Local and Chromosome-wide Regulation of Gene Expression in *Drosophila melanogaster*

ELIZA ARGYRIDOU

München 2017

Local and Chromosome-wide Regulation of Gene Expression in *Drosophila melanogaster*

ELIZA ARGYRIDOU

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

vorgelegt von
ELIZA ARGYRIDOU
aus Strovolos, Zypern

München, den 26. Oktober 2017

1. Gutachter: Prof. Dr. John Parsch
2. Gutachter: Prof. Dr. Peter Becker
Tag der Abgabe: 26. Oktober 2017
Tag der mündlichen Prüfung: 06. März 2018

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 Versicherung

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

Eliza Argyridou

München, den 26. Oktober 2017

Declaration of Author's Contribution

In this dissertation, I present work that was realized in collaboration with other scientists during my doctoral research conducted from November 2013 to January 2017. Two publications stemmed from the results of my thesis, which comprise Chapters 1 and 3, while Chapter 2 is an unpublished manuscript.

Chapter 1

X-linkage Is Not a General Inhibitor of Tissue-Specific Gene Expression in *Drosophila melanogaster*.

Argyridou E, Huylmans AK, Königer A, Parsch J (2017). *Heredity* 119: 27-34.

John Parsch and I conceived the study and its design. I generated the ovary-specific construct, while the Malpighian tubule-specific and accessory-gland specific constructs were generated by Huylmans AK and Königer A, respectively. Huylmans AK also generated 9 of the transgenic fly lines. The rest of the experimental work was performed by me: embryo microinjections, genetic crosses (transgene location and mobilization), fly tissue dissections (testes, ovaries, accessory glands and Malpighian tubules), β-galactosidase assays, tissue staining, and inverse PCR. In addition, I maintained fly stocks and performed the data analysis. John Parsch and I wrote the manuscript.

Chapter 2

Regulation of a Ubiquitously-Expressed, X-Linked Reporter Gene in Male *Drosophila melanogaster*.

Argyridou E, Parsch J. (Unpublished Manuscript).

John Parsch and I conceived the study and its design. The reporter gene construct was generated by John Parsch. The rest of the experimental work was performed by me:

genetic crosses (transgene location and mobilization), fly tissue dissections (heads and testes), β-galactosidase assays, and inverse PCR. In addition, I maintained fly stocks and analyzed the data. I also wrote the manuscript which was revised by John Parsch.

Chapter 3

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

Catalán A, Glaser-Schmitt A, **Argyridou E**, Duchen P, Parsch J (2016). *PLoS Genet* 12: e1005987.

I collected wild *Drosophila melanogaster* from Cyprus (isofemale lines). Using those fly lines I extracted gDNA, performed PCR and determined the *MtnA* 3' UTR allele frequencies in this population. In addition, I determined the RNAi-mediated *MtnA* knockdown efficiency by carrying out RNA extractions and qRT-PCR assays. Furthermore, I performed the copper and oxidative stress tolerance assays using the RNAi-mediated knockdown flies, in collaboration with Glaser-Schmitt A. For the above, I maintained fly stocks and performed genetic crosses. I also helped with manuscript revision.

To my parents

Contents

List of Figures

List of Tables

Abstract

Within animal species, phenotypic variation is striking both between sexes and among individuals of the same sex. Much of this variation can be attributed to diverse gene expression patterns that evolved due to selection pressures specific to the external physical environment or to the internal chromosomal environment. Mechanisms that regulate expression can act on individual genes (local regulation) or on whole chromosomes (chromosome-wide). This dissertation seeks to examine the evolution of local and chromosome-wide expression regulation using the fruit fly *Drosophila melanogaster* as a model organism.

When sex chromosomes determine sex, usually their ploidy differs between the sexes. In the XY system, present in both mammals and *Drosophila*, males are hemizygous for the X chromosome. This disparity between sexes exposes the X chromosome to unique selective forces that cause it to evolve different gene content and regulatory mechanisms from the autosomes. Notably, *D. melanogaster* tissue-specific genes, with the exception of ovary-specific genes, are underrepresented on the X chromosome. Furthermore, in the male soma gene expression is equalized between the single X chromosome and the autosomes through a mechanism known as dosage compensation. In contrast, in the male germline the expression of testis-specific genes residing on the X chromosome is suppressed through a mechanism

known as X suppression. The main aim of this thesis is to investigate these aspects of gene regulation on the X chromosome in *D. melanogaster*.

Chapter 1 and part of Chapter 2 focus on the phenomenon of X suppression. In Chapter 1, using testis- and other tissue-specific constructs, I survey autosomal and X-linked reporter gene expression in whole flies and carcasses with the tissue of interest removed. First, by reanalyzing the expression of the testis-specific reporter genes for which X suppression was initially described, I confirm the occurrence of X suppression in testis. Second, I show that X suppression is not a general property of tissue-specific genes, and that the X chromosome is neither a restrictive nor an unrestrictive environment for the expression of genes expressed specifically in the accessory gland (analogous to the mammalian prostate gland), ovary, or Malpighian tubule (analogous to the mammalian kidney). Moreover, I show that X-linkage has no impact on the tissue-specificity of gene expression. These findings suggest that the observed genomic distribution of tissue-specific genes is not the consequence of a chromosome-wide regulatory mechanism. This is the first study to functionally investigate the effect of X-linkage on the expression of tissue-specific genes, other than those specific to the testis. Chapter 2 investigates further the extent of X suppression. Using a ubiquitously-expressed reporter gene with an exogenous promoter in order to exclude any sex- or tissue-specific effects, I survey the expression of X-linked and autosomal reporter genes in testes and male somatic tissues. As expected, X suppression is absent in somatic tissues, which indicates that X suppression exclusively affects testis-expressed genes. Surprisingly, I find that the exogenous reporter gene, which has a basal level of expression in testis, shows no sign of X suppression in the male germline. This demonstrates that the expression level of a gene, together with its sex- and tissue-specificity, can be a major factor that influences the extent of X suppression. Thus, the present work makes a valuable contribution to the characterization of this newly-discovered regulatory mechanism.

In Chapter 2, I also perform a pilot study regarding the effect the dosage compensation, which is mediated via the dosage compensation complex (DCC), on the chromosomal distribution of sex-biased genes in various tissues. I use X-linked insertions of the above mentioned ubiquitously-expressed reporter gene and correlate

its expression in testis and male somatic tissues (heads and carcasses) with the proximity to different DCC binding sites. I find that the expression level of the X-linked reporter genes is not correlated with their distance to a binding site of DCC components, with the exception of maleless protein (MLE), for which there was a positive correlation between expression level and MLE distance in somatic tissues. Based on my findings, I provide recommendations that will serve as a foundation for a future study of this topic.

Chapter 3 examines a candidate gene for local regulatory adaptation. The *Metallothionein A* (*MtnA*) gene exhibits expression variation in brains of natural populations of *D. melanogaster*. By collecting flies and analyzing the deletion frequency in an additional population (Cyprus), I provide further evidence that this expression variation is associated with a 49-bp deletion in the *MtnA* 3′ untranslated region (UTR), which is present at intermediate frequency in derived populations of the species. These results, supported with population genetic analysis, suggest that the deletion allele has been a target of local adaptation. By performing hydrogen peroxide tolerance assays, I show that the deletion is associated with increased oxidative stress tolerance, which suggests that the deletion (and increased *MtnA* expression) is an adaption to oxidative stress.

General Introduction

"So what hinders the different parts (of the body) from having this merely accidental relation in nature? As the teeth, for example, grow by necessity, the front ones sharp, adapted for dividing, and the grinders flat, and serviceable for masticating the food; since they were not made for the sake of this, but it was the result of accident. And in like manner as to the other parts in which there appears to exist an adaptation to an end. Wheresoever, therefore, all things together (that is all the parts of one whole) happened like as if they were made for the sake of something, these were preserved, having been appropriately constituted by an internal spontaneity, and whatsoever things were not thus constituted, perished, and still perish."

Aristotle (384-322 B.C.)

The immense collection of organisms on earth displays an enormous diversity of phenotypes and functions. Since classical times there have been frequent efforts to comprehend the forces that shape biodiversity. Eventually the evolutionary process was explained thanks to the substantial contributions from numerous philosophers and naturalists, including Aristotle, Gregor Mendel, Charles Darwin, and Theodosius Dobzhansky. Nowadays, we know that the biodiversity is derived from randomly

acquired variation within a common molecular toolkit consisting of DNA, RNA and proteins. The information contained in the genome of each organism serves as a handbook with instructions for development and response to environmental cues. This system has been established over billions of years in all forms of life. While the genomic variation is acquired randomly, it is maintained over time only if it does not contribute negatively to survival or reproduction. In other words, nature's spectacular picture that we now observe has been repainted countless times by virtue of natural selection.

Even within a species, phenotypic variation is often prominent between different populations and between sexes. Typically, this variation is not the result of differences in the coding sequences of genes, but instead is due to variation in the expression of the same genes between different populations (reviewed in Wray *et al.*, 2003) or between sexes (Ellegren and Parsch, 2007). The mechanisms of gene regulation that give rise to such expression variation could affect single genes (local regulation), or in some cases, whole chromosomes (chromosome-wide regulation).

To a large extent, eukaryotic gene expression is regulated at some stage during the transcription of DNA to messenger RNA (Latchman, 2015). The presence or absence of transcription, as well as the rate and timing of transcription, are determined chiefly by *cis*-regulatory elements; promoter (**Figure 1a**) and enhancer/silencer sequences that are targets of *trans*-acting factors such as DNA binding proteins or complexes (sequence-specific, general, or cofactors) (Biggin and Tjian, 2001; Latchman, 2015). The latter can directly interact with the transcriptional machinery and/or induce a change in chromatin structure, leading to subsequent activation (open chromatin) (**Figure 1b**) or inhibition (**Figure 1c**) (closed chromatin) of transcription. The promoter is composed of the core promoter responsible for a basal expression level and the upstream promoter elements (response elements) responsible for the rate of transcription (**Figure 1a**). Enhancer and silencer sequences can be located upstream, downstream or within the gene and can induce increased and suppressed transcription, respectively. These sequences often have tissue-specific activity (Latchman, 2015).

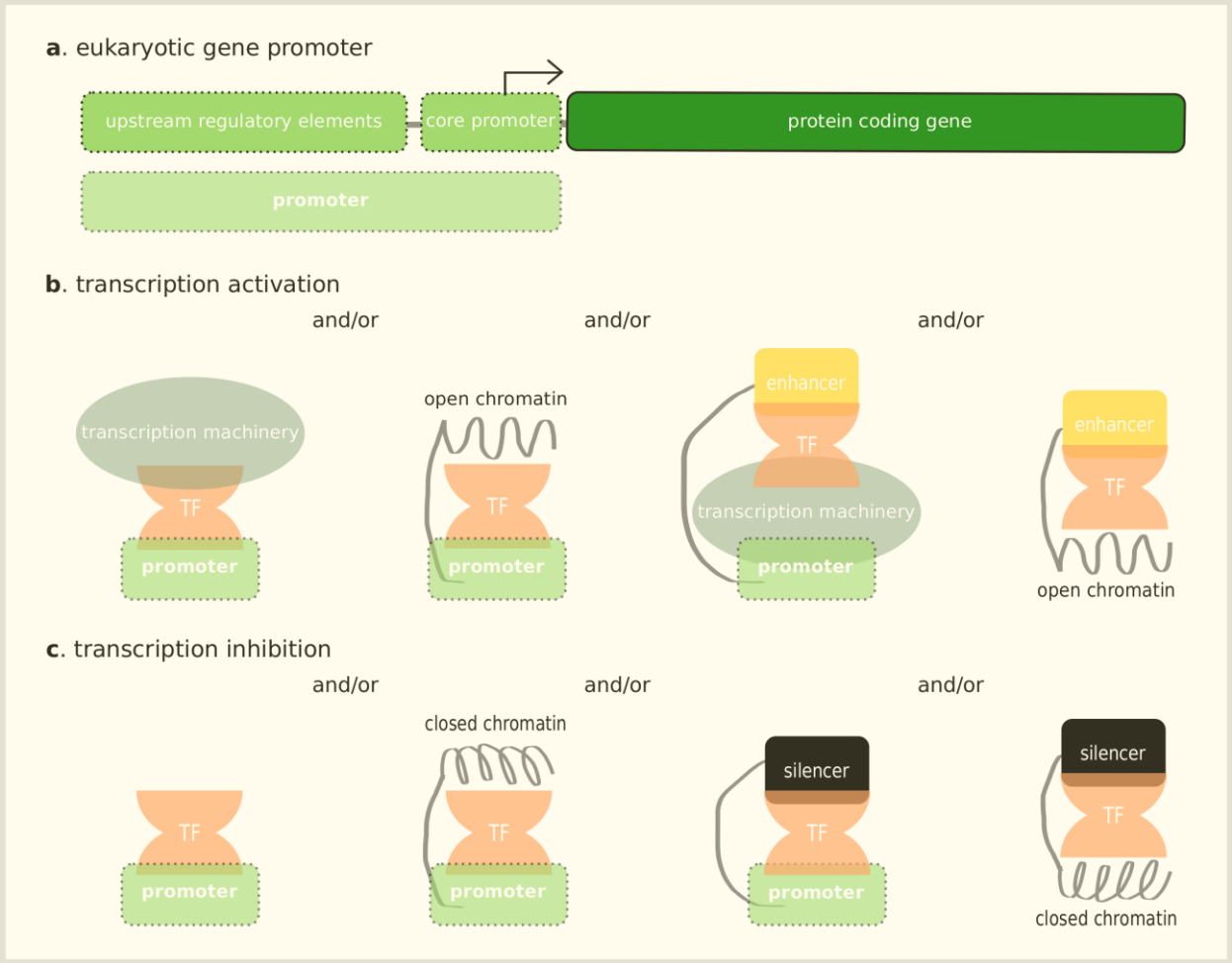


Figure 1. Eukaryotic transcription regulation. a. The eukaryotic gene promoter consists of the core promoter that contains the transcription start site (black arrow), and the upstream regulatory elements. **b**. Transcriptional activation can be induced by the interaction of *cis*-regulatory elements (promoter, enhancer) and *trans*-acting factors (TF) with the transcription machinery and/or the direct interaction between TF and chromatin structure (open chromatin) **c**. Transcriptional inhibition can be induced by the interaction of *cis*-regulatory elements (promoter, silencer) and TF and/or the direct interaction between TF and chromatin structure (closed chromatin).

In light of the big impact of gene expression variation on phenotype, it is important to pinpoint the evolutionary pressures shaping genome regulation. This dissertation is accordingly devoted to the investigation of different types of gene expression regulation using the fruit fly *Drosophila melanogaster* as a model organism.

Drosophila melanogaster as a model organism

The fruit fly has proven to be an excellent tool for studying the eukaryotic genome for more than a hundred years. The use of *D. melanogaster*, a simple invertebrate with a short life cycle, small size, and easy rearing conditions, enabled the research community to gain precious knowledge regarding fundamental genetic, developmental and metabolic processes that have been evolutionarily conserved across much more complex animals, including vertebrates. Indeed, many of the genes and genetic mechanisms discovered in the fly share extensive similarities to those of other animals, including humans (Reiter *et al.*, 2001). For example, sex in the fruit fly is genetically determined by the combination of sex chromosomes, which is the case for the majority of animals (Manolakou *et al.*, 2006) and for some plants (Charlesworth, 2002). More specifically, *Drosophila* has the XY sex determination system, which is also present in mammals. In this system, females bear two copies of the X chromosome, while males bear a single copy of the X and a single copy of the Y chromosome (**Figure 2**).

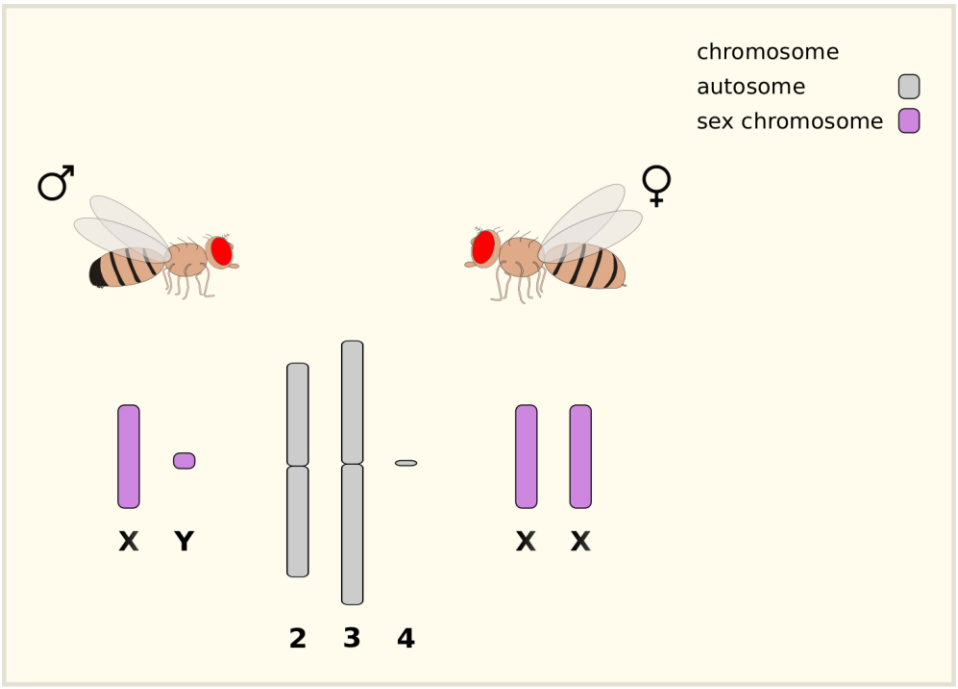


Figure 2. *Drosophila melanogaster* karyotype. There are three pairs of autosomes (2, 3, 4) shared by both sexes and a pair of sex chromosomes distinct between sexes. Males have a single copy of the X and a single copy of the Y chromosome (XY), whereas females have two copies of the X chromosome (XX).

Drosophila melanogaster has also become a model organism for studying adaptation, partly because of its demographic history (Pool *et al.*, 2012). The species originated in sub-Saharan Africa and expanded approximately 15,000–17,000 years ago to Asia and Europe (Li and Stephan, 2006), eventually colonizing all continents except Antarctica. During this expansion, *D. melanogaster* encountered a broad spectrum of habitats. As a cosmopolitan species that now lives in association with humans, *D. melanogaster* faces an assortment of selection pressures within its derived species range that are often profoundly different from those in the ancestral one. Such selection pressures give rise to environmental adaptations (local adaptation), which are characteristics that enhance survival within a certain habitat. Thus, local adaptation can give rise to phenotypic variation between populations.

Local adaptation often stems from differences in gene expression regulation, which is characterised by the differential expression of individual genes between populations. This expression divergence typically results from differences in allele frequencies between populations. A characteristic example is insecticide (DDT) resistance in *D. melanogaster*, a trait attributed to the upregulation of the cytochrome P450 gene *Cyp6g1* (Daborn, 2002). The presence of an *Accord* transposon insertion, including a *cis*-regulatory element (tissue-specific enhancer) in its long terminal repeat, upstream of *Cyp6g1* drives increased expression of *Cyp6g1* in tissues responsible for detoxification, such as the midgut, Malpighian tubules, and the fat body (Chung *et al.*, 2007). The frequency of the *Accord*-insertion allele leading to insecticide resistance varies in natural populations, with high frequencies found in derived (non-African) populations and lower frequencies found in ancestral (African) populations (Catania *et al.*, 2004).

Sex chromosome evolution

Sex chromosomes consist of a pair of heteromorphic chromosomes (XY or ZW) and have evolved many times independently in different taxa (Kaiser and Bachtrog, 2010). They are derived from a pair of homologous autosomes that differentiated over evolutionary time through a dynamic process (reviewed in Wright *et al.*, 2016). Crucial stages of this process include the loss of recombination between the initially homologous pair of chromosomes and the genetic degeneration of the sex-specific chromosome (Y or W), which is present in the heterogametic sex. These processes lead to a reduction of the gene content and size of the sex-specific chromosome. In the end, the heterogametic sex is hemizygous for all genes located on the non-sex-specific sex chromosome. For instance, the Y chromosomes of both fruit flies and humans bear very few genes. The genes that remain Y-linked usually are related to male fertility, while most of the chromosome contains heterochromatic repetitive elements (reviewed in Bachtrog, 2013).

The X is a special chromosome

In contrast to the Y, the X chromosome is practically indistinguishable from the autosomes in terms of its gene density and cytological appearance. Nevertheless, there are differences between the X chromosome and the autosomes regarding sequence divergence, gene content, and patterns of gene expression (Vicoso and Charlesworth, 2006).

Gene content

On average, *Drosophila* genes residing on the X chromosome are more divergent between species with respect to their protein coding sequence (measured as the ratio of non-synonymous to synonymous substitution) than autosomal genes. This phenomenon is known as the 'faster X effect' and is thought to be driven by the selection of beneficial, recessive mutations on the X chromosome (Meisel and Connallon, 2013). Furthermore, the X chromosome is a hotspot for hybrid male

sterility factors (Presgraves, 2008). This explains why the X chromosome has greater influence on hybrid sterility and inviability than the autosomes in introgression analyses, a phenomenon known as the 'large-X effect'. Therefore, the X chromosome makes a greater contribution to the speciation process than an autosome, as the heterogametic hybrid (XY) has a fitness disadvantage. What is more, male-biased genes, that is genes that are expressed at a higher level in males than in females, on the X chromosome have been shown to evolve more rapidly than autosomal ones in *Drosophila* species (Baines *et al.*, 2008; Llopart, 2012). Thus, it is expected that male-biased genes, which tend to have a large impact on male fitness, show the strongest 'faster X effect', since the effects of recessive mutations can be immediately exposed to selection in males (Meisel *et al.*, 2012a).

Sex-biased genes exhibit a non-random genomic distribution. In whole flies and in reproductive tissues, male-biased genes are underrepresented on the X chromosome, whereas female-biased genes are overrepresented (Parisi, 2003; Sturgill *et al.*, 2007). Nevertheless, in head and brain, there is an overrepresentation of both male- and female-biased genes on the X chromosome relative to the autosomes (Chang *et al.*, 2011; Catalán *et al.*, 2012; Huylmans and Parsch, 2015). This feminization and demasculinization of the *Drosophila* X chromosome could be indicative of the X being an unfavorable environment for male-biased genes (Parsch, 2009; Gallach *et al.*, 2011). Consistent with this is the high duplication rate of newly retroposed gene copies with male-biased expression from the X chromosome to the autosomes that has been reported in the *Drosophila* genus and suggests an evolutionary pressure for such genes to 'escape' the X chromosome (Betrán *et al.*, 2002; Vibranovski *et al.*, 2009a).

Tissue-specific genes also exhibit a non-random genomic distribution in the *D. melanogaster* genome. As a null hypothesis, it is expected that each class of tissue-specific genes should be evenly allocated across the genome, including the autosomes and the X chromosome (**Figure 3a**). Interestingly, genes expressed in male-specific tissues such as testis and accessory gland (a somatic tissue analogous to the mammalian prostate gland), as well as genes that are expressed in somatic tissues of both sexes such as the Malpighian tubule (a somatic tissue that is analogous to the

mammalian kidney), are underrepresented on the X chromosome (Mikhaylova and Nurminsky, 2011; Meisel *et al.*, 2012b) (**Figure 3b**). Ovary-specific genes, which are overrepresented on the X chromosome, are an exception to this pattern (**Figure 3b**). Based on these observations, it has been postulated that the X chromosome might be an unfavorable environment for most classes of tissue-specific genes if their specialized expression regulation (activation or repression) is hindered due to a chromosome-wide mechanism (Mikhaylova and Nurminsky, 2011). To date, the selection pressures leading to the observed genomic distributions of tissue-specific genes remain obscure.

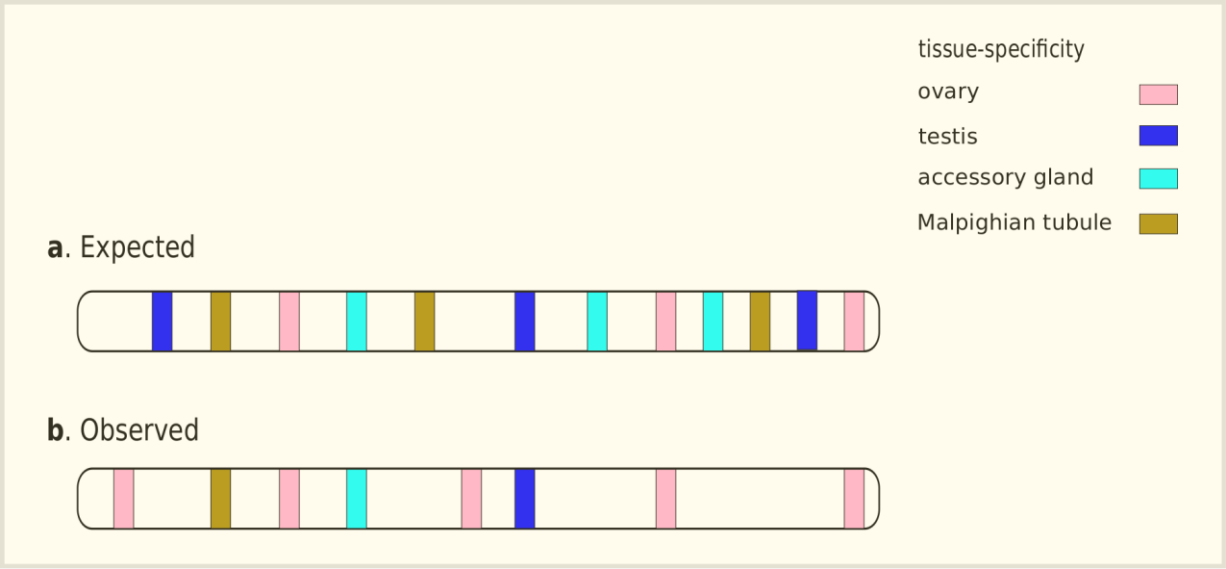


Figure 3. Distribution of tissue-specific genes on the X chromosome. a. If tissue-specific genes are randomly distributed in the genome, then an even representation on the X chromosome and the autosomes is expected. **b**. The observed distribution of tissue-specific genes differs from the random expectation, with ovary-specific genes being overrepresented and testis-, accessory gland- and Malpighian tubule-specific genes being underrepresented.

Gene expression regulation

The evolutionary uniqueness of the X chromosome in *Drosophila* is exemplified by two chromosome-wide mechanisms of gene expression regulation that are present in heterogametic males. Through the first mechanism, dosage compensation (DC), the expression of X-linked genes in somatic tissues is upregulated in males in order to compensate for the lack of a second copy of the X chromosome (**Figure 4a**). Through

the second mechanism, X suppression, the expression of X-linked genes is suppressed in the male germline (**Figure 4b**). The latter mechanism was first demonstrated by comparing the expression of testis-specific reporter genes located on the X chromosome to those located on the autosomes (Hense *et al.*, 2007).

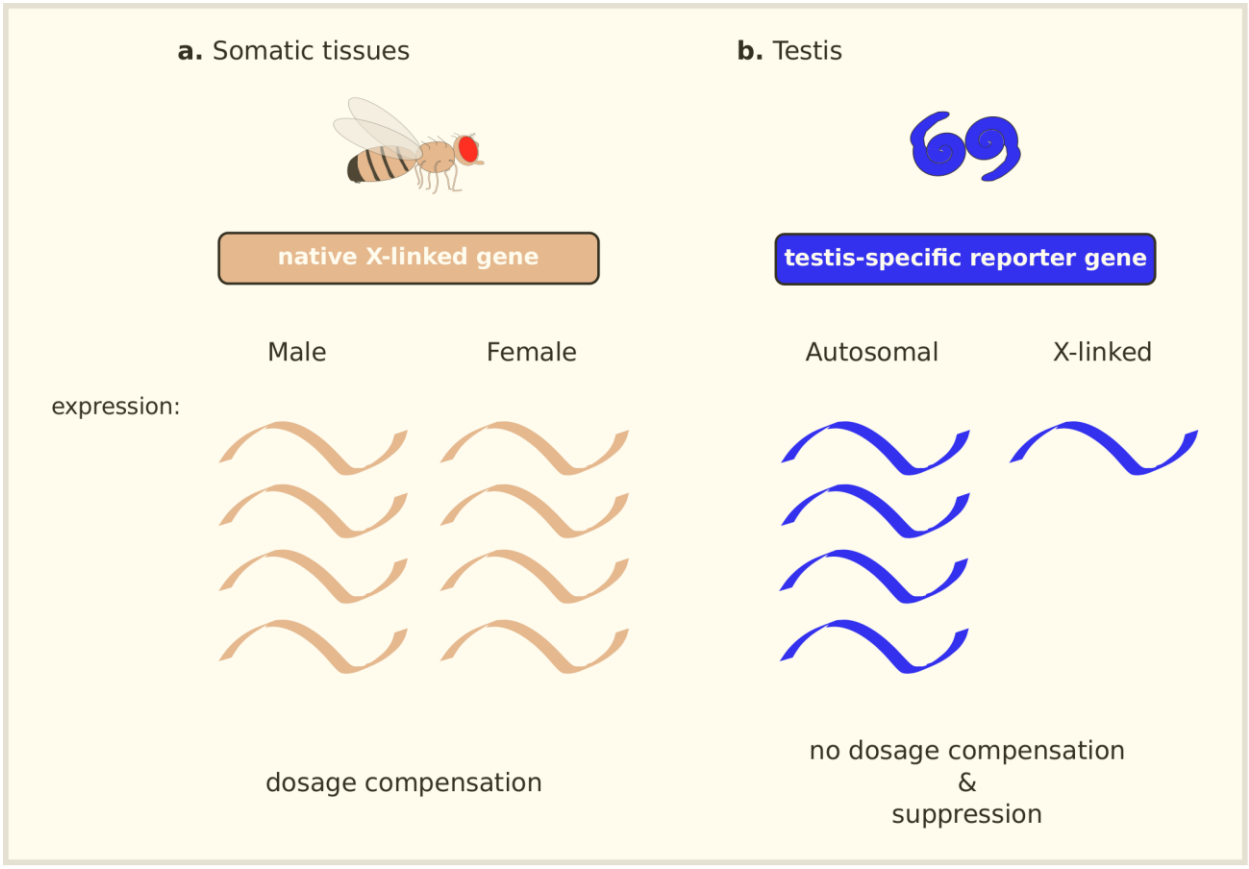


Figure 4. Regulation of the male X chromosome in *D. melanogaster*. a. In somatic tissues, the expression of native X-linked genes is equalized between males (one copy of the X chromosome) and females (two copies of the X chromosome) by DCC-mediated dosage compensation in males that leads to approximately two-fold hypertranscription. **b**. In testis, DCC-mediated dosage compensation is absent and expression of X-linked testis-specific reporter genes is suppressed in comparison to autosomal reporter genes.

Dosage Compensation

The equalization of gene expression between X-linked and autosomal genes, and also between males and females, is designated dosage compensation. For genes encoding RNAs and proteins that are involved in processes requiring interaction with multiple constituents, a balanced stoichiometry is crucial. DC is accomplished through diverse strategies that have evolved in different organisms with sex chromosomes (Julien *et al.*, 2012; Ercan, 2015). In mammals, one of the female X chromosomes is inactivated in the somatic cells during embryogenesis via the formation of the Barr body, while the remaining female and the single male X chromosomes are upregulated (Brockdorff and Turner, 2015).

In male fruit flies, the expression of the single X chromosome is upregulated roughly two-fold. This chromosome-wide upregulation is the consequence of a combination of basal dosage compensation that contributes ~1.5-fold increased expression and X chromosome-specific DC that contributes ~1.35-fold increased expression (Zhang *et al.*, 2010). The X chromosome-specific DC is mediated by a ribonucleoprotein complex, the dosage compensation complex (DCC), also known as the male-specific lethal (MSL) complex, through the targeting of binding sites that are enriched on the X chromosome (**Figure 5a**).

The DCC consists of one long non-coding RNA (rox1 or rox2) with a structural role, and five proteins: MSL1 (scaffolding protein), MSL2 (RING finger protein), MSL3 (chromodomain protein), MLE (RNA helicase), and MOF (acetyltransferase) (reviewed in Conrad and Akhtar, 2012). MSL2, which has ligase activity, recognizes and binds to the high affinity sites (HAS) that are typically localized within or in close proximity to active gene bodies. After initial binding, MSL1, MSL2 and MLE jointly induce the complex assembly. MOF catalyzes local histone acetylation (H4K16ac), resulting in an opened chromatin structure and consequent hypertranscription of the exposed region through increased transcriptional elongation from RNA Pol II (Ferrari *et al.*, 2014). MSL3 is thought to be involved in the spreading of the complex. The current model of the DCC spreading along the X chromosome includes initial binding of the DCC to HAS, and spreading to spatially close sites with lower affinity (Ramirez *et al.*, 2015) (**Figure 5b**). Although DCC-mediated DC exists in the male soma, this is not the case

for the male germline. The absence of the mechanism was revealed by microarray and RNA-sequencing data in which expression of X-linked and autosomal genes in premeiotic cells was measured and found to be similar to expression of those genes in cell lines where *msl2* was knocked down by RNAi (Meiklejohn *et al.*, 2011).

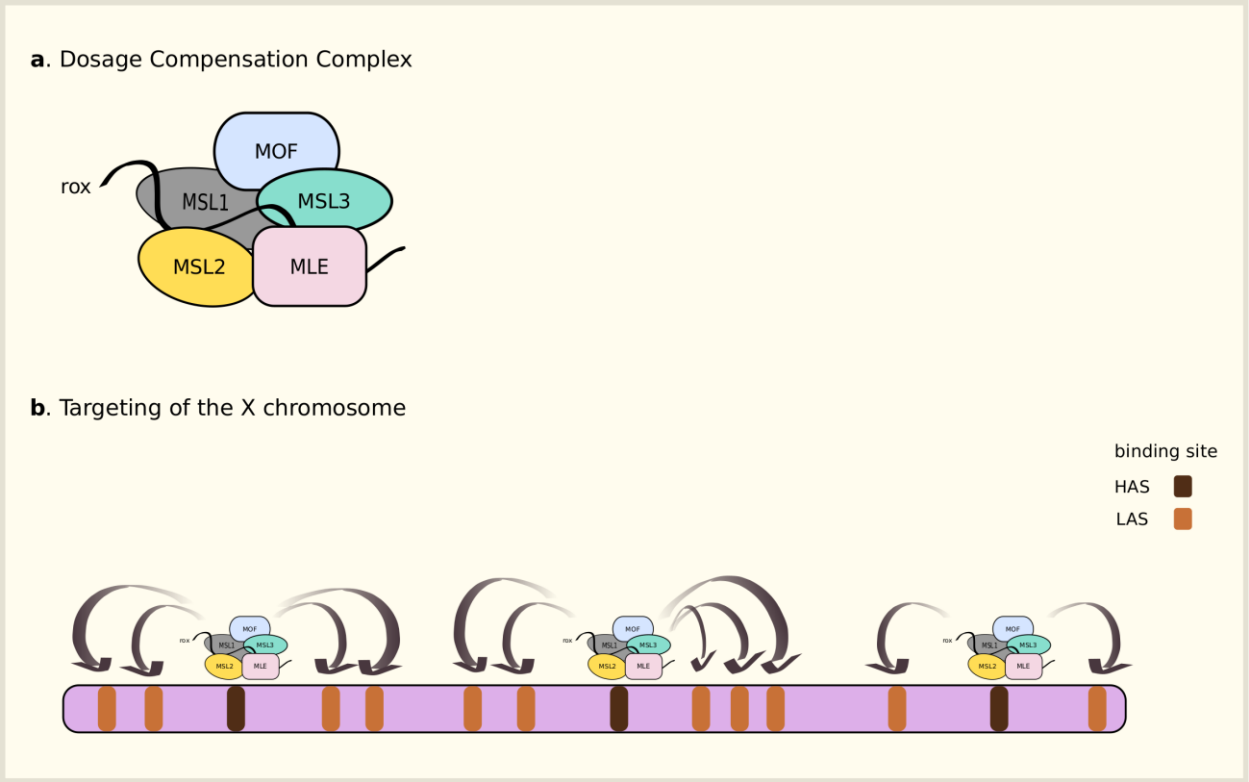


Figure 5. Dosage compensation in the male soma of *D. melanogaster*. a. The dosage compensation complex (DCC) mediates the two-fold upregulation of the X chromosome. The DCC consists of a lncRNA, rox1 or rox2 (rox), and the proteins MSL1, MSL2, MSL3, MLE, MOF. The upregulation is achieved through the activity of MOF, which catalyzes a local chromatin modification and results in an opened chromatin structure and therefore enabling increased transcription. **b**. The DCC targets the X chromosome through initial binding to high affinity sites (HAS) and most likely spreading in *cis* to lower affinity sites (LAS).

It is probable that the DCC-mediated DC has an effect on the distribution of sex-biased genes on the X chromosome, as significant correlations between a gene's male-to-female expression ratio and its distance to the nearest DCC binding site have been detected (Huylmans and Parsch, 2015) (see **Figure 6**). In head and brain, there is a highly significant negative correlation between a gene's male-to-female expression ratio and its distance to the nearest DCC binding site. This suggests that the excess of

X-linked male-biased genes in these tissues is possibly a result of "overcompensation" of genes located close to DCC binding sites. In gonads and whole fly, there is a highly significant positive correlation between a gene's male-to-female expression ratio and distance to the nearest DCC binding site, although it is not clear why this should be the case if dosage compensation does not occur in the male germline (Meiklejohn *et al.*, 2011). The latter correlation is present, but weaker, for other somatic tissues, including gonadectomized flies (containing all somatic tissues) and Malpighian tubules.

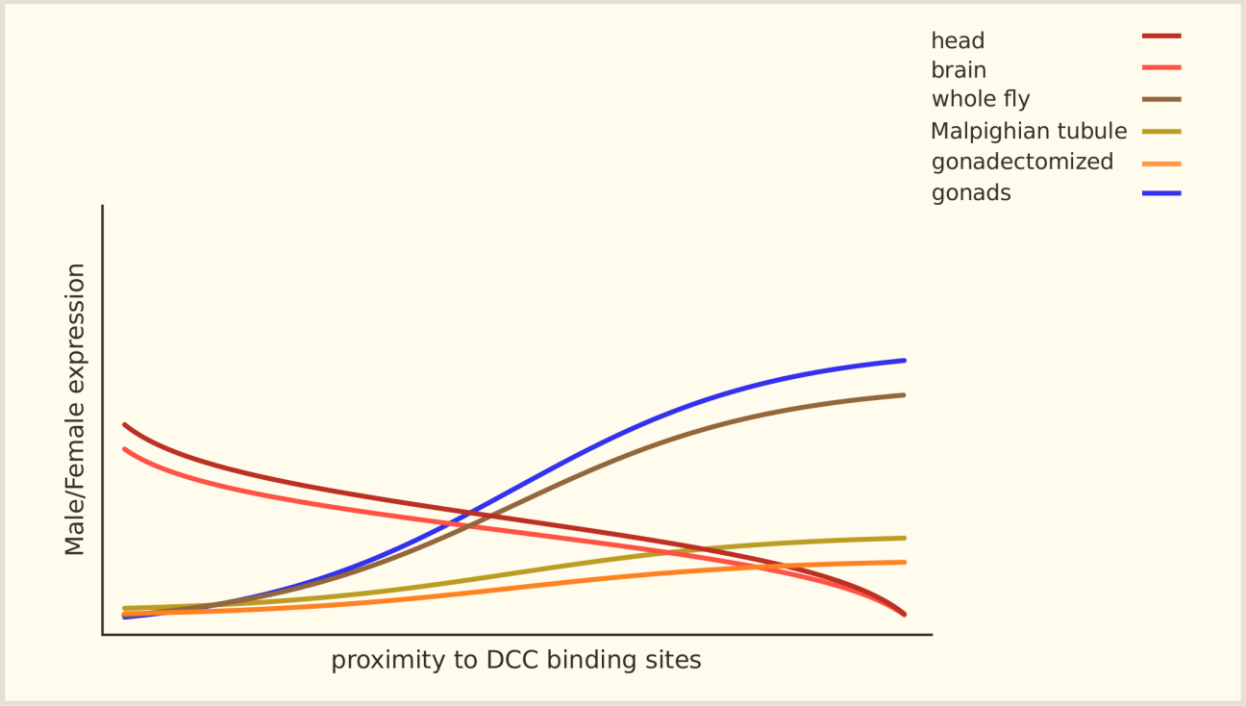


Figure 6. Male-biased gene expression and proximity to DCC binding sites in different tissues/body segments of *Drosophila melanogaster*. Data described by Huylmans & Parsch (2015), are illustrated here as monotonic curves. In head (dark red) and brain (light red), the degree of male-bias is negatively correlated with the proximity to DCC binding sites and, a gene reaches higher expression in males than in females when it is located close to a DCC binding site. This relationship is strongest in head. In whole flies (brown), Malpighian tubules (yellow), gonadectomized flies (orange) and gonads (blue), the degree of male-bias is positively correlated with the proximity to DCC binding sites and, a gene reaches higher expression in males than in females when it is located further away from a DCC binding site. This relationship is strongest in gonads.

Suppression of X-linked genes in the male germline

In contrast to the X chromosomal DC that occurs in the male soma, there appear to be opposing selective pressures in heterogametic males that lead to the suppression, or even silencing, of the X chromosome in the male germline. In mammals, this silencing is known as meiotic sex chromosome inactivation (MSCI). MSCI occurs through the formation of the sex-body, that is, compartmentalization of the sex chromosomes in the periphery of the nucleus, during the pachytene stage of meiosis I (reviewed in Turner, 2007). It is thought that this mechanism ensures prevention of erroneous recombination between the heterologous X and Y chromosomes (Lifschytz and Lindsley, 1972; McKee and Handel, 1993).

In *D. melanogaster*, no obvious suppression of native X-linked genes in testes can be detected. The median autosomal expression in testis is 1.44-fold greater than that of the X chromosome (Gan *et al.*, 2010), but this is likely due to the absence of DCC-mediated DC in the male germline (Meiklejohn *et al.*, 2011). Nonetheless, a number of studies have found evidence of X chromosomal suppression of testis-specific reporter genes. The expression of testis-specific reporter genes driven by four different testis-specific promoters is significantly reduced (3- to 7-fold) when they are located on the X chromosome compared to the autosomes (Hense *et al.*, 2007; Kemkemer *et al.*, 2014). The suppression of X-linked reporter genes, also known as 'X suppression' (Landeen *et al.*, 2016), is not influenced by the position of the transgenes, as a study with one of the testis-specific reporter genes analyzed 107 independent X-linked insertions and showed that all regions of the X chromosome follow the same pattern (Kemkemer *et al.*, 2011). In addition, it has been demonstrated recently that the above observations could not be an artifact of the use of reporter genes in transposable element vectors, as when regions of the X chromosome are transposed to the autosomes, native testis-specific genes also exhibit higher expression levels (Landeen *et al.*, 2016). This was also the case for housekeeping genes in testis, but not in whole males/females or ovaries (Landeen *et al.*, 2016). The same authors tested reporter genes with promoters of two different housekeeping genes in testis and there was also increased expression of autosomal reporter genes compared to the X-linked ones. Since the aforementioned findings consistently compared expression of genes

present in only one copy (reporter or transposed) on the autosomes and X chromosome, the observed X suppression could not be attributed to gene dose or to the absence of dosage compensation in the male germline.

Although a growing body of literature has investigated the phenomenon of X suppression affecting testis- and broadly-expressed genes in the male germline, it is still unknown whether it takes place through a mechanism comparable to mammalian MSCI. In fact, the mechanism of X suppression has been the subject of debate as to whether the timing of X suppression and MSCI coincide (Meiklejohn *et al.*, 2011; Mikhaylova and Nurminsky, 2012; Vibranovski *et al.*, 2012; Vibranovski, 2014). What is more, a microarray-based study found that not only genes expressed specifically in the male germline, but also other tissue-specific genes, are underrepresented on the X chromosome. This suggests that reduced X-linked expression may be common to all tissue-specific genes (Mikhaylova and Nurminsky, 2011). If this is true, then X suppression is not limited to the male germline and must be caused by a mechanism other than MSCI.

Motivation and Objectives

In this dissertation, I examine chromosome-wide and local gene expression regulation. Selection can act on entire chromosomes, such as the X chromosome, and therefore lead to the evolution of chromosome-wide regulatory mechanisms, such as X suppression in the male germline and dosage compensation in the male soma. Alternatively, selection can act on individual genes, and therefore lead to the evolution of a specific adaptive trait (e.g. population-specific environmental adaptations). Understanding the fundamental mechanisms that underlie gene expression regulation, as well as the evolutionary forces responsible for their maintenance in the *Drosophila* genome, and potentially in the genomes of other organisms, is important for addressing a plethora of biological questions. For example, a myriad of common human diseases, including leukemia, diabetes, and Alzheimer's arise through deviations from normal gene expression regulation mechanisms (Theuns and Van

Broeckhoven, 2000; Lee and Young, 2014). If we are able to pinpoint the defective mechanisms, we could design treatment plans more efficiently.

Most of this dissertation focuses on mechanisms of gene expression regulation on the X chromosome of *D. melanogaster.* In the first two chapters, the focus is on either tissue-specific (Chapter 1) or ubiquitously-expressed (Chapter 2) reporter genes. More specifically, each transgenic construct consisted of a *D. melanogaster* tissue-specific promoter (Chapter 1) or the human cytomegalovirous (CMV) promoter (Chapter 2), linked to the coding sequence of the *Escherichia coli lacZ* reporter gene. Each construct was inserted randomly into the *D. melanogaster* genome by *P*-element transposition. With this approach I was able to quantify the expression of the reporter gene when located at different chromosomal positions. Consequently, I could compare the expression of autosomal and X-linked reporter genes in various tissues.

In Chapter 1, I address the open question concerning the extent of the X-linked gene expression suppression that was first observed in the male germline. A major aim of my thesis was to investigate the occurrence of X suppression for various tissue-specific genes, since previous studies have focused exclusively on testis-expressed genes. For this reason, it was not clear whether X suppression was restricted to the male germline, if it occurred in other male-limited tissues, or if it also occurred in the female germline. It was also possible that X suppression was a more general phenomenon that occurred for tissue-specific genes that are expressed in somatic tissues common to both sexes. To test these possibilities, I employ reporter genes expressed specifically in accessory gland, ovary or Malpighian tubule. My research fills a gap in the literature by showing that (i) X suppression is a distinct property of testis-expressed genes and not of tissue-specific genes in general, (ii) the X chromosome is not a restrictive or a favorable environment for the expression of genes expressed in the male-limited accessory gland, the female germline or the somatic Malpighian tubule, and (iii) the underrepresentation of tissue-specific genes on the X chromosome, with the exception of the ovary-specific genes, is not due to the existence of a chromosome-wide mechanism.

In Chapter 2, I further examine regulatory mechanisms acting on the *D. melanogaster* X chromosome. This is the first study where a reporter gene with a

regulatory sequence that is exogenous to *D. melanogaster* is used in order to exclude sex- and tissue- specific effects on X suppression. In the first part of the Chapter, I set out to determine whether X suppression affects a ubiquitously-expressed gene in testis or male somatic tissues. If X suppression is a chromosome-wide mechanism, then this reporter gene also should be affected. As expected, I find no evidence of X suppression in somatic tissues, although I did find evidence of (partial) X chromosome dosage compensation. Contrary to the expectation, I find that X suppression is absent in testis for this reporter gene, which has a basal expression level. It is thus possible that X suppression affects only genes with an expression level above a certain threshold. This knowledge offers us an insight into the nature of the mechanism of X suppression. In the second part of the Chapter, I conduct an exploratory study concerning the potential influence of DCC-mediated DC on male gene expression in different tissues by analyzing the relationship between a gene's expression and its proximity to a DCC binding site. I find a positive correlation between X-linked reporter gene expression in head and carcass, and proximity to one of the binding sites, MLE. Also, based on my results I put forward recommendations for a future study.

In Chapter 3, I contribute to a study of the regulation of the *Metallothionein A* gene, *MtnA*, which exhibits variation in expression within and between natural populations of *D. melanogaster*. The expression of *MtnA* is much higher in derived, non-African populations than in ancestral sub-Saharan African populations. The expression difference is associated with an insertion/deletion polymorphism in the *MtnA* 3' untranslated region (UTR). At the phenotypic level, I show that higher *MtnA* expression is associated with greater oxidative stress tolerance, suggesting that the deletion form of the variant has been the target of local regulatory adaptation.

CHAPTER 1

X-Linkage Is Not a General Inhibitor of Tissue-Specific Gene Expression in *Drosophila melanogaster*

Eliza Argyridou, Ann Kathrin Huylmans, Annabella Königer and John Parsch

Heredity, 2017, **119:** 27–34
doi:10.1038/hdy.2017.12

CHAPTER 2

Regulation of a Ubiquitously-Expressed, X-Linked Reporter Gene in Male *Drosophila melanogaster*

Eliza Argyridou and John Parsch

(Unpublished manuscript)

Regulation of a Ubiquitously-Expressed, X-Linked Reporter Gene in Male *Drosophila melanogaster*

Eliza Argyridou and John Parsch

(Unpublished manuscript)

Abstract

In the XY sex determination system, males are hemizygous for the X chromosome. This exposes the X chromosome to unique selective forces that cause it to evolve differently from the autosomes. In *D. melanogaster*, the uniqueness of the X chromosome is reflected by its sex-biased gene content and its special mechanisms of gene expression regulation, such as dosage compensation and suppression of X-linked expression in the male germline. Here we test if the expression of a ubiquitously-expressed, exogenous reporter gene is suppressed in the male germline when it is located on the X chromosome. In addition, we explore the relationship between the reporter gene's expression level in males and its proximity to binding sites of the dosage compensation complex (DCC), which is responsible for X chromosome dosage compensation. In contrast to previous studies, we did not detect evidence of X chromosomal suppression of gene expression in the male germline. These results suggest that X suppression principally affects genes with high levels of expression in testis, such as testis-specific genes or highly-expressed housekeeping genes, but has less influence on genes with low levels of expression in testis. In general, the expression level of the X-linked reporter genes was not correlated with their distance to a binding site of a DCC component, with the exception of MLE, for which there was a positive correlation between expression level and DCC distance in somatic tissues. Our findings suggest that the regulation of X-linked gene expression in males depends on a complex interaction among expression level, tissue-specificity, and dosage compensation.

Introduction

In most animals, sex is genetically determined by the combination of sex chromosomes. In organisms with the XY system, females have two copies of the X chromosome, while males have only one. As a result, the X chromosome is subjected to unique evolutionary forces and has evolved sex-specific mechanisms of gene regulation (Vicoso and Charlesworth, 2006). Selective and/or regulatory differences between the X chromosome and autosomes are reflected in the non-random genomic distribution of sex-biased genes (i.e. genes expressed at a higher level in one sex than the other). Previous studies of *Drosophila melanogaster* have revealed that in whole flies and in reproductive tissues, male-biased genes are underrepresented on the X chromosome, whereas female-biased genes are overrepresented (Parisi, 2003; Sturgill *et al.*, 2007). However, in the brain and head, there is an overrepresentation of both male- and female-biased genes on the X chromosome relative to the autosomes (Catalán *et al.*, 2012).

One mechanism of gene expression regulation that is unique to the male X chromosome operates in the male germline and suppresses the expression of X-linked, testis-specific genes. It has been shown that testis-specific reporter genes have higher expression levels when located on the autosomes than on the X chromosome (Hense *et al.*, 2007; Kemkemer *et al.*, 2011, 2014). This phenomenon, known as "X suppression" is restricted to testis-specific genes and is not a common property of all tissue-specific genes, as ovary-, accessory gland- and Malpighian tubule-specific reporter genes do not show any signal of X suppression (see **Chapter 1**). Moreover, X-linked housekeeping genes are also transcriptionally suppressed in the male germline, but not in the female germline or male somatic tissues (Landeen *et al.*, 2016). However, the housekeeping genes are suppressed to a lesser extent than testis-specific genes in the male germline. The molecular mechanism responsible X suppression remains unknown. It is also unclear why testis-specific and non-testis-specific genes differ in their magnitudes of X suppression.

In contrast to the X suppression that takes place in the male germline, males upregulate the expression of their single X chromosome approximately two-fold in somatic tissues to compensate for the difference in gene dose between the X

chromosome and the autosomes. The above mechanism known as dosage compensation (DC) is achieved primarily through binding of the dosage compensation complex (DCC), a complex containing proteins and long non-coding RNAs (Conrad and Akhtar, 2012; Ercan, 2015). A component of the DCC, MOF, induces a modification of the local chromatin structure by histone acetylation. The chromatin becomes loose and therefore accessible for the transcriptional machinery, leading to hypertranscription. It is thought that the DCC binds to distinct nucleation sites, the High Affinity Sites (HAS), and then spreads in *cis* from these initiation sites in order to expand on the X chromosome (Ramirez *et al.*, 2015).

It has been demonstrated that a gene's proximity to a DCC binding site has an impact on its expression pattern. In head and brain, there is a highly-significant negative correlation between a gene's male-to-female expression ratio and its distance to the nearest DCC binding site (Chang *et al.*, 2011; Catalán *et al.*, 2012; Huylmans and Parsch, 2015). This is the case mainly for genes that show weak male bias. Thus, the excess of X-linked male-biased genes in these tissues is likely to be caused by "overcompensation" of genes located close to DCC binding sites (Huylmans and Parsch, 2015). Opposed to the pattern in head and brain, in testis, there is a highly-significant positive correlation between a gene's male-to-female expression ratio and distance to the nearest DCC binding site, even though DCC-mediated dosage compensation does not occur in the male germline of *Drosophila* (Meiklejohn *et al.*, 2011).

The main objective of this study was to investigate the occurrence of X suppression for a ubiquitously-expressed reporter gene in different male tissues. For this, we compared the expression of the same autosomal and X-linked reporter genes in testis and somatic tissues. We expected that if X suppression is a chromosome-wide mechanism unique to the male germline, then it would be detectable only in testis, but not in head or carcass. Furthermore, we expected that (at least partial) DC would be evident in the male soma, leading to higher expression of X-linked reporter genes in head and carcass. Our secondary objective was to examine the influence of the DCC on the expression of X-linked reporter genes in males. More specifically, we were

interested in the relationship between the expression level of such X-linked genes and their proximity to binding sites of various DCC components.

To tackle both of the above objectives, we used a reporter gene construct containing a minimal human cytomegalovirus (CMV) promoter fused to the *Escherichia coli lacZ* reporter gene (**Figure 7a**). This approach allowed us to exclude any sex- or tissue-specific regulatory mechanism that might affect reporter gene expression, since both the promoter and the reporter gene were not native to *D. melanogaster*. In addition, the CMV promoter was previously found to drive high, ubiquitous expression in *Drosophila* (Parsch, 2004), which allowed us to compare the expression of the same reporter gene inserted at the same genomic location in multiple tissues. For the current study, we tested reporter gene expression in testis, head, and carcass (whole fly with head and testes removed) (**Figure 7b**).

Materials and Methods

Reporter gene construct

The reporter gene construct we used is part of a transposable element vector that was designed by Parsch (2004). In the construct, terminal sequences of the *P*-element flank two copies of the reporter gene and the *D. melanogaster mini-white* eye color marker gene (**Figure 7a**). The reporter gene consists of a human cytomegalovirus (CMV) promoter alongside the *E. coli lacZ* coding sequence in the same transcriptional orientation. The CMV promoter is capable of driving ubiquitous expression of the reporter gene in flies (Parsch, 2004).

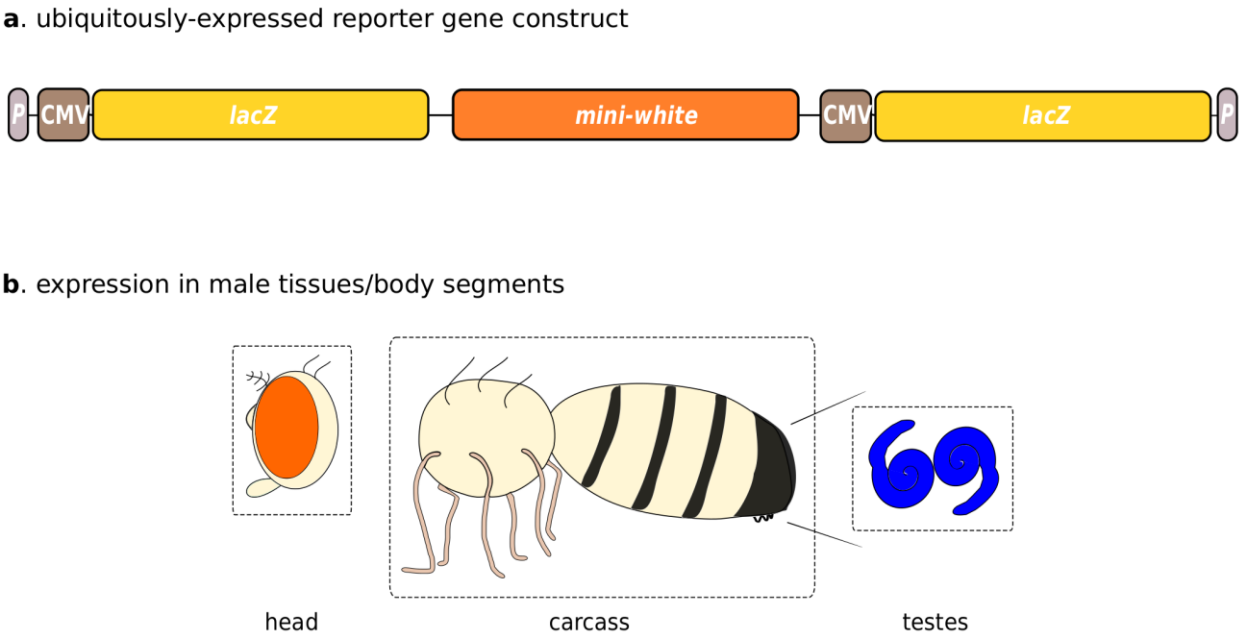


Figure 7. Schematic illustration of the reporter gene construct and the male tissues/body segments tested. a. The construct contains the *mini-white* marker gene and two copies of the *lacZ* reporter gene. The terminal sequences of the *P*-element (*P*) represent the boundaries of the transgene. The human cytomegalovirus promoter (CMV) is able to drive ubiquitous expression of the reporter gene in flies. **b.** The expression of the reporter gene was measured in dissected male heads, carcasses (whole flies with the heads and testes removed) and testes.

Transgene mobilization to new genomic locations

Fly lines with independent reporter gene insertions on an autosome or the X chromosome were created via mobilization of the transgene by genetic crosses as described by Hense *et al.* (2007). The following mating scheme was carried out, with the aim of mobilizing an X-linked transgene identified by the red eye color to a third chromosome containing a transposase gene linked to the stubble bristle phenotypic marker (*yw; Δ2-3, Sb/TM6*) (**Supplementary Figure S1**). The ultimate goal was to then mobilize these new transgene insertions to new random locations on the X chromosome and autosomes (**Supplementary Figure S1**).

Starting with an available X-linked line with a *yw* background (yellow body and white eyes), red-eyed females were mated with *yw; Δ2-3, Sb/TM6* males. Red-eyed male offspring with stubble bristles were collected because they inherited both the X-linked transgene and the transposase gene. These males were individually mated to *yw* females. Detection of male offspring with red eyes indicated transgene movement off

of the X, since only the female offspring should have inherited an X-linked transgene. If the male offspring had wild type bristles, they were mated individually to *yw* females in order to start new autosomal stocks. If all offspring of these crosses had red eyes and stubble bristles, then the transgene must have moved to the third chromosome containing the transposase gene. In this case, red-eyed stubble males were mated individually to *yw* females. Offspring (males or virgin females) that had red eyes and wild-type bristles were mated with *yw* flies of the opposite sex to start a new stock. To distinguish transgenes located on the X chromosome from those on an autosome, crosses of transformed males with *yw* females were performed (**Supplementary Figure S2**). The red eye phenotype from males with an X-linked transgene is passed on exclusively to their female offspring, whereas males with an autosomal transgene pass it on to 50% of their offspring of both sexes.

Mapping insertion locations

Inverse PCR was performed to determine the exact genomic locations of the transgene insertions (Bellen *et al.*, 2004). Genomic DNA of each transformed line was extracted from nine flies using the MasterPure^{TM} DNA Purification Kit (Epicentre, Madison, WI, USA). The genomic DNA was then digested with either *HinP*I or *Hpa*II, which both cut frequently within the *D. melanogaster* genome. Numerous small fragments were produced and were self-ligated with T4 DNA ligase (New England Biolabs, Ipswich, MA, USA). The fragment containing the inserted transgene was amplified by PCR using two primer pairs that matched parts of the of the *pP[wFl]* transformation vector (5'-3'): Plac1-Plac4 (CACCCAAGGCTCTGCTCCCACAAT, ACTGTGCGTTAGGTCCTGTTCATTGTT) and EY.3.F-EY.3.R (CAATAAGTGCGAGTGAAAGG, ACAATCATATCGCTGTCTCAC). The resulting PCR product was sequenced with the primers Sp1 (5'- ACACAACCTTTCCTCTCAACAA -3') and EY.3.F (above) using BigDye v1.1 chemistry on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, CA, USA). The mapping of the flanking genomic sequences to the *D. melanogaster* reference genome (Release 6.09) with a BLAST search (Altschul *et al.*, 1990) led to the identification of the insertion location.

β-galactosidase activity assays

The expression of the *lacZ* reporter gene was measured with β-galactosidase activity assays. Soluble protein was extracted from five heads, testes or carcasses by homogenizing the tissues/body segments in 200 µl of cold buffer (0.1 M Tris-HCl, 1 mM EDTA, 7 mM 2-mercaptoethanol; pH 7.5), incubating the homogenate on ice for 15 minutes, centrifuging at 12,000 rpm for 15 minutes at 4°C, and collecting the supernatant, which included the protein extract. 50 µl of protein extract together with 50 µl of 2 × Assay buffer (200 mM sodium phosphate (pH 7.3), 2 mM $MgCl_2$, 100 mM 2-mercaptoethanol, 1.33 mg/ml o-nitrophenyl-β-D-galactopyranoside) were assayed and considered a technical replicate. One biological replicate of a given sample (derived from five flies providing tissue or body segments) was used for two technical replicates. For each transformed line and for the corresponding male tissues/body segments from the *yw* strain (used as negative control), 2–3 biological replicates were carried out. β-galactosidase activity was measured spectrophotometrically by tracking absorbance for 50 min at 420 nm at 37˚C. The activity units were defined as the change in absorbance per minute (maximum slope).

Maintenance of fly strains

All fly strains were maintained at 22°C on cornmeal-agar-molasses medium with a 14hr light:10hr dark cycle. All flies used for β-galactosidase assays were 4–6 days old, mated, and either heterozygous or hemizygous for the transgene insertion. Thus, all comparisons were of flies carrying a single copy of the reporter gene.

Reporter gene expression analysis

To test for differences in reporter gene activity between X-linked and autosomal inserts in head, testis and carcass, we performed the non-parametric Wilcoxon (Mann-Whitney) test using the mean activity across biological replicates of each transformed line. The mean of a biological replicate was defined as the mean of its two technical replicates.

To determine whether there is a monotonic relationship between the reporter gene expression and its proximity to the nearest DCC binding site (either upstream or

downstream) in head, testis and carcass, we calculated the Spearman's correlation coefficient for each of the DCC components: MLE, MSL2, and MSL3, and for the HAS (defined by the co-localization of MSL2 and MLE). In order to further explore the nature of the relationship, we performed linear regression analysis in the same fashion, i.e. between reporter gene expression and proximity to nearest DCC component binding site. Information about the location of DCC binding sites was obtained from published chromatin immunoprecipitation sequencing (ChIP-seq) experiments (Straub *et al.*, 2013).

Results

Genomic distribution of reporter genes

We obtained a total of 29 transgenic lines, each with a reporter gene insertion at a unique genomic location. Twelve of the lines had autosomal insertions, which were located on all possible chromosomal arms, including the fourth chromosome (1 line) (**Supplementary Table S1**). Seventeen of the lines had an X-linked insert, distributed widely along the X chromosome, except for two inserts that were close together with a distance of only 320 bp between them (**Figure 8**, **Supplementary Table S2**).

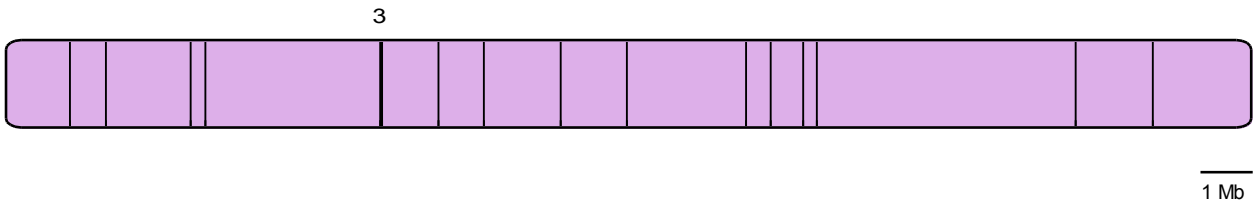


Figure 8. **Map of X-linked transgene insertion locations.** Each vertical line represents an insertion at a unique site as determined by inverse PCR. The thick line represents three insertions that, due to their proximity to each other (within 30 kb), are not distinct at the resolution of the figure.

The X-linked lines were used in the association analysis between reporter gene expression and proximity to the nearest DCC binding site. Therefore, it was important to characterize the frequency of X-linked inserts in the proximity classes (bp) from

each of the four DCC binding site used here (**Figure 9**). The largest distance between a reporter gene and a DCC binding site was 186 kb with the majority of the reporter genes being found within 100 kb of all binding sites. For HAS, MSL2 and MSL3 binding sites, only 3–4 reporter genes were located beyond 100 kb, while for MLE none was located beyond 61 kb.

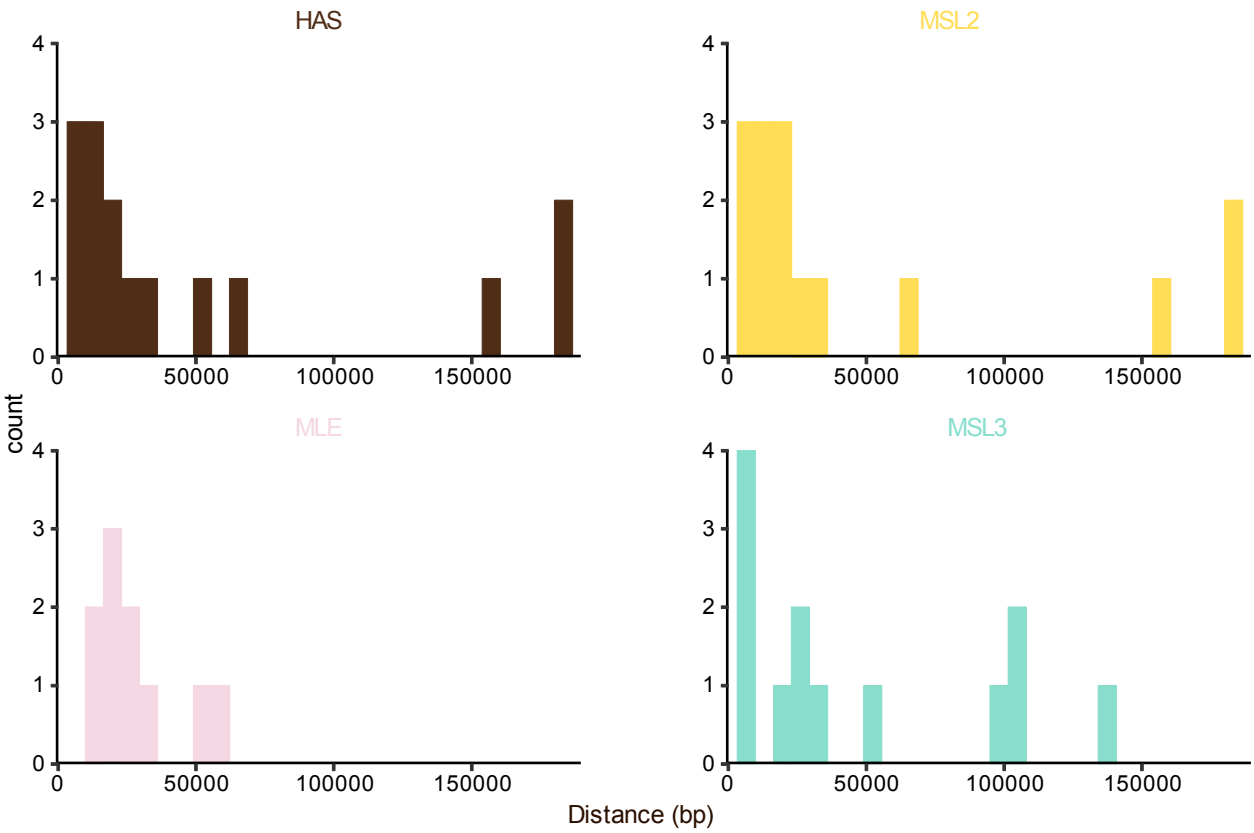


Figure 9. Frequency of X-linked inserts in the nearest distance classes from each of the four DCC binding site used in our analysis. Each color represents a different DCC binding site.

Autosomal *versus* X-linked expression in head, testis, and carcass

We investigated the occurrence of X suppression of a ubiquitously-expressed gene in different tissues by comparing autosomal and X-linked expression of a *lacZ* reporter gene with an exogenous CMV regulatory sequence. The expression of 12 autosomal and 17 X-linked reporter genes was compared in head, testis, and carcass of the same flies.

In testis, the mean (median) β-galactosidase activities of autosomal and X-linked lines were 0.18 (0.17) and 0.25 (0.28) mOD/min, respectively, and did not differ significantly (Wilcoxon test, *P*=0.2) (**Figure 10a**). Hence, there was no evidence for X suppression in the male germline.

In male heads, the mean (median) β-galactosidase activities of autosomal and X-linked lines were 3.45 (3.61) and 4.34 (4.56) mOD/min, respectively, and did not differ significantly (Wilcoxon test, *P*=0.14) (**Figure 10b**). Even after an outlier autosomal line (A39 from **Supplementary Table S1**) with extremely low expression (0.54 mOD/min) compared to the rest of the lines was removed from the analysis, there was no significant difference between autosomal and X-linked expression (Wilcoxon test, *P*=0.24). Hence, there was no evidence for X suppression in male heads.

In male carcasses, the mean (median) β-galactosidase activities of autosomal and X-linked lines were 11.85 (12.52) and 16.36 (15.04) mOD/min, respectively, and did not differ significantly (Wilcoxon test, *P*=0.06) (**Figure 10c**). Even after an outlier autosomal line (A39 from **Supplementary Table S1**) with extremely low expression (0.62 mOD/min) compared to the rest of the lines was removed from the analysis there was no significant difference between autosomal and X-linked expression (Wilcoxon test, *P*=0.11). Hence, there was no evidence for X suppression in male carcasses.

Male expression and proximity to DCC binding sites

We measured the reporter gene expression with β-galactosidase assays in head, testis and carcass of the same male flies of 17 X-linked lines. At first, to investigate possible patterns in the relationship between reporter gene expression and distance to the nearest DCC component binding site, we plotted the observed data and fitted smooth curves (**Figure 11**, **Figure 12** and **Figure 13**). Furthermore, we calculated the Spearman's correlation coefficient and performed a linear regression analysis (see **Table 1**).

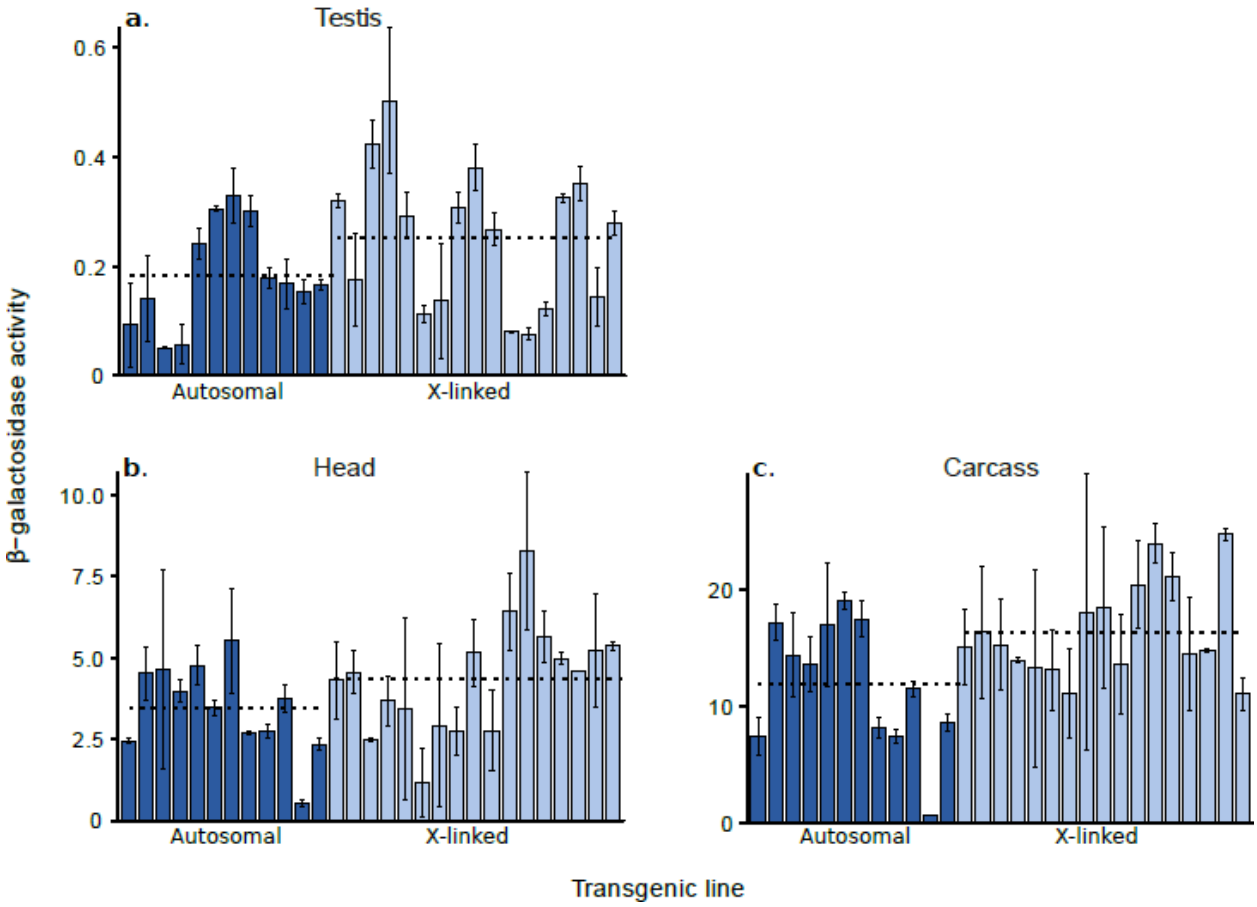


Figure 10. Autosomal and X-linked expression of a ubiquitously-expressed reporter gene in males. a. testis, **b**. head, **c**. carcass. Each bar represents a transformed line with the reporter gene inserted at a unique autosomal (dark blue) or X-linked (light blue) location. Expression was measured spectrophotometrically as β-galactosidase activity in units of mOD/min. Error bars indicate the standard deviation across biological replicates. Dotted lines indicate the average activities of all autosomal or X-linked lines.

Table 1. Statistical measures for the relationship between reporter gene expression and the nearest DCC component binding site in head, testis, and carcass.

DCC binding site	Spearman's correlation[1]		Linear regression[1]		Tissue
	rho	*p-value*	R^2	*p-value*	
HAS	0.22	0.38	0.11	0.10	Testis
MLE	-0.34	0.17	0.13	0.08	
MSL2	0.18	0.48	0.09	0.13	
MSL3	0.04	0.89	0.06	0.18	
HAS	0.17	0.50	-0.07	0.96	Head
MLE	0.47	0.057	0.34	0.008	
MSL2	0.14	0.58	-0.07	0.94	
MSL3	-0.07	0.79	-0.03	0.50	
HAS	0.36	0.15	-0.06	0.87	Carcass
MLE	0.55	0.024	0.35	0.007	
MSL2	0.35	0.16	-0.06	0.83	
MSL3	0.19	0.47	-0.06	0.87	

[1]Shaded cells indicate significant correlations (P<0.05).

In testis, the expression level of the reporter gene was generally low. It was not possible to detect a significant association between the reporter gene expression and the proximity to the DCC binding sites for any of the binding sites (**Figure 11** and **Table 1**).

In head, the expression level of the reporter gene showed a positive correlation only with the distance to the MLE binding site (*rho*=0.47, P=0.057), which was marginally significant. This correlation was highly significant in a linear regression analysis (R^2=0.34, P=0.008), with the expression increasing as the distance from the MLE binding site increases (**Figure 12** and **Table 1**).

In carcass, the expression level of the reporter gene was significantly positively correlated only with the distance to the MLE binding site (*rho*=0.55, P=0.024). This correlation was also was also highly significant in the linear regression analysis (R^2=0.35, P=0.007), with the expression increasing as the distance from the MLE binding increases (**Figure 13** and **Table 1**).

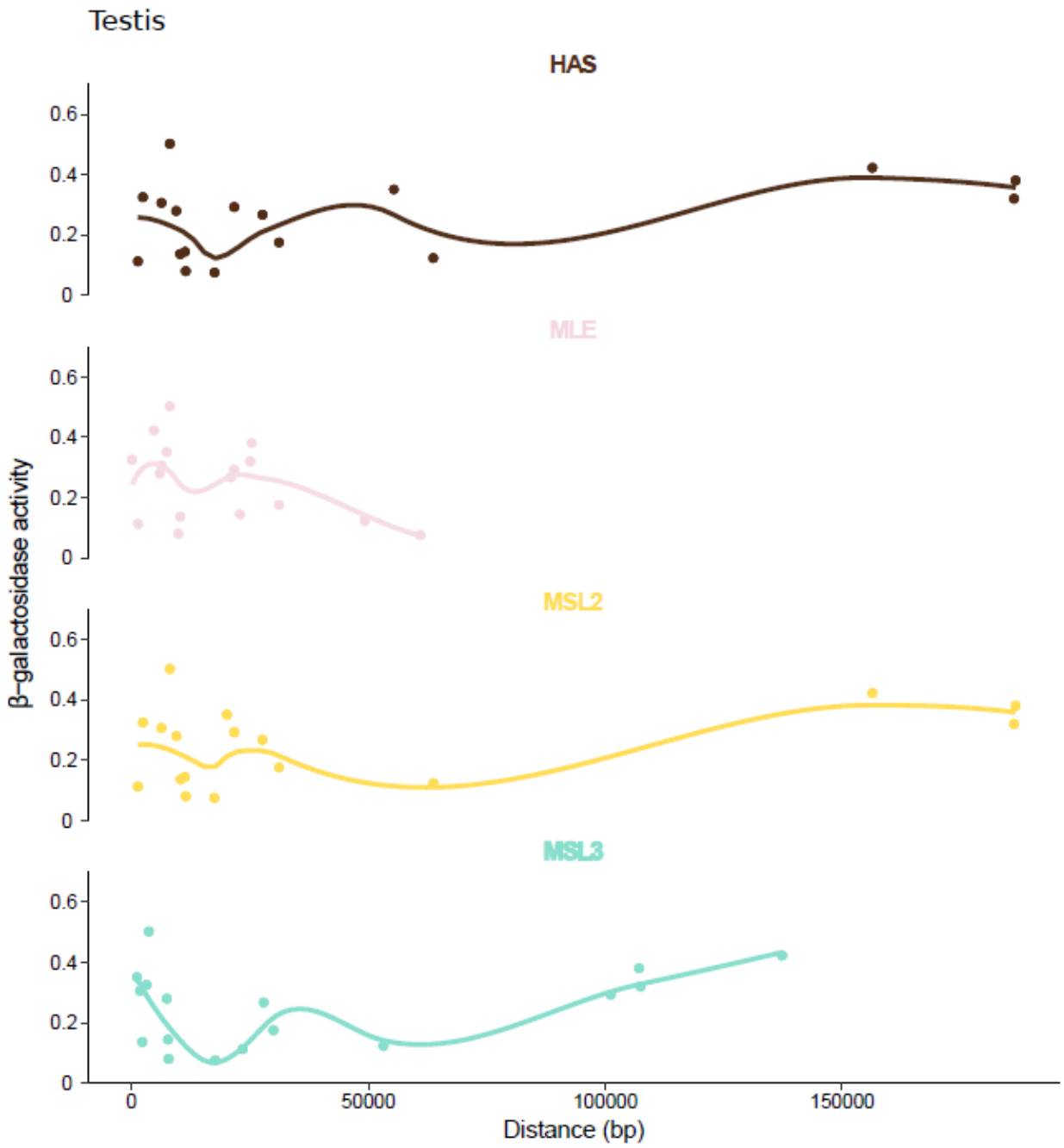


Figure 11. Reporter gene expression and proximity to DCC binding sites in testis. Each data point represents a reporter gene inserted at a unique location on the X chromosome. Expression of the reporter gene (y-axis) was measured spectrophotometrically as β-galactosidase activity in units of mOD/min. Distance to the DCC binding site: HAS, MLE, MSL2, or MSL3, (x-axis) was defined as the nearest distance in base pairs. For visualization, smooth curves were fitted to the data.

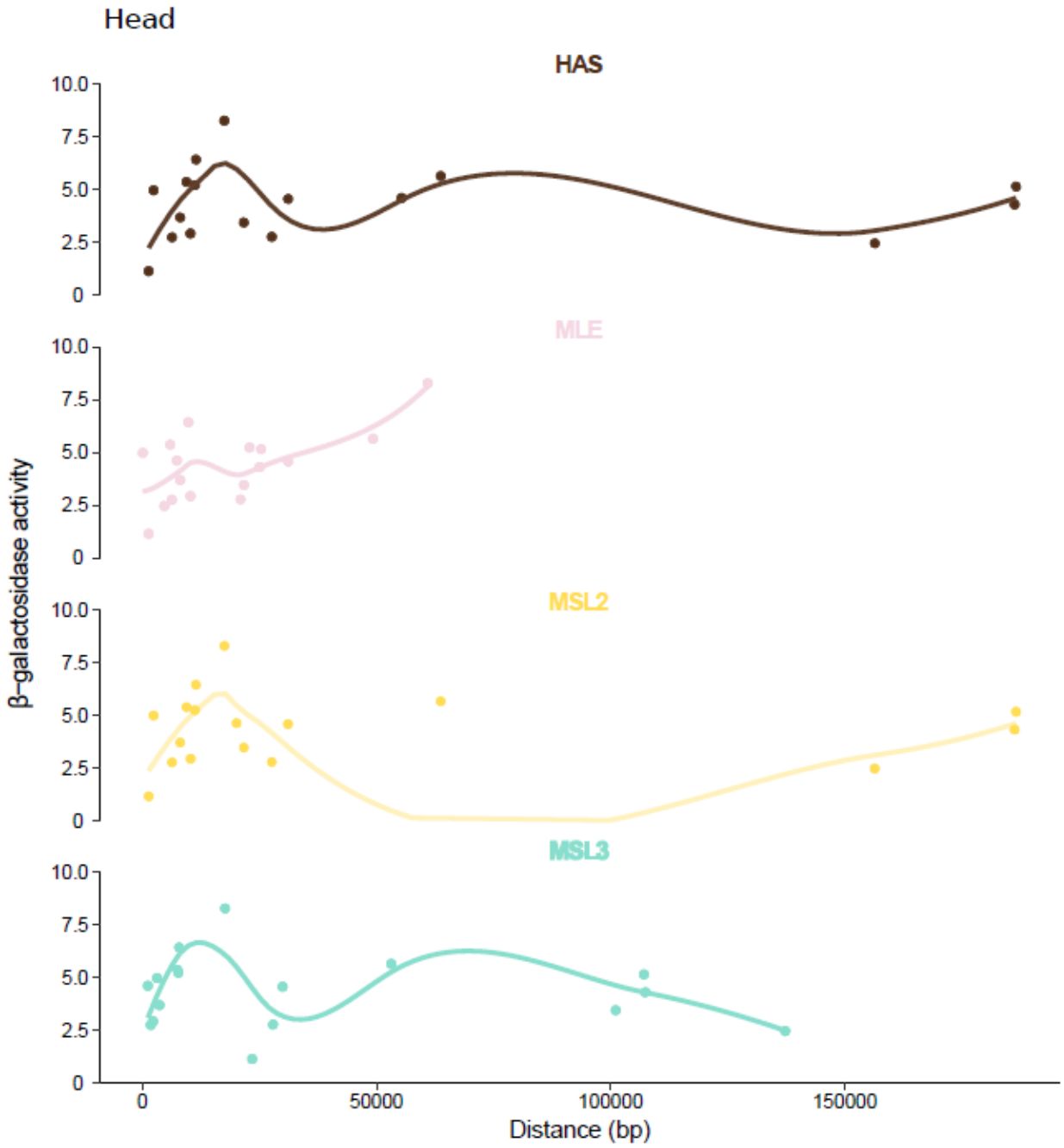


Figure 12. Reporter gene expression and proximity to DCC binding sites in head. Each data point represents a reporter gene inserted at a unique location on the X chromosome. Expression of the reporter gene (y-axis) was measured spectrophotometrically as β-galactosidase activity in units of mOD/min. Distance to the DCC binding site: HAS, MLE, MSL2, or MSL3, (x-axis) was defined as the nearest distance in base pairs. For visualization, smooth curves were fitted to the data.

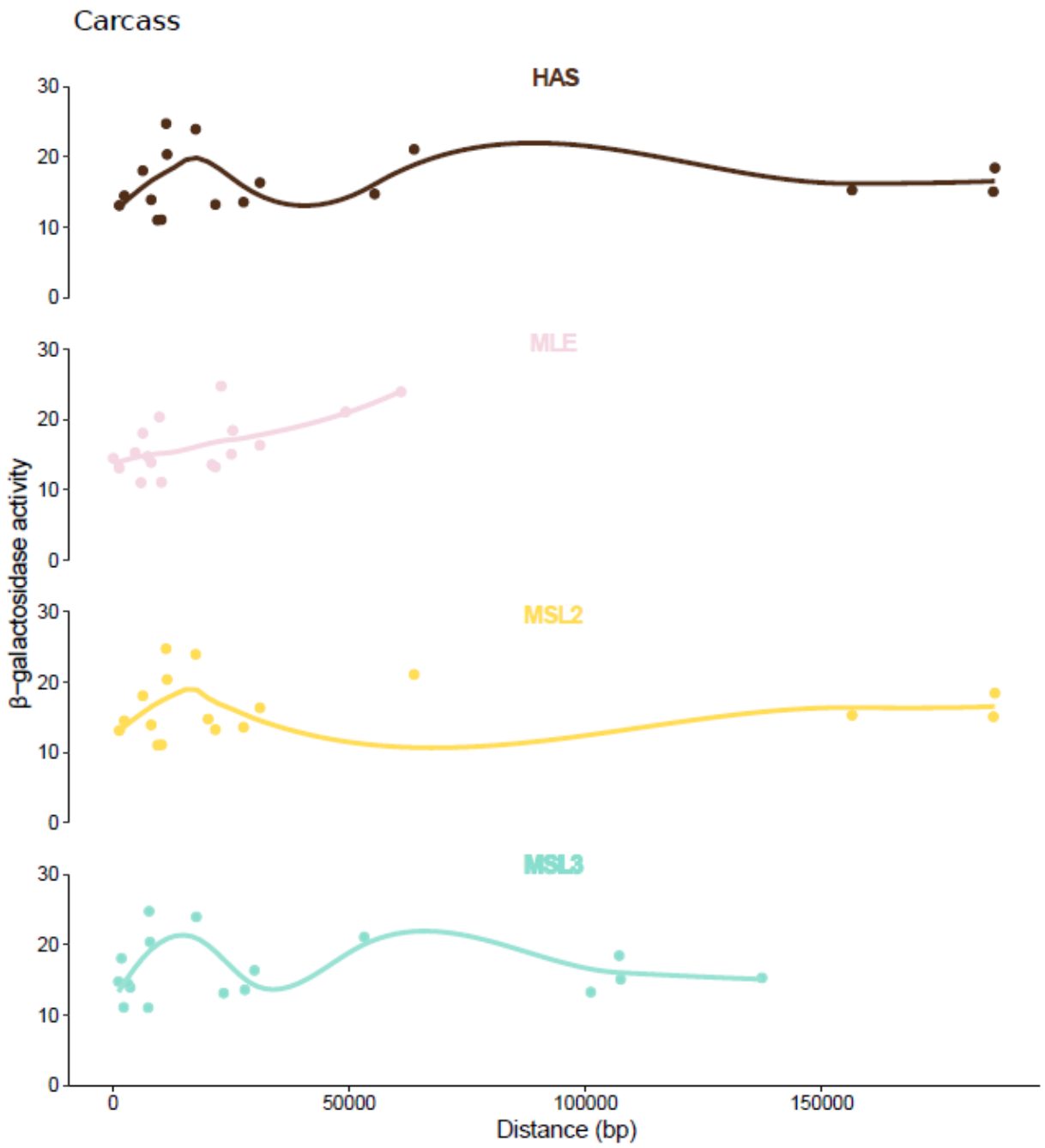


Figure 13. Reporter gene expression and proximity to DCC binding sites in carcass. Each data point represents a reporter gene inserted at a unique location on the X chromosome. Expression of the reporter gene (y-axis) was measured spectrophotometrically as β-galactosidase activity in units of mOD/min. Distance to the DCC binding site: HAS, MLE, MSL2, or MSL3, (x-axis) was defined as the nearest distance in base pairs. For visualization, smooth curves were fitted to the data.

Discussion

In this study, we focused on two regulatory mechanisms of the male X chromosome in *Drosophila*: X suppression in the male germline and dosage compensation. We employed an exogenous, ubiquitously-expressed reporter gene and measured the expression of autosomal and X-linked inserts in testis, male heads, and male carcasses. With this approach, we could exclude the influence of any sex- or tissue-specific regulatory elements on reporter gene expression.

The primary aim of this study was to assess whether suppression of the X chromosome occurs for a non-testis-specific gene in the male germline (or potentially in somatic tissues). The most striking result to emerge from the data is the absence of X suppression in the male germline, which contradicts previous studies of other reporter genes (Hense *et al.*, 2007; Kemkemer *et al.*, 2011, 2014). In the somatic tissues we analyzed, head and carcass, we observed higher expression of the X-linked reporter genes, although the differences were not significant. The average increase of X-linked expression relative to autosomal expression in head and carcass was 1.26-fold and 1.38-fold, respectively, and might be the result of (partial) dosage compensation of these X-linked reporter genes, as expected in male somatic tissues.

In testis, the mean reporter gene expression across both autosomal and X-linked lines was relatively low (0.21 mOD/min) when compared to that measured in head (3.89 mOD/min) or carcass (14.10 mOD/min). Because of this low level of expression, it is possible that a difference between autosomal and X-linked reporter genes was not detectable. In addition, the degree of X suppression may be smaller for genes with lower levels of expression in testis. This has been reported previously for housekeeping genes, which exhibit a smaller decrease in expression when X-linked than highly-expressed, testis-specific genes (Landeen *et al.*, 2016). Thus, the low level of expression of our reporter gene in testis may explain why we did not detect a difference between the X chromosome and the autosomes. It is possible that a chromosome-wide mechanism is responsible for X suppression in the male germline, since it affects both testis-specific and housekeeping genes. If such a mechanism affects all X-linked genes with expression that surpasses a certain threshold, then the magnitude of suppression could be different among highly (e.g. testis-specific),

moderately (e.g. housekeeping) and basally (e.g. our reporter gene) expressed genes. A potential chromosome-wide mechanism for X suppression could rely on silencer elements found along the X chromosome, but not the autosomes, and does interfere with the basal transcription machinery.

The secondary aim of this study was to explore the relationship between the expression of X-linked reporter genes in males and their distance to the binding sites of DCC components (MLE, MSL2, MSL3, and HAS). In testis, no correlation was found between reporter gene expression and any of the four DCC binding sites analyzed here. This observation was not surprising given that dosage compensation does not take place in the male germline (Meiklejohn *et al.*, 2011).

Male expression was positively correlated with the distance to the nearest MLE binding site in head in a linear fashion and in carcass in both a monotonic and linear fashion. Despite the significant correlations observed between male expression and proximity of MLE binding sites in head and carcass, we have to consider the possibility that the fact that the X-linked inserts are, on average, located closer to MLE binding sites than to other DCC component binding sites (**Figure 9**) might be affecting our results. Given that there are more MLE binding sites (508) than HAS (244), it is not surprising that the transgenes tend to be closer to MLE binding sites. Also, the function of the MLE binding site subset that does not coincide with the HAS is not clear. It is possible that those sites are an artifact of the ChIP-seq because MLE which is an RNA helicase can bind on RNA, and might have interfered with the crosslinked DNA-protein complexes. Therefore, a correlation of the expression only with those sites might not be related to dosage compensation *per se*. Nonetheless, all the reporter genes are situated in close proximity of genes (**Supplementary Table S2**), as *P*-element insertions are frequently associated with genes (Bellen *et al.*, 2004). This is an advantage of our methodology, since DCC binding sites are usually found in close proximity to active genes (Straub *et al.*, 2008) and, thus the transgenes should be a good representative for active, endogenous genes.

Finally, a number of limitations to this pilot study should be acknowledged. For future studies, it will be essential to measure expression in females, in addition to males, so that the male-to-female expression ratio for each reporter gene insertion can

be determined in each tissue/body segment. Because male-to-female expression ratio is the measure typically used in studies of sex-biased gene expression, it will allow us to compare our results to previous studies (e.g.Huylmans and Parsch, 2015). Furthermore, additional X-linked lines should be included in the analysis in order to increase statistical power. For example, in order to detect a strong (*rho*=0.5) or moderate (*rho*=0.30) correlation as significant with a power of 80% and a significance level of 5%, sample sizes of 28 or 84 would be needed, respectively (power test 'pwr.r.test' from the pwr package in R). Moreover, setting an upper limit to the distance from the nearest DCC binding site could improve the analysis. It might be more meaningful to take into account DCC binding site proximities only up to a maximum of 50 kb, as the DCC operates on active gene domains of the genome (Schauer *et al.*, 2017), which are typically of this size and are arranged into three dimensional structures (Schwartz and Cavalli, 2017). Besides, in the study where the significant negative correlation between male-biased expression and DCC binding sites in head and brain was described, 95% of the genes analyzed were located within 10kb of any DCC binding site (Huylmans and Parsch, 2015).

In conclusion, this study has contributed to our understanding of the regulation of the X chromosome across various tissues in male *D. melanogaster*. First, we demonstrated that the mechanism leading to suppression of the X chromosome does not affect all genes to the same extent in the male germline. Our results suggest that the chromosome-wide mechanism responsible for the suppression of X-linked, testis-specific genes in the male germline has little or no effect on ubiquitously-expressed genes with low levels of expression in testis. Second, we performed an initial study of the relationship between the expression of X-linked reporter genes and their distance to the nearest DCC binding site. Taken together, our findings point towards a complex interaction among expression level, tissue-specificity, and dosage compensation that regulates the expression of X-linked genes in males.

Acknowledgements

We thank Hilde Lainer for excellent technical assistance. We also thank the master's students Aris Aristodemou and Juanita Gutiérrez for their contribution to this project. This work was supported by Deutsche Forschungsgemeinschaft grant PA 903/6.

Supplementary Material

Supplementary Table S1. Genomic locations of the autosomal transgene insertions.

Line ID	Chr[1]	Cyt. band[2]	Coordinate[3]	Location[4]	Affected gene[5]	Proximal gene[6]	Distal gene[6]
A6	3R	88D5	14817989	intron/exon	btsz		
A10	3L	68E1	11821852	exon/intron, exon	CG5946, CG11597	CG14130, CG42255	Rpl10Ab, CG32095, CycA, CG7264
A13	3R	90C5	17794503	5'UTR	CG43102		pasi1, CG7379, CG17803
A14	3R	93F14	21856246	5'UTR	glec	Gr93d	lsn
A16	2R	55D1	18504287	inter		CG30116	GstE11
A26	2L	30B10	9521294	inter		CG33298	Oatp30B, CG31883
A27	2R	44F6	8913824	intron	Pgi	CG8252, CG30349, CG8258	lin, CG34219, CG8248, Spt
A28	2L	31A2	10057376	5'UTR	Pen	Cpr31A, CG33301	Spn31A, CG44153
A29	2L	35D4	15762783	inter		Gli, l(2)35Df	CG3793, wek, Ku80, CG31826
A33	3L	70A3	13227763	inter			caps
A39	4	102D4	704530	intron	ey	myo	
A40	3L	61C8	699829	inter		CG32483, RabX6, Vti1a, CG13894	CG13895, CkIIalpha-i3

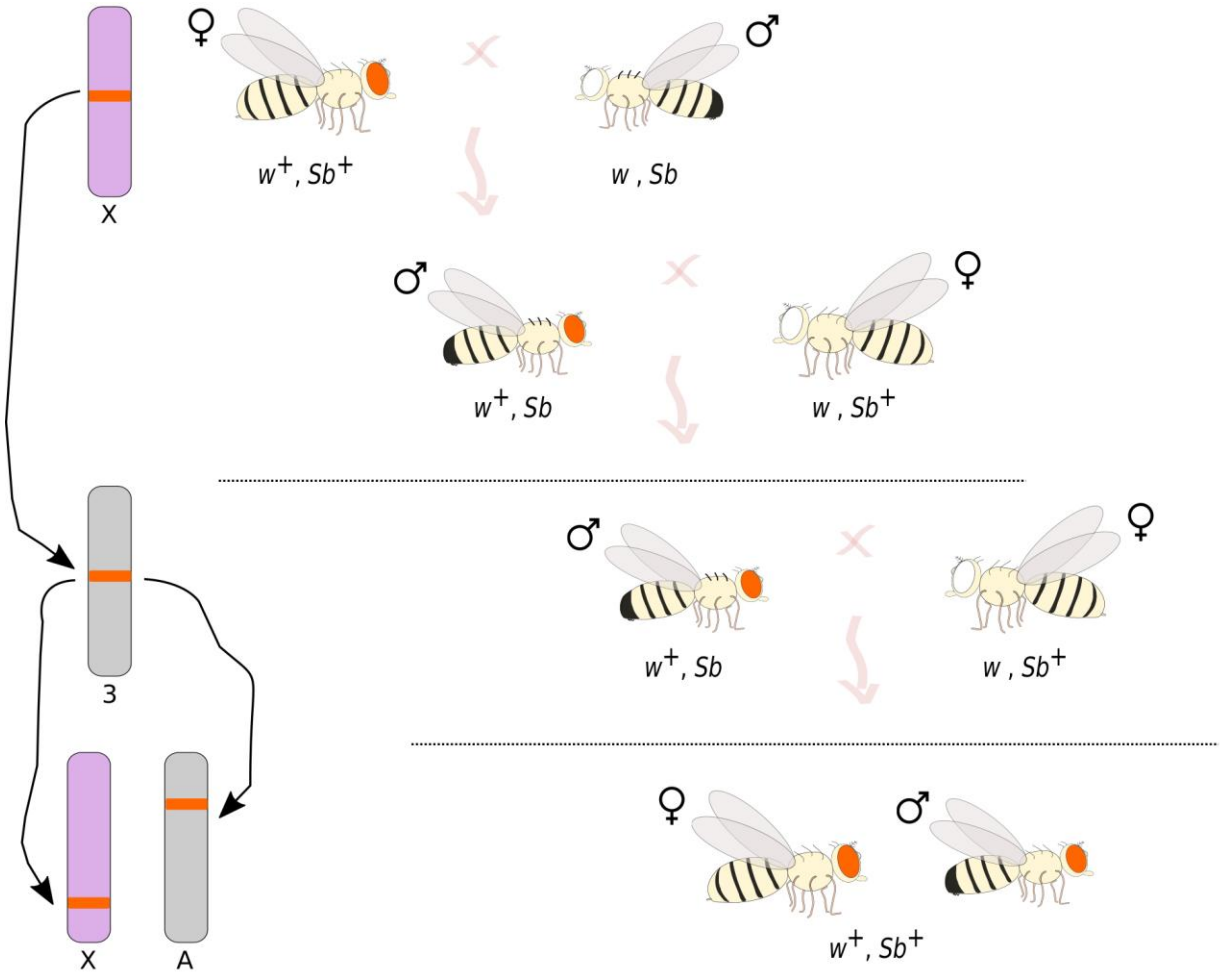
[1] Chromosomal location
[2] Cytological band
[3] Coordinate based on *D. melanogaster* reference genome Release 6.09
[4] Type of genomic location, "inter" indicates intergenic regions
[5] Gene overlapping with the transgene insertion
[6] Gene found within 10kb of the transgene insertion

Supplementary Table S2. Genomic locations of the X-linked transgene insertions.

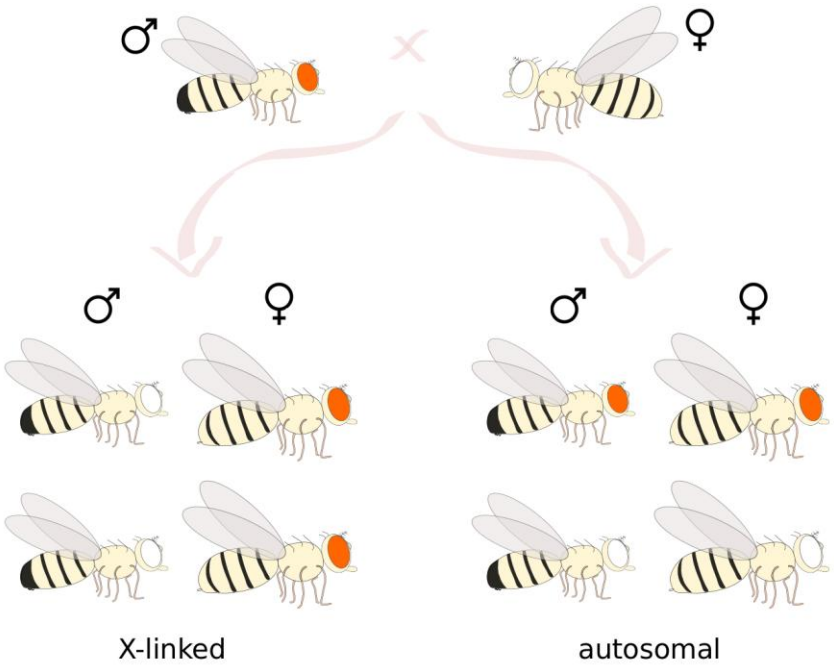
Line ID	Chr[1]	Cyt. band[2]	Coordinate[3]	Location[4]	Affected gene[5]	Proximal gene[6]	Distal gene[6]
2X	X	6E4	6998549	inter		Inx7, ogre	Inx2
X1	X	9B6	10318800	intron	alpha-Man-Ia		CG2909, Gip
X5	X	12F4	14826069	5'UTR	rut		CG14408, CG14411
X8	X	2B13	1873539	exon	Pgam5	CG14803, Pex5, MED18, CG14814	Vsp26, CG14817, CG14805, CG14818, CG14806, trr
X9	X	13A9	15075124	intron	Lsd-2	dob, opm, ND-B18	CG33177, CG33178, CG9065
X11	X	8B6	8894224	intron	Moe		CG1885, Rbm13, e(r)
X15	X	3E6	3720051	inter		Rala	Tlk
X20	X	3D3	3451351	inter		CG12535	
X21	X	6E4	6968661	3'UTR, 5' UTR	CG14431, CG32732		CG4586
X22	X	12C1	13762705	inter		Yp3, Rtc1, CG32625	rdgB
X23	X	10C5	11558520	5'UTR/intron	CG1572	Drak	PGRP-SA, RpII215
X24	X	19F4	21319242	5'UTR	SLIRP1	CG33713, Rpt6, CG1801	Dd, CG1486
X25	X	12E5	14222240	intron	l(1)G0007		CG11674, mRpL38
X26	X	7D5	8057219	intron	fs(1)h		mys
X27	X	6E4	6998869	5'UTR/intron	Inx2	Inx7	

X28	X	18F4	19887117	3'UTR	amn	Hers	Hers
X29	X	1E4	1209204	inter		CG14625, CG11381, CG14624	CG11382, CG11398, CG3638

[1] Chromosomal location
[2] Cytological band
[3] Coordinate based on *D. melanogaster* reference genome Release 6.09
[4] Type of genomic location, "inter" indicates intergenic regions
[5] Gene overlapping with the transgene insertion
[6] Gene found within 10kb of the transgene insertion



Supplementary Figure S1. **Transgene mobilization mating scheme that is enabled by two phenotypic markers.** The red eye wild type phenotype (w^+) is indicative of the presence of the transgene, while the white eye phenotype is indicative of the absence of the transgene (*w*). The transposase gene located on the third chromosome is linked to the stubble bristle phenotype (*Sb*), whereas the wild type bristle phenotype (Sb^+) is indicative of the absence of the transposase. Initially a transformed female with an X-linked transgene insertion is mated to males carrying the transposase. A male offspring that inherits both the transposase and the transgene is mated to females lacking the transpose and the transgene. In case of a mobilization event to the third chromosome, all red-eyed flies have stubble bristles. Such a male is selected and mated to females lacking the transpose and the transgene. Transgenic offspring without the transposase signal independent mobilization events off the third chromosome (to the X chromosome or an autosome).



Supplementary Figure S2. Location cross. To determine whether a transgene is X-linked or autosomal a transformed male (red eyes) is mated to *yw* females lacking the transgene (white eyes). In case the transgene is X-linked it is inherited exclusively by female offspring, whereas if it is autosomal it is inherited by 50% of both male and female the offspring.

CHAPTER 3

An 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 Genet, 2016, **12**: e1005987

doi:10.1371/journal.pgen.1005987

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

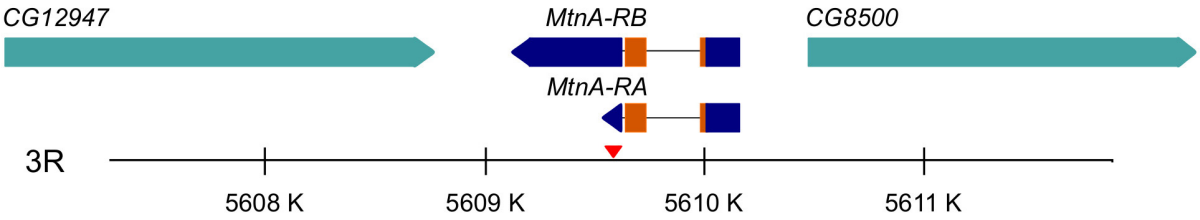


Fig 1. Structure of the *MtnA* locus. Two transcripts that differ only in their 3' UTRs have been annotated for *MtnA* (*MtnA-RA* and *MtnA-RB*). Dark blue boxes represent the UTRs with the arrowheads indicating the direction of transcription. Orange boxes represent the coding exons. The thin lines joining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

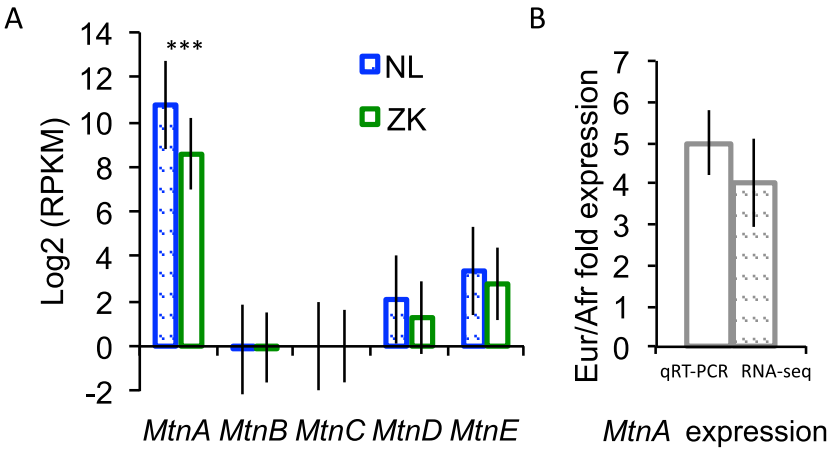


Fig 2. Expression of metallothionein genes in the brain in two populations of *D. melanogaster*. (A) Expression level of *Mtn* paralogs in the brain from RNA-seq data. Expression is reported in reads per kilobase per million mapped reads (RPKM). Only *MtnA* showed a significant difference in expression 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

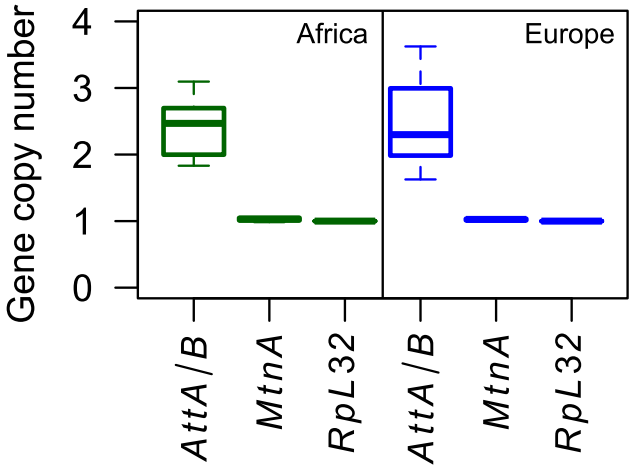


Fig 3. Results of CNV assays. Flies from Africa (Zimbabwe), shown in green, and Europe (the Netherlands), shown in blue, were tested for *MtnA* CNV. The close paralogs *AttA* and *AttB* were used as a positive control for multiple gene copies, while *RpL32* was used as a single-copy reference.

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].

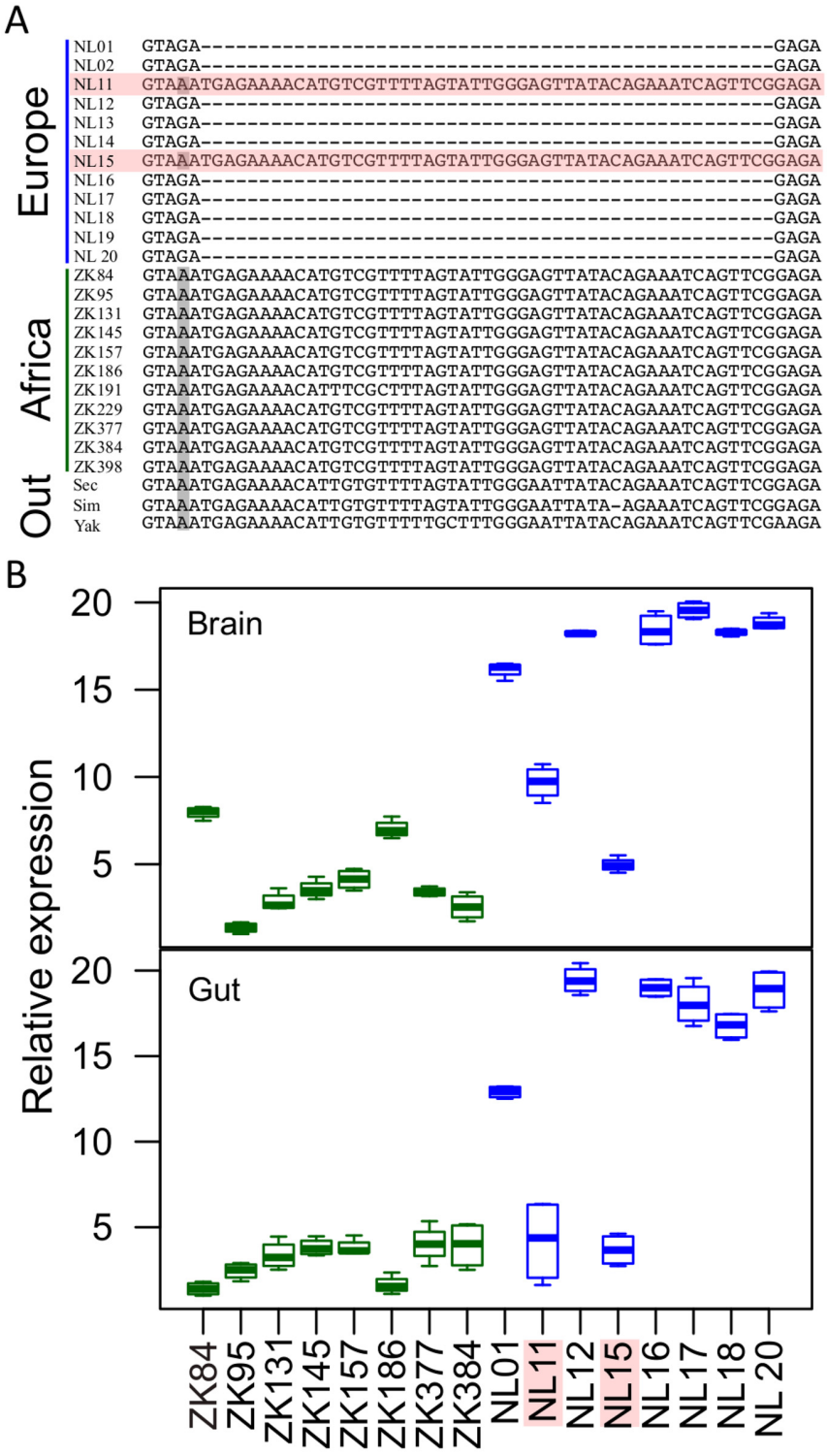


Fig 4. Association between an indel polymorphism in the *MtnA* 3' UTR and gene expression variation. (A) An indel (and a linked SNP marked in gray) in the *MtnA* 3' UTR are the only polymorphisms within the 6-kb *MtnA* region that show a large difference in frequency between an African and a European 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

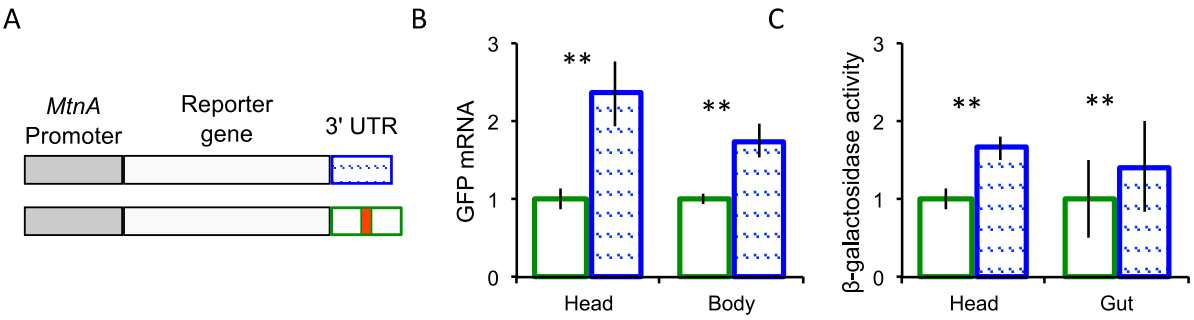


Fig 5. Reporter gene constructs and their expression. (A) The gray boxes represent the *MtnA* promoter, which is identical between the African and European alleles. The white boxes represent the GFP/*lacZ* reporter genes. The blue hatched box represents the *MtnA* 3' UTR with the deletion. The green 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

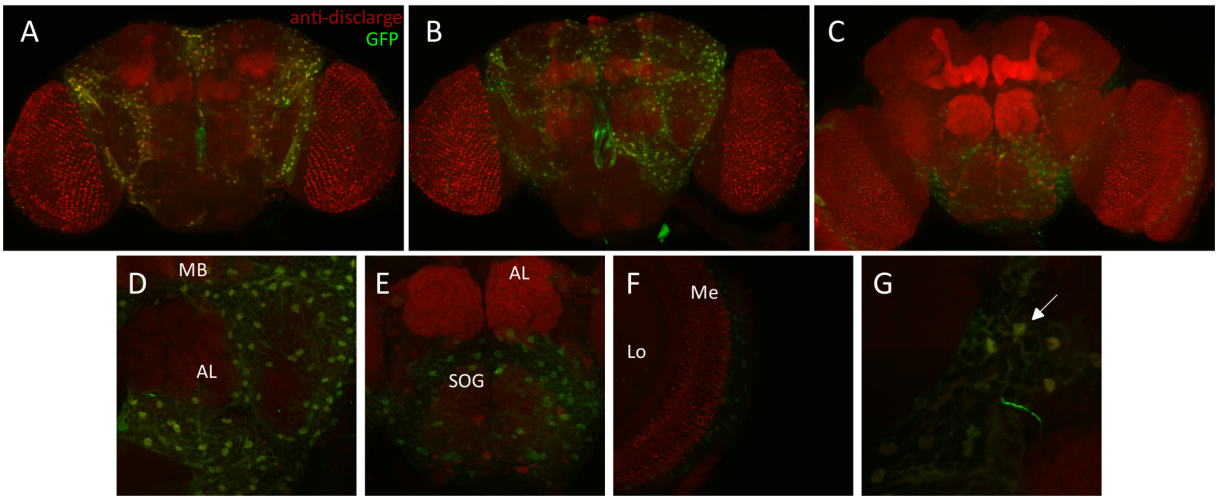


Fig 6. Expression of an *MtnA*-GFP reporter gene in the brain. (A-C) GFP expression driven by the reporter gene construct with the ancestral *MtnA* 3' UTR variant. (D-G) Higher magnification of the brain regions where GFP is expressed. AL: antennal lobe, MB: mushroom bodies, SOG: subesophageal ganglion, 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

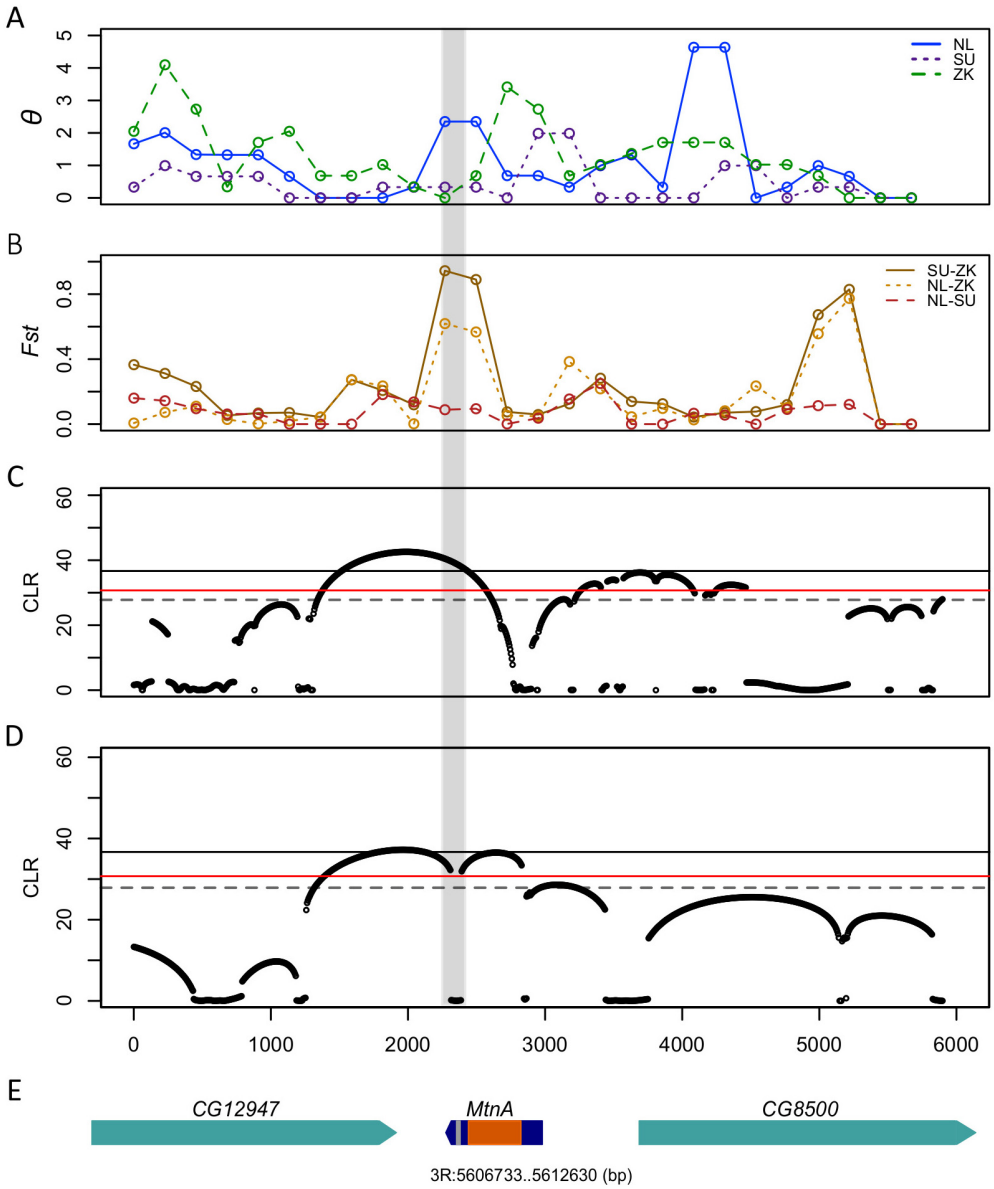


Fig 7. Evidence for positive selection at the *MtnA* locus. (A) Watterson's θ of *D. melanogaster* populations from Zimbabwe (ZK), the Netherlands (NL) and Sweden (SU) calculated in sliding windows of 500 bp with a step size of 250 bp. (B) F_{st} values for pairwise comparisons of ZK, NL and SU calculated 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

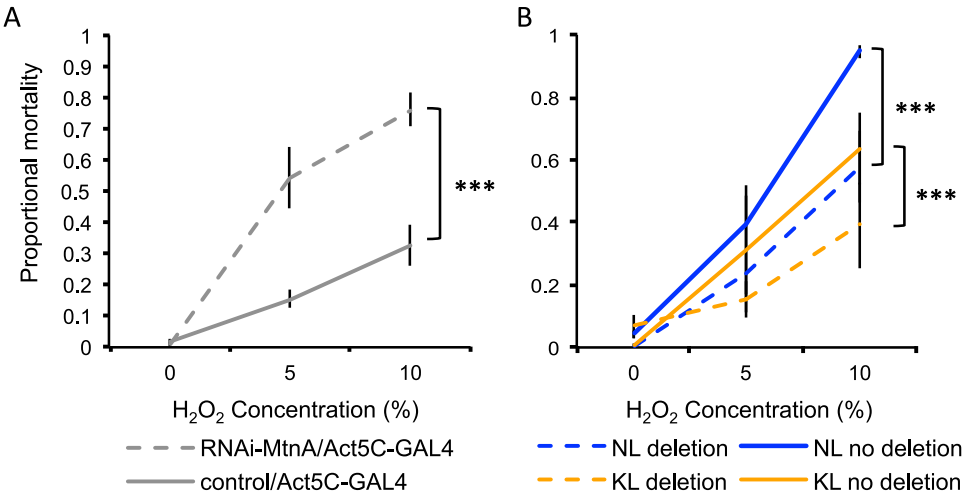


Fig 8. Proportional mortality after oxidative stress. (A) RNAi-mediated *MtnA* knockdown (hatched lines) and control flies (solid lines). *P*-values are shown for within population/background comparisons. (B) Dutch (blue) and Malaysian (orange) flies with the deletion (hatched lines) and without the 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

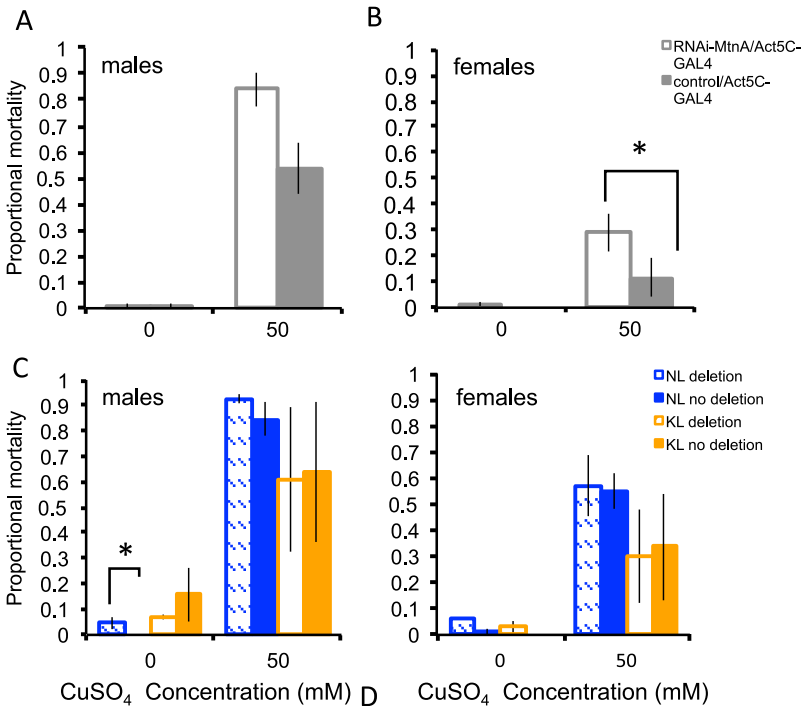


Fig 9. Proportional mortality after copper sulphate exposure. (A,B) Copper tolerance in RNAi-mediated *MtnA* knockdown flies (white, RNAi-*MtnA*/Act5C-GAL4) and control flies expressing normal levels of *MtnA* (solid grey, control/Act5C-GAL4). (C) Male and (D) female flies from the Dutch (NL, blue) and 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.

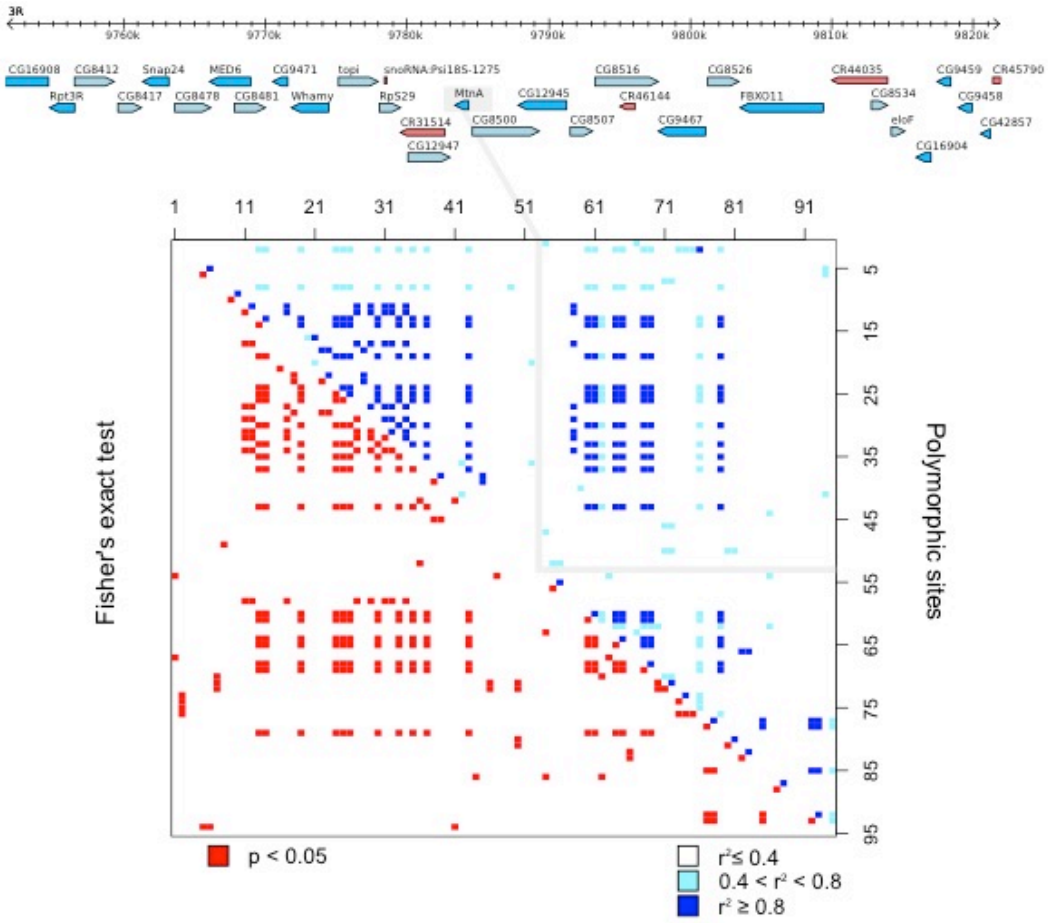
67. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; 25: 1451–2. doi: 10.1093/bioinformatics/btp187 PMID: 19346325

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

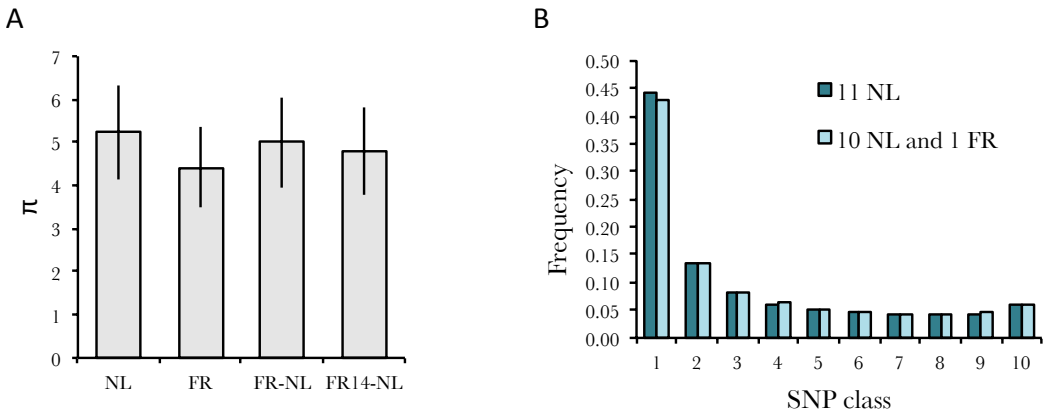
69. Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 2002; 18: 337–338. doi: 10.1093/bioinformatics/18.2.337 PMID: 11847089

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

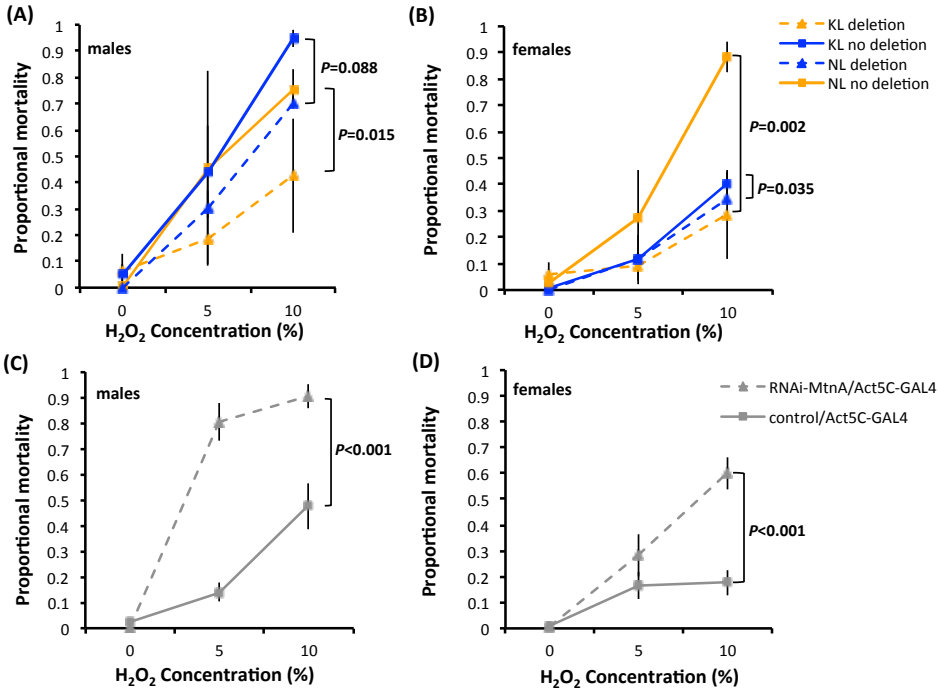
71. Fiston-Lavier A-S, Singh ND, Lipatov M, Petrov DA. *Drosophila melanogaster* recombination rate calculator. Gene. 2010; 463: 18–20. doi: 10.1016/j.gene.2010.04.015 PMID: 20452408



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

The diverse phenotypes found between populations or sexes of the same species are reflected in gene expression variation that is usually achieved through alternative regulation of gene expression. The local and chromosome-wide mechanisms of gene expression regulation could be the results of an assortment of selective forces, including those based upon external physical or internal chromosomal environments.

Chromosome-wide regulation

Tissue-specific genes on the X chromosome

In Chapter 1, I present my research regarding the regulation of tissue-specific genes on the X chromosome. Previously, it was reported that tissue-specific genes, except for ovary-specific genes, are underrepresented on the X chromosome (Mikhaylova and Nurminsky, 2011; Meisel *et al.*, 2012b). Based on that and the observed suppression of X-linked testis-specific genes (Hense *et al.*, 2007; Kemkemer *et al.*, 2014), it was consequently speculated that the expression of tissue-specific genes might be suppressed generally on the X chromosome (Mikhaylova and Nurminsky, 2011). To address this, I used tissue-specific reporter genes that were inserted randomly on the autosomes and on the X chromosome and compared their expression.

Initially, I reanalyzed the expression of a testis-specific reporter gene that was used in the study in which X suppression in the male germline was first demonstrated

(Hense *et al.*, 2007). I compared the expression of autosomal and X-linked reporter genes in whole flies and carcasses with the testes removed and I was able to confirm that X suppression occurs in the male germline (see **Chapter 1**). The X suppression resulted in ~6-fold higher expression on the autosomes than on the X chromosome, which is a difference in expression than would be not be expected even in the absence of DC, as one copy of the reporter gene was tested in each case (**Table 2**). Furthermore, by examining expression in carcasses, I was able to show that there was no difference in the tissue specificity of autosomal and X-linked reporter genes. This was a new result that demonstrated that the X chromosome is not an inherently unfavorable environment for tissue-specific genes.

To further investigate X suppression, I compared the expression of autosomal and X-linked tissue-specific reporter genes (Malpighian tubule, accessory gland, or ovary) in whole flies and carcasses with the tissue of interest removed (see **Chapter 1**). I did not detect any evidence of a chromosome-wide suppression mechanism acting on the X chromosome in any of these tissues. Moreover, I showed that there is no difference in the tissue-specificity of autosomal and X-linked genes, which indicates that promoters of such tissue-specific genes can function equally well on the autosomes or the X chromosome.

In the Malpighian tubule, a somatic tissue present in both sexes that is analogous to the mammalian kidney, I found evidence for partial DC in males (**Table 2**). This finding is not surprising, since DC is known to occur in male somatic tissues. Though, if the expression of Malpighian tubule-specific genes is not suppressed on the X chromosome, it is puzzling why they would be underrepresented on that chromosome, especially if the genes have essential functions in both sexes. Notably, there is an excess of X-linked female-biased genes expressed in the Malpighian tubule, whereas the chromosomal distribution of male-biased genes does not differ from that of genes with unbiased expression (Huylmans and Parsch, 2014). What is more, aside from the expression of genes performing functions expected of this tissue, such as detoxification (Gullan and Cranston, 2010), there is also expression of genes that have sex-specific functions, such as egg production and spermatogenesis (Huylmans and

Parsch, 2014). Likewise, in the mammalian kidney genes with sex-specific functions are also expressed such as hormones involved in sexual maturation.

In accessory gland, a male-limited somatic tissue equivalent to the mammalian prostate gland, not only was there no evidence of X suppression, but instead there was evidence of partial DC (**Table 2**), as the mean expression on the X chromosome was 1.15-fold higher than that on the autosomes. It is not clear if this is the result of DCC-mediated DC, as it is probable that this small degree of increase in expression could be the outcome of basal dosage compensation. Nevertheless, this finding has not been previously described and implies that the strong underrepresentation of X-linked accessory gland-specific genes is not driven by expression restrictions on that chromosome.

In ovary, a female germline tissue, I find neither evidence of X suppression nor higher expression on the X chromosome. This finding implies that the overrepresentation of X-linked ovary-specific genes is not driven by the potential of higher expression levels on that chromosome (**Table 2**).

Table 2. Summary of tissue-specific gene expression regulation on the X chromosome.

Tissue-specificity	Suppression[1]	DC[1]	Representation[3]	Possible explanation
Ovary	no	NA	over	gene-specific sexual antagonism
Accessory gland	no	partial	under	gene-specific sexual antagonism
Malpighian tubule	no	partial	under	gene-specific sexual antagonism
Testis	yes	no	under	chromosome-wide X suppression

[1]Supporting evidence regarding presence (yes), absence (no) or degree (partial) of X suppression and dosage compensation (DC) comes from the different tissue-specific reporter genes used in this dissertation.
[3]Information about the representation (under- or over-) of native *D. melanogaster* tissue-specific genes on the X chromosome was obtained from previous studies (Mikhaylova and Nurminsky, 2011; Meisel *et al.*, 2012b).

A possible explanation for the underrepresentation of accessory gland-specific and Malpighian tubule-specific genes on the X chromosome, as well as the overrepresentation of ovary-specific genes on the X chromosome, could be sexually antagonistic selection in which the fitness effect of an allele differs between males and females. The fact that the X chromosome spends twice as much of its evolutionary time in females as in males may result in the more frequent acquisition of dominant mutations that are beneficial to females and detrimental to males, and more frequent acquisition of recessive mutations that are beneficial to males and detrimental to females (Rice, 1984; Charlesworth *et al.*, 1987; Ellegren and Parsch, 2007) (**Figure 14**). Indeed, there is an enrichment of genes in the *Drosophila* accessory gland that are beneficial for male fertilization success (Innocenti and Morrow, 2010). In the same vein, some of the accessory gland-expressed ACP proteins that are contained in the seminal fluid have been shown to have a negative impact on female fitness (Wolfner, 1997). Characteristically, the gene of which the accessory gland-specific promoter was used here (*Acp26Aa*) acts as a sex peptide that increases egg-laying, but suppresses a female's receptivity to remating.

Similarly, several cases of sexual antagonism have been reported in mammals (reviewed in Mokkonen and Crespi, 2015). A well-studied example is the association of the hormone testosterone and immune function in the bank vole *Myodes glareolus*. The alleles responsible for high testosterone levels in males that are beneficial for reproductive success and social status, and therefore beneficial for their fitness, lead to reduced immune function when they are present in females.

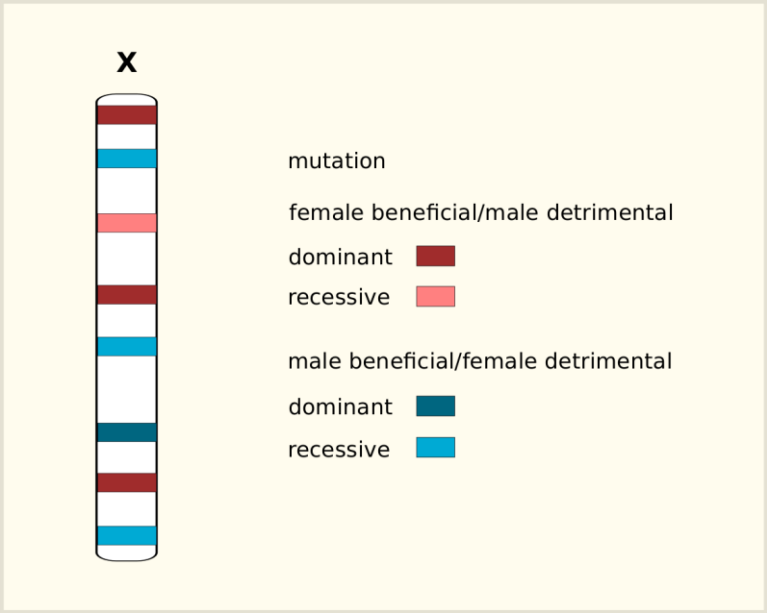


Figure 14. Sexually antagonistic mutations. The mode of inheritance of the X chromosome differs between the sexes. As a consequence, it is expected that dominant mutations (dark) that are female beneficial/male detrimental (red), as well as recessive mutations (light) that are male beneficial/female detrimental (blue), will accumulate on the X chromosome.

Suppression of X-linked genes in the male germline

In Chapter 2, I aimed to better understand the mechanism responsible for the observed X suppression in the male germline. If the entire X chromosome is affected, then a non-native gene that is has at least basal expression in the male germline should also be affected. Additionally, if X suppression is limited to the male germline, then it should not affect the expression of an exogenous gene that is free of tissue-specific and sex-specific regulation in somatic tissues. Surprisingly, I find that when a promoter that is non-native to *Drosophila* is driving the expression of reporter genes in male somatic tissues or the male germline, the reporter gene expression is not affected by the X suppression mechanism that is known to affect other genes in the male germline.

Taken together, these results demonstrate that the X chromosome is suppressed only in the male germline of *D. melanogaster* and not in somatic tissues (Landeen *et al.*, 2016, see **Chapter 1** and **Chapter 2**). Furthermore, X suppression affects testis-

specific genes (Hense *et al.*, 2007; Kemkemer *et al.*, 2014), but not tissue-specific genes as a whole (see **Chapter 1**), and also affects genes with broad expression patterns (Landeen *et al.*, 2016). It can therefore be assumed that X suppression is mediated through a chromosome-wide mechanism present exclusively in testes. I propose a model for the mechanism that leads to X suppression that is based on the interaction of *cis*- and *trans*- elements. The *cis*-regulatory elements must be widely distributed along the X chromosome, as the suppression seems to be global (Kemkemer *et al.*, 2011). Moreover, these elements are likely to be silencers, as they tend to be tissue-specific and can act indirectly to inhibit transcription (see **General Introduction**).

An interesting feature of the X suppression is the varying magnitude of effects that it has on different testis-expressed genes, which is reflected in the findings regarding autosomal and X-linked reporter genes (**Figure 15**). Four testis-specific reporter constructs, one of which one was reanalyzed in this dissertation (*CG7929*) (see **Chapter 1**), show different degrees of X suppression, with their mean expression being 2.8-fold to 5.3-fold higher when autosomal than when X-linked. Additionally, the X suppression mechanism seems to have little or no effect on genes with basal expression, as the reporter gene with a non-native promoter tested here (see **Chapter 2**) shows no sign of X suppression, with its mean autosome-to-X expression ratio being 0.72.

A linear regression analysis (R^2=0.75, P = 0.036) indicated that the magnitude of X suppression could be predicted from the mean autosomal expression level of a gene, which can be used as a proxy of its potential expression level. Put differently, the higher the potential expression level of a gene is, the greater the degree of X suppression. This combination of findings provides support for the hypothesis that the mechanism interferes with genes that exceed a certain expression threshold and reduces their expression level below that threshold (**Figure 16**). The existence of an upper limit of expression on the X chromosome is also hinted at by two observations from an RNA sequencing study in wild-type testis (Gan *et al.*, 2010). First, the median expression level of native X-linked genes is significantly lower when compared to

autosomal genes. Second, the expression of the X-linked gene with the highest expression is still lower than the expression of several autosomal genes.

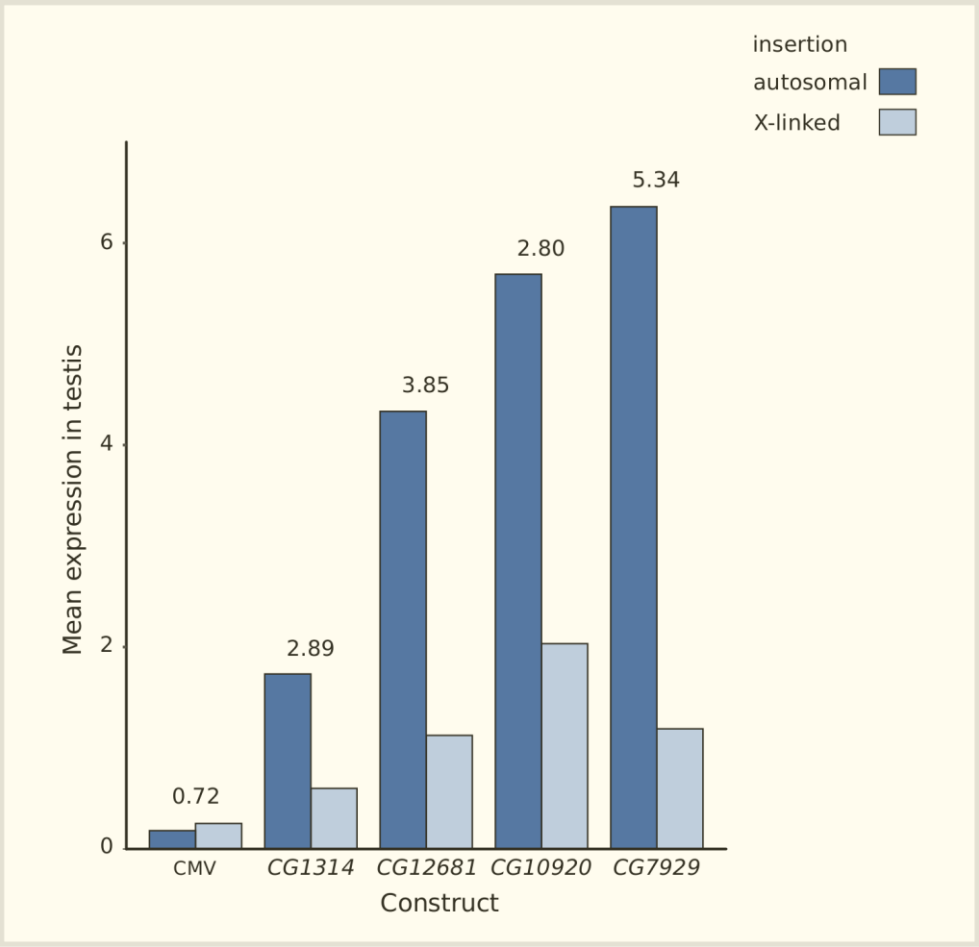


Figure 15. Degree of X suppression in the male germline of *D. melanogaster* as revealed by reporter gene constructs. The mean X-linked (light blue) and autosomal (dark blue) reporter gene expression using male flies or dissected testes (CMV), as well as the autosomal to X-linked expression ratios (number above the bars) are shown. The expression of three previously published testis-specific constructs (*CG1314*, *CG12681*, *CG10920*) (Kemkemer *et al.*, 2014) is shown after normalization (6 flies were used per biological replicate, in contrast to 5 flies for the other constructs). An additional testis-specific construct used in a previous study (Hense *et al.*, 2007) was reanalyzed here (*CG7929*). The ubiquitously-expressed construct (CMV), used in this dissertation, is also shown.

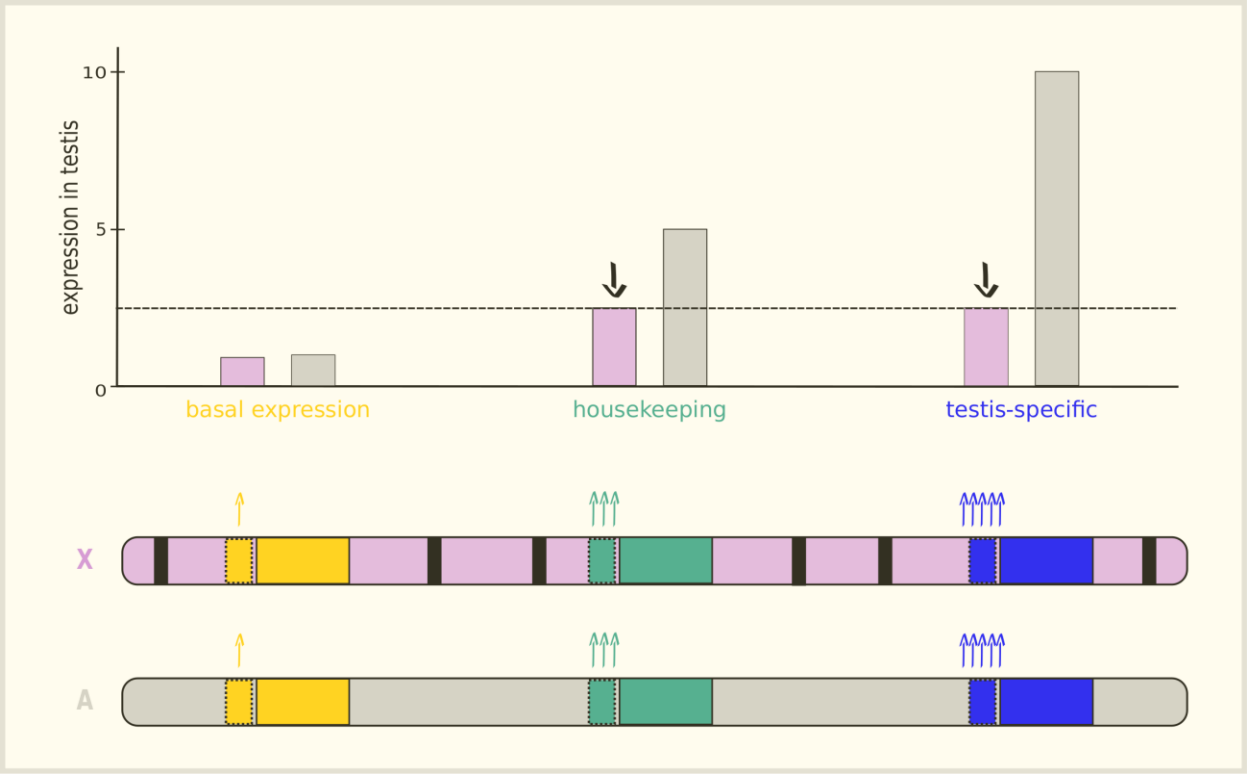


Figure 16. Proposed model for the regulation of the X chromosome in the male germline of *D. melanogaster*. Testis-specific silencer elements (black boxes) enriched along the X chromosome, but not on the autosomes, lead to suppression (black arrows) of genes that are expressed above a certain threshold (dashed line). The promoters (boxes with dotted outline) of housekeeping (green boxes) and testis-specific (blue boxes) genes have the potential to drive moderate and high expression on the autosomes. Genes that are not testis-specific or housekeeping and therefore have only basal expression in testis are not affected by the X suppression, as they do not reach the expression threshold.

It is common for species with heteromorphic sex chromosomes to face selective pressures for inactivation of the sex chromosomes in the gonads of the heterogametic sex. MSCI is present in species that have different sex chromosomes such as mammals with XY in males (see General Introduction), chicken with ZW females (Schoenmakers et al., 2009), and grasshopper (Cabrero et al., 2007) and nematode with X0 males (Bean et al., 2004). Also, it has been suggested that MSCI is a conserved mechanism in vertebrates (Namekawa and Lee, 2009).

X suppression in *Drosophila* is similar to MSCI in that it is present in the male gonads and is a chromosome-wide expression inhibition mechanism. Yet, the two mechanisms are not completely analogous, because they inhibit expression in different extents. MSCI is marked by a massive switch-off of X-linked genes (although some escape inactivation) leading to complete inactivation (da Cruz *et al.*, 2016), while when X suppression occurs the X-linked genes are still expressed but in lower levels than the autosomal genes. Also, it is possible that those mechanisms act at different developmental stages of spermatogenesis. Even though 40% of X-linked genes are downregulated in meiotic cells in *Drosophila* testis (Vibranovski *et al.*, 2009b), the onset of X suppression is not known. It is likely that it is already established in premeiotic cells, including germline and somatic cells, and maintained in meiotic cells (Meiklejohn *et al.*, 2011), while MSCI occurs in prophase of the first meiosis (Turner, 2007). Accordingly, it is not clear whether those mechanisms evolved to serve a similar purpose. In view of the observed patterns, it is plausible to wonder: Why is there a necessity of evolving a spectrum of mechanisms for the suppression of the sex chromosomes in the heterogametic sex? Among several hypotheses that have been put forth to provide evolutionary explanations include non-homologous recombination prevention, sexual antagonism, and as a defense against selfish meiotic drive elements.

Meiotic recombination is essential for the introduction of genetic variability upon which selection for adaptive traits can act. In spite of this, in the heterogametic sex recombination events are commonly suppressed between the non-homologous sex chromosomes as predicted by the Haldane-Huxley rule (Bell, 1982). This is thought to happen as a protective mechanism (e.g. MSCI) to avoid the potentially deleterious

outcomes of aneuploidy, such as developmental and mental retardation in humans (Hassold and Hunt, 2001). However, in *D. melanogaster* recombination of homologous chromosomes does not occur during meiosis in males (reviewed in McKee *et al.*, 2012). So, if the risk of erroneous recombination does not have to be eliminated through the action of suppression mechanisms, this could not provide a good explanation for the existence of X suppression. However, it is possible that the X suppression is an evolutionary leftover. In line with this scenario is the finding that in *Drosophila* testis-specific genes have evolved strong promoters, presumably to compensate for X suppression (Landeen *et al.*, 2016).

Sexual antagonism, namely the situation where traits have different fitness optima among the sexes, could be the driving force of the germline inactivation in the heterogametic sex, known as the SAXI (Sexual Antagonism and X Inactivation) hypothesis (Wu and Xu, 2003). Since the X chromosome spends two thirds of its evolutionary time in females, it could become a favorable environment for genes whose expression is beneficial in females, but deleterious to males. This is supported by the genomic distribution of sex-biased genes (Connallon and Knowles, 2005). Hence, suppression mechanisms could evolve, including X suppression, for genes that are detrimental to male fitness. Notably, some form of X suppression as well as an underrepresentation of X-linked male-biased genes that are mainly testis-expressed has been reported for the mosquito *Anopheles gambiae* (Magnusson *et al.*, 2012). In contrast, even though the mammalian male-biased genes are underrepresented on the X chromosome, those with testis expression are overrepresented (Wang *et al.*, 2001).

Meiotic drive is the phenomenon where an allele of a selfish gene can induce destruction of the portion of gametes that do not carry it. As a result, the frequency of the driver allele can increase in the population, although it decreases the fitness of the individuals who carry it, as half of the gametes are not able to contribute to the reproductive success of those individuals. For instance, in stalk-eyed flies the presence of a driving X chromosome in males leads to eradication of the gametes bearing the Y chromosome (Hurst and Pomiankowski, 1998). Consequently, the progeny has a sex ratio that does not follow Mendelian inheritance with almost all of the offspring being female. If, however, the father has a suppressing Y chromosome the effect of the

driving X chromosome is avoided. Meiotic drive can be harmful for the sperm and the population if there is an unequal sex ratio. Thus, due to selective pressures for sperm survival and for balancing of the sex ratio to 50:50, a protective mechanism could have evolved in which the expression of genes, including of those with driver properties, is suppressed. In agreement with this hypothesis is that the X chromosome is a hotspot for male sterility factors (Presgraves, 2008).

Dosage Compensation

In Chapter 2, I performed a pilot study regarding the influence of DCC-mediated DC on male gene expression in *D. melanogaster*. I show that the proximity of a reporter gene with an exogenous promoter to the HAS, MSL2, and MSL3 binding sites of the DCC does not correlate with the expression in male heads, testes or carcasses containing the rest of the somatic tissues. Nevertheless, distance to the nearest MLE binding site seems to have a monotonic and linear relationship with male somatic expression, with the expression being increased the further away the binding site is located.

Local regulation

When a species expands through colonization of diverse niches, its local subpopulations face distinct selection pressures. These pressures can lead to local adaptation that are often the result of expression variation involving individual genes (Wray *et al.*, 2003). To date, several studies in *Drosophila* detected remarkable expression variation between African and European populations in various tissues (Hutter *et al.*, 2008; Müller *et al.*, 2011; Catalán *et al.*, 2012; Huylmans and Parsch, 2014). Nevertheless, even though many genes can be considered candidates for adaptive regulatory divergence, there is still a lack of specific, well-documented adaptations such as insecticide resistance (Daborn, 2002) and xenobiotic resistance (Mateo *et al.*, 2014).

In Chapter 3, I investigated a case of local adaptation. The *Metallothionein A* gene (*MtnA*) that shows expression variation in natural populations of *D. melanogaster* has an indel polymorphism in its 3’ untranslated region (UTR). The 3' UTR of a gene is

a region containing elements that are often involved in post-transcriptional regulation of expression (reviewed in Barrett et al., 2012). The derived allele has a deletion of 49-bp in the 3' UTR and shows a clinal distribution across Africa and Europe, as well as across North America and Australia, with very low frequencies in sub-Saharan Africa and very high frequencies in northern Europe. We found that the deletion allele leads to 4-fold higher *MtnA* expression compared to the ancestral sequence. In addition, we found that the deletion is associated with increased oxidative stress tolerance. This is suggestive of local adaptation, with the deletion and associated high expression of *MtnA* being favored in temperate, northern environments. Another study provides evidence for increased oxidative stress tolerance in derived *D. melanogaster* populations. In this case, the increased tolerance is associated with the insertion of *Bari-Jheh* transposable element in a *cis*-acting element of *Jheh* genes (Guio *et al.*, 2014). This insertion is also under selection (González *et al.*, 2008). Therefore, it is seems that oxidative stress have been a selective constraint on the colonization of new environments. Oxidative stress can be caused by cytotoxic reactive oxygen species (ROS) which are produced through regular metabolic processes but also through environmental factors such as radiation, UV light or exposure to toxins. Given the wide range of stressors that could lead to an increased production of ROS, it is difficult to determine which one (or perhaps more than one) was the key stressor in the *MtnA* case. A potential stressor that is more prominent within some regions, such as northern Europe, could be the exposure to cold temperatures or thermal fluctuations which has been shown to induce oxidative stress in insects (Lalouette *et al.*, 2011).

One of the well-described examples for human local adaptations is also characterized by population variation of allele frequencies. The lactase persistence trait, i.e. the ability to digest lactose as an adult, achieved through expression of the LCT gene encoding the enzyme lactase, is commonly found in populations that have traditionally domesticated cattle. LCT expression is regulated by *cis*-regulatory elements and some variants have been linked to persistence of expression in adulthood (Tishkoff *et al.*, 2007). The frequency of these regulatory variants is very high (>90%) in northern European and some African populations that regularly consume milk. In contrast, the variants are found in intermediate frequencies in

southern Europe and the Middle East and in low frequencies in Asian and African populations that have different dietary habits.

Although expression variation is often caused by regulation of expression at the transcriptional level via variation in the proximal promoter (see **General Introduction**), this is not the case for the *MtnA* gene as the large deletion in the 3'UTR is likely to affect regulation at the posttranscriptional level, for example through deletion of a micro RNA binding site. It is therefore essential to study individual candidate genes since local regulation might be achieved by very diverse means.

Outlook

The analysis of native tissue-specific and non-native promoters in *D. melanogaster* undertaken here has extended our knowledge about the phenomenon of X suppression in the male germline. Nonetheless, the genes responsible for X suppression and the precise molecular mechanism through which it is achieved remain to be elucidated. Future research could approach this topic by focusing on determining genes that have *trans* effects in the mechanism. A promising methodology could involve mutagenesis of a fly line carrying an X-linked testis-specific reporter gene with a known expression pattern. Random point mutations or deletions in the genome could easily be induced, for example with the use of a mutagenic agent (Phadnis *et al.*, 2016). If a resulting mutation is introduced within *trans*-acting factors and the suppression pathway is interrupted, then this will be manifested on the expression level of the reporter gene by reaching expression levels similar to those of autosomal reporter genes. The identification of the mutated regions will be the first step for uncovering the molecular mechanism of X suppression.

My research on the impact of the DCC-mediated DC on the genomic distribution of male-biased genes provides a framework for the exploration of the relationship between sex-biased gene expression and genomic location. A future study with an experimental design akin to the one applied here could test male-to-female expression ratios of reporter genes in multiple somatic tissues. To have sufficient statistical power, this should be carried out with approximately 80 fly lines. In addition, an upper

threshold of 50 kb for the gene proximity to DCC binding site should be set so that the results can be compared with previous studies. Lastly, the numerous MLE binding sites that do not coincide with the HAS (colocalization of MSL2 and MLE) might not be informative or even misleading with regard to the DCC binding, as their function is unknown. For that reason, I recommend that in a future study only MSL2 and HAS binding sites be used, as they are likely to be the best indicators of DCC binding.

Final Remarks

The following conclusions can be drawn from this dissertation.

- Local regulation of gene expression can be achieved via variation in different types of functional genetic elements, such as upstream regulatory sequences or untranslated regions. Therefore, in order to understand the environmental pressures and the underlying molecular mechanisms of local adaptation, it is important to inspect individual candidate genes.
- Chromosome-wide regulation of gene expression can differ between males and females, as well as across different tissues. Furthermore, it can affect the gene content of specific chromosomes, which is evident from the genomic distribution of tissue-specific and sex-biased genes.
- Patterns of gene expression on the X chromosome of *D. melanogaster* are the result of a complex interaction between sex, tissue and gene expression level.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Bachtrog D (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* **14**: 113–124.

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.

Barrett LW, Fletcher S, Wilton SD (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci* **69**: 3613–3634.

Bean CJ, Schaner CE, Kelly WG (2004). Meiotic pairing and imprinted X chromatin assembly in *Caenorhabditis elegans*. *Nat Genet* **36**: 100–105.

Bell G (1982). *The Masterpiece of Nature: the Evolution and Genetics of Sexuality*. University of California Press: Berkeley.

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

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

Biggin MD, Tjian R (2001). Transcriptional regulation in *Drosophila*: The post-genome challenge. *Funct Integr Genomics* **1**: 223–234.

Brockdorff N, Turner BM (2015). Dosage compensation in mammals. *Cold Spring Harb Perspect Biol* **7**: a019406.

Cabrero J, Teruel M, Carmona FD, Jiménez R, Camacho JPM (2007). Histone H3 lysine 9 acetylation pattern suggests that X and B chromosomes are silenced during entire male meiosis in a grasshopper. *Cytogenet Genome Res* **119**: 135–142.

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

Catania F, Kauer MO, Daborn PJ, Yen JL, Ffrench-Constant RH, Schlötterer C (2004). World-wide survey of an *Accord* insertion and its association with DDT resistance in *Drosophila melanogaster*. *Mol Ecol* **13**: 2491–2504.

Chang PL, Dunham JP, Nuzhdin S V, Arbeitman MN (2011). Somatic sex-specific transcriptome differences in *Drosophila* revealed by whole transcriptome sequencing. *BMC Genomics* **12**: 364.

Charlesworth D (2002). Plant sex determination and sex chromosomes. *Heredity (Edinb)* **88**: 94–101.

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

Chung H, Bogwitz MR, McCart C, Andrianopoulos A, Ffrench-Constant RH, Batterham P, *et al.* (2007). *Cis*-regulatory elements in the *Accord* retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics* **175**: 1071–1077.

Connallon T, Knowles LL (2005). Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends Genet* **21**: 495–499.

Conrad T, Akhtar A (2012). Dosage compensation in *Drosophila melanogaster*: epigenetic fine-tuning of chromosome-wide transcription. *Nat Rev Genet* **13**: 123–134.

da Cruz I, Rodríguez-Casuriaga R, Santiñaque FF, Farías J, Curti G, Capoano CA, *et al.* (2016). Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to postmeiotic-related processes at pachytene stage. *BMC Genomics* **17**: 294.

Daborn PJ (2002). A single P450 allele associated with insecticide resistance in *Drosophila*. *Science* **297**: 2253–2256.

Ellegren H, Parsch J (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet* **8**: 689–698.

Ercan S (2015). Mechanisms of X chromosome dosage compensation. *J Genomics* **3**: 1–19.

Ferrari F, Alekseyenko AA, Park PJ, Kuroda MI (2014). Transcriptional control of a whole chromosome: emerging models for dosage compensation. *Nat Struct Mol Biol* **21**: 118–125.

Gallach M, Domingues S, Betrán E (2011). Gene duplication and the genome distribution of sex-biased genes. *Int J Evol Biol* **2011**: 1–20.

Gan Q, Chepelev I, Wei G, Tarayrah L, Cui K, Zhao K, *et al.* (2010). Dynamic regulation of alternative splicing and chromatin structure in *Drosophila* gonads revealed by RNA-seq. *Cell Res* **20**: 763–783.

González J, Lenkov K, Lipatov M, Macpherson JM, Petrov DA (2008). High rate of recent transposable element-induced adaptation in *Drosophila melanogaster*. *PLoS Biol* **6**: e251.

Guio L, Barrón MG, González J (2014). The transposable element *Bari-Jheh* mediates oxidative stress response in *Drosophila*. *Mol Ecol* **23**: 2020–2030.

Gullan PJ, Cranston P s. (2010). *The Insects: An Outline of Entomology*, 4th edn. Wiley-Blackwell.

Hassold T, Hunt P (2001). To err (meiotically) is human: the genesis of human aneuploidy. *Nat Rev Genet* **2**: 280–291.

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

Hurst LD, Pomiankowski A (1998). The eyes have it. *Nature* **391**: 223–224.

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.

Huylmans AK, Parsch J (2014). Population- and sex-biased gene expression in the excretion organs of *Drosophila melanogaster*. *G3* **4**: 2307–2315.

Huylmans AK, Parsch J (2015). Variation in the X:autosome distribution of male-biased genes among *Drosophila melanogaster* tissues and its relationship with dosage compensation. *Genome Biol Evol* **7**: 1960–1971.

Innocenti P, Morrow EH (2010). The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biol* **8**: e1000335.

Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schütz F, *et al.* (2012). Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol* **10**: e1001328.

Kaiser VB, Bachtrog D (2010). Evolution of sex chromosomes in insects. *Annu Rev Genet* **44**: 91–112.

Kemkemer C, Catalán A, Parsch J (2014). 'Escaping' the X chromosome leads to increased gene expression in the male germline of *Drosophila melanogaster*. *Heredity (Edinb)* **112**: 149–155.

Kemkemer C, Hense W, Parsch J (2011). Fine-scale analysis of X chromosome inactivation in the male germline of *Drosophila melanogaster*. *Mol Biol Evol* **28**: 1561–1563.

Lalouette L, Williams CM, Hervant F, Sinclair BJ, Renault D (2011). Metabolic rate and oxidative stress in insects exposed to low temperature thermal fluctuations. *Comp Biochem Physiol Part A Mol Integr Physiol* **158**: 229–234.

Landeen EL, Muirhead CA, Wright L, Meiklejohn CD, Presgraves DC (2016). Sex chromosome-wide transcriptional suppression and compensatory *cis*-regulatory evolution mediate gene expression in the *Drosophila* male germline. *PLoS Biol* **14**: e1002499.

Latchman DS (2015). *Gene Control*, 2nd edn. Garland Science.

Lee TI, Young RA (2014). Transcriptional regulation and its misregulation in disease. *Cell* **152**: 1237–1251.

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

Lifschytz E, Lindsley DL (1972). The role of X-chromosome inactivation during spermatogenesis. *Proc Natl Acad Sci* **69**: 182–186.

Llopart A (2012). The rapid evolution of X-linked male-biased gene expression and the large-X effect in *Drosophila yakuba*, *D. santomea*, and their hybrids. *Molecular Biology and Evolution* **29**: 3873–3886.

Magnusson K, Lycett GJ, Mendes AM, Lynd A, Papathanos P-A, Crisanti A, *et al.* (2012). Demasculinization of the *Anopheles gambiae* X chromosome. *BMC Evol Biol* **12**: 69.

Manolakou P, Lavranos G, Angelopoulou R (2006). Molecular patterns of sex determination in the animal kingdom: a comparative study of the biology of reproduction. *Reprod Biol Endocrinol* **4**: 59.

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

McKee BD, Handel MA (1993). Sex chromosomes, recombination, and chromatin conformation. *Chromosoma* **102**: 71–80.

McKee BD, Yan R, Tsai J-H (2012). Meiosis in male *Drosophila*. *Spermatogenesis* **2**: 167–184.

Meiklejohn CD, Landeen EL, Cook JM, Kingan SB, Presgraves DC (2011). Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. *PLoS Biol* **9**: e1001126.

Meisel RP, Connallon T (2013). The faster-X effect: Integrating theory and data. *Trends Genet* **29**: 537–544.

Meisel RP, Malone JH, Clark AG (2012a). Faster-X evolution of gene expression in *Drosophila*. *PLoS Genet* **8**: e1003013.

Meisel RP, Malone JH, Clark AG (2012b). Disentangling the relationship between sex-biased gene expression and X-linkage. *Genome Res* **22**: 1255–1265.

Mikhaylova LM, Nurminsky DI (2011). Lack of global meiotic sex chromosome inactivation, and paucity of tissue-specific gene expression on the *Drosophila* X chromosome. *BMC Biol* **9**: 29.

Mikhaylova LM, Nurminsky DI (2012). No severe and global X chromosome inactivation in meiotic male germline of *Drosophila*. *BMC Biol* **10**: 50.

Mokkonen M, Crespi BJ (2015). Genomic conflicts and sexual antagonism in human health: Insights from oxytocin and testosterone. *Evol Appl* **8**: 307–325.

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

Namekawa SH, Lee JT (2009). XY and ZW: Is meiotic sex chromosome inactivation the rule in evolution? *PLoS Genet* **5**: e1000493.

Parisi M (2003). Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* **299**: 697–700.

Parsch J (2004). Functional analysis of *Drosophila melanogaster* gene regulatory sequences by transgene coplacement. *Genetics* **168**: 559–561.

Parsch J (2009). X chromosome: Expression and escape. *PLoS Genet* **5**: e1000724.

Phadnis N, Baker EP, Cooper JC, Frizzell KA, Flor A, Cruz A De, *et al.* (2016). An essential cell cycle regulation gene causes hybrid inviability in *Drosophila*. *Science* **350**: 1552–1555.

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

Presgraves DC (2008). Sex chromosomes and speciation in *Drosophila*. *Trends Genet* **24**: 336–343.

Ramirez F, Lingg T, Toscano S, Lam KC, Georgiev P, Chung H-R, *et al.* (2015). High-affinity sites form an interaction network to facilitate spreading of the MSL complex across the X chromosome in *Drosophila*. *Mol Cell* **60**: 146–162.

Reiter LT, Potocki L, Chien S, Gribskov M, Bier E (2001). A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res* **11**: 1114–25.

Rice WR (1984). Sex chromosomes and the evolution of sexual dimorphism. *Evolution (N Y)* **38**: 735–742.

Schauer T, Ghavi-Helm Y, Sexton T, Albig C, Regnard C, Cavalli G, *et al.* (2017). Chromosome topology guides the *Drosophila* dosage compensation complex for target gene activation. *EMBO Rep*: e201744292.

Schoenmakers S, Wassenaar E, Hoogerbrugge JW, Laven JSE, Grootegoed JA, Baarends WM (2009). Female meiotic sex chromosome inactivation in chicken. *PLoS Genet* **5**: e1000466.

Schwartz YB, Cavalli G (2017). Three-dimensional genome organization and function in *Drosophila*. *Genetics* **205**: 5–24.

Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB (2008). The chromosomal high-affinity binding sites for the *Drosophila* dosage compensation complex. *PLoS Genet* **4**: e1000302.

Straub T, Zabel A, Gilfillan GD, Feller C, Becker PB (2013). Different chromatin interfaces of the *Drosophila* dosage compensation complex revealed by high-shear ChIP-seq. *Genome Res* **23**: 473–485.

Sturgill D, Zhang Y, Parisi M, Brian O (2007). Demasculinization of X chromosomes in the *Drosophila* genus. *Nature* **450**: 238–241.

Theuns J, Van Broeckhoven C (2000). Transcriptional regulation of Alzheimer's disease genes: implications for susceptibility. *Hum Mol Genet* **9**: 2383–2394.

Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, *et al.* (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nat Genet* **39**: 31–40.

Turner JMA (2007). Meiotic sex chromosome inactivation. *Development* **134**: 1823–1831.

Vibranovski MD (2014). Meiotic sex chromosome inactivation in *Drosophila*. *J Genomics* **2**: 104–117.

Vibranovski MD, Lopes HF, Karr TL, Long M (2009). Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. *PLoS Genet* **5**: e1000731.

Vibranovski MD, Zhang YE, Kemkemer C, Lopes HF, Karr TL, Long M (2012). Re-analysis of the larval testis data on meiotic sex chromosome inactivation revealed evidence for tissue-specific gene expression related to the *Drosophila* X chromosome. *BMC Biol* **10**: 49; author reply 50.

Vibranovski MD, Zhang Y, Long M (2009). Out of the X chromosomal gene movement in the *Drosophila* genus. *Genome Res* **19**: 897–903.

Vicoso B, Charlesworth B (2006). Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet* **7**: 645–653.

Wang PJ, McCarrey JR, Yang F, Page DC (2001). An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* **27**: 422–426.

Wolfner MF (1997). Tokens of love: Functions and regulation of *Drosophila* male accessory gland products. *Insect Biochem Mol Biol* **27**: 179–192.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman M V., *et al.* (2003). The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**: 1377–1419.

Wright AE, Dean R, Zimmer F, Mank JE (2016). How to make a sex chromosome. *Nat Commun* **7**: 12087.

Wu CI, Xu EY (2003). Sexual antagonism and X inactivation - The SAXI hypothesis. *Trends Genet* **19**: 243–247.

Zhang Y, Malone JH, Powell SK, Periwal V, Spana E, MacAlpine DM, *et al.* (2010). Expression in aneuploid *Drosophila* S2 cells. *PLoS Biol* **8**: e1000320.

Acknowledgments

First, I would like to thank John for his guidance during my PhD studies. His hands-off supervision together with the knowledge that his office door is always open for any questions made me a better researcher and a tougher person. I consider myself lucky for having such a great supervisor while surrounded from nice people at the workplace, namely the Parsch girls. Thank you Ina, Amanda, Annabella, Sonja, and Ana, it has been an amazing experience to work with you, and have fun with you!

I am grateful to Hilde for the excellent help in the lab, but also for the psychological support! I would also like to thank the members of the evolutionary biology group: especially Soumya and Myriam for the 'group therapy' coffee breaks, and Dirk for all the helpful advice and fun discussions.

I thank my friends Andri, Rebekka, Amanda and Kasia, as well as my brother, for their online-support and not letting the distance become an obstacle. Thank you for always reminding me about the big picture and having faith in my abilities.

Last but not least, I express my deep gratitude to my beloved Haris for his unlimited patience and support throughout the last four years. Thank you for believing in me and standing by me all the way. I doubt that this dissertation would have been completed without your support!

This dissertation is dedicated to my beloved parents because their generosity over the years has been priceless. I appreciate all the sacrifices you made in your lives to ensure the best future possible for me. You taught me how to love and respect nature which was a milestone in becoming a Biologist.

Curriculum Vitae

Date of birth	27/06/1989
Place of birth	Strovolos, Cyprus
Citizenship	Cypriot

Education

11/2013 - present	**PhD Candidate in Biology** Ludwig-Maximilians-Universität (LMU) Munich, Germany Supervisor: Prof. Dr. John Parsch Topic: Local and Chromosome-wide Regulation of Gene Expression in *Drosophila melanogaster*
10/2011 - 10/2013	**Master of Science in Evolution, Ecology and Systematics** LMU Munich, Germany Thesis: The effect of X-linkage on ovary-specific gene expression in *Drosophila melanogaster*
09/2007 - 06/2011	**Bachelor of Science in Biological Sciences** University of Cyprus, Cyprus Thesis: Investigation of the genetic basis of a melanistic phenotype in *Sylvia melanocephala* species within Paphos' populations

Work and Teaching Experience

Sum. Semester 2017	**Evolutionary Biology Tutor** Bachelor's level practical, LMU Munich, Germany
04/2014 - 09/2016	**Student Supervision** 3 Master's theses, 6 Master's research projects, 1 Bachelor's thesis
01/2012 - 08/2013	**Research Laboratory Assistant** Evolutionary & Functional Genomics, LMU Munich, Germany
06/2010 - 08/2010	**Field & Laboratory Assistant** AP Marine Environmental Consultancy Ltd, Cyprus macrofauna sorting, recording fish morphological parameters

06/2009 - 07/2009	**Medical Laboratory Assistant** Histopathology & Cytology Medical Centre Dr. T. Lyssiotis, Cyprus histopathological routine; tissue biopsy, Pap test

Skills and Qualifications

Computing	MS Office, Graphics Applications (Photoshop, Gimp, Inkscape), CLC Main Workbench, Perl and R programming (basic)
Laboratory	RNA, DNA and protein extraction (blood and tissue), PCR, sequencing, genotyping, fragment analysis (STR), inverse PCR, qRT-PCR, enzymatic activity assay, molecular cloning, embryo microinjections, *P*-element transgenesis, fly tissue dissection
Field	bird mist netting and ringing, bird blood sampling and recording morphological parameters, *Drosophila* sampling and species identification
Languages	Greek (native), English (fluent), German (advanced), Spanish (intermediate)

Conferences

Poster presentation	**EMBO Meeting**, Mannheim, Germany 2016
Poster presentation	**ESEB Conference**, Lausanne, Switzerland 2015
Poster presentation	DFG FOR 1078 Final Meeting and Symposium 'Population and Evolutionary Genomics', Tutzing, Germany 2015

Publications

Argyridou E, Huylmans AK, Königer A, Parsch J. X-linkage is not a general inhibitor of tissue-specific gene expression in *Drosophila melanogaster*. *Heredity* 2017; 119: 27-34.

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.

Voluntary Activity

Bavarian Red Cross	Paramedic