed with CO_2 and placed in pre-weighed 1.5 mL Eppendorf tubes on ice for 5 minutes before being weighed on a Mettler H51 scale (d = 0.01 mg, error = 0.05 mg). The weight of a fly was then calculated as the weight of the fly and tube – the weight of the tube. For each fly, the right wing was then dissected and the wing area estimated as described above. For each line and sex, wing loading was measured for 5 flies for population comparisons and 10-15 flies for RNAi-*CG9509*/Act5C-GAL4 and *CG9509*$^{-}$ lines as well as their respective control lines. Significance was assessed using a *t*-test.

Larval growth assays

To assess larval growth rate, larval volume was measured in the following stages: 2nd instar (48 hr AEL), early 3rd instar (72 hr AEL), early wandering 3rd instar (110 hr AEL), and late wandering 3rd instar (116 hr AEL). Larvae were staged as described above. Before imaging, larvae were placed on ice for at least 5 minutes. Larvae were photographed using a Nikon D5100 camera and compound microscope and images were analyzed in ImageJ (Schneider et al 2012). A piece of millimeter paper was included in all images for scale. Larval volume was calculated as $4/3\pi(L/2)^2(d/2)$, where L=length and d=diameter (Colombani et al 2005). For each stage and line, larval volume was measured in 15–20 larvae for RNAi-*CG9509* /*Act5C*-GAL4 and *CG9509*$^-$ lines as well as their respective controls. Significance at each larval stage was assessed using a *t*-test and a Bonferroni multiple test correction was applied.

Developmental timing assays

As a measure of developmental timing, time from the 1st instar larval stage to pupariation and duration of the wandering stage were measured. As described in the larval staging section above, flies were allowed to lay eggs for 12 hours and 1st instar larvae were picked. Larvae were transferred in groups of 50 to cornmeal-molasses medium and allowed to mature. In order to measure time from the 1st instar larval stage to pupariation, pupariation was recorded every 2 hours for 25-110 larvae per line. In order to measure duration of the wandering stage, larvae were screened for onset of wandering behavior every hour and transferred individually to a petri dish containing moistened filter paper.

Pupariation was recorded every hour for 10-50 larvae per line. Both assays were performed at 25°C to prevent fluctuations in developmental timing due to temperature.

Tolerance assays

DDT, malathion, ethanol, and cold tolerance assays were performed using RNAi-*CG9509*/*Act5C*-GAL4 and UAS^{-}/Act5C-GAL4 flies. For DDT, malathion, and ethanol tolerance assays, for each line, sex, and concentration, 6–7 tolerance chambers with 20 flies were exposed to four concentrations of a compound and mortality was measured as the number of flies dead or unable to move after 30 minutes (malathion), 2 hours (DDT), or 48 hours (ethanol). For ethanol tolerance assays, tolerance chambers consisted of a plastic vial (diameter = 25 mm, height = 95 mm) with compressed cotton at the bottom containing 2.5 ml ethanol solution supplemented with 5% sucrose and sealed with a cork. For DDT and malathion assays, tolerance chambers consisted of glass vials (h = 5 cm, r = 1.65 cm) in which 200 ml of DDT (Dr. Ehrenstorfer, Augsburg, Germany) or malathion (Dr. Ehrenstorfer, Augsburg, Germany) diluted in acetone (Sigma Aldrich, St.Louis, Missouri, USA) was swirled until the acetone dried, allowed to dry an additional hour before addition of flies, and sealed with compressed cotton soaked in 5% sucrose solution. For all assays, 2-3 control chambers containing only 5% sucrose solution were also tested. The data for each assay was fit to a generalized linear model using concentration, line, and sex as factors (unless sex was not significant, in which case, it was removed from the model) and a quasi-binomial distribution using the glm function in R (R core team 2008). For cold tolerance assays, for each line and sex, 25 groups of 5 flies were exposed to an ice water bath for 5 hours and the time in minutes until each fly had recovered from chill coma (able to stand upright again) was

recorded. The mean recovery time for each vial was calculated and a *t*-test was applied to assess significance.

Acknowledgements

We thank Hilde Lainer, Hedwig Gebhart, Simone Lange, Annabella Königer, and Eliza Argyridou for technical assistance in the lab. We also thank the following bachelor and master's students for their contribution to this project: Lisa Marie Keitel, Sandra Petrone Mendoza, and Anastacia Martirosyan. 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.

Supporting Information

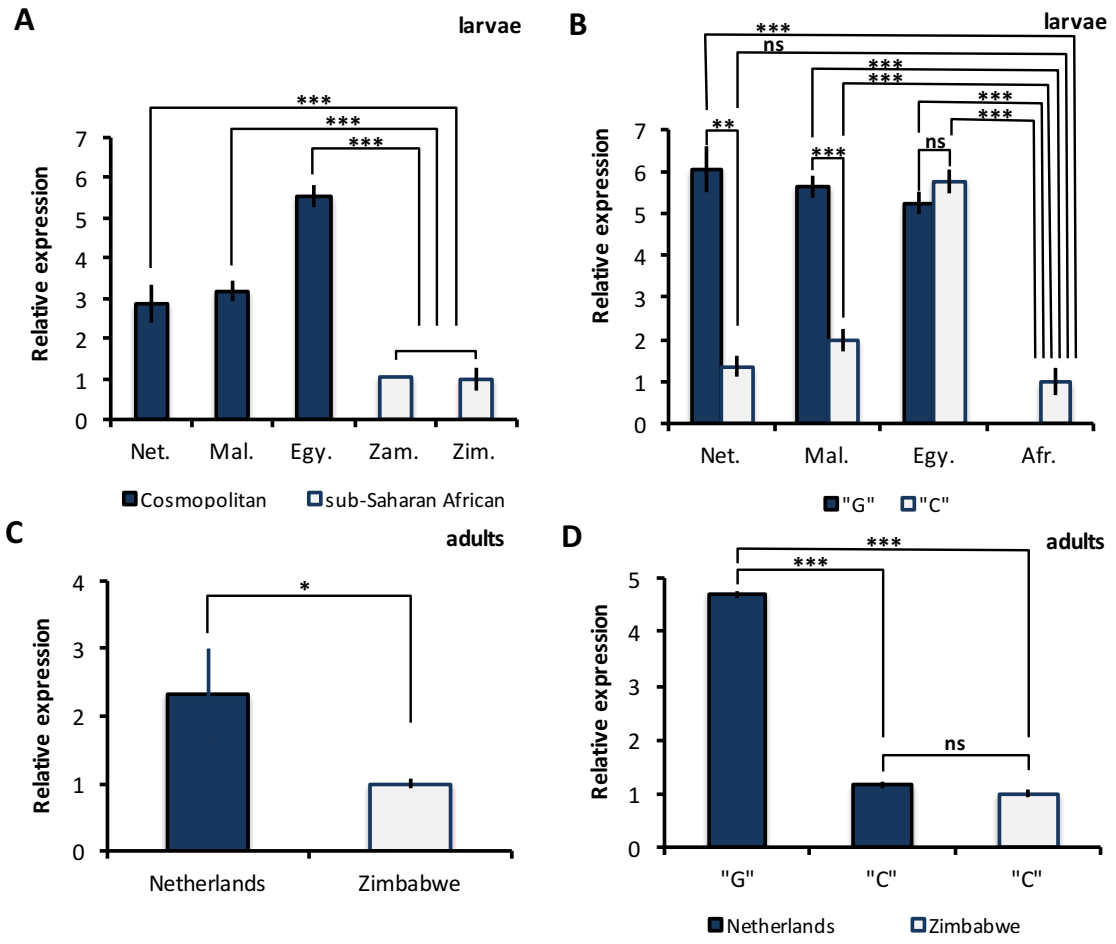


Figure S1: *CG9509* expression. Relative expression in (A, B) late wandering 3rd instar larvae and (C, D) adult males. (A) Larval expression in the Netherlands (Net.), Malaysia (Mal.), Egypt (Egy.), Zimbabwe (Zim.), and Zambia (Zam.). (B) Larval expression with populations separated according to the variant at position 67. (B) Adult expression in the Netherlands and Zimbabwe. (B) Adult expression with populations separated according to the variant at position 67. Blue bars represent cosmopolitan populations and white bars represent sub-Saharan African populations. Expression is calculated relative to the sub-Saharan African populations. Error bars indicate standard error of the mean. Differences between populations were tested by a *t*-test. ns, not significant **P*<0.05, ***P*<0.01, ****P*<0.005.

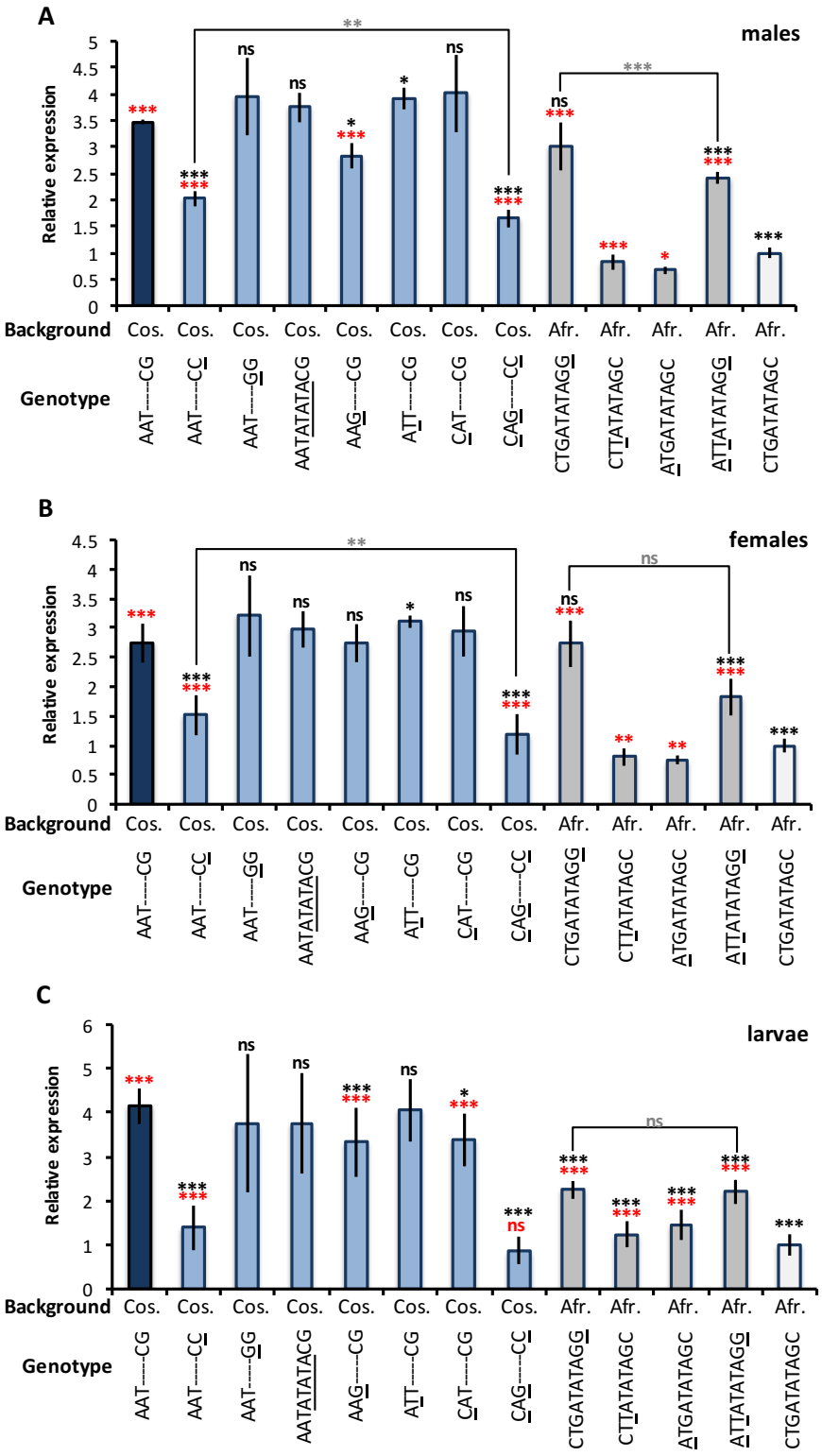


Figure S2: Transgenic reporter gene expression of all constructs. *CG9509* reporter gene expression for all constructs in adult (A) males, (B) females, and (C) larvae. Expression of the *LacZ* reporter gene was measured as the change in absorbance at 420 nm. Relative expression, which was calculated in comparison to expression of the sub-Saharan African enhancer, is shown. Blue bars indicate

expression driven by a wild-type cosmopolitan (Cos.) enhancer, while white bars show expression driven by a wild-type sub-Saharan African (Afr.) enhancer. Light blue bars indicate expression driven by the enhancer after mutations were introduced into a cosmopolitan background. Dark gray bars show expression driven by the enhancer after mutations were introduced in a sub-Saharan African background. "Genotype" indicates the nucleotides at positions 1174, 1155, 1063, 821-817, 765, and 67 before the *CG9509* start codon, respectively. Mutated sites are underlined. Significance was assessed using a *t*-test and a Bonferroni multiple test correction was applied. Significance is represented in black for comparisons to the cosmopolitan enhancer, red for the sub-Saharan African enhancer, and gray for comparison between mutated enhancers. ns, not significant **P*<0.05, ***P*<0.01, ****P*<0.005.

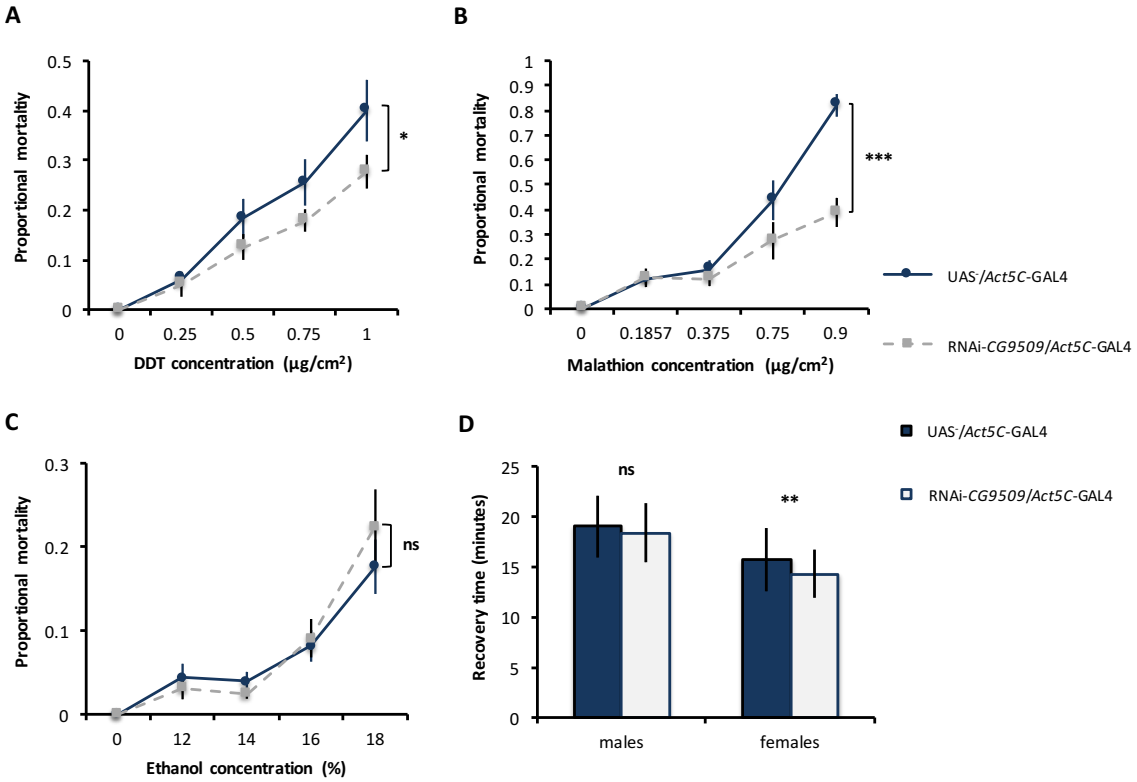


Figure S3: Effect of *CG9509* expression on insecticide, ethanol, and cold tolerance. Adult (A) DDT, (B) malathion, (C) ethanol, and (D) cold tolerance assay results in RNAi-*CG9509*/*Act5C*-GAL4 (gray hatched lines or white bars) and UAS^{-}/*Act5C*-GAL4 (blue lines or bars) flies. Error bars represent standard error of the mean. (A-C) Significance was assessed using a generalized linear model with a quasi-binomial distribution. (D) Significance was assessed using a *t*-test. ns, not significant *P<0.05, **P<0.01, ***P<0.005.

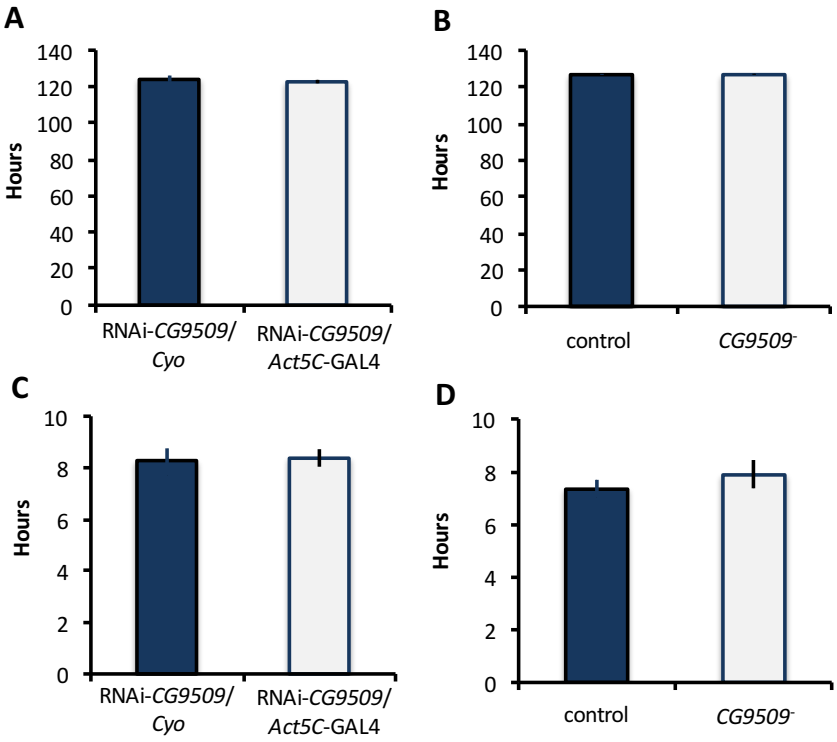


Figure S4: Effect of *CG9509* expression on developmental timing. (A, B) Time from first instar larval stage to pupariation in (A) RNAi-*CG9509*/*Act5C*-GAL4 (white) and RNAi-*CG9509*/*Cyo* (blue) flies and (B) *CG9509*$^{-}$ (white) and control (blue) flies. (C, D) Duration of wandering stage in (C) RNAi-*CG9509*/*Act5C*-GAL4 (white) and RNAi-*CG9509*/*Cyo* (blue) flies and (D) *CG9509*$^{-}$ (white) and control (blue) flies. Error bars represent standard error of the mean. Significance was assessed via *t*-test.

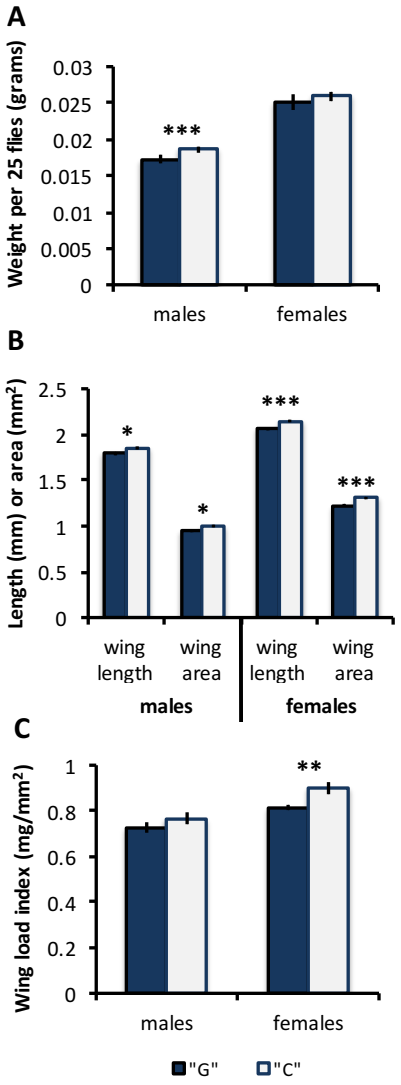


Figure S5: Effect of *CG9509* expression on body size and wing loading. (A) Body weight per 25 flies, (B) wing length and wing area, and (C) wing loading in a Dutch population according to variant at position 67. Low *CG9509* expression, sub-Saharan African "C" variants are shown in white and high *CG9509* expression, cosmopolitan "G" variants are shown in blue. Error bars represent standard error of the mean. Significance was assessed via *t*-test. **P*<0.05, ***P*<0.01, ****P*<0.005.

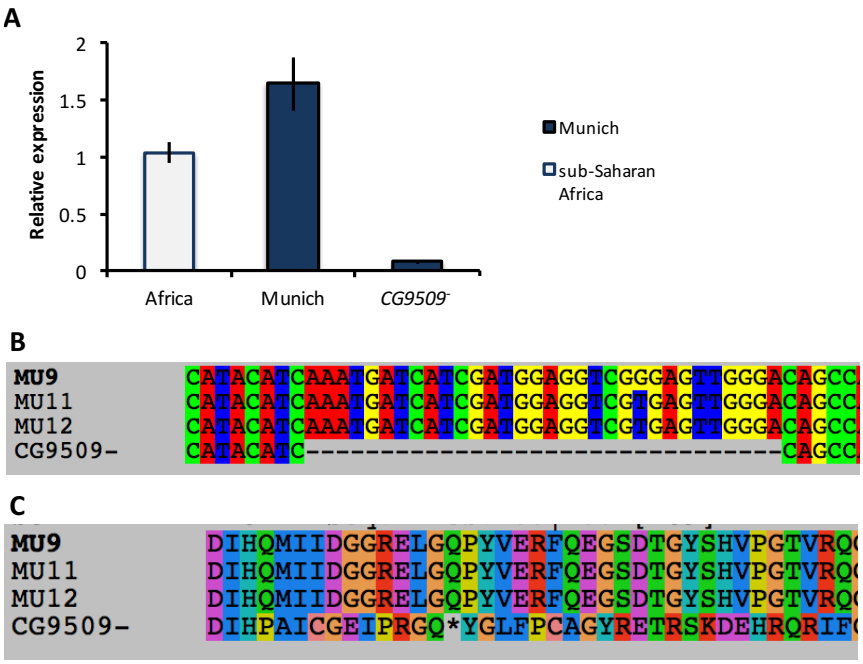


Figure S6: *CG9509*⁻ line details. (A) Relative *CG9509* expression in *CG9509*⁻ line as determined by qRT-PCR. Relative expression was determined in comparison to sub-Saharan African flies (Africa). For comparison, the average expression of the population in which the *CG9509*⁻ line was discovered (Munich) is shown. Error bars represent standard error of the mean. (B) Sequence alignment of deletion within *CG9509*⁻ coding region. Sequences of flies from the *CG9509*⁻ source population are shown for comparison (MU). (C) Amino acid alignment of frameshift within *CG9509*⁻ coding region. Sequences of flies from the *CG9509*⁻ source population are shown for comparison (MU). * Stop codon.

S1 Table. Site directed mutagenesis primers

Pos.[a]	Primer pair (5'->3')	Background[b]	Overlap[c]
1174	CGGGCACGCGTTTTAATTACTTTGTAAAGC GTAATTAAAACGCGTGCCCGAAAAGGCGC	Cos.	partial
1174	GGGCAAGCGTTTTAATTACTTTGTTAAGC GTAATTAAAACGCTTGCCCAAAAAGGCGC	Afr.	partial
1155	CGGGCAAGCGTTTTAATTACTTTGTTAAGCTGCATTTTTG CAAAAATGCAGCTTAACAAAGTAATTAAAACGCTTGCCCG	Cos.	full
1063	GCCGTCTTAATGTGTGTTTGTGTCGAGCCAAGTGC CGACACAAACACACATTAAGACGGCAAAAAAAATC	Cos.	partial
1063	CTTAATGTTTGTTTGTGTCGAGCCAAGTGC CGACACAAACAAACATTAAGACAGC	Afr.	partial
765	CAATTTTGTTATTTTTTAAATCTATGCTTTGATTTTAG GCATAGATTTAAAAAATAACAAAATTGTTTTTAAAATTTTATAAC	Cos.	partial
67	CGCGACTGGGCCTCAGAGTCAAATAG CTGAGGC**C**CAGTCGCGGCTGAAGATTCGC	Afr.	partial

[a]Positon (pos.) in bp before *CG9509* start codon. [b]Mutations were introduced into either a Cosmopolitan (Cos.) or sub-Saharan African background (Afr.). [c]Overlap indicates degree of primer pair overlap.

Chapter 3

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

Ana Catalán, Amanda Glaser-Schmitt, Eliza Argyridou, Pablo Duchen, and John Parsch

PLoS Genetics 2016; 12 (4): e1005987

PLOS | GENETICS

RESEARCH ARTICLE

An Indel Polymorphism in the *MtnA* 3' Untranslated Region Is Associated with Gene Expression Variation and Local Adaptation in *Drosophila melanogaster*

Ana Catalán[1,2]*, Amanda Glaser-Schmitt[1], Eliza Argyridou[1], Pablo Duchen[3], John Parsch[1]*

1 Faculty of Biology, Ludwig-Maximilians-Universität München, Planegg, Germany, **2** Department of Ecology and Evolutionary Biology, University of California, Irvine, Irvine, California, United States of America, **3** Department of Biology and Biochemistry, University of Fribourg, Fribourg, Switzerland

* ana.catalan@gmail.com (AC); parsch@bio.lmu.de (JP)

OPEN ACCESS

Citation: Catalán A, Glaser-Schmitt A, Argyridou E, Duchen P, Parsch J (2016) An Indel Polymorphism in the *MtnA* 3' Untranslated Region Is Associated with Gene Expression Variation and Local Adaptation in *Drosophila melanogaster*. PLoS Genet 12(4): e1005987. doi:10.1371/journal.pgen.1005987

Editor: David J. Begun, University of California Davis, UNITED STATES

Received: September 7, 2015

Accepted: March 22, 2016

Published: April 27, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files except for the sequences which are available at GenBank/EMBL database under the accession numbers KT008059–KT008093.

Funding: This work was supported by Deutsche Forschungsgemeinschaft grants PA 903/5 and PA 903/6 (http://gepris.dfg.de/gepris/OCTOPUS/;jsessionid=GFThHN1bZ8512PbSCTK4WktLp716MWR2NVt6pymwlTzhnTgmtGDz!1709242108!1200488219358?task=doSearchSimple&context= projekt&keywords_criterion=PA+903%2F5+&task=doSearchSimple). The

Abstract

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 frequency of the deletion increases with latitude across multiple continents and approaches 100% in northern Europe. Flies with the deletion have more than 4-fold higher *MtnA* expression than flies with the ancestral sequence. Using reporter gene constructs in transgenic flies, we show that the 3' UTR deletion significantly contributes to the observed expression difference. Population genetic analyses uncovered signatures of a selective sweep in the *MtnA* region within populations from northern Europe. We also find that the 3' UTR deletion is associated with increased oxidative stress tolerance. These results suggest that the 3' UTR deletion has been a target of selection for its ability to confer increased levels of *MtnA* expression in northern European populations, likely due to a local adaptive advantage of increased oxidative stress tolerance.

Author Summary

Although molecular variation is abundant in natural populations, understanding how this variation affects organismal phenotypes that are subject to natural selection remains a major challenge in the field of evolutionary genetics. Here we show that a deletion mutation in a noncoding region of the *Drosophila melanogaster Metallothionein A* gene leads to a significant increase in gene expression and increases survival under oxidative stress. The deletion is in high frequency in three distinct geographic regions: in northern

funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

European populations, in northern populations along the east coast of North America, and in southern populations along the east coast of Australia. In northern European populations the deletion shows population genetic signatures of recent positive selection. Thus, we provide evidence for a regulatory polymorphism that underlies local adaptation in natural 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 [1–3]. Several genomic studies concluded that *D. melanogaster* underwent a population expansion around 60,000 years ago within Africa that set the ground for an out-of-Africa expansion 13,000–19,000 years ago and the subsequent colonization of Europe and Asia 2,000–5,000 years ago [4–6]. 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* [7–9].

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 [10,11], as well as the dissected brains and Malpighian tubules of each sex [12,13]. 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 (Fig 1) and belongs to a gene family of five members that also includes *MtnB*, *MtnC*, *MtnD* and *MtnE* [14,15]. Metallothioneins are present in all eukaryotes and have also been identified in some prokaryotes [16]. In general, MTs are cysteine-rich proteins, a feature that makes them thermostable, and have a strong affinity to metal ions, especially zinc and copper ions [17]. 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 [18]. In natural isolates of *D. melanogaster*, increased *MtnA* expression has been linked to copy number and insertion and deletion (indel) variation and is associated with increased tolerance to heavy metals [19,20].

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). Outside of sub-Saharan Africa, the deletion shows a latitudinal cline in frequency across multiple continents, reaching very high frequencies in northern Europe. Using transgenic reporter genes, we show that the indel polymorphism in the 3' UTR contributes to the expression difference observed between populations. Furthermore, we use hydrogen peroxide tolerance assays to show that

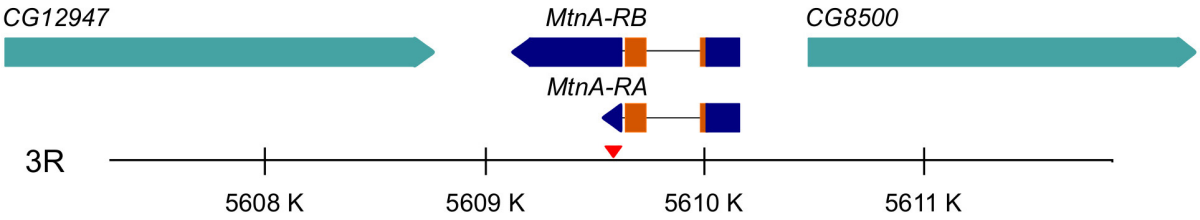


Fig 1. Structure of the *MtnA* locus. Two transcripts that differ only in their 3' UTRs have been annotated for *MtnA* (*MtnA-RA* and *MtnA-RB*). Dark blue boxes represent the UTRs with the arrowheads indicating the direction of transcription. Orange boxes represent the coding exons. The thin lines joining 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 coding genes flanking *MtnA* (CG12947 and CG8500), only the whole gene model is shown.

doi:10.1371/journal.pgen.1005987.g001

the deletion is associated with increased oxidative stress tolerance. 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*-regulatory polymorphism in the *MtnA* 3' UTR has undergone recent positive selection to increase *MtnA* expression and oxidative stress tolerance in derived northern populations of *D. melanogaster*.

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) [12]. Of the members of the *Mtn* gene family, only *MtnA* showed high levels of expression and a significant difference in expression between populations (Fig 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 previously [12]. With this approach, we found *MtnA* to have 5-fold higher expression in the European population than in the African population (Fig 2B).

The 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 (Fig 1). The *MtnA-RA* transcript completely overlaps with that of *MtnA-RB* and contains no unique sequence. The *MtnA-RB* transcript, however, contains an extra 371 bp at the 3' end that can be used to assess isoform-specific expression. Using RNA-seq data [12], 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 (S1 Table).

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 [19,20]. To determine if CNV could explain the observed expression difference between the European and the African populations, we assayed *MtnA* copy number in flies of both populations by quantitative PCR. We found no evidence for CNV within or between the populations (Fig 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* [21],

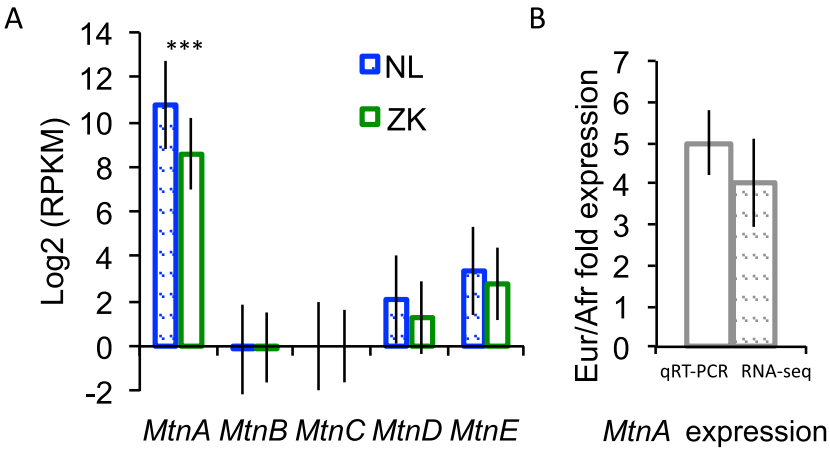


Fig 2. Expression of metallothionein genes in the brain in two populations of *D. melanogaster*. (A) Expression level of *Mtn* paralogs in the brain from RNA-seq data. Expression is reported in reads per kilobase per million mapped reads (RPKM). Only *MtnA* showed a significant difference in expression between a European (the Netherlands), shown in blue, and an African (Zimbabwe), shown in green, population (adjusted $P < 10^{-7}$ in the RNA-seq analysis [12]). Expression of *MtnC* was not detected. (B) *MtnA* expression in the brains of European and African flies, as determined by qRT-PCR. The expression difference between populations is highly significant (*t*-test, $P = 5x10^{-5}$). In both panels, the error bars indicate the standard error of the mean.

doi:10.1371/journal.pgen.1005987.g002

which can be co-amplified by the same PCR primers and serve as a positive control. These results indicate that CNV cannot account for the observed variation in *MtnA* gene expression.

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 (Fig 1) in 12 lines from the Netherlands (NL) and 11 lines from Zimbabwe

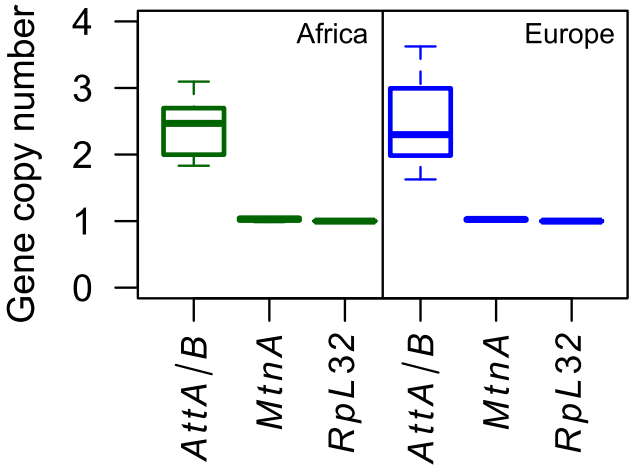


Fig 3. Results of CNV assays. Flies from Africa (Zimbabwe), shown in green, and Europe (the Netherlands), shown in blue, were tested for *MtnA* CNV. The close paralogs *AttA* and *AttB* were used as a positive control for multiple gene copies, while *RpL32* was used as a single-copy reference.

doi:10.1371/journal.pgen.1005987.g003

(ZK). 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, being this deletion present in 10 of the 12 European lines, but absent in Africa (Fig 4A). This indel was previously observed to segregate in natural populations from North America [20]. A comparison with three outgroup species (*D. sechellia*, *D. simulans*, and *D. yakuba*) indicated that the deletion was the derived variant. 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 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 (Fig 4B).

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 either a green fluorescent protein (GFP) or *lacZ* reporter gene. Two versions of each 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 (Fig 5A). The reporter genes were then introduced into the *D. melanogaster* genome by PhiC31 site-specific integration [22,23].

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 (Fig 4B). This is further supported by the expression of the reporter gene constructs. For the GFP reporter gene, the presence of the 3' UTR deletion led to increased expression in both the brain and body (Fig 5B), with the difference in expression being 2.3-fold and 1.75-fold, respectively. A similar result was found for the *lacZ* reporter gene, where the 3' UTR deletion led to 1.7-fold and 1.4-fold higher expression in the head and gut, respectively (Fig 5C).

MtnA expression in the brain

MtnA shows high expression in most *D. melanogaster* organs, including the fat body, digestive system, Malpighian tubule, and brain [24]. Although it has been documented that *MtnA* and its paralogs are involved in heavy metal homeostasis 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 the expression of the GFP reporter gene by confocal imaging of dissected brain tissue (Fig 6).

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 (Fig 6). *MtnA* does not appear to be expressed at a discernible level 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 are glia, which provide neurons with developmental, structural and trophic support as well as with protection against toxic elements [25–27]. In a genome-wide expression profiling study it was found that *MtnA* is expressed in the astrocyte glial cells of larvae and adults of *D. melanogaster* [28]. Although we cannot be certain that *MtnA* expression is limited to the glia in the brain, 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 [29].

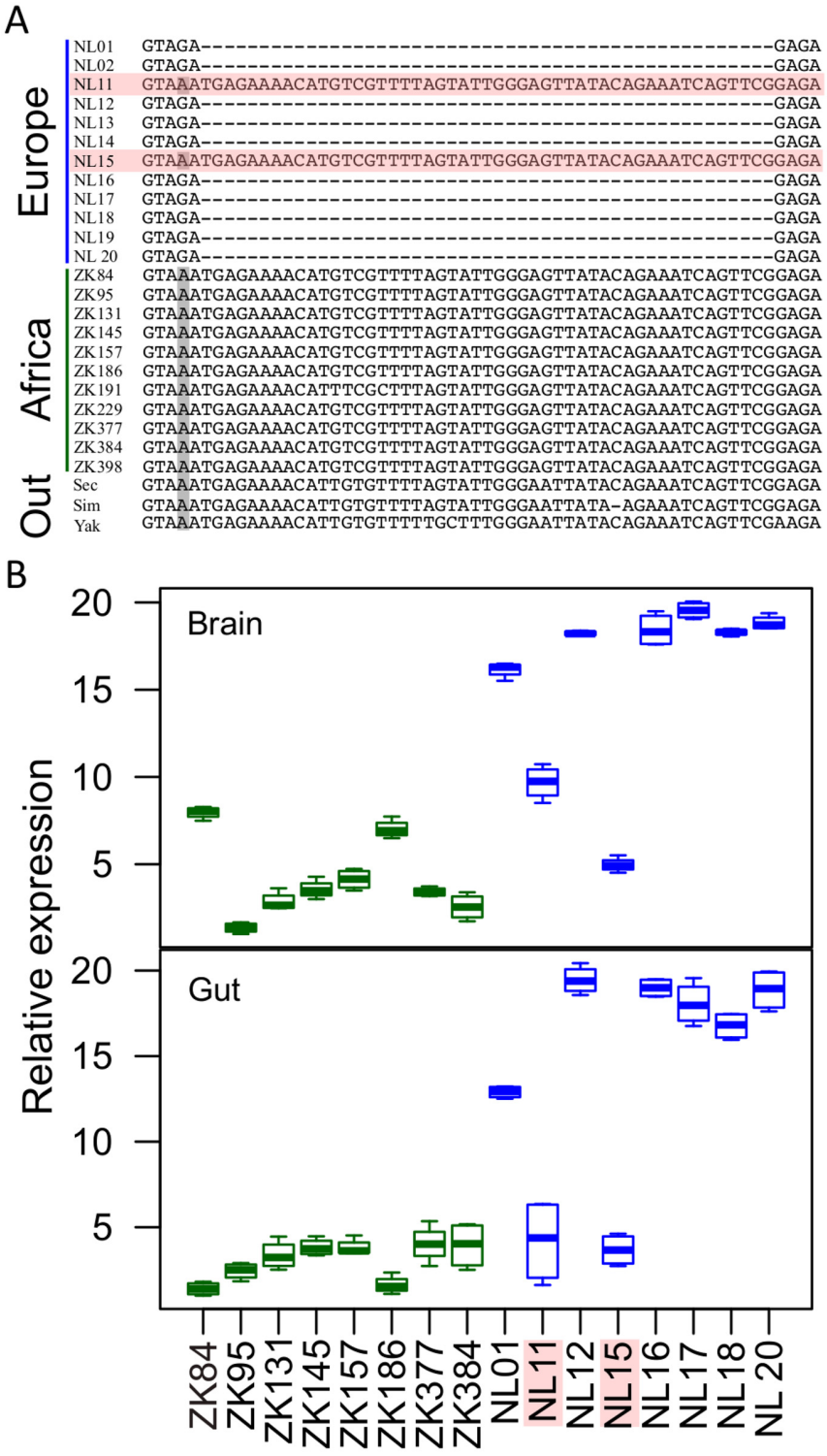


Fig 4. Association between an indel polymorphism in the *MtnA* 3' UTR and gene expression variation. (A) An indel (and a linked SNP marked in gray) in the *MtnA* 3' UTR are the only polymorphisms within the 6-kb *MtnA* region that show a large difference in frequency between an African and a European population of *D. melanogaster*. A comparison 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 (NL) lines, shown in blue, and eight African (ZK) lines, shown in green. The two European lines lacking the deletion, *NL11* and *NL15*, show lower *MtnA* expression than those with the deletion.

doi:10.1371/journal.pgen.1005987.g004

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 ten additional *D. melanogaster* populations across a latitudinal range spanning from tropical sub-Saharan Africa to northern Europe (Table 1). We found that the deletion was at very low frequency in sub-Saharan Africa, but nearly fixed in populations from northern Europe. This suggests that, at least outside of the ancestral species range, there is a latitudinal cline in the deletion frequency. Indeed, when the sub-Saharan populations are excluded, there is a highly significant correlation between latitude and deletion frequency (linear regression; $R = 0.95$, $P = 0.0004$). This correlation still holds when the sub-Saharan populations are included (using the absolute value of latitude), but is weaker ($R = 0.80$, $P = 0.001$).

To investigate if the clinal distribution of the *MtnA* 3' UTR deletion is present on other continents, we analyzed pooled sequencing (pool-seq) data from North America and Australia [30,31]. In North America, there is a significant correlation between latitude and deletion frequency ($R = 0.94$, $P = 0.005$) (Table 2). A similar pattern was seen in Australia, although data from only two populations were available. The deletion is at a frequency of 42% in Queensland (latitude 16 S) and 61% is Tasmania (latitude 42 S). The difference in deletion frequency between the two populations is significant (Fisher's exact test, $P = 0.02$).

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) population samples. 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 1). 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

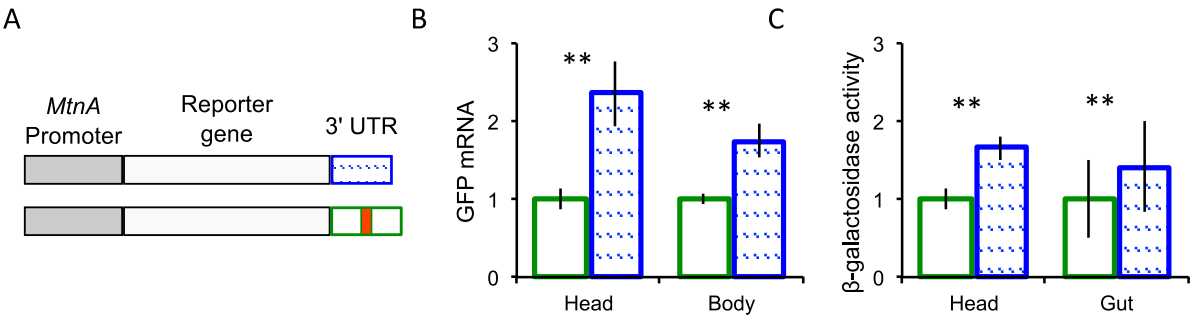


Fig 5. Reporter gene constructs and their expression. (A) The gray boxes represent the *MtnA* promoter, which is identical between the African and European alleles. The white boxes represent the GFP/*lacZ* reporter genes. The blue hatched box represents the *MtnA* 3' UTR with the deletion. The green box represents the *MtnA 3'* UTR with the additional 49 bp marked in red. The same color scheme applies to the bar plots. (B) The two versions of the GFP reporter gene differ significantly in expression in heads (*t*-test, $P = 0.0019$) and bodies (*t*-test, $P = 0.0046$), as assayed by qRT-PCR. (C) The two versions of the *lacZ* reporter gene differed significantly in expression in heads (*t*-test, $P = 0.0006$) and guts (*t*-test, $P = 0.0001$) as measured by β-galactosidase enzymatic activity. The error bars represent the standard error of the mean.

doi:10.1371/journal.pgen.1005987.g005

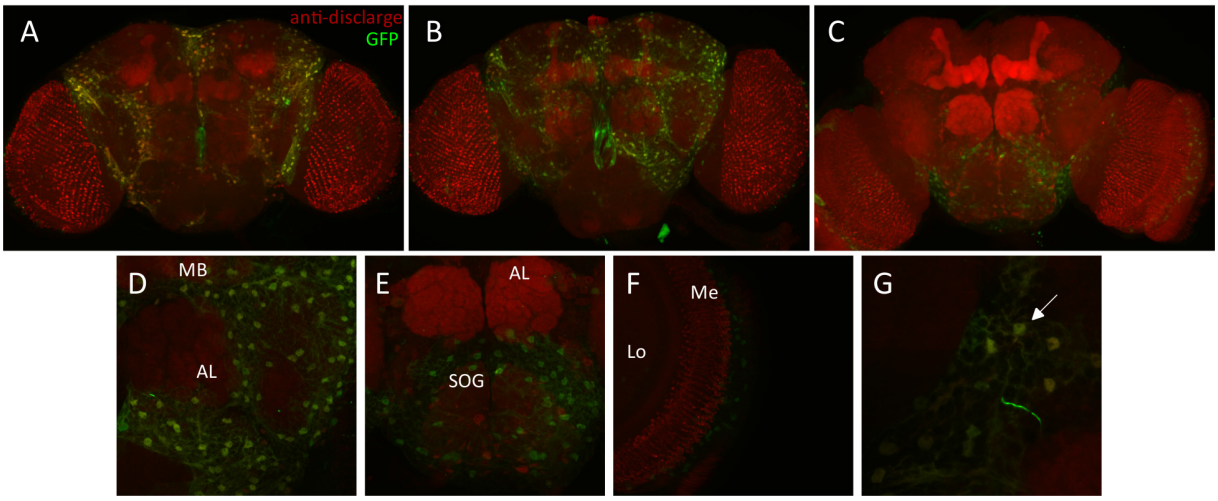


Fig 6. Expression of an *MtnA*-GFP reporter gene in the brain. (A-C) GFP expression driven by the reporter gene construct with the ancestral *MtnA* 3' UTR variant. (D-G) Higher magnification of the brain regions where GFP is expressed. AL: antennal lobe, MB: mushroom bodies, SOG: subesophageal ganglion,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

doi:10.1371/journal.pgen.1005987.g006

negative selection at this locus, but could also be caused by demographic factors, such as population expansion.

A sliding window analysis was performed to determine the distribution of nucleotide diversity (θ) (Fig 7A) and population differentiation (F_{st}) (Fig 7B) 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 for the deletion (two of the 12 lines have the ancestral state). This leads to higher nucleotide diversity than in the Swedish population, because the ancestral, non-deletion alleles contain more SNPs than the derived, deletion alleles. On average, Sweden and Zimbabwe

Table 1. Frequency of the *MtnA* 3' UTR deletion in different populations.

Population	*N*	Latitude	Frequency of deletion [95% CI]
Sweden	12	63.8 N	1.00 [0.86–1.00]
Denmark	12	55.7 N	0.96 [0.80–1.00]
The Netherlands	12	52.2 N	0.83 [0.64–0.94]
Germany	11	48.1 N	0.91 [0.73–0.98]
France	12	45.8 N	0.92 [0.75–0.98]
Cyprus	10	35.1 N	0.65 [0.43–0.83]
Egypt	14	30.1 N	0.60 [0.42–0.77]
Cameroon	6	6.3 N	0.00 [0.00–0.26]
Malaysia	12	3.1 N	0.45 [0.27–0.65]
Rwanda	12	2.5 S	0.08 [0.02–0.25]
Zambia	10	16.5 S	0.05 [0.01–0.24]
Zimbabwe	11	17.3 S	0.00 [0.00–0.15]

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

doi:10.1371/journal.pgen.1005987.t001

Table 2. Frequency of the *MtnA* 3' UTR deletion in North American populations.

Population	*N*[a]	Latitude	Total reads[b]	*MtnA* 3' UTR reads[c]	Deletion reads[d]	Frequency of deletion [95% CI]
Maine	322	45.5 N	125.8	301	171	0.57 [0.51–0.62]
Pennsylvania	900	40.0 N	593.9	1400	743	0.53 [0.50–0.56]
North Carolina	92	35.5 N	47.1	67	32	0.48 [0.36–0.60]
South Carolina	96	33.0 N	81.8	255	107	0.42 [0.36–0.48]
Georgia	102	30.9 N	96.9	246	101	0.41 [0.35–0.47]
Florida	174	25.5 N	103.7	225	76	0.34 [0.28–0.40]

[a] Number of autosomes in the pooled sample (including all replicates)

[b] Number of paired reads for the whole genome (in millions)

[c] Number of reads that mapped to the *MtnA* 3' UTR

[d] Number of reads that matched the *MtnA* 3' UTR deletion allele

doi:10.1371/journal.pgen.1005987.t002

showed the greatest population differentiation, with F_{st} reaching a peak in the 3' UTR of *MtnA*, whereas values of F_{st} were lowest for the comparison of the Dutch and Swedish populations, indicating that there is very little differentiation between them (Fig 7b).

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 the Hudson's haplotype test (HHT) [36] using three different demographic models of the *D. melanogaster* out-of-Africa bottleneck for neutral simulations. Under the model of Werzner et al. [6], HHT was significant ($P = 0.031$). Under the models of Thornton and Andolfatto [35] and Duchen et al. [5], HHT was marginally significant ($P = 0.076$ and $P = 0.094$, respectively). 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 [37,38]. Within the Dutch population, the CLR statistic shows a peak in the region just adjacent to the *MtnA* 3'

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 estimate of nucleotide diversity (per 100 sites) [32]; *π*, mean pairwise nucleotide diversity (per 100 sites) [33]; *TajD*, Tajima's *D* [34]; *nHap*, number of haplotypes.

*$P < 0.05$.

doi:10.1371/journal.pgen.1005987.t003

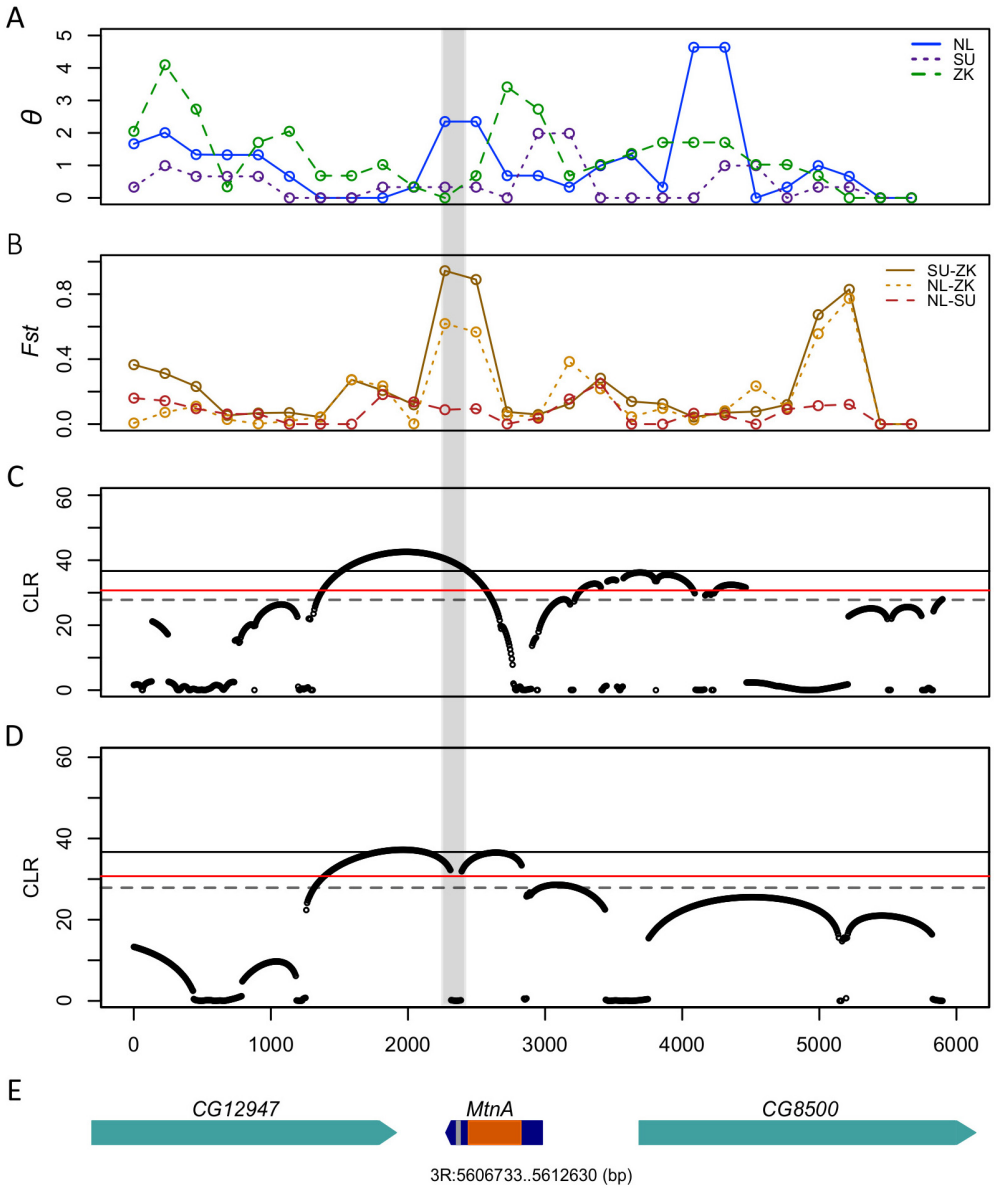


Fig 7. Evidence for positive selection at the *MtnA* locus. (A) Watterson's θ of *D. melanogaster* populations from Zimbabwe (ZK), the Netherlands (NL) and Sweden (SU) calculated in sliding windows of 500 bp with a step size of 250 bp. (B) F_{st} values for pairwise comparisons of ZK, NL and SU calculated in sliding windows of 500 bp with a step size of 250 bp. (C) Selective sweep *(SweepFinder)* analysis of the Netherlands population showing the composite likelihood ratio (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 black line indicates the 5% significance threshold calculated using the demographic model of Duchen et al. [5] for neutral simulations. The red line indicates the 5% significance threshold calculated using the demographic model of Werzner et al. [6] for neutral simulations and the gray dashed line indicates the 5% significance threshold using the model of Thornton and Andolfatto [35]. (E) Gene models for the 6-kb region analyzed. The gray highlighted region indicates the position of the 49-bp indel polymorphism in the *MtnA* 3' UTR.

doi:10.1371/journal.pgen.1005987.g007

UTR deletion (Fig 7C). This peak was significant when the demographic models of Duchen et al. [5], Werzner et al. [6], and Thornton and Adolfatto [35] were 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 (Fig 7D), where the CLR statistic was above the 5% significance threshold determined from all three of

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

To test the possibility that the deletion in the *MtnA* 3' UTR might have risen to high frequency as a result of hitchhiking with another linked polymorphism, we examined linkage disequilibrium (LD) across a 100 kb region flanking the *MtnA* locus in the Netherlands population (S1 Fig). The degree of linkage disequilibrium, r^2 [39], was calculated between all pairs of SNPs present in the 100 kb region, excluding singletons. The SNP corresponding to the indel polymorphism (Fig 4a), position 53 of the linkage disequilibrium matrix, is not in significant LD with any of the 94 SNPs present along the 100 kb region analyzed (S1 Fig). These results indicate that the high frequency of the *MtnA* 3' UTR deletion cannot be explained by linkage with another positively selected locus.

Association of the *MtnA* 3' UTR deletion with increased oxidative stress tolerance

MtnA expression has been linked to increased heavy metal tolerance [19,20,40] and metallothioneins in general have been associated with protection against oxidative stress [18,41]. To test if *MtnA* plays a role in oxidative stress and/or heavy metal tolerance, we used RNA interference (RNAi) to knockdown *MtnA* expression; these flies, along with their respective controls, were exposed to either hydrogen peroxide or copper sulfate. A knockdown in *MtnA* expression was significantly associated with increased mortality in the presence of hydrogen peroxide ($P < 0.001$; Fig 8A) and copper sulphate ($P = 0.026$; Fig 9A and 9B), although for the latter, this decrease was only significant in females.

To further test if the deletion in the *MtnA* 3' UTR could be associated with an increase in oxidative stress and/or heavy metal tolerance, a subset of *D. melanogaster* lines from the Dutch and Malaysian populations, either with or without the deletion, were exposed to hydrogen peroxide and copper sulfate. The 3' UTR deletion was associated with a significant increase in survival in the presence of hydrogen peroxide in both the Dutch ($P = 0.001$; Fig 8B) and Malaysian ($P = 0.001$; Fig 8B) populations. The 3' UTR deletion had no significant effect on survival

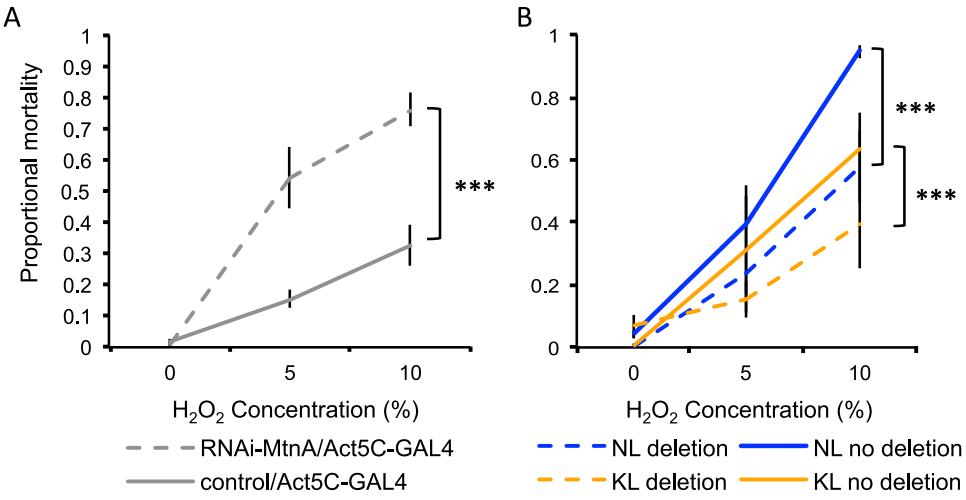


Fig 8. Proportional mortality after oxidative stress. (A) RNAi-mediated *MtnA* knockdown (hatched lines) and control flies (solid lines). *P*-values are shown for within population/background comparisons. (B) Dutch (blue) and Malaysian (orange) flies with the deletion (hatched lines) and without the deletion (solid lines) in the *MtnA* 3' UTR. Error bars indicate the standard error of the mean. $*P < 0.05$, $**P < 0.01$, $***P < 0.005$.

doi:10.1371/journal.pgen.1005987.g008

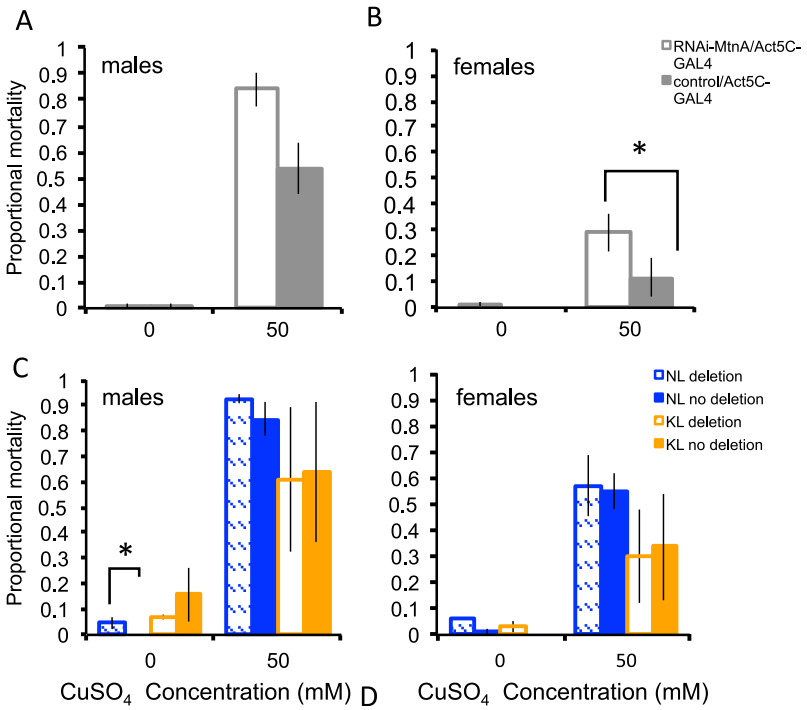


Fig 9. Proportional mortality after copper sulphate exposure. (A,B) Copper tolerance in RNAi-mediated *MtnA* knockdown flies (white, RNAi-*MtnA*/Act5C-GAL4) and control flies expressing normal levels of *MtnA* (solid grey, control/Act5C-GAL4). (C) Male and (D) female flies from the Dutch (NL, blue) and the Malaysian (KL, orange) population with the deletion (hatched) and without the deletion (solid). *P*-values are shown for within population/background comparisons. Error bars indicate the standard error of the mean. **P* < 0.05, ***P* < 0.01, ****P* < 0.005.

doi:10.1371/journal.pgen.1005987.g009

in the presence of copper sulfate in Dutch and Malaysian females ($P = 0.976$ and $P = 0.732$ respectively; Fig 9D) or males ($P = 0.578$ and $P = 0.904$ respectively; Fig 9C). Thus, the deletion in the *MtnA* 3' UTR was associated with increased oxidative stress tolerance, but not increased heavy metal tolerance.

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 [12]. 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 this indel polymorphism. The ancestral state of the 3' UTR contains a 49-bp sequence that is deleted in a derived allele that is present in worldwide populations. The deletion is nearly absent from sub-Saharan Africa, but present in frequencies >80% in northern Europe (Table 1). The deletion is present at intermediate frequency in Egypt (60%), Cyprus (65%) and Malaysia (45%). These findings suggest that positive selection has favored the 3' UTR deletion, at least within northern European populations. This interpretation is supported by population genetic analyses that indicate a recent selective sweep at the *MtnA* locus in populations from the Netherlands and Sweden (Fig 7). Furthermore, a clinal relationship between deletion

frequency and latitude is also seen in North America and Australia, suggesting that there is a common selection gradient affecting all populations outside of sub-Saharan Africa.

Although chromosome arm 3R is known to harbor inversion polymorphisms that vary in frequency with latitude in cosmopolitan populations [42], we can rule out linkage to a segregating inversion as a cause for the clinal pattern seen for the *MtnA* 3' UTR deletion. A previous analysis of the same Dutch population used in our study found that only one of the isofemale lines harbored an inversion on 3R, *In(3R)P* [43]. This was line *NL13*, which is one of the 10 lines with the *MtnA* 3' UTR deletion (Fig 4A). Thus, there is no evidence for linkage between the inversion and the deletion. Moreover, the *MtnA* gene lies 7 Mb outside of the nearest breakpoint of *In(3R)P*.

Using hydrogen peroxide tolerance assays, we found evidence that knocking down *MtnA* expression decreases oxidative stress tolerance (Fig 8B). The association of the deletion in the *MtnA* 3' UTR with increased survival in the presence of hydrogen peroxide (Fig 8A) suggests that the deletion has been selectively favored in some environments because it confers increased tolerance to oxidative stress. While cytotoxic reactive oxygen species (ROSs) are generated by natural metabolic processes, they can also be introduced via abiotic factors in the environment, such as radiation, UV light or exposure to toxins. The significant correlation between the frequency of the 3' UTR deletion and latitude, coupled with its association with increased oxidative stress tolerance suggests that environmentally induced oxidative stress may vary clinally, with greater stress in northern European environments. Regulation of the oxidative stress response usually occurs via upregulation of antioxidant protective enzymes in response to the binding of a *cis*-acting *antioxidant-responsive element* (ARE), which contains a characteristic sequence to which stress-activated transcription factors can bind [41]. A recent example of adaptation to oxidative stress in *Drosophila* is the insertion of the *Bari-Jheh* transposable element into the intergenic region of *Juvenile Hormone Epoxy Hydrolase* (*Jheh*) genes, which adds additional AREs that upregulate two downstream *Jheh* genes and was associated with increased oxidative stress tolerance [44]. Interestingly, the *Bari-Jheh* insertion also shows evidence for a partial selective sweep in non-African *D. melanogaster* [45], suggesting that oxidative stress may have imposed an important selective constraint on the colonization of Europe. However, the *MtnA* 3' UTR deletion cannot mediate its associated increase in oxidative stress tolerance in a similar way, since it does not add any new AREs.

Due to their high inducibility in response to heavy metals, metallothioneins have traditionally been thought to play a role as detoxifiers specifically of heavy metals. However, this view has come into question recently, and metallothioneins are now thought to be a part of the general stress response and may function as scavengers of free radicals [41]. The association of the *MtnA* 3'UTR deletion with increased oxidative stress tolerance (Fig 8A) is in line with this more recent view of the role of metallothioneins, while the observed increased mortality after copper exposure in females in which *MtnA* expression has been knocked down (Fig 9D) is in keeping with the more traditional view. However, we found no association between the presence of the deletion and copper tolerance. This may be because the RNAi knockdown results in an *MtnA* expression level that is much lower than that of naturally occurring alleles, and copper tolerance is only affected when *MtnA* expression falls below a minimal threshold. The precise mechanisms of how metallothioneins interact with other metal processing systems after their initial binding and help remove excess of heavy metals, remain unclear [41].

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 (S1 Table), 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,

Table 4. Top scoring miRNAs 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 in base pairs 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). The energetic score for the probability and stability of the binding is denoted by ddG. The more negative the score is, the more probable is the interaction between the 3' UTR and the miRNA.

doi:10.1371/journal.pgen.1005987.t004

non-coding RNAs that modulate the expression of genes by inhibiting transcription or inducing mRNA degradation [46]. 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 [46–48]. 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 [49]. 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 with 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.

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 [50–52]. 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 similar expression pattern is also seen in the gut (Fig 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* [24]. However, tissue-specific effects of the difference in *MtnA* expression cannot be ruled out. As shown in Fig 6, GFP expression driven by the *MtnA* promoter in the brain is limited to what seems only one cell type, which according to their morphological and anatomical characteristics, could correspond to glia. It has been reported that glia cells protect neurons and other brain cells from ROS damage caused by oxidative stress [53,54] and the fact that *MtnA* has been found to be expressed in the astrocyte glial cells in larva and adult flies [28], suggests that *MtnA* expression in glia could serve as neuronal protection against environmental factors, such as exposure to xenobiotics, that trigger an oxidative stress response [29,55–57]. Our functional experiments showing an association between genetic variation in *MtnA* and oxidative stress tolerance are consistent with *MtnA* expression in glia providing protection against oxidative stress, which may be especially important in the brain, as neurons are highly susceptible to ROS damage.

Materials and Methods

Fly strains

This study used isofemale lines from 12 populations of *D. melanogaster*, including: Zimbabwe (Lake Kariba), Zambia (Lake Kariba), Rwanda (Gikongoro), Cameroon (Oku), Egypt (Cairo), Cyprus (Nicosia), 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 [10–12]. Flies from Germany were collected from different locations in the greater Munich area. Flies from Cyprus were collected from a single location near Nicosia. 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 Genomics Project [8] and were kindly provided by John Pool and Charles Langley (University of California, Davis).

Flies expressing hairpin RNA targeted against *MtnA* mRNA under the control of the GAL4/UAS system (RNAi-*MtnA*; transformant ID: 105011) and the host line used in their creation (control; transformant ID: 60100) were obtained from the Vienna *Drosophila* Stock Center [58] *Act5C*/*Cyo* flies expressing GAL4 under the control of an *Act5C* driver were kindly provided by Ilona Grunwald Kadow. For tolerance assays, *Act5C*/*Cyo* females were crossed to RNAi-*MtnA* and control males and the progeny (RNAi-*MtnA*/*Act5C*-GAL4 and control/*Act5C*-GAL4) were used in tolerance assays. Using qRT-PCR as described below, *MtnA* expression was confirmed to be knocked down by 90.03% in males and 87.58% in females in RNAi-*MtnA*/*Act5C*-GAL4 flies in comparison to control/*Act5C*-GAL4. 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 [12] 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 (*ZK84*, *ZK95*, *ZK131*, *ZK145*, *ZK157*, *ZK186*, *ZK191*, *ZK229*, *ZK377*, *ZK384*, *ZK398*) and five brains were dissected from males and females of each of the 12 lines from the Netherlands (*NL01*, *NL02*, *NL11*, *NL12*, *NL13*, *NL14*, *NL15*, *NL16*, *NL17*, *NL18*, *NL19*, *NL20*). The dissected brains of each population and sex were pooled following the RNA-seq strategy previously described [12]. The above procedure was repeated in two biological replicates for each sex and population. To compare the *MtnA* expression of individual lines within populations, subsets of eight lines were chosen from Zimbabwe (*ZK84*, *ZK95*, *ZK131*, *ZK145*, *ZK157*, *ZK186*, *ZK377*, *ZK384*) and the Netherlands (*NL01*, *NL02*, *NL11*, *NL12*, *NL15*, *NL16*, *NL17*, *NL18*). Thirty whole brains and digestive tracts (from foregut to hindgut) were dissected per line. Two biological replicates of each line (each consisting of 30 brains or guts) were processed. 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* (Dm02362764_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 processed for each sample. The ΔΔCt method was used to compute the normalized expression of *MtnA* using the ribosomal protein gene *RpL32* as the reference [59].

CNV assays

The paralogous genes *AttacinA* (*AttA*) and *AttacinB* (*AttB*) were used as positive controls for CNV assays, because they share 97% nucleotide identity [21] and can be co-amplified with the same primer set. The sequences for *AttA* and *AttB* were downloaded from FlyBase [60] and aligned using the ClustalW2 algorithm implemented in SeaView (version 4) [61]. Primers were designed for 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 is not known to show CNV, was used as a negative control. The *RpL32*-specific primers were: forward (5'-GACAATCTCCTTGCGCTTCT-3') and reverse (5'-AGCTGGAGGTCCTGCTCAT-3'). The primers specific 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 (Fig 1). The following primer pairs were used (all 5' to 3'): GATGGTGGAATACCCTTTGC and AAAGCGGGTTTACCAGTGTG; GTTGG CCTGGCTTAATAACG and ACTGGCACTGGAGCTGTTTC; GCTCTTGCTAGCCAT TCTGG and AGAACCCGGCATATAAACGA; GATATGCCCACACCCATACC and GTA GAGGCGCTGCATCTTGT; CACTTGACCATCCCATTTCC and CAAGTCCCCAAAGTG GAGAA; CTTGATTTTGCTGCTGACCA and ATCGCCACGATTATGATTGC; CAGGA CAATCAAGCGGAAGT and TTATGAAGCGCAGCACCAGT; GACCCACTCGAATCCG TATC 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) [61]. All sequences have been submitted to GenBank/EMBL under the accession numbers KT008059–KT008093.

MtnA indel polymorphism screening and latitude 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 (35 cycles of 98° for 5 sec. and 60° for 10 sec.) using the following primers: forward (5'-GCCGCAGACCAATTGATTA-3') and reverse (5'-TTCTTTCCAGGATGCAAATG-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 [62]. The strength and significance of the correlation between the frequency of the deletion and latitude was determined using linear regression.

To determine the frequency of the *MtnA* 3' UTR deletion on other continents, raw pool-seq reads from North America [30] and Australia [31] were downloaded from the National Center for Biotechnology Information (NCBI) short read archive (SRA). The reads were mapped to either the ancestral or derived (with 49-bp deletion) version of the *MtnA* 3' UTR using NextGenMap [63]. Only reads spanning the site of the indel were considered informative. The deletion frequency was estimated as the proportion of informative reads that matched the deletion allele. The 95% confidence interval was estimated using the probit method in R [55].

Cloning and transgenesis

To test whether the indel polymorphism found in the *MtnA* 3' UTR can account for the difference in expression observed between the Dutch and the Zimbabwean populations, we constructed transgenic flies using the phiC31 transgenesis system [23]. Two expression vectors containing a green fluorescent protein (GFP) reporter gene were constructed. *MtnA* 3' UTR sequences from the Netherlands (line *NL20*) and Zimbabwe (line *ZK84*), corresponding to chromosome arm 3R coordinates 5,607,448–5,611,691, were PCR-amplified with forward (5'-TTTCCTCGAACTTGTTCACTTG -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 Dutch and the Zimbabwean populations, was also PCR amplified and cloned separately into the p*CR2.1-TOPO* vector using forward (5'-GCCGCAGACCAATTGATTA-3') and reverse (5'-TTCTTTCCAGGATGCAAATG-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 [23]. For the *lacZ* constructs, the *MtnA* promoter was excised from the p*CR2.1-TOPO* vector with *Eco*RI and ligated into the *Eco*RI site 5' of the *lacZ* coding sequence in the p*CMV-SPORT-βgal* plasmid (Life Technologies). PCR primers with overhangs containing restriction sites for *Xho*I and *Xba*I (forward 5'- GGTC CGACTCGAGGCGAAATACGGGCAGACATG -3' and reverse 5'- GGTGCTCTAGAGCTC CATAGAAGACACCGGGAC -3') were used to amplify the *MtnA* promoter/*lacZ* fragment and the product was ligated into the *Xho*I–*Xba*I sites just upstream of the *MtnA* 3' UTR fragment in the p*CR2.1-TOPO* vector. Finally, the whole construct was excised using *Xba*I and *Kpn*I and ligated into the *Xba*I–*Kpn*I sites of the *pattB* vector (Fig 5). 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. The *M{vas-int.Dm}ZH-2A*, *M{3xP3-RFP.attP}ZH-51D* line was used for embryo microinjections. 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

GFP 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 ten heads of each transformant line using the RNA extraction and reverse transcription protocols described above. Thirteen biological replicates (six male and seven female) were processed for each line, each with two technical replicates. qRT-PCR was performed as described above using a custom Taqman probe for GFP (Applied Biosystems; forward primer: 5'-GAGCGCACCATCTTCTT CAAG-3', reverse primer: 5'-TGTCGCCCTCGAACTTCAC-3', FAM-labeled primer: 5'-AC GACGGCAACTACA-3') and a probe for *RpL32* (Dm02151827_g1), which was used as an endogenous control. The data analysis was also performed as described above for *MtnA* gene expression. A *t*-test was performed to assess significance.

β-galactosidase assays. β-galactosidase activity was measured in groups of 30 heads or eight guts of homozygous transformant flies. Proteins were extracted from the tissues and the β-galactosidase activity assay was performed as described in [64] with the following exceptions. After dissection in cold PBS, tissues were frozen with liquid nitrogen and homogenized before the addition of 150 μL of the 0.1 M Tris-HCl, 1 mM EDTA, and 7 mM 2-mercaptoethanol buffer (pH 7.5). For each assay, two technical replicates of 60 μL of the supernatant containing the soluble proteins were combined with 50 μl of the 200 mM sodium phosphate (pH 7.3), 2 mM $MgCl_2$, 100 mM 2-mecaptoethanol assay buffer. β-galactosidase activity was measured spectrophotometrically by following the change in absorbance at 420 nm at 37° Celsius. Four to five biological replicates were performed per tissue and per sex. Significance was assessed using a *t*-test.

Brain confocal imaging

Brain tissue was dissected in ice-cold 1X PBS and fixed with PLP (8% paraformaldehyde in NaOH and PBS with lysine (1)-HCl) for one hour at room temperature as described in [65]. After fixation, the tissue was washed twice for 15 minutes with PBS-0.5% Triton X and then incubated for one 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° Celsius in blocking solution. After washing twice with PBS-0.5% Triton X, the tissue was incubated with the secondary antibody, 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 [66].

Population genetic analysis and tests for selection

Summary statistics, including the number of segregating sites (S), number of haplotypes and Tajima's D [34] were calculated using DnaSP v.5.10.1 [67]. The mean pairwise nucleotide diversity (π) [33], Watterson's [32] estimate of nucleotide diversity (θ) and F_{st} [68] were calculated as described in [5]. Hudson's haplotype test (HHT) was carried out using *ms* [69] to perform coalescent simulations and *psubs* [70] to calculate the probability of observing a subset of n sequences containing p or fewer polymorphic sites. The demographic models of Thornton and Andolfatto [35], Duchen et al. [5], and Werzner et al. [6] were used to simulate the out-of Africa bottleneck.

To test for a selective sweep, a *SweepFinder* analysis was performed using the *SweeD* software [38]. 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 [8]. 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 (S2 Table, S2a Fig). Furthermore, the inclusion of a French line did not lead to a skew in the background SFS (S2b Fig). For the Swedish population, the background SFS of chromosome arm 3R was determined from 12 whole genome sequences from the Umeå population (S3 Table). In order to increase the power of the test, the invariant sites in the alignment were also included [37]. To assess the significance of the composite likelihood ratio (CLR) statistic, neutral simulations were performed using *ms* [69]. In the neutral simulations three demographic models were taken into account [5,6,35]. 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. 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. [5] infer this bottleneck to have occurred around 19,000 years ago, Thornton and Andolfatto [35] around 16,000 years ago, and Werzner et al. [6] around 13,000 years ago. However, the 95% confidence intervals of the estimates are very wide, ranging from 7,359–43,000 years ago. Thus, the three estimates are not incompatible with each other. The recombination rate of the *MtnA* genomic region was obtained from the *D. melanogaster* recombination rate calculator [71]. 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. Linkage disequilibrium was calculated between all pairs of SNPs present 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 [39]. A fragment of ~100 kb flanking the *MtnA* locus (3R: 9,732,746..9,835,406) was analyzed, with singletons excluded. A Fisher's exact test was used to assess significance of the r^2 values.

Copper and oxidative stress tolerance assays

Copper sulfate and hydrogen peroxide tolerance assays were performed using five *D. melanogaster* lines containing the *MtnA* 3' UTR deletion (two Dutch and three Malaysian lines) and three lines without the deletion (two Malaysian and one Dutch line), as well as an *MtnA* knockdown line (RNAi-*MtnA*/*Act5C*-GAL4 and its control (control/*Act5C*-GAL4). Assays were performed at 25°C in tolerance chambers consisting of a plastic vial (diameter = 25 mm, height = 95 mm) with compressed cotton at the bottom containing 2.5 ml copper sulfate (Sigma Aldrich) or hydrogen peroxide (Sigma Aldrich) solution supplemented with 5% sucrose and sealed with a cork. Four to six day-old flies were separated by sex and tested in groups of 20. For each assay, one concentration of copper sulfate (50 mM) or two concentrations of hydrogen peroxide (5 or 10%) were tested with 5–7 replicates per sex and concentration. A control solution containing only sucrose was also tested with 3–5 (10–15 for *Act5C*-GAL4 background) replicates per sex for each assay. Mortality was recorded as the number of dead flies after 48 ± 1 hours. To determine the effect of the deletion, lines with and without the deletion were compared within each population or background. For copper sulfate assay analysis, *t*-tests were performed to assess significance. In order to account for potential differences in mortality inherent among the lines, proportional mortality data was scaled by mortality at 0 mM using the formula mortality/(1 + mean mortality at 0mM). For hydrogen peroxide assay analysis, the data for each assay and population was fit to a generalized linear model (GLM) using concentration, line, sex, and presence of the deletion as factors and a quasibinomial distribution using the glm function in R [62]. The tolerance results for each sex (S3 Fig) and the GLM coefficients (S4–S12 Tables) are provided as supporting information.

Supporting Information

S1 Fig. Linkage disequilibrium matrix. Linkage disequilibrium was assessed from a genomic region of 100 kb with *MtnA* at the center. The upper and right axes show the SNPs found in the 100 kb fragment, excluding singletons, and each cell represents r^2 values. The left and bottom part of the matrix shows the results of a Fisher's exact test for each pair of SNPs. Red boxes indicate significant *P*-values.
(PDF)

S2 Fig. 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 included with the NL lines to calculate the background site frequency spectrum.
(PDF)

S3 Fig. Oxidative stress tolerance by sex. Proportional mortality of *D. melanogaster* males (A, C) and females (B, D) after exposure to hydrogen peroxide for 48 hours in (A,B) flies with (hatched lines) and without (solid lines) the deletion in the *MtnA* 3' UTR and (C,D) RNAi-mediated *MtnA* knockdown (hatched lines) and control (solid lines) flies (C,D). (A,B) The Dutch (NL) population is shown in blue and the Malaysian (KL) population in orange. Legends are provided to the right of each row. *P*-values are shown for within population/background and sex comparisons. Error bars represent standard error of the mean.
(PDF)

S1 Table. Isoform-specific expression of *MtnA* in the brain.
(PDF)

S2 Table. Average pairwise differences per kb between French (FR) and Dutch (NL) lines.
(PDF)

S3 Table. Site frequency spectrum (SFS) of the Swedish population drawn from the whole 3R chromosome arm.
(PDF)

S4 Table. Oxidative stress tolerance glm coefficients for Malaysian population.
(PDF)

S5 Table. Oxidative stress tolerance glm coefficients for Dutch population.
(PDF)

S6 Table. Oxidative stress tolerance glm coefficients for *MtnA* knockdown and control lines.
(PDF)

S7 Table. Male oxidative stress tolerance glm coefficients for Malaysian population.
(PDF)

S8 Table. Male oxidative stress tolerance glm coefficients for Dutch population.
(PDF)

S9 Table. Male oxidative stress tolerance glm coefficients for *MtnA* knockdown and control lines.
(PDF)

S10 Table. Female oxidative stress tolerance glm coefficients for Malaysian population.
(PDF)

S11 Table. Female oxidative stress tolerance glm coefficients for Dutch population.
(PDF)

S12 Table. Female oxidative stress tolerance glm coefficients for *MtnA* knockdown and control lines.
(PDF)

Acknowledgments

We thank Hilde Lainer, Hedwig Gebhart, Angélica Cuevas and Annabella Königer for excellent assistance in the lab. We also thank Ilona Grunwald Kadow for sharing fly lines and primary/secondary antibodies, as well as for helpful comments on the confocal images. We further thank Ricardo Wilches and Wolfgang Stephan for sharing fly lines and sequences. Finally we thank Sebastian Hoehna and Kyle McCulloch for comments on the manuscript. This work is part of the dissertation of Ana Catalán.

Author Contributions

Conceived and designed the experiments: AC AGS JP. Performed the experiments: AC AGS EA. Analyzed the data: AC AGS PD JP. Wrote the paper: AC AGS JP.

References

1. David J, Capy P. Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. 1988; 4: 106–111. doi: 10.1016/0168-9525(88)90098-4 PMID: 3149056
2. Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M. Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol Biol. Plenum Press; 1988; 22: 159–225. Available: http://link.springer.com/chapter/10.1007/978-1-4613-0931-4_4
3. Stephan W, Li H. The recent demographic and adaptive history of *Drosophila melanogaster*. Heredity (Edinb). 2007; 98: 65–8. doi: 10.1038/sj.hdy.6800901
4. Laurent SJY, 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: 2041–51. doi: 10.1093/molbev/msr031 PMID: 21300986
5. Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S. Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. Genetics. 2013; 193: 291–301. doi: 10.1534/genetics.112.145912 PMID: 23150605
6. Werzner A, Pavlidis P, Ometto L, Stephan W, Laurent S. Selective sweep in the *Flotillin-2* region of European *Drosophila melanogaster*. PLoS One. Public Library of Science; 2013; 8: e56629. doi: 10.1371/journal.pone.0056629
7. 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: 1167–74. doi: 10.1093/molbev/msr284 PMID: 22101416
8. Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, et al. Population genomics of sub-saharan *Drosophila melanogaster*: African diversity and non-African admixture. Malik HS, editor. PLoS Genet. Public Library of Science; 2012; 8: e1003080. doi: 10.1371/journal.pgen.1003080
9. Glaser-Schmitt A, Catalán A, Parsch J. Adaptive divergence of a transcriptional enhancer between populations of *Drosophila melanogaster*. Philos Trans R Soc Lond B Biol Sci. 2013; 368: 20130024. doi: 10.1098/rstb.2013.0024 PMID: 24218636

10. 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. doi: 10.1186/gb-2008-9-1-r12 PMID: 18208589
11. Müller L, Hutter S, Stamboliyska R, Saminadin-Peter SS, Stephan W, Parsch J. Population transcriptomics of *Drosophila melanogaster* females. BMC Genomics. 2011; 12: 81. doi: 10.1186/1471-2164-12-81 PMID: 21276238
12. Catalán A, Hutter S, Parsch J. Population and sex differences in *Drosophila melanogaster* brain gene expression. BMC Genomics. 2012; 13: 654. doi: 10.1186/1471-2164-13-654 PMID: 23170910
13. Huylmans AK, Parsch J. Population- and sex-biased gene expression in the excretion organs of *Drosophila melanogaster*. G3 (Bethesda). 2014; 4: 2307–15. doi: 10.1534/g3.114.013417
14. Egli D, Selvaraj A, Yepiskoposyan H, Zhang B, Hafen E, Georgiev O, et al. Knockout of "metal-responsive transcription factor" MTF-1 in *Drosophila* by homologous recombination reveals its central role in heavy metal homeostasis. EMBO J. 2003; 22: 100–8. doi: 10.1093/emboj/cdg012 PMID: 12505988
15. Pérez-Rafael S, Kurz A, Guirola M, Capdevila M, Palacios O, Atrian S. Is *MtnE*, the fifth *Drosophila* metallothionein, functionally distinct from the other members of this polymorphic protein family? Metallomics. The Royal Society of Chemistry; 2012; 4: 342–9. doi: 10.1039/c2mt00182a
16. Guirola M, Naranjo Y, Capdevila M, Atrian S. Comparative genomics analysis of metallothioneins in twelve *Drosophila* species. J Inorg Biochem. 2011; 105: 1050–1059. Available: http://www.sciencedirect.com/science/article/pii/S016201341100119X doi: 10.1016/j.jinorgbio.2011.05.004 PMID: 21726767
17. Capdevila M, Bofill R, Palacios Ò, Atrian S. State-of-the-art of metallothioneins at the beginning of the 21st century. Coord Chem Rev. 2012; 256: 46–62. Available: http://www.sciencedirect.com/science/article/pii/S0010854511001937
18. Nath R, Kumar D, Li T, Singal PK. Metallothioneins, oxidative stress and the cardiovascular system. Toxicology. 2000; 155: 17–26. Available: http://www.sciencedirect.com/science/article/pii/S0300483X00002730 PMID: 11154793
19. Maroni G, Wise J, Young JE, Otto E. Metallothionein gene duplications and metal tolerance in natural populations of *Drosophila melanogaster*. Genetics. 1987; 117: 739–44. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1203245&tool=pmcentrez&rendertype=abstract PMID: 2828157
20. Lange BW, Langley CH, Stephan W. Molecular evolution of *Drosophila* metallothionein genes. Genetics. 1990; 126: 921–32. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1204289&tool=pmcentrez&rendertype=abstract PMID: 1981765
21. Lazzaro BP, Clark AG. Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the *Attacin* genes of *Drosophila melanogaster*. Genetics. 2001; 159: 659–671. Available: http://www.genetics.org/content/159/2/659 PMID: 11606542
22. Groth CA, Fish M, Nusse R, Calos PM. Construction of transgenic *Drosophila* by using the site-specific integrase from phage C31. Genetics. 2004; 166: 1775–1782. doi: 10.1534/genetics.166.4.1775 PMID: 15126397
23. Bischof J, Maeda RK, Hediger M, Karch F, Basler K. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. Proc Natl Acad Sci U S A. 2007; 104: 3312–7. doi: 10.1073/pnas.0611511104 PMID: 17360644
24. Chintapalli VR, Wang J, Dow JAT. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat Genet. 2007; 39: 715–20. PMID: 17534367
25. Hartenstein V, Spindler S, Pereanu W, Fung S. The development of the *Drosophila* larval brain. Adv Exp Med Biol. 2008; 628: 1–31. doi: 10.1007/978-0-387-78261-4_1 PMID: 18683635
26. Edwards TN, Meinertzhagen IA. The functional organisation of glia in the adult brain of *Drosophila* and other insects. Prog Neurobiol. 2010; 90: 471–97. doi: 10.1016/j.pneurobio.2010.01.001 PMID: 20109517
27. Hartenstein V. Morphological diversity and development of glia in *Drosophila*. Glia. 2011; 59: 1237–52. doi: 10.1002/glia.21162 PMID: 21438012
28. Huang Y, Ng FS, Jackson FR. Comparison of larval and adult *Drosophila* astrocytes reveals stage-specific gene expression profiles. G3 (Bethesda). 2015; 5: 551–8. doi: 10.1534/g3.114.016162
29. Egli D, Domènech J, Selvaraj A, Balamurugan K, Hua H, Capdevila M, et al. The four members of the *Drosophila* metallothionein family exhibit distinct yet overlapping roles in heavy metal homeostasis and detoxification. Genes Cells. 2006; 11: 647–58. doi: 10.1111/j.1365-2443.2006.00971.x PMID: 16716195

30. Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. PLoS Genet. Public Library of Science; 2014; 10: e1004775. doi: 10.1371/journal.pgen.1004775

31. Reinhardt JA, Kolaczkowski B, Jones CD, Begun DJ, Kern AD. Parallel geographic variation in *Drosophila melanogaster*. Genetics. 2014; 197: 361–373. doi: 10.1534/genetics.114.161463 PMID: 24610860

32. Watterson GA. On the Number of segregating sites in genetical models without recombination. Theor Popul Biol. 1975; 276: 256–276. Available: http://www.sciencedirect.com/science/article/pii/0040580975900209

33. Tajima F. Evolutionary relationship of DNA sequences in finite populations. Genetics. 1983; 105: 437–60. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1202167&tool=pmcentrez&rendertype=abstract PMID: 6628982

34. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. Genetics Society of America; 1989; 123: 585–95. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1203831&tool=pmcentrez&rendertype=abstract

35. Thornton K, Andolfatto P. Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. Genetics. 2006; 172: 1607–19. doi: 10.1534/genetics.105.048223 PMID: 16299396

36. Hudson RR. The how and why of generating gene genealogies. Takahata N, Clark AG, editors. MA: Sinauer Associates, Sunderland, MA.; 1993.

37. Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. Genome Res. 2005; 15: 1566–75. doi: 10.1101/gr.4252305 PMID: 16251466

38. Pavlidis P, Živković D, Stamatakis A, Alachiotis N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. Mol Biol Evol. 2013; 30: 2224–34. doi: 10.1093/molbev/mst112 PMID: 23777627

39. Lewontin RC. The interaction of selection and linkage. I. General considerations; Heterotic models. Genetics. 1964; 49: 49–67. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1210557&tool=pmcentrez&rendertype=abstract PMID: 17248194

40. Maroni G, Otto E, Lastowski-Perry D. Molecular and cytogenetic characterization of a metallothionein gene of *Drosophila*. Genetics. 1986; 112: 493–504. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1202760&tool=pmcentrez&rendertype=abstract PMID: 3007277

41. Straalen van NM, Roelofs D. An introduction to ecological genomics [Internet]. 2nd Edition. OUP Oxford; 2012. Available: https://books.google.com/books?hl=en&lr=&id=VUAfAQAAQBAJ&pgis=1

42. Kapun M, Van Schalkwyk H, McAllister B, Flatt T, Schlötterer C. Inference of chromosomal inversion dynamics from Pool-Seq data in natural and laboratory populations of *Drosophila melanogaster*. Mol Ecol. 2014; 23: 1813–1827. doi: 10.1111/mec.12594 PMID: 24372777

43. 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. doi: 10.1534/genetics.107.074922 PMID: 17660560

44. Guio L, Barrón MG, González J. The transposable element *Bari-Jheh* mediates oxidative stress response in *Drosophila*. Mol Ecol. 2014; 23: 2020–30. doi: 10.1111/mec.12711 PMID: 24629106

45. González J, Macpherson JM, Petrov DA. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. Mol Biol Evol. 2009; 26: 1949–61. doi: 10.1093/molbev/msp107 PMID: 19458110

46. Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet. 2007; 8: 93–103. PMID: 17230196

47. Berezikov E. Evolution of microRNA diversity and regulation in animals. Nat Rev Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2011; 12: 846–60.

48. Flynt AS, Lai EC. Biological principles of microRNA-mediated regulation: shared themes amid diversity. Nat Rev Genet. 2008; 9: 831–42. doi: 10.1038/nrg2455 PMID: 18852696

49. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. The role of site accessibility in microRNA target recognition. Nat Genet. 2007; 39: 1278–84. PMID: 17893677

50. Carroll SB. Endless Forms. Cell. 2000; 101: 577–580. doi: 10.1016/S0092-8674(00)80868-5 PMID: 10892643

51. Wray GA. The evolutionary significance of *cis*-regulatory mutations. Nat Rev Genet. Nature Publishing Group; 2007; 8: 206–16.

52. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008; 134: 25–36. doi: 10.1016/j.cell.2008.06.030 PMID: 18614008

53. Gandhi S, Abramov AY. Mechanism of oxidative stress in neurodegeneration. Oxid Med Cell Longev. 2012; 2012: 428010. doi: 10.1155/2012/428010 PMID: 22685618

54. Shih AY, Johnson DA, Wong G, Kraft AD, Jiang L, Erb H, et al. Coordinate regulation of glutathione biosynthesis and release by Nrf2-expressing glia potently protects neurons from oxidative stress. J Neurosci. 2003; 23: 3394–3406. Available: http://www.jneurosci.org/content/23/8/3394.short PMID: 12716947

55. Gruenewald C, Botella JA, Bayersdorfer F, Navarro JA, Schneuwly S. Hyperoxia-induced neurodegeneration as a tool to identify neuroprotective genes in *Drosophila melanogaster*. Free Radic Biol Med. 2009; 46: 1668–1676. Available: http://www.sciencedirect.com/science/article/pii/S0891584909001841 doi: 10.1016/j.freeradbiomed.2009.03.025 PMID: 19345730

56. Logan-Garbisch T, Bortolazzo A, Luu P, Ford A, Do D, Khodabakhshi P, et al. Developmental ethanol exposure leads to dysregulation of lipid metabolism and oxidative stress in *Drosophila*. G3 (Bethesda). 2014; 5: 49–59. doi: 10.1534/g3.114.015040

57. Hosamani R. Acute exposure of *Drosophila melanogaster* to paraquat causes oxidative stress and mitochondrial dysfunction. Arch Insect Biochem Physiol. 2013; 83: 25–40. doi: 10.1002/arch.21094 PMID: 23564607

58. Dietzl G, Chen D, Schnorrer F, Su K-C, Barinova Y, Fellner M, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature. 2007; 448: 151–156. PMID: 17625558

59. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001; 29: e45. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=55695&tool=pmcentrez&rendertype=abstract PMID: 11328886

60. Dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, et al. FlyBase: introduction of the *Drosophila melanogaster* release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res. 2015; 43: D690–7. doi: 10.1093/nar/gku1099 PMID: 25398896

61. Gouy M, Guindon S, Gascuel O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010; 27: 221–4. doi: 10.1093/molbev/msp259 PMID: 19854763

62. R Core Team. R: A language and environment for statistical computing. [Internet]. Viena Austria; 2015. Available: http://www.r-project.org.

63. Sedlazeck FJ, Rescheneder P, Von Haeseler A. NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. Bioinformatics. 2013; 29: 2790–2791. doi: 10.1093/bioinformatics/btt468 PMID: 23975764

64. Hense W, Baines JF, Parsch J. X chromosome inactivation during *Drosophila* spermatogenesis. Noor MAF, editor. PLoS Biol. Public Library of Science; 2007; 5: e273. doi: 10.1371/journal.pbio.0050273

65. Cayirlioglu P, Kadow IG, Zhan X, Okamura K, Suh GSB, Gunning D, et al. Hybrid neurons in a microRNA mutant are putative evolutionary intermediates in insect CO_2 sensory systems. Science. 2008; 319: 1256–60. doi: 10.1126/science.1149483 PMID: 18309086

66. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2012; 9: 671–675.

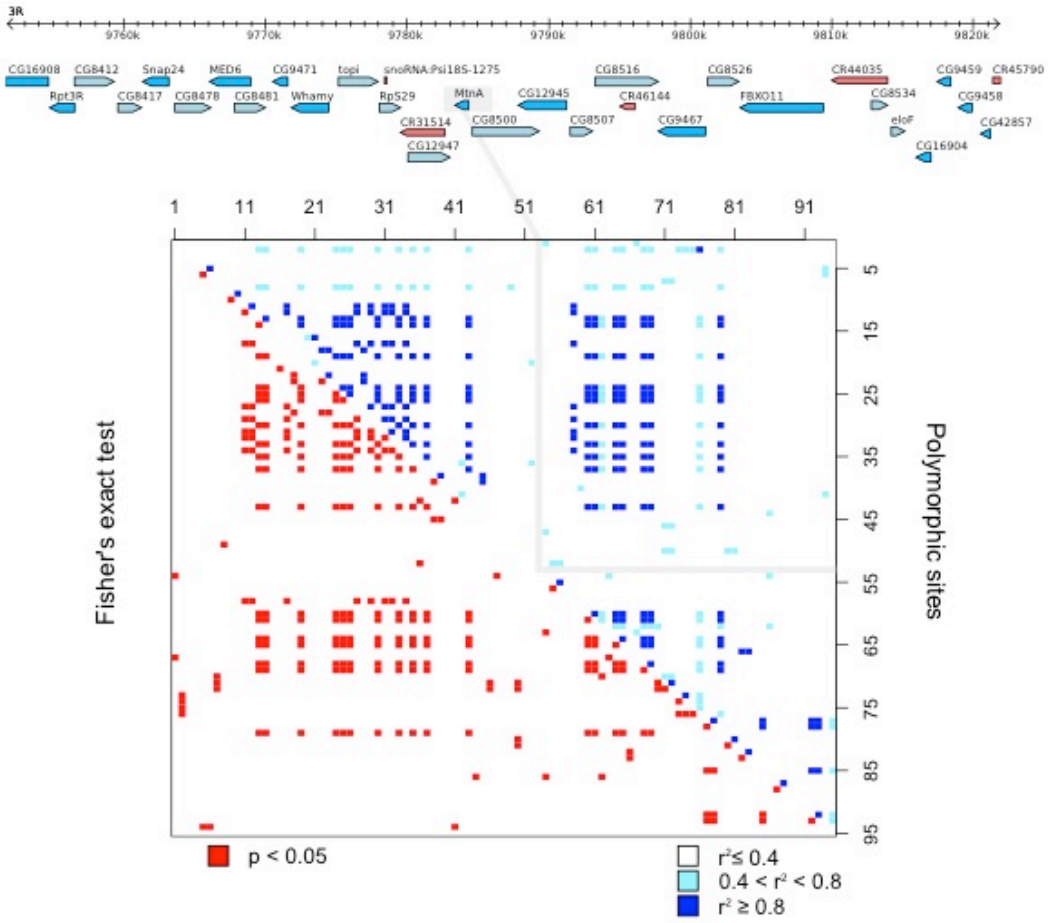
67. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25: 1451–2. doi: 10.1093/bioinformatics/btp187 PMID: 19346325

68. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. Genetics. 1992; 132: 583–589. Available: http://www.genetics.org/content/132/2/583.abstract PMID: 1427045

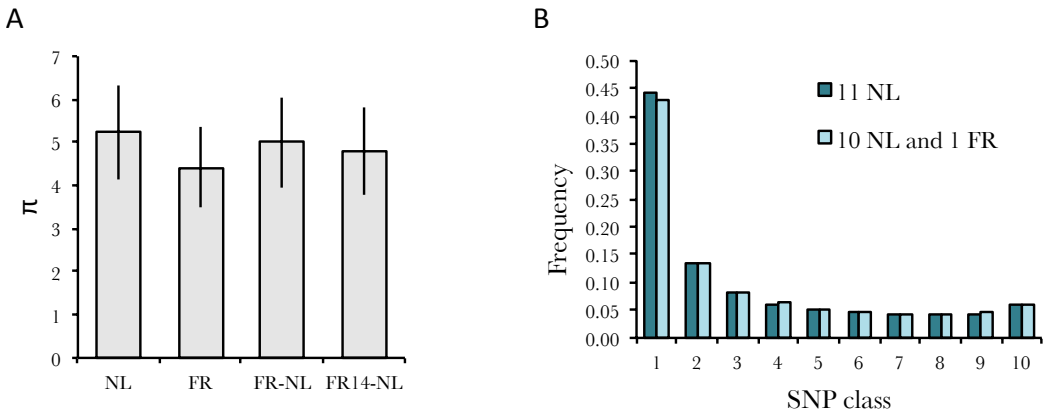
69. Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 2002; 18: 337–338. doi: 10.1093/bioinformatics/18.2.337 PMID: 11847089

70. Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ. Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. Genetics. 1994; 136: 1329–1340. Available: http://www.genetics.org/content/136/4/1329.abstract PMID: 8013910

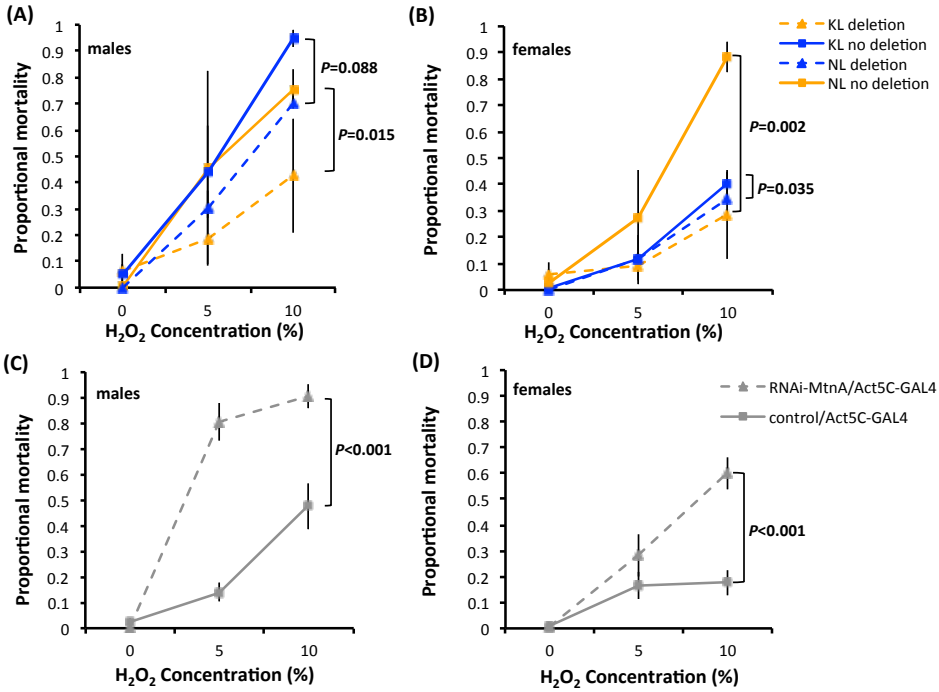
71. Fiston-Lavier A-S, Singh ND, Lipatov M, Petrov DA. *Drosophila melanogaster* recombination rate calculator. Gene. 2010; 463: 18–20. doi: 10.1016/j.gene.2010.04.015 PMID: 20452408



S1 Fig. Linkage disequilibrium matrix. Linkage disequilibrium was assed from a genomic region of 100 kb with *MtnA* at the center. The upper and right axes show the SNPs found in the 100 kb fragment not taking into account singletons and each bin represent r^2 values. The left and bottom part of the matrix show the Fisher exact test for each of the r^2 values. Red boxes indicate significant p-values.



S2 Fig. 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 included with the NL lines to calculate the background site frequency spectrum.



S3 Fig. Oxidative stress tolerance by sex. Proportional mortality of *D. melanogaster* males (A, C) and females (B, D) after exposure to hydrogen peroxide for 48 hours in (A,B) flies with (hatched lines) and without (solid lines) the deletion in the *MtnA* 3' UTR and (C,D) RNAi-mediated *MtnA* knockdown (hatched lines) and control (solid lines) flies (C,D). (A,B) The Dutch (NL) population is shown in blue and the Malaysian (KL) population in orange. Legends are provided to the right of each row. *P*-values are shown for within population/background and sex comparisons. Error bars represent standard error of the mean.

S1 Table. Isoform-specific expression of *MtnA* in the brain.

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

S2 Table. 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

S3 Table. Site frequency spectrum (SFS) of the Swedish population drawn from the whole 3R chromosome arm.

Class	Frequency
1	0.799963
2	0.086063
3	0.025863
4	0.015205
5	0.011744
6	0.009859
7	0.008616
8	0.007853
9	0.007445
10	0.007420
11	0.008246
12	0.011711

S4 Table. Oxidative stress tolerance glm coefficients for Malaysian population

	Estimate	Std. Error	t value	*P*-value
Intercept	4.16137	0.49944	8.332	6.13E-14
Concentration	-0.3314	0.04201	-7.888	7.43E-13
Deletion present	2.91704	0.88944	3.28	0.001307
Line KL02	-1.11336	0.40784	-2.73	0.007138
Line KL10	-3.71588	0.88483	-4.2	4.70E-05
Line KL11	-3.03958	0.89059	-3.413	0.000837
sex male	-1.18672	0.29168	-4.069	7.81E-05

S5 Table. Oxidative stress tolerance glm coefficients for Dutch population

	Estimate	Std. Error	t value	*P*-value
Intercept	3.38682	0.52533	6.447	6.86E-09
Concentration	-0.4679	0.06073	-7.705	2.36E-11
Deletion present	1.5757	0.44937	3.506	0.000732
Line NL17	-0.0708	0.43904	-0.161	0.872275
sex male	-1.15475	0.36258	-3.185	2.03E-03

S6 Table. Oxidative stress tolerance glm coefficients for *MtnA* knockdown and control lines

	Estimate	Std. Error	t value	*P*-value
Intercept	-5.55789	0.53808	-10.329	2.00E-16
Concentration	0.42612	0.04458	9.559	4.99E-15
Line	2.08414	0.33802	6.166 2	4.30E-09
sex male	1.54003	0.30396	5.067	2.41E-06

S7 Table. Male oxidative stress tolerance glm coefficients for Malaysian population

	Estimate	Std. Error	t value	*P*-value
Intercept	3.4931	0.6327	5.521	5.72E-07
Concentration	-0.3786	0.0614	-6.166	4.35E-08
Deletion present	3.0115	1.2059	2.497	0.01494
Line KL02	-1.8723	0.6116	-3.062	0.00315
Line KL10	-3.8161	1.2074	-3.161	2.35E-03
Line KL11	-3.2141	1.208	-2.661	0.00972

S8 Table. Male oxidative stress tolerance glm coefficients for Dutch population

	Estimate	Std. Error	t value	*P*-value
Intercept	2.537163	0.62857	4.036	2.31E-04
Concentration	-0.477495	0.085953	-5.555	1.85E-06
Deletion present	1.162638	0.664607	1.749	0.087711
Line NL17	0.003915	0.649791	0.006	0.995222

S9 Table. Male oxidative stress tolerance glm coefficients for *MtnA* knockdown and control lines

	Estimate	Std. Error	t value	*P*-value
Intercept	-5.1622	0.8194	-6.3	2.21E-07
Concentration	0.5323	0.0824	6.46	1.34E-07
Line	2.9362	0.6192	4.742	2.96E-05

S10 Table. Female oxidative stress tolerance glm coefficients for Malaysian population

	Estimate	Std. Error	t value	*P*-value
Intercept	3.59488	0.61971	5.801	1.82E-07
Concentration	-0.28988	0.05879	-4.931	5.42E-06
Deletion present	2.73793	1.27415	2.149	0.03516
Line KL02	-0.30477	0.56493	-0.539	0.59129
Line KL10	-3.51212	1.25831	-2.791	6.79E-03
Line KL11	-2.7662	1.27759	-2.165	0.03384

S11 Table. Female oxidative stress tolerance glm coefficients for Dutch population

	Estimate	Std. Error	t value	*P*-value
Intercept	3.00525	0.68272	4.402	7.78E-05
Concentration	-0.45054	0.08932	-5.044	1.03E-05
Deletion present	2.03235	0.62754	3.239	0.00242
Line NL17	-0.16871	0.62394	-0.27	0.78824

S12 Table. Female oxidative stress tolerance glm coefficients for *MtnA* knockdown and control lines

	Estimate	Std. Error	t value	*P*-value
Intercept	-4.56657	0.52753	-8.657	5.69E-11
Concentration	0.35188	0.04982	7.063	1.04E-08
Line	1.46324	0.38049	3.846	3.92E-04

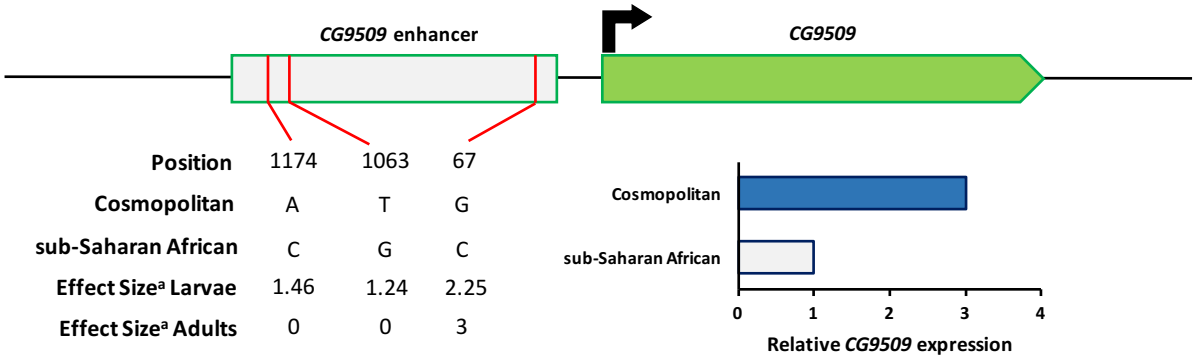
General Discussion

Elucidating specific cases of adaptive *cis*-regulatory divergence is important to our understanding of phenotypic evolution as a whole. Identifying the organismal phenotype(s) affected by *cis*-regulatory divergence can help us elucidate and understand the selection pressures a species faces as it encounters new environments or changes in its current environment. Furthermore, the identification of cases of adaptive *cis*-regulatory divergence is especially important because *cis*-regulatory changes can affect expression in a myriad of ways, including changes affecting transcription rate, transcription initiation, and/or post-transcriptional mechanisms. Thus, each individual case is unique with *cis*-regulatory mutations changing expression and affecting interactions with *trans*-acting factors in a multitude of ways. Although changes in *cis*-regulation can only be studied in-depth on an individual basis, these individual cases can provide insight into the mechanisms and prevalence of *cis*-regulatory changes as a whole.

Identification of causal mutations

In chapter 2, I present a functional analysis of a previously identified case of adaptive *cis*-regulatory divergence in the *CG9509* enhancer region. We identified the causal mutations in the *CG9509* enhancer driving the adaptive expression divergence between cosmopolitan and sub-Saharan African populations of *D. melanogaster* (Figure 13). Interestingly, three SNPs can almost fully account for the observed expression change, with

the SNP with the largest effect on expression showing signs of recent balancing selection, while the other two have a smaller effect on expression and appear to be the targets of a selective sweep. These results are consistent with the model of Sellis et al (2011), which predicts that advantageous regulatory mutations with small effects will quickly sweep to fixation, while mutations with large effects will likely overshoot the fitness optimum in the homozygous state and are therefore likely to be subject to balancing selection and remain polymorphic in the population (Figure 14). However, in order to confirm that the large effect SNP is truly a target of overdominant balancing selection, further tests are needed. One possible option is to test for a departure from Hardy-Weinberg equilibrium by searching for an excess of heterozygotes in fresh-caught flies of a natural, cosmopolitan population. If an excess of heterozygotes is observed, it would suggest that they have a higher fitness than either homozygote.



[a] Fold-change increase in expression when cosmopolitan variant is introduced into sub-Saharan African background

Figure 13: Schematic of sites contributing to *CG9509* expression divergence. Sites (red lines) in *CG9509* enhancer element (gray box) contributing to the observed *CG9509* expression divergence between cosmopolitan (blue bar) and sub-Saharan (white bar) *D. melanogaster*. Position indicates basepairs before *CG9509* (green) start codon. Effect size indicates fold-change increase in expression when a cosmopolitan variant is introduced into a sub-Saharan African background.

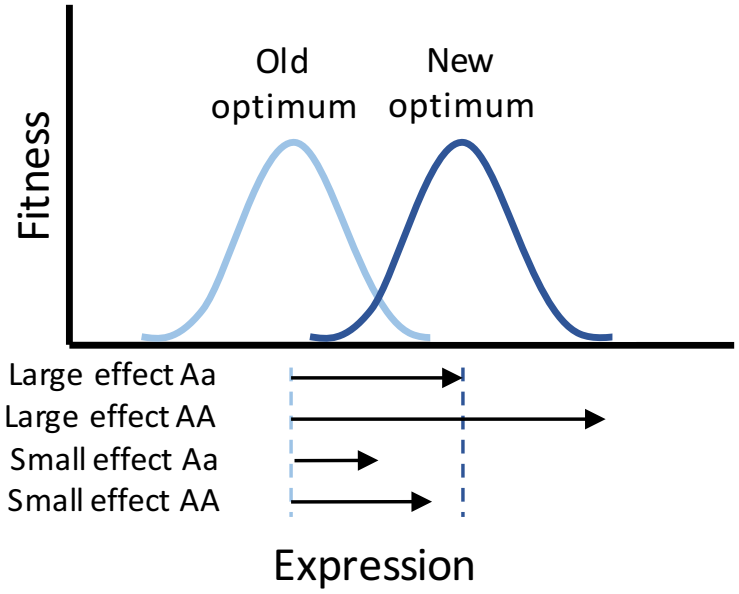


Figure 14: Model detailing effects of regulatory mutations on gene expression and fitness. The model (Sellis et al 2011) shows the differences in expression between homozygotes (AA) and heterozygotes (Aa) when the allele "A" is an advantageous, co-dominant mutation with either a large or a small effect on expression, as well as the relationship between expression and fitness. The "A" allele moves expression away from the old fitness optimum (light blue) and towards the new fitness optimum (dark blue). When "A" has a large effect, heterozygotes will meet the new the new fitness optimum but homozygotes will overshoot it, leading to overdominant balancing selection. On the other hand, when "A" has a small effect, expression will move toward the fitness optimum, but may not reach it, even in the homozygous state. This latter scenario is expected to lead to a "classical" selective sweep.

By identifying the underlying causal mutations behind an instance of adaptive *cis*-regulatory divergence (Figure 13), we were able to provide empirical evidence for a theoretical model of adaptive regulatory evolution (Figure 14), helping to contribute to our general understanding of how expression evolves. Indeed, it has often been argued that the identification of the precise genetic and molecular mechanisms, including the causal

mutations, involved in phenotypic variation is important to furthering our understanding of phenotypic evolution as a whole (Wittkopp and Kalay 2012, Hoekstra 2014). However, a few have questioned the efficacy and value of allocating the time and resources needed for such in-depth analyses, arguing that elucidating the genetic and molecular mechanisms behind individual cases of *cis*-regulatory divergence actually does little to help answer the more general questions these studies are supposed to address (Rockman 2012, Travisano and Shaw 2013).

The mechanisms through which the identified causal mutations affect *CG9509* transcript abundance remain unknown. Since the *CG9509* enhancer is located so close to the *CG9509* coding region, it is likely that these mutations affect transcription factor binding, although other causes, such as changes in mRNA processing, binding of RNA-binding proteins, and/or nucleosome positioning, cannot be ruled out. Many recent studies have focused on the evolution of transcription factor binding sites, revealing large-scale turnover of transcription factor binding sites and that most sites are subject to only weak selection (Moses et al 2006, Paris et al 2013, Arbiza et al 2013). Much attention has also been paid to the improved prediction of transcription factor binding sites (Cheng et al 2012, Yáñez-Cuna et al 2012, Mathelier and Wasserman 2013). However, even in well-studied model species, such as *D. melanogaster*, information on sequence binding motifs is not available for all transcription factors, although the availability of such information continues to grow (ENCODE Project Consortium 2012, Mathelier et al 2014). Moreover, causal mutations do not have to occur within a transcription factor binding site to disrupt binding, as has been shown in yeast where a two basepair deletion upstream of *ERG28* between the binding sites of two transcription factors disrupted binding of both transcription factors (Chang et al 2013). Further complicating matters, not all instances of a sequence binding motif will

necessarily be bound and for those that will be bound, transcription factor binding can also be very context specific (Yáñez-Cuna et al 2012), making transcription factor binding site prediction even more difficult.

The potential for adaptive *cis*-regulatory divergence

Although they did not contribute to the observed *CG9509* expression divergence, in chapter 2 we found that many of the tested sequence variants within the *CG9509* enhancer had background-, sex-, and/or stage-specific effects (Table 4), suggesting that many sequence polymorphisms may affect expression in a developmental stage-, background-, and/or sex-specific manner. Similarly, a study of the association of sequence variants with gene expression across multiple time points during *C. elegans* development revealed that many sequence variants have a context-dependent influence on gene expression, with many variants influencing gene expression only at particular time points, or sometimes even showing opposite effects on expression at different time points (Francesconi and Lehner 2014). Thus, it appears that similar to transcription factor binding, the effects of individual DNA sequence variants on gene expression can also be context-dependent. Furthermore, potential *cis*-regulatory targets on which natural selection can act are pervasive across organisms and development.

Table 4: Background-, sex-, and stage-specific effects in the *CG9509* enhancer

		Fold change in expression[a] (pos_m) for position[b]:					
Background[c]	**Stage/Sex**	**1174**	**1155**	**1063**	**821-817**	**765**	**67**
Cosmopolitan	Adult males	–	1.13	0.82	–	–	0.58
Cosmopolitan with mutation at position 67	Adult males	0.82 (1063)	NT	0.82 (1174)	NT	NT	NA
sub-Saharan African	Adult males	0.67	NT	0.82	NT	NT	3.01
sub-Saharan African with mutation at position 67	Adult males	0.80 (1063)	NT	0.80 (1174)	NT	NT	NA
Cosmopolitan	Adult females	–	1.13	–	–	–	0.56
Cosmopolitan with mutation at position 67	Adult females	0.78 (1063)	NT	0.78 (1174)	NT	NT	NA
sub-Saharan African	Adult females	0.76	NT	0.80	NT	NT	2.74
sub-Saharan African with mutation at position 67	Adult females	0.44 (1063)	NT	0.44 (1174)	NT	NT	NA
Cosmopolitan	Larvae	0.81	NT	0.80	–	–	0.34
sub-Saharan African	Larvae	1.46	NT	1.24	NT	NT	2.25

[a]Fold change in expression was calculated in comparison to background expression. Numbers greater than one indicate an increase in expression while numbers less than one indicate a decrease in expression [b]Postion in basepairs before *CG9509* start codon. [c]Background in which mutation was introduced. Cosmopolitan variants were introduced into a sub-Saharan African background and vice versa. In some cases, the effects of a mutation on expression were measured in conjunction with a mutation at another position. In these cases, the positon of the mutation, pos_m, that fold change in expression was measured in conjunction with is indicated in parentheses. NT not tested. NA not applicable. – no significant effect on expression. Expression changes that contribute to *CG9509* expression divergence are shown in red.

This observation is confirmed by studies utilizing transcriptomic technologies, the advent of which has greatly improved the rate at which scientists are able to identify *cis*-regulatory polymorphisms. For instance, one can compare expression between populations to first identify expression divergence and then screen for signs of selection in the immediate gene region to identify potential instances of adaptive *cis*-regulatory divergence (Hutter et al 2008, Müller et al 2011, Catalán et al 2012), which is how both cases of adaptive *cis*-regulatory divergence analyzed in this dissertation were identified. Alternatively, one can combine expression data with whole genome sequence data for individual isofemale strains to identify *cis*-expression quantitative trait loci (*cis*-eQTLs) across the whole genome for a given population (Massouras et al 2012). Another method, which can be used for intra- or inter-species data, is to compare the expression of parental strains as well as hybrids, using the relative abundance of allele-specific, in this case, population- or species-specific, transcripts in hybrids to infer *cis*-regulatory variants (Wittkopp et al 2008). Large datasets, such as these, have revealed the pervasiveness of individual sequence variants affecting gene expression and all of these variants represent potential targets upon which selection can act.

Identification of beneficial mutations

Adaptation does not proceed through the emergence of the perfect mutation in response to a particular selective pressure; rather, natural selection acts upon newly emerged or standing mutations within a population that can bring the population closer to the fitness optimum for a particular trait. Furthermore, the potential beneficial mutations that can arise within a population can be dependent upon the mutations that have come before them, i.e. beneficial mutations are dependent upon the evolutionary trajectory of a

population. Epistatic interactions within a particular background can modify the size of the fitness effects of potential beneficial mutations, causing both the effect and appearance rate of potential beneficial mutations to vary during adaptation (Wieinreich et al 2006, Rokyta et al 2011, Khan et al 2011, Chou et al 2011). For example, a study of adaptation to high temperature in experimentally evolved populations of *Escherichia coli* revealed extensive positive epistasis, with certain mutations only becoming beneficial once a particular set of precursor mutations had already occurred (Tenaillon et al 2012). Furthermore, once an evolutionary trajectory was set by the occurrence of a particular beneficial mutation, other mutations that in another background could have been potentially beneficial were no longer beneficial in that background (negative epistasis) and in some cases were actually deleterious in that background (Tenaillon et al 2012). Thus, whether a mutation has the potential to become beneficial is strongly dependent upon the background in which it occurs and the mutations that have come before it.

Identification of beneficial regulatory mutations can be further complicated by the pleiotropic effects of many genes, making disentangling the phenotype under selection from other phenotypes influenced by a genes expression difficult. For instance, in chapter 2, I propose that selection occurred on the *CG9509* enhancer for increased *CG9509* expression resulting in reduced wing loading outside of sub-Saharan Africa. This proposal is rather counter-intuitive for several reasons. Wing loading is usually thought to be reduced via an increase in wing area (Stalker 1980, Azevedo et al 1998, Gilchrist and Huey 2004). It is believed, and there is some evidence supporting, that reductions in wing loading are generally associated with an overall increase in body size (Starmer and Wolf 1989, Gilchrist and Huey 2004). Indeed, many studies have documented the occurrence of clines in both body size and wing loading in *Drosophila* across multiple continents, with body size

increasing and wing loading decreasing with increasing latitude, and these clines are believed to be maintained by natural selection (James and Partridge 1995, James et al 1995, Azevedo et al 1998, Gilchrist et al 2000, Robinson and Partridge 2001, de Jong and Bochdanovits 2003, Gilchrist and Huey 2004, Bhan et al 2014). However, in chapter 2 I show that increased *CG9509* expression, which was found across all of the surveyed cosmopolitan populations, is associated with a decrease in both wing area and body weight as well as a reduction in wing loading. Thus, *CG9509*'s effect on phenotype is in the opposite direction of the cline maintained by selection for body size, but in keeping with the cline for wing loading. This juxtaposition likely occurs because it is *CG9509* expressions effect on proportional body size that natural selection appears to have acted upon, leading to the observed reduced wing loading and, likely, better flight ability at cool temperatures outside of sub-Saharan Africa (Figure 15).

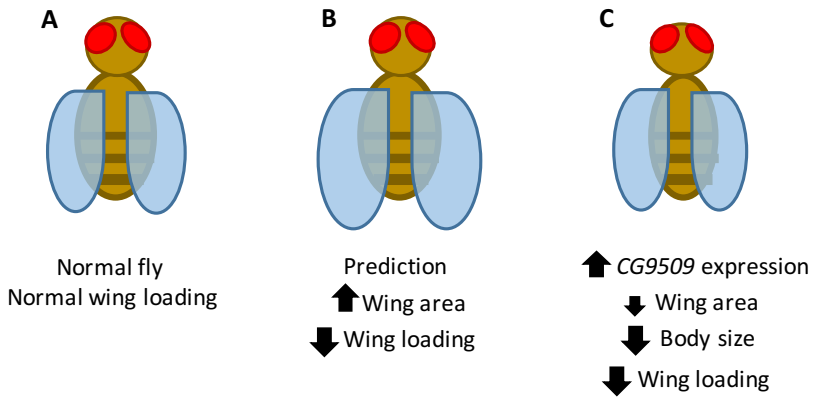


Figure 15: Schematic of *CG9509* expressions effect on wing loading. (A) Illustration of a normal fly with normal wing loading. (B) Illustration of how wing loading is usually thought to be reduced (Stalker 1980, Azevedo et al 1998, Gilchrist and Huey 2004). (C) Illustration of how increased *CG9509* expression leads to a reduction in wing loading. Arrows indicate an increase (up) or reduction (down) in a particular trait, with the size of the arrow indicating the relative size of the effect.

Environmental adaptation in *Drosophila melanogaster*

As *D. melanogaster* expanded out of Africa, it encountered many new environmental factors to which it has had to adapt, such as cooler temperatures, new food sources, new parasites and diseases, and/or new natural and xenobiotic compounds. Since the differences in climate between Europe and Africa are so remarkable, adaptation to cold temperature, in particular, has received much attention. Many studies using a variety of methods have identified multiple genes involved in cold tolerance (Morgan and Mackay 2006, Norry et al 2008, Clowers et al 2010, Svetec et al 2011). The measure of cold tolerance most studies have focused on is chill coma recovery time (CCRT) in which flies are exposed to freezing or near-freezing temperatures for a preset amount of time, during which they fall into a chill coma, losing both their mobility and ability to eat (David et al 1998). CCRT is then measured as the time until a fly has recovered. While a good measure of tolerance of near-freezing temperatures, CCRT does not measure other potential adaptations to cold such as quiescence, overwintering, or improved performance of basic functions at cold temperatures, such as metabolic processes, behavioral responses, reproductive processes, and locomotor skills.

In chapter 2, I show that selection on the *CG9509* enhancer as *D. melanogaster* moved out of Africa likely occurred for reduced wing loading and, therefore, improved flight performance at cold temperatures. For small ectotherms, such as *Drosophila*, cold can have severe consequences for flight dynamics. Indeed, flight performance and flight muscle mechanical power output increase with ambient temperature in *D. melanogaster*, so that flight performance and power output are highest around approximately 30°C; while, at 15°C, power output is just at the minimum requirements for hovering, severely limiting flight ability, and at 10°C flight ability is almost completely non-existent (Lehmann 1999). Reduced

wing loading is thought to help improve this diminished flight ability at low temperatures (Stalker 1980, Gilchrist and Huey 2004), and clinal variation in wing loading, with reduced wing loading found in higher latitude, more temperate populations, has been shown to occur across multiple continents (Azevedo et al 1998, Gilchrist et al 2000, Gilchrist and Huey 2004, Bhan et al 2014). Taken together, it appears that the need for improved flight at cooler temperatures during winter must have imposed an important selective constraint as *D. melanogaster* expanded its range out of Africa, especially as it colonized Europe, where the climate is more temperate and the winters are more severe.

In chapter 3, we present another case of environmental adaptation in *D. melanogaster*, showing that selection on the *MtnA* gene likely occurred for increased oxidative stress tolerance in a European population. Another recent example of adaptation to oxidative stress outside of Africa in *D. melanogaster* is the insertion of a transposable element into the intergenic region of *Juvenile Hormone Epoxy Hydrolase* (*Jheh*) genes, leading to the upregulation of two downstream *Jheh* genes (Guio et al 2014). This insertion was associated with increased oxidative stress tolerance (Guio et al 2014) and shows evidence for a partial selective sweep in non-African *D. melanogaster* (González et a 2009). When these two cases are considered together, this suggests that oxidative stress may have imposed an important selective constraint as *D. melanogaster* colonized Europe. Cytotoxic reactive oxygen species (ROSs) are natural byproducts of aerobic metabolism, but they can also be introduced via environmental stressors, such as radiation, UV light or exposure to toxins, resulting in oxidative stress (van Straalen and Roelofs 2012). However, since so many sources of environmental stress can introduce high levels of ROSs, it is difficult to determine the precise environmental stressor in Europe that may be imposing such strong selective constraint without further testing.

Final Remarks

This thesis focused on adaptive *cis*-regulatory divergence in two genes: *CG9509* and *MtnA*, revealing several interesting similarities as well as differences between the two cases. For instance, both appear to be cases of environmental adaption. In the case of *CG9509*, we hypothesize that cooler temperatures in the derived species range led to selection for reduced wing loading. In the case of *MtnA*, we hypothesize that adaptation occurred in response to increased oxidative stress in the derived species range, although the precise ecological factor responsible for this oxidative stress remains unknown. In the case of *CG9509*, selection occurred upon multiple SNPs within an upstream *cis*-regulatory element, while in the case of *MtnA*, selection acted upon a moderately sized (49-bp) deletion within the 3'UTR. The locations of these adaptive *cis*-regulatory polymorphisms differ greatly; thus, the mechanisms through which they affect gene expression likely differ as well. The proximity of the causal SNPs in the *CG9509* enhancer to the *CG9509* start codon suggests that they likely affect transcription initiation, or possibly transcription rate. On the other hand, the location of the deletion in the *MtnA* 3'UTR suggests that the mechanism through which it affects expression is most likely post-transcriptional, perhaps through the deletion of one or more microRNA binding sites. Furthermore, there are several distinct differences and similarities in the signatures of selection in these two cases. Selection upon the *CG9509* enhancer appears to have occurred in two phases: initial selection upon the two SNPs with smaller effects on *CG9509* expression, leaving a signature of the "classical" hard selective sweep, followed by what is likely overdominant balancing selection on the SNP with a larger effect on expression, since the SNP is maintained at intermediate frequency across multiple cosmopolitan populations. Similar to the high expression causal SNP in the *CG9509* enhancer, the deletion in the *MtnA* 3'UTR also does not show signs of the "classical" hard

selective sweep, has a large effect on expression, and is polymorphic in several of the surveyed populations, suggesting it may be under balancing selection in some of the populations. However, the selection on the *MtnA* 3'UTR deletion is unlikely to be overdominant selection, since it is found at very high frequency in the Northern European populations.

Works Cited

Angilletta MJ, Steury TD, Sears MW. Temperature, growth rate, and body size in ectotherms: fitting pieces of a life-history puzzle. *Integrative and Comparative Biology*. 2004; 44(6): 498-509.

Arbiza L, Gronau I, Aksoy BA, Hubisz MJ, Gulko B, Keinan A, et al. Genome-wide inference of natural selection on human transcription factor binding sites. *Nat Genet*. 2013; 45(7): 723-9.

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(1573): 1655-62.

Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, et al. Systems genetics of complex traits in *Drosophila melanogaster. Nat Genet*. 2009; 41(3): 299-307.

Bischof J, Maeda RK, Hediger M, Karch F, Basler K. An optimized transgenesis system for *Drosophila* using germ-line-specific ϕC31 integrases. *Proc Natl Acad Sci U S A*. 2007; 104(9): 3312-7.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. *Nature*. 2011; 478(7369): 343-8.

Broughton SJ, Piper MD, Ikeya T, Bass TM, Jacobson J, Driege Y, et al. Longer lifespan, altered metabolism, and stress resistance in *Drosophila* from ablation of cells making insulin-like ligands. *Proc Natl Acad Sci U S A*. 2005; 102(8): 3105-10.

Caldwell PE, Walkiewicz M, Stern M. Ras activity in the *Drosophila* prothoracic gland regulates body size and developmental rate via ecdysone release. *Curr Biol*. 2005; 15(20): 1785-95.

Carroll SB. Endless Forms: The evolution of gene regulation and morphological diversity. *Cell*. 2000; 101(6): 577-80.

Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*. 2008; 134(1): 25-36.

Catalán A, Glaser-Schmitt A, Argyridou E, Duchen P, Parsch J. An indel polymorphism in the *MtnA* 3' untranslated region is associated with gene expression variation and local adaptation in *Drosophila melanogaster. PLoS Genet*. 2016; 12(4): e1005987.

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

Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH, Schlotterer C. World-wide survey of an *Accord* insertion and its association with DDT resistance in *Drosophila melanogaster. Mol Ecol*. 2004; 13(8): 2491-504.

Cavalieri D, Townsend JP, Hartl DL. Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc Natl Acad Sci U S A*. 2000; 97(22): 12369-74.

Chang J, Zhou Y, Hu X, Lam L, Henry C, Green EM, et al. The molecular mechanism of a *cis*-regulatory adaptation in yeast. *PLoS Genet*. 2013; 9(9): e1003813.

Charlesworth D. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet*. 2006; 2(4): e64.

Cheng C, Alexander R, Min R, Leng J, Yip KY, Rozowsky J, et al. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Res*. 2012; 22(9): 1658-67.

Consortium CSaA. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*. 2005; 437(7055): 69-87.

Chintapalli VR, Wang J, Dow JA. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet*. 2007; 39(6): 715-20.

Chou HH, Chiu HC, Delaney NF, Segrè D, Marx CJ. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science*. 2011; 332(6034): 1190-2.

Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, et al. *Cis*-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics*. 2007; 175(3): 1071-7.

Clowers KJ, Lyman RF, Mackay TF, Morgan TJ. Genetic variation in *senescence marker protein-30* is associated with natural variation in cold tolerance in *Drosophila*. *Genet Res*. 2010; 92(2): 103-13.

Colombani J, Bianchini L, Layalle S, Pondeville E, Dauphin-Villemant C, Antoniewski C, et al. Antagonistic actions of ecdysone and insulins determine final size in *Drosophila*. *Science*. 2005; 310(5748): 667-70.

Cooper TF, Rozen DE, Lenski RE. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2003; 100(3): 1072-7.

Curtsinger JW, Laurie-Ahlberg CC. Genetic variability of flight metabolism in *Drosophila melanogaster*. I. Characterization of power output during tethered flight. *Genetics*. 1981; 98(3): 549-64.

Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, et al. A single P450 allele associated with insecticide resistance in *Drosophila*. *Science*. 2002; 297(5590): 2253-6.

David JR, Allemand R, Capy P, Chakir M, Gibert P, Pétavy G, et al. Comparative life histories and ecophysiology of *Drosophila melanogaster* and *D. simulans*. *Genetica*. 2004; 120(1-3): 151-63.

David RJ, Gibert P, Pla E, Petavy G, Karan D, Moreteau B. Cold stress tolerance in *Drosophila*: analysis of chill coma recovery in *D. melanogaster*. *Journal of Thermal Biology*. 1998; 23(5): 291-9.

Delanoue R, Slaidina M, Léopold P. The steroid hormone ecdysone controls systemic growth by repressing *dMyc* function in *Drosophila* fat cells. *Dev Cell*. 2010; 18(6): 1012-21.

Dietzl G, Chen D, Schnorrer F, Su KC, Barinova Y, Fellner M, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*. 2007; 448(7150): 151-6.

Duchen P, Zivkovic D, Hutter S, Stephan W, Laurent S. Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics*. 2013; 193(1): 291-301.

Enard W, Khaitovich P, Klose J, Zöllner S, Heissig F, Giavalisco P, et al. Intra- and interspecific variation in primate gene expression patterns. *Science*. 2002; 296(5566): 340-3.

Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.

Francesconi M, Lehner B. The effects of genetic variation on gene expression dynamics during development. *Nature*. 2014; 505(7482): 208-11.

Fraser HB, Levy S, Chavan A, Shah HB, Perez JC, Zhou Y, et al. Polygenic *cis*-regulatory adaptation in the evolution of yeast pathogenicity. *Genome Res*. 2012; 22(10): 1930-9.

Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993; 133(3): 693-709.

Gibbs AG, Matzkin LM. Evolution of water balance in the genus *Drosophila*. *J Exp Biol*. 2001; 204(Pt 13): 2331-8.

Gilchrist AS, Azevedo RBR, Partridge L, O'Higgins P. Adaptation and constraint in the evolution of *Drosophila melanogaster* wing shape. *Evolution & Development*. 2000; 2(2): 114-24.

Gilchrist AS, Partridge L. A comparison of the genetic basis of wing size divergence in three parallel body size clines of *Drosophila melanogaster*. *Genetics*. 1999; 153(4): 1775-87.

Gilchrist GW, Huey RB. Plastic and genetic variation in wing loading as a function of temperature within and among parallel clines in *Drosophila subobscura*. *Integr Comp Biol*. 2004; 44(6): 461-70.

Glaser-Schmitt A, Catalán A, Parsch J. Adaptive divergence of a transcriptional enhancer between populations of *Drosophila melanogaster*. *Philos Trans R Soc Lond B Biol Sci*. 2013; 368(1632): 20130024.

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(3): 1269-78.

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature*. 2005; 433(7025): 481-7.

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(10): e251.

González J, Macpherson JM, Petrov DA. A recent adaptive transposable element insertion near highly conserved developmental loci in *Drosophila melanogaster*. *Mol Biol Evol*. 2009; 26(9): 1949-61.

Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM. A molecular basis for classic blond hair color in Europeans. *Nat Genet*. 2014; 46(7): 748-52.

Guio L, Barrón MG, González J. The transposable element *Bari-Jheh* mediates oxidative stress response in *Drosophila*. *Mol Ecol*. 2014; 23(8): 2020-30.

Hahn MW, Rausher MD, Cunningham CW. Distinguishing between selection and population expansion in an experimental lineage of bacteriophage T7. *Genetics*. 2002; 161(1): 11-20.

Han J, Kraft P, Nan H, Guo Q, Chen C, Qureshi A, et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet*. 2008; 4(5): e1000074.

Hense W, Baines JF, Parsch J. X chromosome inactivation during *Drosophila* spermatogenesis. *PLoS Biol.* 2007; 5(10): e273.

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009; 106(23): 9362-7.

Hoekstra H. The secret of a natural blond. *Nat Genet*. 2014; 46(7): 660-1.

Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of adaptation. *Evolution*. 2007; 61(5): 995-1016.

Holloway AK, Lawniczak MK, Mezey JG, Begun DJ, Jones CD. Adaptive gene expression divergence inferred from population genomics. *PLoS Genet*. 2007; 3(10): 2007-13.

Hutter S, Saminadin-Peter SS, Stephan W, Parsch J. Gene expression variation in African and European populations of *Drosophila melanogaster*. *Genome Biol*. 2008; 9(1): R12.

Huylmans AK, Parsch J. Population- and sex-biased gene expression in the excretion organs of *Drosophila melanogaster*. *G3* (Bethesda). 2014; 4(12): 2307-15.

Iida K, Cox-Foster DL, Yang X, Ko WY, Cavener DR. Expansion and evolution of insect GMC oxidoreductases. *BMC Evol Biol*. 2007; 7: 75.

Ingram CJE, Mulcare CA, Itan Y, Thomas MG, Swallow DM. Lactose digestion and the evolutionary genetics of lactase persistence. *Human Genetics*. 2008; 124(6): 579-91.

James AC, Azevedo RB, Partridge L. Cellular basis and developmental timing in a size cline of *Drosophila melanogaster*. *Genetics*. 1995; 140(2): 659-66.

James AC, Partridge L. Thermal evolution of rate of larval development in *Drosophila melanogaster* in laboratory and field populations. *Journal of Evolutionary Biology*. 1995; 8(3): 315-30.

Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics*. 2005; 170(3): 1401-10.

Jong G, Bochdanovits Z. Latitudinal clines in *Drosophila melanogaster*: body size, allozyme frequencies, inversion frequencies, and the insulin-signalling pathway. *Journal of Genetics*. 2003; 82(3): 207-23.

Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature*. 2010; 468(7325): 811-4.

Karim FD, Thummel CS. Ecdysone coordinates the timing and amounts of *E74A* and *E74B* transcription in *Drosophila*. *Genes Dev*. 1991 ;5(6): 1067-79.

Keller A. *Drosophila melanogaste*r's history as a human commensal. *Curr Biol*. 2007; 17(3): R77-81.

Kenny EE, Timpson NJ, Sikora M, Yee MC, Moreno-Estrada A, Eng C, et al. Melanesian blond hair is caused by an amino acid change in *TYRP1*. *Science*. 2012; 336(6081): 554.

Khaitovich P, Enard W, Lachmann M, Pääbo S. Evolution of primate gene expression. *Nat Rev Genet*. 2006; 7(9): 693-702.

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, et al. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*. 2005; 309(5742): 1850-4.

Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*. 2011; 332(6034): 1193-6.

Kimura M. The neutral theory of molecular evolution. Cambridge: Cambridge University Press; 1983.

King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. *Science*. 1975; 188 (4184): 107-16.

Koshikawa S, Giorgianni MW, Vaccaro K, Kassner VA, Yoder JH, Werner T, et al. Gain of *cis*-regulatory activities underlies novel domains of *wingless* gene expression in *Drosophila*. *Proc Natl Acad Sci U S A*. 2015; 112 (24): 7524-9.

Kwasnieski JC, Mogno I, Myers CA, Corbo JC, Cohen BA. Complex effects of nucleotide variants in a mammalian *cis*-regulatory element. *Proc Natl Acad Sci U S A*. 2012; 109(47): 19498-503.

Lachaise D, Silvain JF. How two Afrotropical endemics made two cosmopolitan human commensals: The *Drosophila melanogaster-D. simulans* palaeogeographic riddle. *Genetica*. 2004; 120(1-3): 17-39.

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(7): 2041-51.

Layalle S, Arquier N, Léopold P. The TOR pathway couples nutrition and developmental timing in *Drosophila*. *Dev Cell*. 2008; 15(4): 568-77.

Lehmann FO. Ambient temperature affects free-flight performance in the fruit fly *Drosophila melanogaster*. *J Comp Physiol B*. 1999; 169(3): 165-71.

Li H, Stephan W. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet*. 2006; 2(10): e166.

Massouras A, Waszak SM, Albarca-Aguilera M, Hens K, Holcombe W, Ayroles JF, et al. Genomic variation and its impact on gene expression in *Drosophila melanogaster*. *PLoS Genet*. 2012; 8(11): e1003055.

Mateo L, Ullastres A, González J. A transposable element insertion confers xenobiotic resistance in *Drosophila*. *PLoS Genet*. 2014; 10(8): e1004560.

Mathelier A, Wasserman WW. The next generation of transcription factor binding site prediction. *PLoS Comput Biol*. 2013; 9(9): e1003214.

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2014; 42(Database issue): D142-7.

McQuilton P, St Pierre SE, Thurmond J, Consortium F. FlyBase 101--the basics of navigating FlyBase. *Nucleic Acids Res*. 2012; 40(Database issue): D706-14.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci U S A*. 2003; 100(17): 9894-9.

Mirth CK, Riddiford LM. Size assessment and growth control: how adult size is determined in insects. *Bioessays*. 2007; 29(4): 344-55.

Mirth CK, Shingleton AW. Integrating body and organ size in Drosophila: recent advances and outstanding problems. *Front Endocrinol* (Lausanne). 2012; 3:49.

Morgan TJ, Mackay TF. Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. *Heredity* (Edinb). 2006; 96(3): 232-42.

Morrison-Graham K, Takahashi Y. *Steel factor* and *c-kit receptor*: from mutants to a growth factor system. *Bioessays*. 1993; 15(2): 77-83.

Moses AM, Pollard DA, Nix DA, Iyer VN, Li XY, Biggin MD, et al. Large-scale turnover of functional transcription factor binding sites in *Drosophila*. *PLoS Comput Biol*. 2006; 2(10): e130.

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

Nègre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, et al. A *cis*-regulatory map of the *Drosophila* genome. *Nature*. 2011; 471(7339): 527-31.

Norry FM, Scannapieco AC, Sambucetti P, Bertoli CI, Loeschcke V. QTL for the thermotolerance effect of heat hardening, knockdown resistance to heat and chill-coma recovery in an intercontinental set of recombinant inbred lines of *Drosophila melanogaster*. *Mol Ecol*. 2008; 17(20): 4570-81.

Ohta T. The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematic*s. 1992; 23: 263-86.

Oleksiak MF, Churchill GA, Crawford DL. Variation in gene expression within and among natural populations*. Nat Genet*. 2002; 32(2): 261-6.

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(10): 2119-30.

Paris M, Kaplan T, Li XY, Villalta JE, Lott SE, Eisen MB. Extensive divergence of transcription factor binding in *Drosophila* embryos with highly conserved gene expression. *PLoS Genet*. 2013; 9(9): e1003748.

Pavlidis P, Metzler D, Stephan W. Selective sweeps in multilocus models of quantitative traits. *Genetics*. 2012; 192(1): 225-39.

Pfaffl MW. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Research*. 2001; 29(9): e45.

Pool JE, Corbett-Detig RB, Sugino RP, Stevens KA, Cardeno CM, Crepeau MW, et al. Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. *PLoS Genet*. 2012; 8(12): e1003080.

Preuss TM, Cáceres M, Oldham MC, Geschwind DH. Human brain evolution: insights from microarrays. *Nat Rev Genet*. 2004; 5(11): 850-60.

Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, et al. Repeated morphological evolution through *cis*-regulatory changes in a pleiotropic gene. *Nature*. 2006; 440(7087): 1050-3.

Prud'homme B, Gompel N, Carroll SB. Emerging principles of regulatory evolution. *Proc Natl Acad Sci U S A*. 2007; 104: 8605-12.

Quinn L, Lin J, Cranna N, Lee AJE, Mitchell N, Hannan R. Steroid hormones in *Drosophila*: how ecdysone coordinates developmental signalling with cell growth and division. In: Abduljabbar PH, editor. Steroids - Basic Science. online: InTech; 2012.

Team RDC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science*. 2003; 300(5626): 1742-5.

Ricardo BRA, James AC, McCabe J, Partridge L. Latitudinal variation of wing: thorax size ratio and wing-aspect ratio in *Drosophila melanogaster*. *Evolution.* 1998; 52(5): 1353-62.

Robinson SJW, Partridge L. Temperature and clinal variation in larval growth efficiency in *Drosophila melanogaster*. *Journal of Evolutionary Biology*. 2001; 14(1): 14-21.

Rockman MV. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*. 2012; 66(1): 1-17.

Rockman MV, Wray GA. Abundant raw material for *cis*-regulatory evolution in humans. *Mol Biol Evol*. 2002; 19(11): 1991-2004.

Rokyta DR, Joyce P, Caudle SB, Miller C, Beisel CJ, Wichman HA. Epistasis between beneficial mutations and the phenotype-to-fitness Map for a ssDNA virus. *PLoS Genet*. 2011; 7(6): e1002075.

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(4): 1167-74.

Santos M, Fowler K, Partridge L. Gene-environment interaction for body size and larval density in *Drosophila melanogaster*: an investigation of effects on development time, thorax length and adult sex ratio. *Heredity*. 1994; 72(5): 515-21.

Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012; 9(7): 671-5.

Schoville SD, Barreto FS, Moy GW, Wolff A, Burton RS. Investigating the molecular basis of local adaptation to thermal stress: population differences in gene expression across the transcriptome of the copepod *Tigriopus californicus*. *BMC Evol Biol*. 2012; 12: 170.

Sellis D, Callahan BJ, Petrov DA, Messer PW. Heterozygote advantage as a natural consequence of adaptation in diploids. *Proc Natl Acad Sci U S A*. 2011; 108(51): 20666-71.

Stalker HD. Chromosome studies in wild populations of *Drosophila melanogaster*. II. Relationship of inversion frequencies to latitude, season, wing-loading and flight activity. *Genetics*. 1980; 95(1): 211-23.

Starmer WT, Wolf LL. Causes of variation in wing loading among *Drosophila* species. *Biological Journal of the Linnean Society*. 1989; 37(3): 247-61.

Stephan W, Wiehe THE, Lenz MW. The effect of strongly selected substitutions on neutral polymorphism: Analytical results based on diffusion theory. *Theoretical Population Biology*. 1992; 41(2): 237-54.

Stevenson RD, Josephson RK. Effects of operating frequency and temperature on mechanical power output from moth flight muscle. *Journal of Experimental Biology*. 1990; 149(1): 61-78.

Storey JD, Madeoy J, Strout JL, Wurfel M, Ronald J, Akey JM. Gene-expression variation within and among human populations. *Am J Hum Genet*. 2007; 80(3): 502-9.

Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, et al. Genome-wide associations of gene expression variation in humans. *PLoS Genet*. 2005; 1(6): e78.

Sturm RA. Molecular genetics of human pigmentation diversity. *Hum Mol Genet*. 2009; 18(R1): R9-17.

Sucena É, Stern DL. Divergence of larval morphology between *Drosophila sechellia* and its sibling species caused by *cis*-regulatory evolution of *ovo*/*shaven-baby*. *Proc Natl Acad Sci U S A*. 2000; 97(9): 4530-4.

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, et al. Two newly identified genetic determinants of pigmentation in Europeans. *Nat Genet*. 2008; 40(7): 835-7.

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Magnusson KP, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet*. 2007; 39(12): 1443-52.

Svetec N, Werzner A, Wilches R, Pavlidis P, Alvarez-Castro JM, Broman KW, et al. Identification of X-linked quantitative trait loci affecting cold tolerance in *Drosophila melanogaster* and fine mapping by selective sweep analysis. *Mol Ecol*. 2011; 20(3): 530-44.

Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, et al. The molecular diversity of adaptive convergence. *Science*. 2012; 335(6067): 457-61.

Tennessen JM, Thummel CS. Coordinating growth and maturation - insights from *Drosophila*. *Curr Biol*. 2011; 21(18): R750-7.

Townsend JP, Cavalieri D, Hartl DL. Population genetic variation in genome-wide gene expression. *Mol Biol Evol*. 2003; 20(6): 955-63.

Travisano M, Shaw RG. Lost in the map. *Evolution*. 2013; 67(2): 305-14.

Tzika AC, Helaers R, Schramm G, Milinkovitch MC. Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent *Sauropsida* lineages and the phylogenetic position of turtles. *EvoDevo*. 2011; 2(1): 1-19.

Unwin DM, Corbet SA. Wingbeat frequency, temperature and body size in bees and flies. *Physiological Entomology*. 1984; 9(1): 115-21.

Van Straalen NM, Roelofs D. An introduction to ecological genomics: Oxford University Press; 2012.

Voolstra C, Tautz D, Farbrother P, Eichinger L, Harr B. Contrasting evolution of expression differences in the testis between species and subspecies of the house mouse. *Genome Res*. 2007; 17(1): 42-9.

Wagner GP, Lynch VJ. The gene regulatory logic of transcription factor evolution. *Trends Ecol Evol*. 2008; 23(7): 377-85.

Weinreich DM, Delaney NF, Depristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*. 2006; 312(5770): 111-4.

Whitehead A, Crawford DL. Neutral and adaptive variation in gene expression. *Proc Natl Acad Sci U S A*. 2006; 103(14): 5425-30.

Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet*. 2008; 40(3): 346-50.

Wittkopp PJ, Kalay G. *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet*. 2012; 13(1): 59-69.

Wray GA. The evolutionary significance of *cis*-regulatory mutations. *Nat Rev Genet*. 2007; 8(3): 206-16.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, et al. The Evolution of Transcriptional Regulation in Eukaryotes. *Molecular Biology and Evolution*. 2003; 20(9): 1377-419.

Yamanaka N, Rewitz KF, O'Connor MB. Ecdysone control of developmental transitions: lessons from *Drosophila* research. *Annu Rev Entomol*. 2013; 58: 497-516.

Yáñez-Cuna JO, Dinh HQ, Kvon EZ, Shlyueva D, Stark A. Uncovering *cis*-regulatory sequence requirements for context-specific transcription factor binding. *Genome Res.* 2012; 22(10): 2018-30.

Zhang M, Song F, Liang L, Nan H, Zhang J, Liu H, et al. Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. *Hum Mol Genet*. 2013; 22(14): 2948-59.

Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature*. 2007; 450(7167): 233-7.

Zheng L, Baumann U, Reymond J-L. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*. 2004; 32(14): e115-e.

Acknowledgements

I would like to thank my supervisor, John Parsch, for his excellent support, guidance, and insight throughout my PhD studies. I would also like to thank all of the technicians, Gisela Brinkmann, Gabriele Kumpfmüller and Anica Vrljic, and especially Hilde Lainer, Hedwig Gebhart, and Simone Lange, for their exceptional and tireless assistance in the lab. I also thank Ingrid Kroiß for her help in navigating paperwork and other administrative tasks. For all the interesting discussions and fun times, I'd like to thank my officemates Ina, Liza, Ana, and Annabella and I want to especially thank Liza and Annabella for all of their help when I went into the hospital. I would also like to thank the rest of the Evolutionary Biology group, all of whom it has been a pleasure to get to know. Last but not least, I would like to thank my family for their support throughout the years, and especially Markus, for his endless patience and understanding on all those evenings and weekends where "almost finished" lasted several hours.

Amanda Glaser-Schmitt

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

Language skills: English (native speaker), German (intermediate)

Education

Ludwig-Maximilians-Universität, München 2011–2016
PhD candidate, Evolutionary and Functional Genomics Group
Supervisor: Prof. Dr. John Parsch

Ludwig-Maximilians-Universität, München 2009–2011
Master's in Evolution, Ecology, and Systematics

Iona College, New Rochelle, New York 2004–2008
Bachelors of Science in Biology, Minor: English
Honors program, *summa cum laude*

Research Experience

"Gene regulatory divergence between populations of *Drosophila melanogaster*" 2011–2016
PhD Thesis, Ludwig-Maximilians-Universität, Munich, Germany
Supervisor: Prof. Dr. John Parsch
Description: Using a combination of genetic and functional approaches, I study potentially adaptive regulatory divergence in *D. melanogaster*. My goals are to two-fold: to link expression variation with genetic variation and to correlate phenotype with genetic variation.

High-throughput yeast one-hybrid system training November 2012
Skills training, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Supervisor: Prof. Bart Deplancke
Description: I trained in a high-throughput automated yeast one-hybrid platform for the protein-DNA interaction screening of *Drosophila* regulatory elements.

"The effects of host genotype on virulence in a host-parasite system:
A case study using *Beauveria bassiana* and *Drosophila*" 2010
Master Thesis, Ludwig-Maximilians-Universität, Munich, Germany
Supervisors: Prof. Dr. Wolfgang Stephan, Dr. Stephan Hutter, Dr. Aurèlien Tellier
Description: At both the population and species level, I serial passaged *B. bassiana* through one population or species and studied the changes in virulence in comparison to the ancestral parasite in all species/populations.

"Test of Local Adaptation in Host-parasite Coevolution: An example of
Beauveria bassiana and *Drosophila melanogaster* interactions" Summer Semester 2010
Research Project, Ludwig-Maximilians-Universität, Munich, Germany
Supervisors: Prof. Dr. Wolfgang Stephan, Dr. Stephan Hutter, Dr. Aurèlien Tellier
Description: I tested genotype*genotype*environment interactions in a limited framework to study possible local adaptation.

"miRNA-mediated Alcohol Dehydrogenase down-regulation by
mir-14 in *Drosophila melanogaster*" Winter Semester 2009
Research Project, Ludwig-Maximilians-Universität, Munich, Germany
Supervisor: Dr. John Parsch
Description: I studied the effect of the miRNA *mir-14* on the Alcohol Dehydrogenase content in *Drosophila melanogaster*.

"Characterization of T-DNA Mutants in *Arabidopsis thaliana* and
Determination of Glucosinolate Content" 2006–2007
Bachelor Thesis, Iona College, New Rochelle, NY
Supervisor: Dr. Yourha Kang
Description: I studied the function of the *Aop1* gene through quantitative and qualitative analysis of the glucosinolate content in mutant and wild-type *Arabidopsis thailiana*.

Publications

Catalán A, **Glaser-Schmitt A**, Argyridou E, Duchen P, Parsch J. An Indel Polymorphism in the *MtnA* 3' Untranslated Region Is Associated with Gene Expression Variation and Local Adaptation in *Drosophila melanogaster*. *PLoS Genetics* 2016; 12 (4): e1005987.

Glaser-Schmitt A, Catalán A, Parsch J. Adaptive divergence of a transcriptional enhancer between populations of *Drosophila melanogaster*. *Phil. Trans. R. Soc. B*. 2013; 368 (1632): 20130024

Work and Teaching Experience

Student Supervison 2011–2015
Ludwig-Maximilians-Universität, Munich, Germany
Description: I supervised projects for 6 Masters students and 1 Bachelor student.

Evolutionary Genetics Tutor Winter Semester 2010
Ludwig-Maximilians-Universität, Munich, Germany
Description: I ran the Evolutionary Genetics Tutorial.

Manuscript Proofreader 2010
Ludwig-Maximilians-Universität, Munich, Germany
Description: I proofread manuscripts for Dr. Volker Witte's research group.

Research assistant 2007
Iona College, New Rochelle, NY
Description: I assisted in the field collection of wild *Asclepias tuberosa* (Butterfly Weed) specimens as well as analyzed the genetic diversity of commercially available populations of butterfly weed.

Grants, Societies, and Awards

LSM Travel Grant to Society for Molecular Biology and Evolution Conference	2013
LSM Travel Grant to Lausanne, Switzerland	2012
EES Research Grant	2011
Iona College Departmental Award for Biology	2008
Delta Epsilon Sigma Honor Society	2007–2008
Beta Beta Beta (Biological Honor Society)	2007–2008
NFCA Scholar Athlete	2005–2008
Dean's List	2004–2008
Iona College Presidential Scholarship	2004–2008
National Honor Society	2002–2004

Conference and Event Presentations

Oral Presentation. EES Conference. October 2014
Ludwig-Maximilians-Universität , Munich, Germany
"*CG9509*: An evolutionary and functional analysis of gene regulatory evolution"

Oral presentation. DFG Forshergruppe 1078 Meeting. January 2014
Ludwig-Maximilians-Universität , Munich, Germany
"Identifying the mechanisms and functional consequences behind a *cis*-regulatory adaptation"

Poster Presentation. Society for Molecular Biology and Evolution Conference. July 2013
University of Chicago, Chicago, Illinois, USA
"Adaptive divergence of a transcriptional enhancer between populations of *Drosophila melanogaster*"

Oral presentation. DFG Forshergruppe 1078 Meeting. March 2013
Ludwig-Maximilians-Universität , Munich, Germany
"Linking function with adaptive *cis*-regulatory divergence"

Oral Presentation. Frontiers of Population Genetics Mini Symposium. May 2012
Insitute of Population Genetics, Vienna, Austria
"The X chromosome contributes to the difference in ethanol tolerance between European and African *Drosophila melanogaster*"

Oral presentation. DFG Forshergruppe 1078 Meeting. April 2012
Ludwig-Maximilians-Universität , Munich, Germany
"The functionality behind adaptive gene regulatory divergence"

Oral presentation. EES Conference. October 2011
Ludwig-Maximilians-Universität , Munich, Germany
"The effects of host genotype on virulence during serial passaging within a host-parasite system: A case study in *Beauveria bassiana* and *Drosophila*"

Poster Presentation. EES Conference. October 2010
Ludwig-Maximilians-Universität , Munich, Germany
"Test of Local Adaptation in Host-parasite Coevolution: An example of Beauveria Bassiana and Drosophila melanogaster interactions"