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

Functional analysis of X-chromosomal gene expression in *Drosophila melanogaster*

Claus Kemkemer

aus

Neu-Ulm, Deutschland

2011-03-31

Erklärung:

Diese Dissertation wurde im Sinne von § 12 der Promotionsordnung von Prof. Dr. John 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.

Ehrenwörtliche Versicherung:

Ich versichere ferner hiermit ehrenwörtlich, dass die vorgelegte Dissertation von mir selbstständig, ohne unerlaubte Hilfe angefertigt wurde.

München, den 2011-03-31

Claus Kemkemer

1. Gutachter: Prof. Dr. John Parsch
2. Gutachter: Prof. Dr. Susanne Renner

Dissertation eingereicht am: 2011-03-31

Datum der Disputation: 2011-05-18

1. Table of contents

2. Note

In this dissertation I present my doctoral research, all of which has been done by myself. Prof. Dr. John Parsch assisted with writing the research article cited below that served as the basis for a portion of this dissertation. In addition, Dr. Winfried Hense provided reagents (plasmids and *Drosophila* stocks) that were used in the portion of my research described in the publication cited below.

The results from my dissertation have contributed to the following publication:

Kemkemer C, Hense W, Parsch J. Fine-scale analysis of X chromosome inactivation in the male germline of *Drosophila melanogaster*. Mol Biol Evol. 2010 Dec 30. [Epub ahead of print]

3. List of abbreviations

Abbreviation	Description
A	Autosome
BLAST	Basic local alignment search tool
bp	Base pair
C. elegans	*Caenorhabditis elegans*
CLR	Composite likelihood ratio
D.	*Drosophila*
Δ2-3	*Δ2-3 transposase* fragment, used for *P* element transformation
DCC	Dosage compensation complex
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
h	Dominance factor
Mb / Kb	Mega basepair / Kilo basepair
mRNA	Messenger ribonucleic acid
MSCI	Meiotic sex chromosome inactivation
MSL	Male-specific lethal
MWW	Mann-Whitney-Wilcoxon
M_x / m_x	Sexual antagonistic gene beneficial in males & detrimental in females
N_e	Effective population size
ocn	*ocnus* gene, *CG7929*
PCR	Polymerase chain reaction
qt-PCR	Quantitative reverse transcription polymerase chain reaction
RNA	Ribonucleic acid
sb	Stubble bristle phenotype, bristles on the back
SNP	Single nucleotide polymorphism

Su^F / f	Female sterility gene
TM6	Balancer chromosome
UCSC	University of California, Santa Cruz
UTR	Untranslated region
w	white phenotype, white eyes
WT	Wild type
X	X chromosome
Xist	X-inactive specific transcript
Y	Y chromosome
y	yellow phenotype, yellow body color
ZH-68E	ϕC31 landing site, 3rd chromosome
ZH-86Fb	ϕC31 landing site, 3rd chromosome

4. Figure and table list

5. Zusammenfassung

Die Geschlechtsbestimmung mittels XY Chromosomen findet sich in vielen Organismen wieder, wie zum Beispiel *Drosophila* und Säugetieren und ist ein weit verbreiteter Mechanismus zur Bestimmung des Geschlechts. Einer der wichtigsten Merkmale ist, dass sich das X Chromosom im männlichen Individuum im hemizygoten Zustand befindet. Dieser Unterschied zwischen den weiblichen Geschlecht (XX) und dem männlichen Geschlecht (XY) bewirkt, dass sich das X Chromosom im Vergleich zu den Autosomen evolutionär unterschiedlich entwickelt. Zum Beispiel, wurde ein signifikanter Überschuss an retrotransponierten Genen gefunden, die sowohl in den Testes exprimiert sind, als auch vom X Chromosom zu den Autosomen transponiert wurden. Zusätzlich besitzt das X Chromosomen einen Mangel an männlich-spezifischen Genen. Eine mögliche Erklärung für diese Beobachtungen ist die X Inaktivierungs-Hypothese. Diese Hypothese sagt vorher, dass Gene die spät in der Spermatogenese exprimiert werden, einen Vorteil erlangen, wenn sie vom X Chromosomen weg transponiert werden. Die Transposition erlaubt es den männlich-spezifischen Genen der Inaktivierung der Expression des X Chromosomen in der Keimbahn zu entkommen. Aufgrund der „Flucht“ weg vom X Chromosomen, wird es den testes-spezifisch exprimierten X-chromosomalen Gene ermöglich eine höhere Expression zu erreichen, was einen adaptiven Vorteil mit sich bringen kann. Dieser Vorteil wird durch die neue Umgebung der Autosomen erzielt, welche keine meiotische Geschlechtschromosomen X Inaktivierung besitzen. Des Weiteren, bietet das X Chromosom eine einzigartige Umgebung hinsichtlich Selektion und Expression an. Anhand früherer Resultate unserer Arbeitsgruppe wurden X-chromosomale Gene identifiziert, welche eine unterschiedliche Expression zwischen einer europäischen Population und einer afrikanischen Population von *D. melanogaster* zeigten. Die Kolonisierung Europas durch die einwandernde ursprüngliche afrikanische Population könnte Spuren der Adaption an die neue europäische Umgebung im europäischen Genom hinterlassen. Im Speziellen, könnten veränderte Expressionsmuster und positiv selektionierte *cis*-regulatorische Sequenzen betroffen sein. Die mutmaßlichen Promotoren wurden auf Anzeichen positiver Selektion untersucht.

Um die X Inaktivierung in *Drosophila melanogaster* zu testen, benutzte ich den autosomalen Promoter des testes-spezifischen Gens *ocnus*. Der Promotor wurde zur Regulierung der Expression des Reportergens *lacZ* verwendet. Dieses Promotor Reportergen-Konstrukt wurde

in einen transposablen Elementvektor eingefügt und an eine zufällig Position im *D. melanogaster* Genom transponiert. Die Reportergen Expression war signifikant höher für autosomale Insertionen im Vergleich zu X-chromosomalen Insertionen. Dieses Ergebnis ist in Übereinstimmung mit der X-chromosomalen Inaktivierungs-Hypothese in der männlichen Keimbahn. Im Verlauf dieser Arbeit kartierte ich 112 unabhängige X-chromosomale Reportergene, alle zeigten ein geringeres Expressionslevel. Der durchschnittliche Abstand zwischen zwei Insertionen betrug in etwa 200 Kb. Die Expressionswerte aller 112 Reportergene zeigten, dass die X Inaktivierung eine globale Eigenschaft des X Chromosomen ist und keine Region auf dem X Chromosom der Inaktivierung entkommen kann. Des Weiteren konnte ich beweisen, dass die Anhäufung von neu entwickelten testes-spezifischen Genen in der zytologischen Bande 19 des X Chromosomen ihre Ursache nicht in *cis*-regulatorische Sequenzen besitzt. Diese *cis*-regulatorischen Sequenzen würden es den Genen in der zytologischen Bande 19 erlauben, die transkriptionelle Inaktivierung zu überwinden.

Der oben beschriebene Ansatz wurde benutzt um die Reportergen Expression von drei verschiedenen testes-spezifischen X-chromosomalen Genen (*CG10920*, *CG12681*, *CG1314*) zu untersuchen. In allen Fällen war die Expression X-chromosomaler Insertionen im Vergleich zur Expression autosomaler Insertionen signifikant erniedrigt. Dies beweist, dass die Transposition weg vom X Chromosomen einen Vorteil hinsichtlich des Levels der Genexpression mit sich bringen kann und in Übereinstimmung mit den Vorhersagen der X Inaktivierungs-Hypothese ist. Diese Hypothese erklärt den Überschuss an X Chromosom zu Autosom Transpositionen. Die meiotische Geschlechtschromosomen X-Inaktivierung wurde erstmal in Säugetieren beschrieben. Der Mechanismus, welcher in Säugetieren vorhanden ist, kann nicht vollständig zur Erklärung der von mir gefundenen Ergebnisse herangezogen werden. Durch die Analyse von stadiumsspezifischen Expressionsmustern konnte ich zeigen, dass die X-chromosomale Inaktivierung auch in den mitotischen Zellen vorhanden ist und dies im Widerspruch zur gefundenen X-chromosomalen Inaktivierung ist, wie sie in Säugetieren gefunden wurde. In Säugetieren betrifft die X-chromosomale Inaktivierung ausschließlich die meiotischen Zellen der Keimbahn. Die Schlussfolgerung aus den beschriebenen Ergebnissen ist, dass sich ein unabhängiger Mechanismus zur X-chromosomalen Inaktivierung in *Drosophila* entwickelt hat, der Ähnlichkeiten mit dem Mechanismus in Säugetieren hat, wie zum Beispiel die Inaktivierung der meiotischen Zellen der Keimbahn.

Durch die Suche nach positiv selektionierten *cis*-regulatorischen Sequenzen zwischen europäischen und afrikanischen *Drosophila* Populationen wurde ein Kandidatengen (*CG9509*) gefunden. Dieses Gen zeigte eine höhere Expression in der europäischen Population, als auch Hinweise für positive Selektion der *cis*-regulatorischen Sequenz in der europäischen Population. Um den Nachweis zu erbringen, dass die *cis*-regulatorische Sequenz aus der europäischen Population für die Expressionsunterschiede verantwortlich ist, als auch für das gefundene Selektionsmuster, habe ich beide mutmaßlichen Promotorregionen, welche mit dem Reportergen *lacZ* verknüpft wurden, in einem genetisch uniformen Hintergrund getestet. Die Experimente zeigten einen signifikant höhere Expression für den europäischen Promotor im Vergleich zum afrikanischen Promotor. Diese höhere Expression des europäischen Promotors ist ausschließlich durch eine veränderte europäische *cis*-regulatorische Sequenz erklär bar, weil außer den jeweils populationsspezifischen Promotoren ein genetisch uniformer Hintergrund bestand. Die Expressionsergebnisse erklärten auch das in der europäischen Population gefundene Selektionsmuster.

6. Abstract

Like mammals, *Drosophila* has XY sex determination with the X chromosome hemizygous in males. This difference between the sexes may cause the X chromosome to evolve differently than the autosomes. For example, there is a significant excess of retroposed genes, many of which are expressed in testis, that have moved from the X chromosome to the autosomes. Furthermore, transcriptomic studies have shown that genes with male-biased expression are underrepresented on the X chromosome. A possible explanation for these observations is the X-inactivation hypothesis, which proposes that genes with functions late in spermatogenesis benefit from "escaping" the X chromosome, because otherwise their expression would be limited by male germline X-inactivation. The testis-expressed genes that escape the X chromosome may thus gain a selective advantage due to the increased expression of the new environment of the autosomes, which are not subject to MSCI (meiotic sex chromosome X inactivation). The X chromosome also offers a unique environment in terms of selection and expression. The colonization of Europe by the ancestral migrating African *D. melanogaster* population is expected to have left traces of adaptation to the new European environment in the European genome, including altered expression patterns and positively selected *cis*-regulatory sequences. Previous studies of gene expression and DNA sequence polymorphism identified an X-linked gene (*CG9509*) that appears to have been the target of a selective sweep in the European population.

To investigate X chromosome inactivation in *Drosophila*, I used the promoter of the autosomal testis-specific gene *ocnus* to drive expression of the *lacZ* gene. This promoter reporter construct was inserted into a transposable element vector and inserted randomly into the *D. melanogaster* genome. Reporter gene expression was significantly higher for autosomal inserts than for X-linked inserts, which is consistent with X chromosome inactivation hypothesis in the male germline. I mapped 112 independent reporter gene insertions on the X chromosome, all of which showed very low levels of expression. The average spacing between the X-linked insertions was ~200 Kb. This suggests that the silencing of gene expression is a global property of the X chromosome and that no regions escape inactivation. Furthermore, I found that the hotspot of newly-evolved testis expressed genes at cytological band 19 on the X chromosome was not due to this region of the genome escaping X chromosome inactivation in the male germline.

The above approach was also used to test reporter gene expression driven by the promoters of three different X-linked testis expressed genes (*CG10920, CG12681, CG1314*). In all cases, autosomal inserts showed significantly higher expression than X-linked inserts. This demonstrates that escape from the X chromosome can provide a direct advantage with respect to gene expression levels in testis and is consistent with the predictions of the X-inactivation hypothesis to explain the observed excess of duplicate genes that have moved from the X chromosome to the autosomes. However, I found that MSCI, which was first described in mammals, cannot completely explain the reduced expression of X-linked inserts compared to autosomal inserts, as the difference is present even in pre-meiotic stages of spermatogenesis. This suggests that the suppression of X-linked gene expression in the male germline occurs through different mechanisms in *Drosophila* and mammals.

Statistical analysis of DNA sequence polymorphism on the X chromosome revealed evidence for positive selection in the region containing the gene *CG9509*. This gene shows higher expression in the European population than in the African population and its upstream regulatory sequence appears to have been the target of a selective sweep in the European population. To determine if the putative promoter region is responsible for the observed expression difference between the European and African populations, I tested both promoter variants, which were linked to the reporter gene *lacZ*, in a uniform genetic background. The European promoter drove significantly higher expression than the African promoter. This higher expression for the European promoter indicates that the higher expression in the European population is due to the altered European *cis*-regulatory sequence and suggests that positive selection acted to increase *CG9509* expression in Europe.

7. Introduction

7.1 Sex chromosomes

Sex-chromosome systems have evolved independently many times and are present in many diverse taxa including mammals, insects, birds and plants. Two different sex chromosome systems are distinguished by the chromosome complement of the heterogametic sex. First, when the female is the heterogametic sex, the sex chromosomes are designated Z and W, as is the case in birds. Second, when the male is the heterogametic sex, the sex chromosomes are designated X and Y, as is the case in mammals and *Drosophila*. The evolution of sex chromosomes appears to follow a standard process (Figure 1). It starts with the formation of a sex-determining region linked to a sterility gene on an ordinary chromosome (autosome). To maintain the location of the sex-determining region, this region is not allowed to recombine (Nei 1969) and the continuation of this process leads to the decline of recombination in this region and perhaps in the surrounding regions (Charlesworth *et al.* 2005). The newly-formed proto-sex chromosome with the sex-determining region accumulates mutations that are beneficial for one sex, but detrimental for the other sex (e.g. for the proto-Y, male beneficial/female detrimental mutations). This accumulation extends the decline of recombination outside of the sex-determining region and eventually leads to the loss of recombination on the entire sex chromosome. The final step in this process is the genetic degeneration of the sex chromosome due to the lack of recombination and the accumulation of deleterious mutations and, possibly, transposable elements. This degeneration drives the Y/W chromosome to a reduction in gene content and often in size.

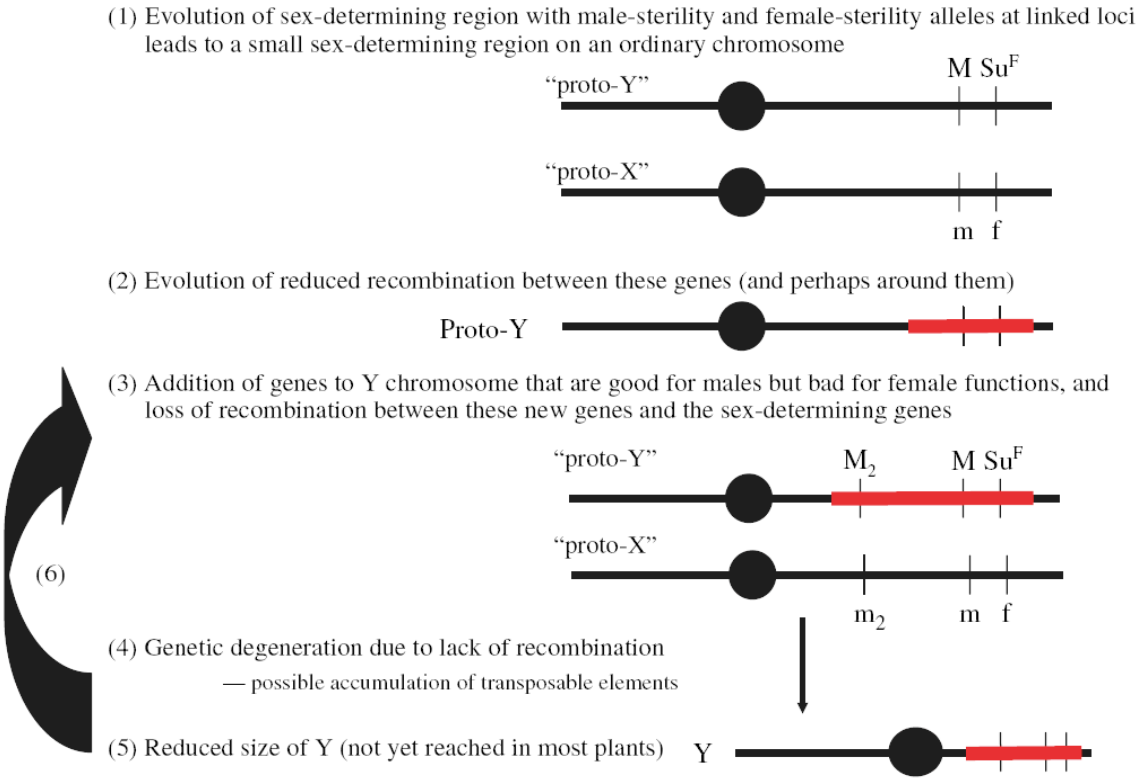

Figure 1: Stages in sex chromosome evolution. The figure shows how proto-sex chromosomes, carrying just two genes a male beneficial/female detrimental (m/M) and female (f/Su^F) sterility genes on the proto-X and proto-Y. By adding further genes good for one sex (M_2) and bad for the other sex the reduced recombination region extends. The genetic degeneration starts and possible accumulation of transposable elements due to lack of recombination. The last step is a reduced size of the Y chromosome in consequence of the degeneration (Figure from Charlesworth *et al.* 2005).

The formation of sex chromosomes presents a major problem due to the hemizygosity of genes and the reduced gene dose in the heterogametic sex. A general mechanism to maintain the gene dose between the heterogametic and the homogametic sex does not exist. Instead, many independent mechanisms have evolved to address the problem of dosage compensation. In humans and other mammals, one of the female X chromosomes is randomly inactivated in each cell (Lyon 1961) and only the genes on the active X chromosome are expressed. This decline in expression of alleles on one X chromosome in the homogametic sex (female, XX) corresponds to the expression of the hemizygous genes to the heterogametic sex (male, XY). In this system a major locus, *Xist*, initiates the transcriptional silencing of the X chromosome (Brown *et al.* 1991). In Drosophila, the female (XX) does not down-regulate the expression of X-linked genes to equalize the gene expression between sexes. Instead, the male up-regulates X-linked gene expression about twofold to compensate for the difference in gene dosage

(Bridges 1925). The exact mechanism responsible for this up-regulation is not known, but molecular factors associated with the up-regulation have been identified, including the male-specific lethal (MSL) dosage compensation complex (DCC; Kuroda *et al.* 1991; Palmer *et al.* 1993) and two noncoding RNAs, *roX1* and *roX2* (Amrein and Axel 1997; Meller *et al.* 1997). The DCC controls the H4 acetylation of the chromatin (Smith *et al.* 2001), which is associated with the up-regulation of the male X chromosome. In birds, a general mechanism of dosage compensation has not been detected (Itoh *et al.* 2010), which suggests that female birds (ZW) have only about half as much Z-linked gene expression as male birds (ZZ).

7.2 Sex chromosomes and speciation

The sex chromosomes play an important role in the process of speciation. Almost one hundred years ago, Haldane observed the preferential sterility or inviability of hybrids of the heterogametic sex (Haldane 1922). In hybrid crosses of recently diverged species in an XY sex chromosome system, the XY hybrids are often sterile or inviable, whereas their XX siblings are not. This observation is known as Haldane's rule. It was supposed that the occurrence of the Y chromosome and the hemizygosity of the X chromosome in the heterogametic sex, in comparison to the homogametic sex, was responsible. However, because the Y chromosome contains only a few functional genes, it could be excluded as a common cause of the observed male sterility. For this reason, the X chromosome was considered to be more important in causing hybrid sterility and inviability. The molecular basis of Haldane' rule has not been identified. However, several explanations have been proposed, including: dominance theory (heterogametic hybrids are affected by all X-linked alleles, both recessive and dominant, involved in incompatibilities, while homogametic hybrids are only affected by the dominant ones), the faster-male theory (genes involved in male reproduction evolve faster than those involved in female reproduction due to sexual selection, leading to more reproductive incompatibilities in males), cryptic sex-ratio meiotic drive (the X-chromosome may violates the Mendelian law of equal segregation by interfering with the transmission of the Y, which is counter by a species-specific suppressor (Sandler 1957)), or male germline X inactivation (the transcriptional silencing of the X chromosome during spermatogenesis, which may differ mechanistically between closely-related species).

The above postzygotic barriers seem to be involved in the reproductive isolation of many recently diverged species (Presgraves 2002; Price and Bouvier 2002).

The second role of sex chromosomes in speciation is referred to as the large X effect. The large X effect is the disproportionately large contribution of the X chromosome versus the autosomes in backcross genetic analyses of hybrid sterility and inviability. The reason for the higher contribution of the X chromosome is a supposed higher density of hybrid male sterility alleles. Evidence for the large X-effect comes from a wide range of taxa, including mouse, birds and Lepidoptera (Coyne 1992). One prominent example is the work of (Masly and Presgraves 2007), where 142 introgressions of *D. mauritania* genome fragments into the *D. sechilllia* genome were investigated in a backcross genetic experiment. The result of this study provided strong evidence for the higher density of male sterility alleles on the X chromosome.

7.3 Sex chromosomes and selection

The uneven distribution of sex chromosomes between the sexes leads to some differences in the selection process of sex chromosomes in comparison to the rest of the genome. The Y/W chromosome tends to degenerate by losing functional genes and accumulates transposable elements (Steinemann and Steinemann 2000; Steinemann and Steinemann 2001). Selection is only possible in males for the few remaining Y/W-linked genes. The consequence is that the contribution of the Y chromosome to the genome is relatively low due to its few remaining functional genes. In contrast, the X chromosome comprises many genes and is not degenerating. Considering an XY system, the X chromosome spends 2/3 of its evolutionary history in females and 1/3 in males. The autosomes spend equal time in the two sexes. The consequence of this difference in residence time, and the resulting difference in the effective population size, drives the X chromosome to evolve differently from the rest of the genome (Rice 1984; Charlesworth *et al.* 1987; Vicoso and Charlesworth 2006). When a recessive mutation arises on one of the autosomes, this mutation is mostly in the heterozygous state, because it is in low frequency in the population. Thus, it will be masked by the ancestral allele. The result is that the new allele cannot be affected by selection unless it is in a

homozygous individual. When a recessive mutation arises on the X chromosome, this mutation is immediately subject to selection in the heterogametic sex (XY, ZW). Therefore, recessive mutations are more efficiently selected on the X/Z chromosome than on the autosomes. Additionally, the difference in a chromosome's residence time in the two sexes has an effect on the mutation process itself. In spermatogenesis, more cell divisions are required to form the gametes and the process of mutation is coupled to the number of cell divisions. Thus, the mutation rate could be higher in males than in females (Haldane 1947). This leads to a lower mutation rate on the X chromosome of mammals (Hurst and Ellegren 1998; Li *et al.* 2002). However, such a mutational difference has not been observed in *Drosophila* (Bauer and Aquadro 1997). Another prediction for the selection on the sex chromosome is the so-called faster X effect. Taking special population genetic conditions into account ($Ne_X > 0.75\ Ne_A$; $h < 0.5$), the X chromosome accumulates beneficial mutations at a faster rate than the autosomes (Charlesworth *et al.* 1987; Vicoso and Charlesworth 2009). Evidence for faster X evolution has been reported for several taxa, including mammals and *Drosophila* (Charlesworth *et al.* 1987; Orr and Betancourt 2001; Torgerson and Singh 2003; Wang and Zhang 2004; Khaitovich *et al.* 2005; Baines *et al.* 2008). If mutations have an antagonistic effect on the sexes, these mutations and the affected genes will be also differently selected on the X chromosome in comparison to the rest of the genome (Rice 1984). If mutations are in general recessive, the X chromosome tends to accumulate male beneficial/female detrimental alleles, because in the male the allele is hemizygous and immediately available for selection (Rice 1984). In females, this mutation is masked by the ancestral allele. The X chromosome, may also accumulates dominant mutations, when the mutations are female beneficial/male detrimental, because the X chromosome spends 2/3 of the time in females and only 1/3 of the time in males.

7.4 Sex chromosome gene expression and gene content

With the appearance of new techniques, such as microarrays, it was possible to measure the entire transcriptome of a species. Several studies investigated the expression of the genome in several organisms, including human (Su *et al.* 2004), mouse (Khil *et al.* 2004), *Drosophila* (Parisi *et al.* 2003; Ranz *et al.* 2003), chicken (Kaiser and Ellegren 2006; Itoh *et al.* 2007) and

C. elegans (Reinke *et al.* 2004). Of particular interest were the expression differences between the sexes. To investigate this difference, male-biased and female-biased genes were defined. Male-biased genes are genes that are exclusively or predominantly expressed in males. Female-biased genes show the opposite pattern of expression. Unbiased genes are equally expressed in the two sexes. One of the first observations was that the distribution of male-biased genes was not random. In *Drosophila*, an under-representation of male-biased genes on the X chromosome was reported (Parisi *et al.* 2003; Ranz *et al.* 2003) (Figure 2).

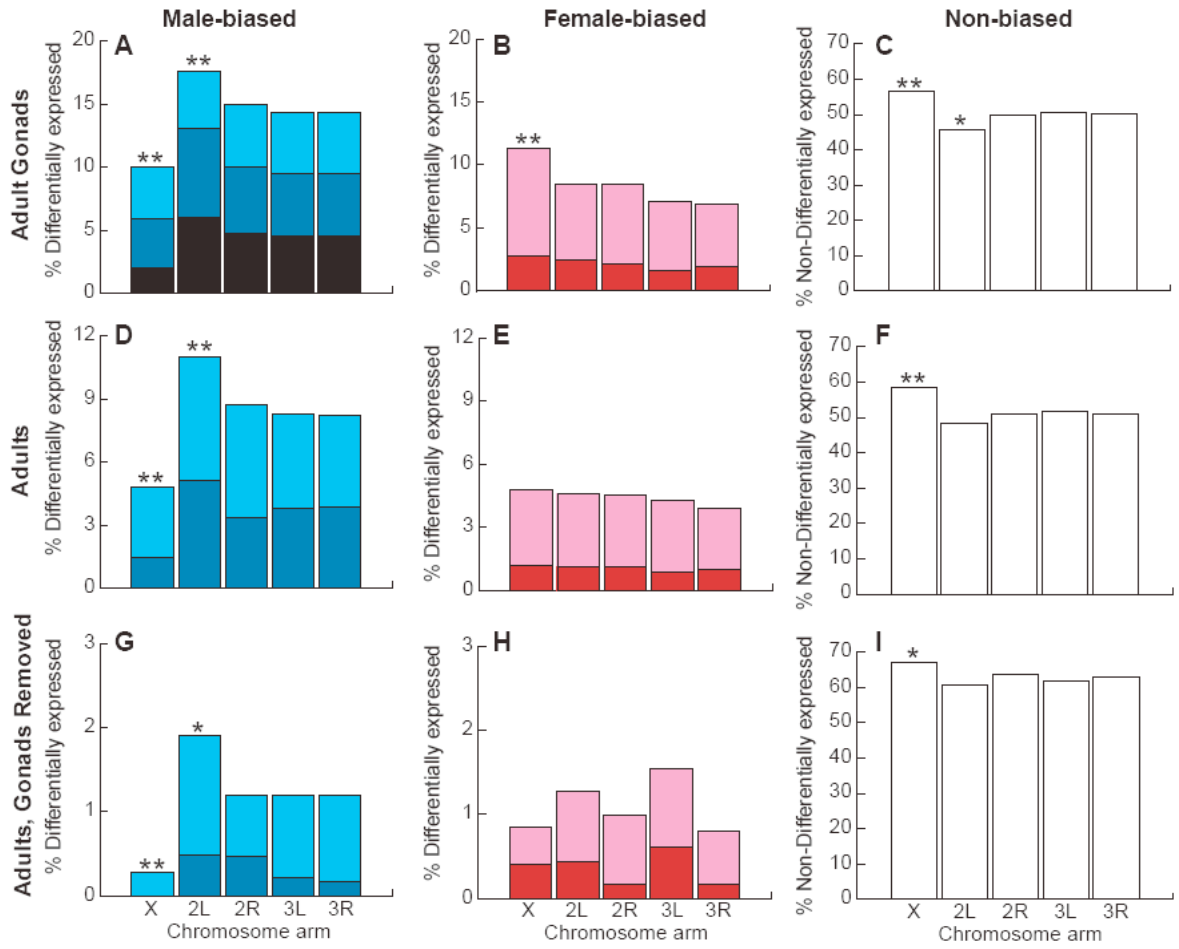

Figure 2: Gene expression for major chromosome arms in *Drosophila melanogaster*. Further the distributions of male-biased, female-biased and unbiased genes on this chromosome arms are depicted for certain thresholds of differently expression. Gene expression was measured in adult gonads, whole flies (adult) and flies with dissected gonads (Figure from Parisi *et al.* 2003).

This under-representation of male-biased genes was also found in other species, including *C. elegans* (Reinke *et al.* 2004), mouse (Khil *et al.* 2004) and in birds for female-biased genes on the Z chromosome (Kaiser and Ellegren 2006). In birds the female is the heterogametic sex (ZW). However, in birds the expression differences of Z-linked genes could be a result of the

lack of dosage compensation in females (see above). Several explanations for the under-representation of male-biased genes on the X chromosome have been proposed. The first explanation is sexual antagonism. The observed demasculinization of the X chromosome requires that most of the sexually antagonistic mutations are dominant. The consequence will be that female beneficial/male detrimental mutations will accumulate and male beneficial/female detrimental mutations will be eliminated (Rice 1984). The result of this mutation/selection process is a demasculinized X chromosome. The second explanation is based on the dosage compensation mechanism. In detail, this means that male-biased genes evolve by increasing their level of expression of existing genes in males. In contrast to the autosomes, a higher expression level could be harder to achieve on the already hyperactive X chromosome, if the rate of mRNA transcription is limited due to dosage compensation. The last explanation is male germline X inactivation (also referred as meiotic sex chromosome inactivation (MSCI); Lifschytz and Lindsley 1972; Betran *et al.* 2002). Male germline X inactivation causes the X chromosome to be transcriptionally silenced during spermatogenesis. Especially genes expressed late in spermatogenesis (meiosis) will be affected. The result of the X inactivation is that male-biased testis-expressed X-linked genes are not expressed or are expressed only at a low level. To avoid this reduction of expression in the testis, genes often escape the X chromosome and move to the autosomes either through the mechanism of retrotransposition or gene duplication. The new environment of the autosomes, with no expression inactivation, allows the re-located copies to be expressed at a higher level in the male germline. Such escape from the X-chromosome was observed in mouse (Emerson *et al.* 2004) and *Drosophila* (Vibranovski *et al.* 2009b). In the study of (Vibranovski *et al.* 2009b) the entire *Drosophila* clade was screened for duplicated genes that re-located either through the mechanism of gene duplication or retrotransposition. The expectation of gene movement inside the *Drosophila* genomes was compared to the observed movement (Figure 3).

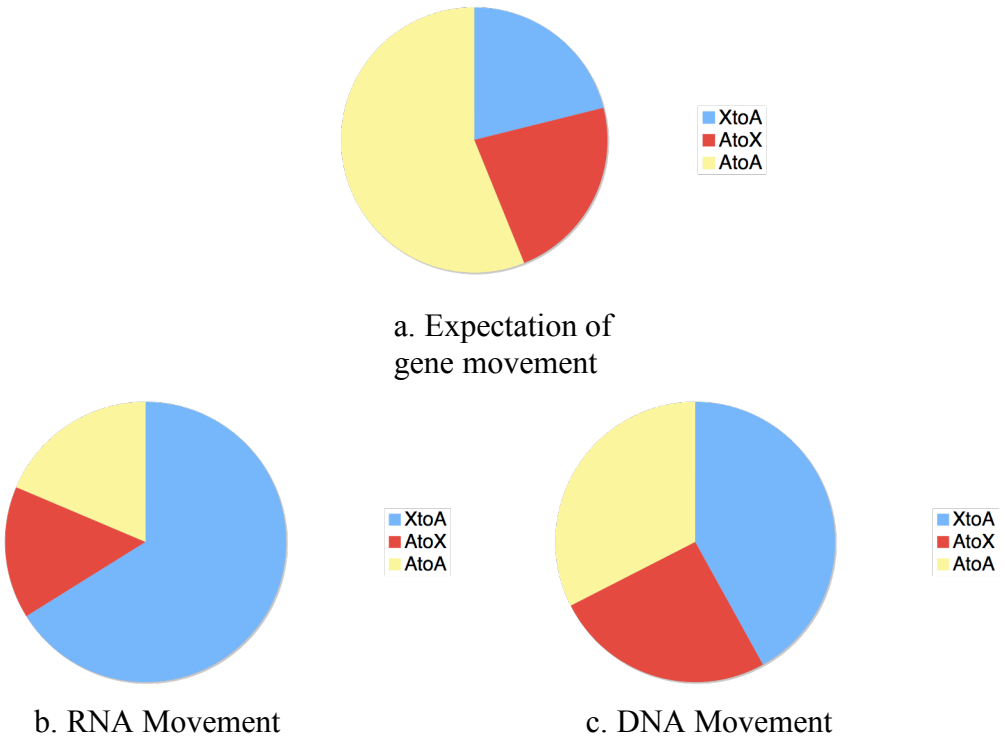

Figure 3: Expected (a) and observed (b+c) gene movement in the *Drosophila* clade. In particular X to autosome, autosome to X and autosome to autosome movement. Retrotransposition (b) and gene duplication (c) were measured separately (Figure from Vibranovski *et al.* 2009b).

The result of this study was that, in the *Drosophila* clade, more X-to-autosome movement was observed than expected. This out-of-X movement bias was detected for both retrotransposition and gene duplication. The escaping genes often show testis expression. In accordance with this, autosomal mutations for *Drosophila* male sterility genes often affect late spermatogenesis (Castrillon *et al.* 1993). These observations suggest that the new testis-biased genes escape from male germline X inactivation. The new autosomal copies would be able to be expressed at a higher level and at later stages during spermatogenesis. These changes in the expression profile of the male-biased genes would be not possible on the inactivated X chromosome. If the changes in the male-biased expression profile are beneficial for the organism, the new copies would be more often retained than other types of gene duplication.

7.5 Male germline X inactivation

Male germline X inactivation (or meiotic sex chromosome inactivation, MSCI) was first proposed by (Lifschytz and Lindsley 1972). In this process, the X chromosome in males is presumed to be heterochromatinized during the first meiotic prophase (Figure 4).

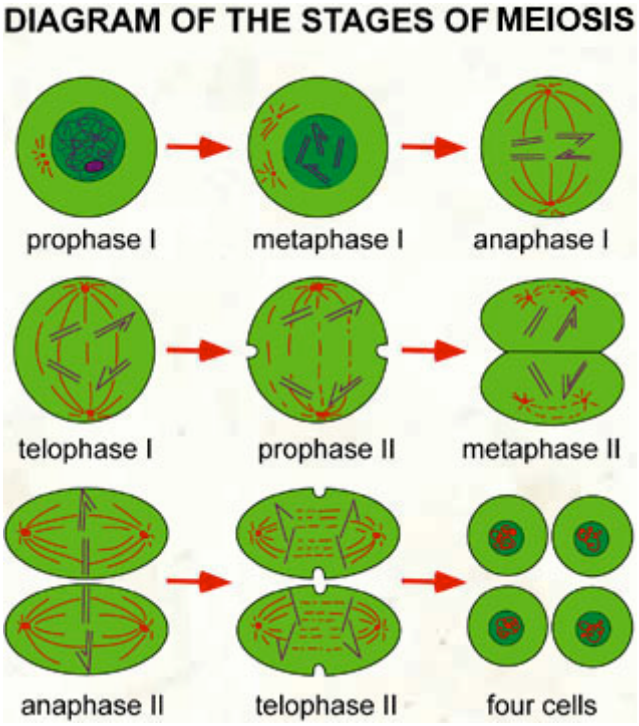

Figure 4: Cell division and segregation of the chromosomes during meiosis. First the stages of meiosis I; prophase I (DNA exchange between homologous chromosomes), metaphase I (attachment of microtubule to the kinetochores), anaphase I (chromosome pair separation to opposite cell poles) and telophase I (complete separation of chromosome pairs and cell division), stages of meiosis II, similar to meiosis I. (http://www.infovisual.info/01/021_en.html)

Further, the X chromosome becomes transcriptionally inactivated and almost no expression is possible in male reproductive cells. One explanation for the presence of X inactivation is that the lack of pairing of the X and Y chromosome is responsible for the meiotic silencing of unsynapsed chromatin or unpaired DNA. This inactivation may be an ancient genome defence mechanism that silences sequences without pairing partners (Shiu *et al.* 2001). Another explanation is given by sexual antagonism. As mentioned above, the X chromosome tends to become feminized over the course of its evolution. The feminized X chromosome will harbor many female beneficial/male detrimental alleles. These alleles may adversely affect spermatogenesis (Wu and Xu 2003) and to avoid the effect of these antagonistic genes the X chromosome is transcriptionally silenced during spermatogenesis.

Empirical results to support the MSCI were found in a variety of species, including mammals (Richler *et al.* 1992; Handel *et al.* 1994; Turner 2007), *C. elegans* (Fong *et al.* 2002; Kelly *et*

al. 2002) and *D. melanogaster* (Hense *et al.* 2007; Vibranovski *et al.* 2009a). The latter two studies in *Drosophila* are of particular relevance to this dissertation. Hense *et al.* (2007) showed that autosomal insertions of a transgenic construct containing the promoter of the testis-specific *ocnus* (*ocn*) gene fused to the *lacZ* reporter gene had a significantly higher expression than X-linked insertions of the same construct (Figure 5). In the study by Vibranovski *et al.* (2009a), dissected parts of the testis from *Drosophila*, corresponding to the pre-meiotic, meiotic and post-meiotic phases of spermatogenesis, were transcriptionally analyzed using microarrays. The result of the transcriptomic study showed that the X chromosome was under-represented for male-biased genes showing higher expression in meiosis compared to mitosis. Both studies are consistent with the expectation of testis gene expression being reduced by X inactivation.

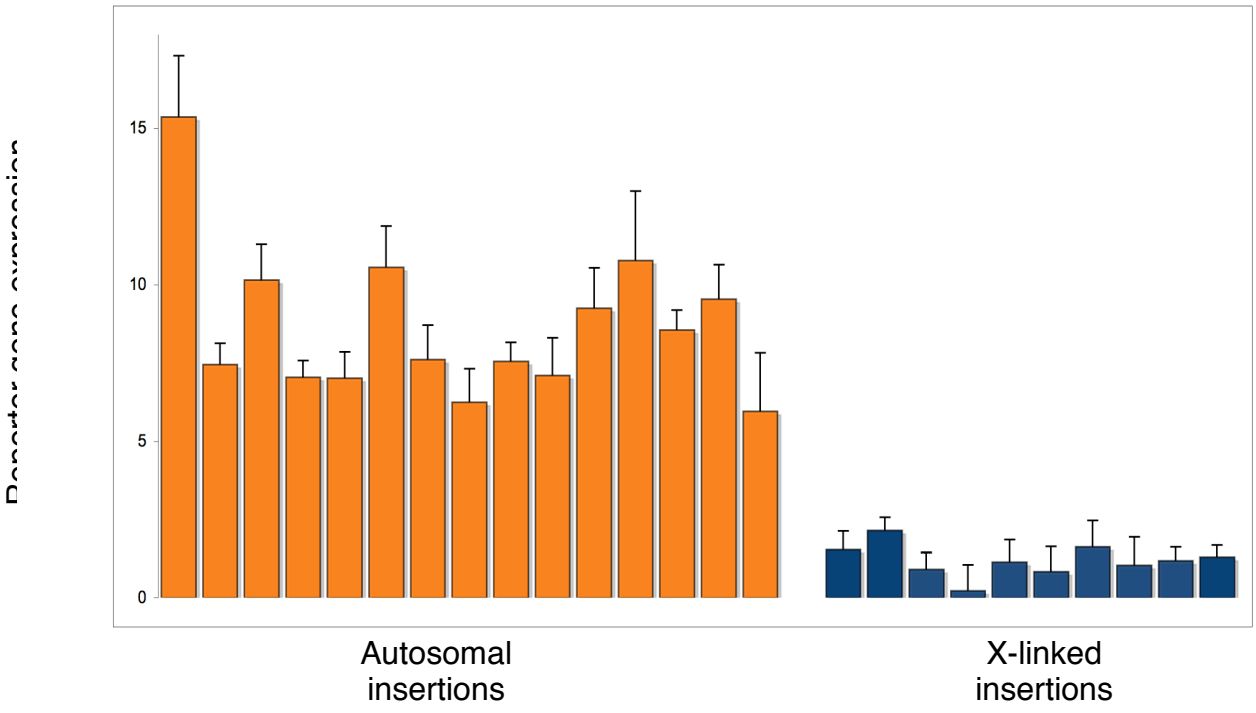

Figure 5: Average ß-galactosidase-activity of adult male flies with the insertion of the P*[wFl-ocn-lacz]* construct. Each bar represents an independent and unique autosomal or X-linked insertion of the construct. (Figure from Hense *et al.* 2007).

However, it has been proposed that the region around cytological band 19, which appears to be a hotspot for new gene evolution, may escape inactivation (Chen *et al.* 2007) This region shows a general enrichment of testis-expressed genes (Boutanaev *et al.* 2002), including the newly evolved genes *Sdic*, *CG15323*, and *hydra* (Nurminsky *et al.* 1998; Levine *et al.* 2006; Chen *et al.* 2007) (Figure 6 + 7). The orthologous region in *D. yakuba* also appears to be a hotspot for *de novo* gene evolution (Begun *et al.* 2007)

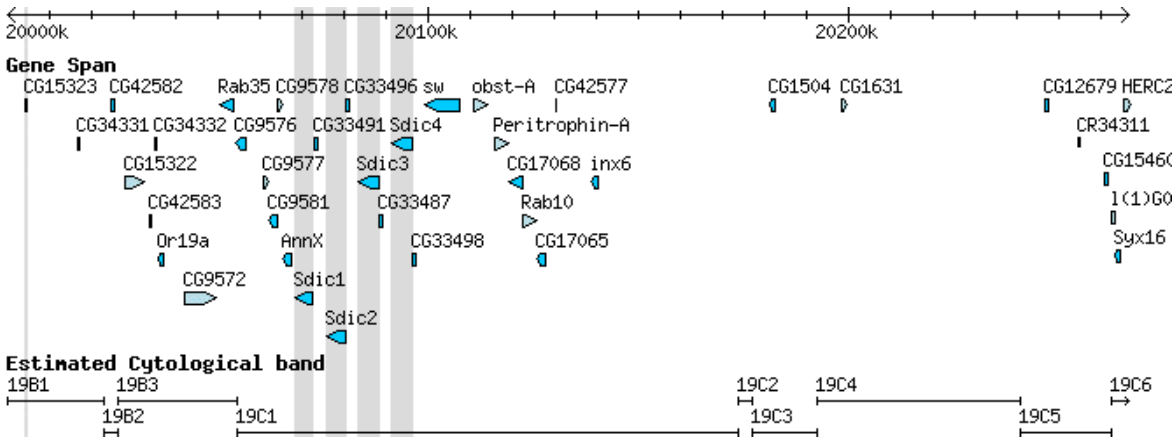

Figure 6: Genes in the cytological bands 19B-C on the *D. melanogaster* X chromosome (X:20,000,000–20,266,000 bp). Newly-evolved, testis expressed genes are highlighted. (Figure from Flybase; Tweedie *et al.* 2009).

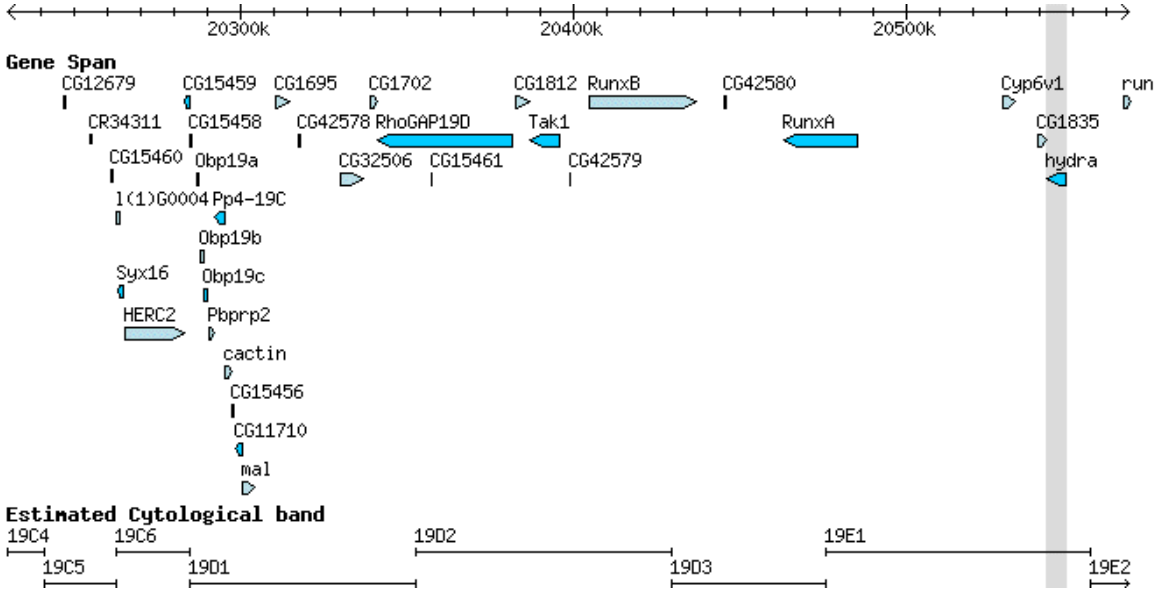

Figure 7: Genes in the cytological bands 19C-E on the *D. melanogaster* X chromosome (X:20,233,000–20,566,833 bp). Newly-evolved, testis expressed genes are highlighted. (Figure from Flybase; Tweedie *et al.* 2009).

There are still several open questions regarding male germline X inactivation, including:

1. Does male germline X inactivation affect the entire X chromosome?
   The study of Hense *et al.* (2007) only demonstrated X inactivation with 10 X-linked insertions, but did not have coverage of the entire X chromosome.
2. Does cytological region 19 on the X chromosome escape inactivation?

Several studies identified genes that show testis expression and are located on the X chromosome. Many of these genes cluster in a region of the X chromosome at cytological band 19, suggesting that this region might escape X inactivation.

3. Do X-linked male-biased genes gain higher testis expression through *cis*-regulatory sequences that help them avoid X inactivation?

   One possibility for the presence of male-biased genes on the X chromosome could be the presence of *cis*-regulatory sequences, which allows these genes to gain higher expression despite male germline X inactivation.

4. Does escaping the X chromosome provide an expression advantage in the male germline?

   No study to date has reported direct experimental evidence to support the X inactivation hypothesis, which has been proposed to explain the excess gene movement from the X chromosome to the autosomes.

To address these questions, I performed two approaches. In the first approach, the *ocnus* construct from Hense *et al.* (2007) was mobilized to additional locations on the X chromosome. I generated a high density of insertions along the X chromosome and was able to map over 100 insertions with an average distance of roughly 200 Kb between insertions. No region on the X chromosome showed evidence for elevated expression in the male germline, indicating that the entire X chromosome is transcriptional silenced and that no chromosomal region escapes inactivation. In the second approach, I examined three promoters from three different X-linked genes. By transforming reporter gene constructs into different X-linked and autosomal locations, I was able to show that there is a selective advantage by increased expression in the male germline associated with escape from the X chromosome. The *cis*-regulatory sequences from testis-expressed, X-linked genes are shown to drive higher testis expression when relocated to the autosomes.

7.6 Sex chromosome gene expression variation

Protein variation makes an important contribution to the phenotypic variation observed between and within species (Kreitman and Hudson 1991; Clark *et al.* 2007). However, it has been proposed that variation in gene control elements, rather than the protein themselves, is likely to be more important in adaptive evolution (King and Wilson 1975). It has recently become possible to measure global gene expression variation between and within species with microarray techniques. Differences in the expression level of genes between populations are of particular interest. These expression differences may underlie the local adaptation of populations to the environment. In a study by Hutter *et al.* (2008), gene expression variation in African and European populations of *Drosophila melanogaster* was analyzed. The African population is the ancestral population. After a slight population expansion within Africa, *D. melanogaster* colonized Europe (Lachaise *et al.* 1988). This study revealed that X-linked genes have consistently less expression polymorphisms than autosomal genes in both populations (Table 1).

Table 1: Expression polymorphism (Average percentage of pairwise differences) on the X chromosome and autosomes. Deviation from 1:1 expectations for the X/A ratios was tested with a two-tailed Fisher's exact test.

Population	X chromosome	Autosomes	X/A ratio	P-value
Overall	2.02	2.90	0.697	0.040
Europe	1.77	2.68	0.661	0.014
Africa	1.86	2.64	0.705	0.017
Between	2.20	3.11	0.708	0.035

This unequal distribution of expression polymorphisms within the population appears to be a result of the unequal genomic distribution of sex-biased genes (under-representation of male-biased genes on the *Drosophila* X chromosome, see above). The cause of the expression variation is still unclear and the contribution of *cis*- and *trans*-regulatory elements to gene expression variation remains controversial. However, several studies reported that changes in *cis*-regulatory sequences contribute to the gene expression variation within (Rockman and Wray 2002) and between species (Wittkopp *et al.* 2008). To investigate the cause of

expression variation on the X chromosome within species, I selected a X-linked gene (*CG9509*), which showed high expression difference between the African and European population, with greater than twofold higher expression in Europe (Meiklejohn *et al.* 2003; Hutter *et al.* 2008) (Figure 8).

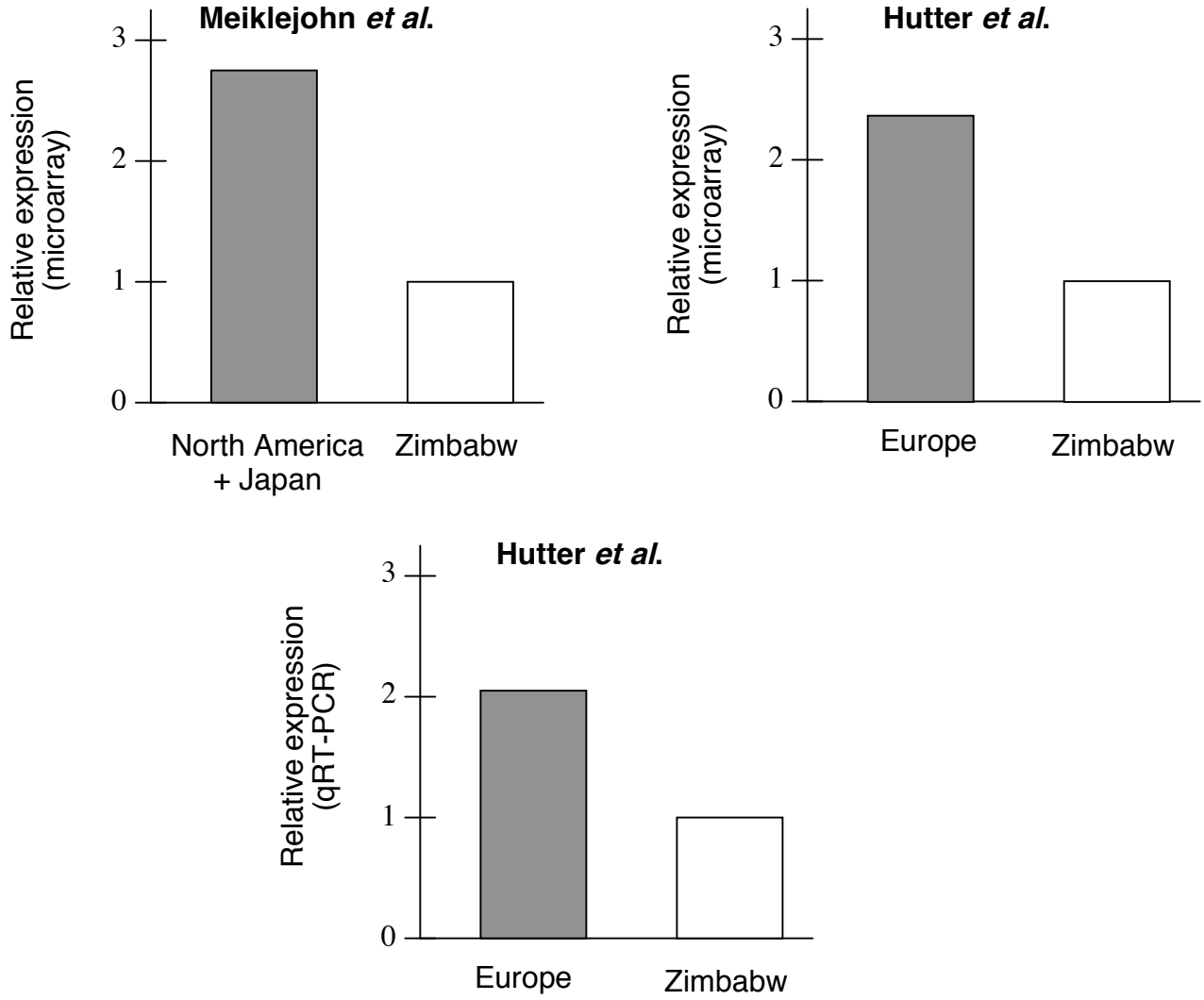

Figure 8: Expression differences of the gene *CG9509* between African and Cosmopolitan/European populations (Meiklejohn *et al.* 2003; Hutter *et al.* 2008). The expression differences were measured either with the microarray technique or qRT-PCR.

Further, this gene showed evidence for adaptive gene evolution in the putative promoter region in a previous study (Saminadin-Peter 2008). The goal of my study was to determine if *cis*-acting variation within the putative promoter region was responsible for the expression difference of *CG9509* observed between populations. I experimentally determined the level of reporter gene expression driven by the European and African versions in an otherwise

uniform genetic background. The results indicate that the entire expression difference can be attributed to variation within the promoter region. Thus, I have uncovered a selective sweep associated with an X-linked *cis*-regulatory variant of a European population of *D. melanogaster*.

8. Material and Methods

8.1 Genome sequences and BLAST search

Genome sequences were obtained from the UCSC browser (http://genome.ucsc.edu) using the *Drosophila* genome release 5.30. The BLAST searches were performed with the BLAST search option on Flybase (http://flybase.org; Tweedie *et al.* 2009) *Drosophila* genome release 5.30.

8.2 Primer sequences for amplification of putative promoters

Putative promoter sequences of three X-linked genes (*CG10920*, *CG12681*, and *CG1314*) and the autosomal gene (*ocnus*) were PCR-amplified from genomic DNA of the *Canton S* strain of *D. melanogaster*. The *CG10920* promoter corresponds to bases 7,748,179–7,748,758 of the X chromosome (FlyBase release 5.30; Tweedie *et al.* 2009). The *CG12681* promoter corresponds to bases 4,769,051–4,769,815 (X chromosome), the *CG1314* promoter corresponds to bases 20,740,370–20,740,877 (X chromosome) and the *ocnus* promoter corresponds to bases 25,863,383–25,863,532 of chromosome 3R. All of the amplified sequences lie just upstream of their respective coding sequences and end at base -28 (*CG10920*), -10 (*CG12681*), -4 (*CG1314*), and -16 (*ocnus*) relative to the start codon. The amplified promoter sequences have sizes of 580 bp (*CG10920*), 765 bp (*CG12681*), 508 bp (*CG1314*) and 150 bp (*ocnus*).

To amplify the promoter sequences, I used the following primer pairs: the *CG10920* promoter was amplified with the “cg10920prom-fw” primer (5’-TATTTATGGCTAGGCAGGTC-3’) and the “cg10920prom-rev” primer (5’-AATTTCAATTCGCCAAAAG-3’), the *CG12681*

promoter sequence was amplified with the "cg12681prom-fw" primer (5'-CAAATTACGTTTCATTACGC-3') and the "cg12681prom-rev" primer (5'-CAAATTTCCGTACTTAATGC-3'), the *CG1314* promoter sequence was amplified with the "cg1314prom-fw" primer (5'-CAGTCCTAGTCCGACTGTTG-3') and the "cg1314prom-rev" primer (5'-GGAATTTTTAAGAAAATGTCG-3'), the *ocnus* promoter sequence was amplified with the "OCNPROFOR" primer (5'-GAATGATCACATGTGCTCCG-3') and the "OCNPROREV" primer (5'-ATCGATGGAAAACGCACTGGAATT-3').

The putative promoter sequence of the X-linked gene (*CG9509*) was amplified from genomic DNA of the African strain (Zimbabwe 82) and the European strain (Europe 12) (Glinka *et al.* 2003). The *CG9509* promoter corresponds to bases 14,803,041–14,804,227 of the X chromosome (*D. melanogaster* genome; FlyBase release 5.30; Tweedie *et al.* 2009). The amplified sequence lies just upstream of their respective coding sequences and end at base -2 relative to the start codon. The amplified promoter sequences have a size of 1174 bp for the African population and 1186 bp for the European population. The *CG9509* promoter sequence for the European population was amplified with the "CG9509Le12" primer (5'-GCCGTCTTAATGTTTGTTTGTG-3'), the promoter sequence for the African population was amplified with the "CG9509Lz82" primer (5'-GCCGTCTTAATGTGTGTTTGTG-3') and the opposite primer for both populations was the "CG9509Right" primer (5'-GCGTTTTGCTTTTCCGTTAG-3').

8.3 DNA extraction

For the isolation of genomic DNA, 15 flies (females and/or males) were used. These 15 flies were homogenized in 400 µl Buffer A (0.1 M Tris HCl, pH7.5; 0.1 M EDTA, pH 8.0; 0.1 M NaCl; 0.5 % SDS). The solution was incubated for 30 min at 65°C with soft shaking. Afterward, 800 µl LiCl/KAc solution (1.4 M KAc; 4.3 M LiCl) was added and incubated for 10 min on ice. The solution was centrifuged for 15 min at 10,000 g and the supernatant was retained. To the supernatant 800 µl of isopropanol was added and the solution was again centrifuged for 15 min at 10,000 g. The supernatant was discarded and the remaining pellet was washed in 500 µl 70% ethanol. After centrifuging the pellet for 15 min at 10,000 g, the

supernatant was discarded and the pellet was dried at room temperature and resuspended in 75 µl H_2O.

8.4 Restriction endonuclease digest

Restriction enzymes from NEB (New England Biolabs; www.neb.com) were used. The reaction volume was in total 20 µl. Each reaction contained 0.1–1 U of the restriction enzyme I. When necessary, restriction enzyme II was used at the same concentration. Depending on the enzyme, the corresponding buffer system (NEB-buffer I-IV) was used (2 µl of 10X NEB-Buffer). DNA in a concentration range of 100 ng–2 µg was cleaved and the reaction was incubated for 1 h at 37°C. Following digestion, the enzymes were heat inactivated at 60°C for 20 min. The following enzymes were used: *Xho*I, *BamH*I, *Xba*I, *Not*I, and *Spe*I.

8.5 Ligation

The ligation was performed with the T4-DNA-Ligase from NEB (New England Biolabs; www.neb.com). A total of 200 U of the ligase was used and the reaction was performed in 20 µl containing the DNA-fragments (10 ng–1 µg) and 2 µl of 10X NEB-Buffer. The reaction was performed at room temperature for 1 h or overnight.

8.6 Polymerase chain reaction

For the amplification of DNA fragments the Taq-polymerase from Peqlab (www.peqlab.de) was used (1 U per reaction). The DNA concentration was in the range of 100 ng–2 µg, the dNTP concentration was 10 mM, the primer concentration was 0.2 pmol/µl and 2.5 µl of 10X

PCR buffer (high yield, or high specificity) was used. The total volume was 25 µl. The protocol to amplify DNA-fragments included the following steps: 95°C for 2 min, a cycle for 39 times (95°C for 0.5 min, primer melting temperature for 0.5 min and 72°C for 1.5 min) and a final step of 72°C for 5 min.

8.7 Sequencing

Before the sequencing reaction was performed, every PCR-reaction was treated with ExoSAP-IT™ (Amersham; www.ge.com) for 30 min at 37°C. Afterwards the ExoSAP enzyme was heat inactivated at 80°C for 15 min. The sequencing reaction included the following components: 2 µl Big Dye v1.1 seq mix (ABI, www.appliedbiosystems.com), 1 µl of 5X sequencing buffer (ABI; www.appliedbiosystems.com), 3 pmol/µl primer, 2 µl PCR-product and 2 µl H_2O. The cycling conditions were 96°C for 1 min followed by 25 cycles of (96°C for 10 s, 50°C for 15 s and 60°C for 4 min). The sequence reaction was diluted with 10 µl of H_2O and analyzed on an ABI 3730 (ABI; www.appliedbiosystems.com) sequencing machine.

8.8 RNA extraction

RNA was extracted from 30 male and/or female flies. These flies were homogenized in 800 µl of Trizol (Invitrogen; www.invitrogen.com) and incubated for 5 min at room temperature. The homogenate was centrifuged for 10 min at 4°C and 12,000 g. The supernatant was retained and mixed with 200 µl of chloroform. The solution was vortexed for 15 sec and centrifuged for 10 min at 4°C and 12,000 g. The supernatant was retained and 500 µl of isopropanol was added. This solution was centrifuged for 10 min at 4°C and 12,000 g. The supernatant was discarded and the pellet was washed in 70% ethanol. The ethanol solution with the RNA-pellet was centrifuged for 10 min at 4°C and 12,000 g. The supernatant was

discarded and RNA pellet was dried at room temperature. The dried RNA pellet was resuspended in 30 µl H_2O.

8.9 Bacterial Transformation

The transformation was performed with One Shot TOP 10 electrocompetent or chemically competent cells (Invitrogen; www.invitrogen.com). For each transformation, 100 µl of cell suspension was mixed with 10 ng–100 ng plasmid DNA. For the chemical transformation and the electro transformation, the manufacture's instruction was followed.

8.10 Plasmid extraction

Overnight cultures of plasmid containing bacteria in LB-media (5 g/l yeast extract, 10 g/l tryptone, 10 g/l NaCl and 60 ng/ml ampicillin) were isolated either using the QIAprep Spin Miniprep Kit (QIAGEN; http://www.qiagen.com) and following the manufacture's instruction or the method described below. 1.5 ml of the overnight culture was centrifuged for 2 min at 10,000 g. The supernatant was discarded and the cell pellet was resuspended in 100 µl solution 1 (9.9 g/l glucose; 25 mM Tris-HCl, pH 8.0; 10 mM EDTA, pH 8.0). 100 µl of solution 2 (1% SDS; 0.2 M NaOH) was added and incubated for 5 min at room temperature. 100 µl of solution 3 (294.4 g/l potassium actetat, 115 ml/l glacial acetic acid) was then added. The cell solution was centrifuged for 15 min at 10,000 g. The supernatant was retained and 700 µl of 100% ethanol was added. This solution was centrifuged for 15 min at 10,000 g and the supernatant was discarded. The plasmid pellet was washed in 500 µl 70% ethanol and again centrifuged for 15 min at 10,000 g. The supernatant was discarded and the plasmid pellet was dried at room temperature. The dried plasmid pellet was resuspended in 50 µl H_2O.

8.11 Agarose gel electrophoresis

The standard electrophoresis buffer was TAE (50 mM EDTA, pH8.0; 242 g/l Tris base; 57.1 ml/l glacial acetic acid). The separation of DNA fragments was performed in 0.5–1.5 % agarose gels depending on the size range of the DNA fragments. The electrophoresis condition was constant 100 V. The size standard was 1 Kb ladder from Invitrogen (www.invitrogen.com) and the loading buffer contained 0.25% bromphenol blue, 0.25% xylene cyanol FF and 30% glycerol.

For cloning, DNA-containing bands were cut out of agarose gels. These DNA bands were then purified with the QIAquick Gel Extraction Kit from QIAGEN; http://www.qiagen.com) following the manufacture's protocol.

8.12 LB-media plates

The selection and reproduction of bacteria were performed on LB-media plates (5 g/l yeast extract, 10 g/l tryptone, 10 g/l NaCl, 15 g/l agar, and 60 ng/ml ampicillin).

8.13 Fly food

All flies used for this PhD thesis were reared at standard condition at 20–25°C on fly food containing 4 g/l agar, 3.8% sugar syrup, 28.5 g/l yeast extract, 38.5 g/l maize polenta, 4.6 ml/l propionic acid, and 1.2 g/l Nipagin (methyl 4-hydroxybenzoate).

8.14 Transformation vector construction for *P*-element transformation

The amplified PCR products were cloned directly into the pCR2.1-TOPO vector (Invitrogen; http://www.invitrogen.com). The identity and orientation of the PCR fragments were confirmed by restriction analysis. A 3.6-kb *Not*I fragment of the pCMV-SPORT-βgal plasmid (Invitrogen; http://www.invitrogen.com) containing the *E. coli lacZ* coding region was cloned into the *Not*I site of the promoter-containing plasmid. Afterward, I performed restriction analysis to ensure that both the promoter and *lacZ* coding sequence were in the same transcriptional orientation. In a final step, an *Spe*I/*Xba*I fragment containing both the promoter and the *lacZ* coding sequence was ligated into the p*P[wFl]* transformation vector (Siegal and Hartl 1996). This vector is derived from the *P* transposable element and contains the *D. melanogaster white* (*w*) gene as a selectable marker (Figure 9).

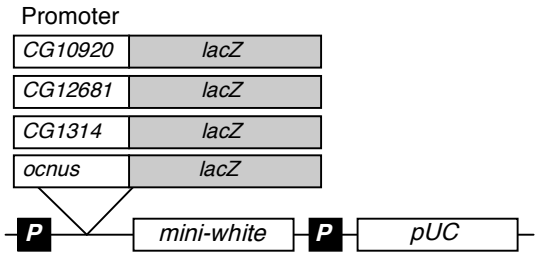

Figure 9: Schematic diagram of the promoter-*lacZ* expression constructs. The promoters of interests were fused to the reporter gene *lacZ* and inserted into the p*P[wFl]* transformation vector. The transformation vector contains the *white* gene (*mini-white*) as a selectable marker. The boundaries of the DNA inserted into the *Drosophila* genome are indicated by "P". The backbone of the vector used for the replication in *E. coli* is labeled "pUC".

8.15 Transformation vector construction for *Φ*C31 transformation

The amplified PCR products were cloned directly into the pCR2.1-TOPO vector (Invitrogen; http://www.invitrogen.com). The identity and orientation of the PCR fragments were confirmed by restriction analysis. A 3.6-kb *Not*I fragment of the pCMV-SPORT-βgal plasmid (Invitrogen; http://www.invitrogen.com) containing the *E. coli lacZ* coding region was cloned into the *Not*I site of the promoter-containing plasmid. Afterward, I performed restriction

analysis to ensure that both the promoter and *lacZ* coding sequence were in the same transcriptional orientation. In a final step, a *BamH*I/*Xba*I fragment containing both the promoter and the *lacZ* coding sequence was ligated into the p*attB* transformation vector (Bischof *et al.* 2007). This vector contains an attB-site, which is homologous to the attP-landing-site in the fly genome and used for the integration of the reporter gene construct into a precise landings site with the aid of the ΦC31 integrase. The transformation vector also contains the *D. melanogaster white* (*w*) gene as a selectable marker (Figure 10).

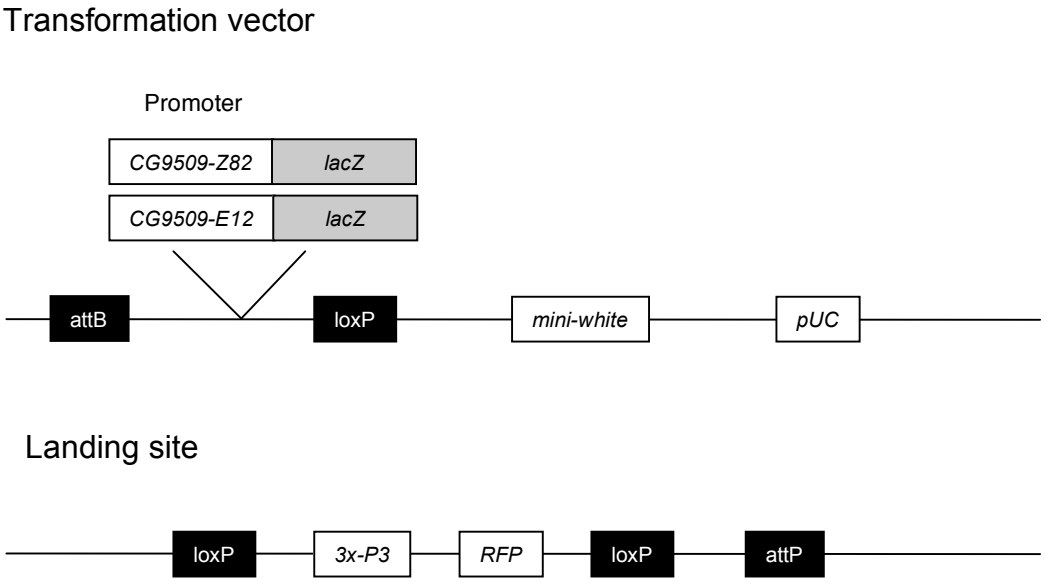

Figure 10: Schematic diagram of the promoter-*lacZ* expression constructs and the corresponding landing site in the *Drosophila* genome. The promoters of interests were fused to the reporter gene *lacZ* and inserted into the p*attB* transformation vector. The transformation vector contains the *white* gene (*mini-white*) as a selectable marker. The attB-site of the transformation vector and the homologous attP-site in the *Drosophila* genome are depicted. The backbone of the vector used for the replication in *E. coli* is labeled "pUC". The *red fluorescent protein* (*RFP*) gene serves as a selectable marker for the presence of the landings site. The *3xP3* promoter drives the expression of the *RFP* gene. The recombinase recognition sites are labeled "loxP".

8.16 Germline transformation for *Φ*C31 transformation

All transformation vectors were purified with the QIAprep Spin Miniprep Kit (QIAGEN; http://www.qiagen.com) and eluted from the column with injection buffer (0.1 mM Sodium Phosphate, pH 6.8; 5 mM KCl). Vector DNA at a concentration of 200 ng/µl was used for

microinjection of early-stage embryos of the strain *ZH-attP-86Fb* (location of landing site: 3rd chromosome cytological band 86F) and the strain *ZH-attP-68E* (location of landing site: 3rd chromosome cytological band 68E). The *w* mutation is associated with eye color and changes the eye color from the wild-type red to white. The stable genomic *Φ*C31 integrase on the X chromosome served to facilitate the integration of the reporter gene construct into the landing site. After microinjection, all surviving flies were crossed to an *yw* strain to remove the integrase source and establish stable lines. The offspring of this cross were screened for red eye color (imparted by the wild-type w^+ gene of the vector), which was diagnostic for stable germline transformants (Bischof *et al.* 2007).

8.17 Germline transformation for *P*-element transformation

All transformation vectors were purified with the QIAprep Spin Miniprep Kit (QIAGEN; http://www.qiagen.com) and eluted from the column with injection buffer (0.1 mM Sodium Phosphate pH 6.8; 5 mM KCl). Vector DNA at a concentration of 200 ng/µl was used for microinjection of early-stage embryos of the strain *yw*; *Δ2-3, sb/TM6*. The *w* mutation is associated with eye color and changes the eye color from the wild-type red to white. The stable genomic *P* element transposase *Δ2-3* on the third chromosome served as source of transposase. After microinjection, all surviving flies were crossed to an *yw* strain to remove the transposase source and establish stable lines. The offspring of this cross were screened for red eye color (imparted by the wild-type w^+ gene of the vector), which was diagnostic for stable germline transformants (Rubin and Spradling 1982; Spradling and Rubin 1982). Additional mobilizations of transgenes to and from the X chromosome were carried out through genetic crosses with a *Δ2-3* transposing-containing stock. Transformed females were mated to *yw*; *Δ2-3, sb/TM6* males and the male offspring carrying both the transgene and *Δ2-3* transposase were mated to *yw* females. From this cross, I selected male offspring carrying the transgene (which could not be on the X chromosome inherited from the mother). These males were mated to *yw* females to establish stable transformed lines with new autosomal or X-linked insertions of the transgene.

8.18 Insertion mapping

The chromosomal location of each transgene (X or autosome) was mapped initially by genetic crosses. Transformed males were mated to *yw* females and inheritance of the w^+ marker was observed in the next generation. Transformed lines with X-linked insertions were identified as those producing only daughters that carry the w^+ allele. Subsequently, the exact chromosomal position of each transgene insertion was determined by inverse PCR (Bellen *et al.* 2004). Briefly, genomic DNA was digested with *Hpa*II or *Hinp*1I and the resulting fragments were self-ligated with T4 DNA-Ligase (NEB; http://www.neb.com). The target sequence, the inserted expression construct, was amplified with one of two primer pairs either Pry1 (5'-CCTTAGCATGTCCGTGGGGTTTGAAT-3') and Pry2 (5'-CTTGCCGACGGGACCACCTTATGTTATT-3') or Plac1 (5'-CACCCAAGGCTCTGCTCCCACAAT-3') and Plac4 (5'-ACTGTGCGTTAGGTCCTGTTCATTGTT-3'). The resulting PCR-products were sequenced using the above primers and BigDye v1.1 chemistry on an ABI 3730 automated sequencer (Applied Biosystems; www.appliedbiosystems.com). DNA sequences were used for a BLAST search of the *D. melanogaster* genome (FlyBase release 5.30, Tweedie *et al.* 2009) to determine the exact position of transgene insertion.

8.19 β–galactosidase assay and staining

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

For each enzymatic assay, six flies (*CG10920*, *CG12681*, and *CG1314* promoters) or five flies (*ocnus*, *CG9509* promoters) were homogenized in 150 µl of a buffer containing 0.1 M

Tris-HCl, 1 mM EDTA and 7 mM 2-mercaptoethanol at pH 7.5. The homogenate was kept on ice for 15 min, then centrifuged at 12000 g for 15 min at 4° C. Enzymatic assays were performed using 50 µl of supernatant and 50 µl of assay buffer (200 mM sodium phosphate, pH 7.4; 2 mM $MgCl_2$; 100 mM 2-mercaptoethanol) containing 1.33 mg/ml o-nitro-phenyl-β-D-galactopyranoside. β-galactosidase activity was measured spectrophotometrically at a wavelength of 420 nm over a period of 45 min at 25°C. The slope of the absorbance in relation to the incubation time was used to determine the amount of β-galactosidase and the relative expression between the autosomal and X-linked insertions. For each transformed line, β-galactosidase activity was measured for three biological replicates, each with two technical replicates.

In order to visualize β-galactosidase activity in whole tissues, dissected testes were incubated in the above buffer containing 1 mg/ml ferric ammonium citrate and 1.8 mg/ml of S-GAL sodium salt (Sigma-Aldrich; www.sigmaaldrich.com) for either 4 h or 8 h at 37°C.

8.20 Quantitative reverse transcription polymerase chain reaction

Total RNA was extracted from flies heterozygous (or hemizygous) for the transgene insertion using Trizol (Invitrogen; www.invitrogen.com) and following the manufacturer's protocol. Beginning with 5 µg of total RNA, *DNase*I treatment was carried out for 1 h at room temperature. Afterward, the RNA was reverse transcribed using the Superscript II reverse transcriptase and random hexamer primers (Invitrogen; www.invitrogen.com). A custom-designed TaqMan probe (Applied Biosystem; www.appliedbiosystems.com; forward primer: 5'-GCTGGGATCTGCCATTGTCA-3'; reverse primer: 5'-CAGCGCAGACCGTTTTCG-3'; FAM-labeled primer: 5'-CCCCGTACGTCTTCC-3') was used to quantify relative *lacZ* mRNA abundance using a Bio-Rad CFX 96 real-time PCR machine (Bio-Rad; www.bio-rad.com). As an internal reference, a probe to the ribosomal protein gene *RpL32* (probe number Dm 02151827_g1) was used. Relative transcript abundance was measured as the difference in threshold cycle (ΔC_t) between the target and the reference gene. The difference in transcript abundance between lines with X-linked and autosomal transgene insertions was measured as the average difference in ΔC_t among lines ($\Delta\Delta$Ct).

Stage-specific profiling of transcript abundance was performed using the above procedure, with the exception that the starting material consisted of dissected apical or proximal regions of 50 testes from each transformed line. The apical and proximal regions were defined according to (Vibranovski *et al.* 2009a). The measurement of the malpighian tubule was performed using the above procedure, with the exception that the starting material consisted of ten dissected tubule from each transformed line.

9. Results

9.1 Fine-scale mapping of additional insertions of the *ocnus* reporter gene construct

To test for regions of the X chromosome that escape MSCI, I used the approach of Hense *et al.* (2007) to generate a large number of independent insertions of a testis-specific reporter gene construct on the *D. melanogaster* X chromosome and create a fine-scale map of X chromosome inactivation in the male germline. In particular, I used genetic crosses to a transposase-expressing stock to produce 107 new independent X-chromosomal insertions. Additionally five previously mapped insertions of the *P[wFl-ocn-lacZ]* reporter gene construct, which contains the promoter of the *D. melanogaster* testis-specific *ocnus* gene fused to the *lacZ* gene of *E. coli* (Hense *et al.* 2007) were used. The precise chromosomal location of each insertion was determined by inverse-PCR (Bellen *et al.* 2004) (Appendix A).
To compare the X-linked expression to the autosomal expression, I mapped seven new autosomal insertions in this study and used the 15 previously mapped autosomal insertions of Hense *et al.* (2007) (Appendix B). For two of the 15 previously mapped autosomal insertions I was not able to determine the exact position inside the *D. melanogaster* genome. It was only possible to infer that the landing sites were associated with autosomal inheritance by following the inheritance of the *mini-white* gene (red eye color).

The first analysis included the comparison of autosomal and X-linked insertions. In particular, I compared the distribution of landing sites within and between classes of landing sites of insertions (Table 2).

Table 2: Comparison of X-linked and autosomal insertion sites. Expression was measured as mean units of β-galactosidase activity.

Location	X-linked insertions	X-linked expression	Autosomal insertions	Autosomal expression
5' UTR	65	2.34	9	9.77
Coding-exon	6	2.36	1	9.15
Intron	12	2.18	1	9.54
Intergenic	29	2.52	9	8.36
Unknown	0	–	2	7.57
Total	112	2.37	22	8.96

First, I distinguished two different types of landing sites: those in which the inserted construct was associated with genes, and those associated with intergenic regions. Further, if the insertions were associated with genes, I subdivided these landing sites into landings sites inside the 5'UTR, in coding exonic or intronic sequences. The last class consists of insertions for which exact position of the landing site could not be determined. 65 of the X-linked insertions were in the 5'UTR, six in coding exonic sequences, 29 in intronic sequences and 29 in intergenic regions. For the autosomal insertions, there were nine in the 5'UTR, one in coding exonic sequence, one in intronic sequence and nine in intergenic sequence. For autosomal and X-linked insertions I observed that the majority of insertions were associated with transcriptional units, including 12 out of 20 mapped autosomal insertions and 83 out of 112 mapped X-linked insertions. From the 12 autosomal insertions and the 83 X-linked insertions, nine autosomal and 65 X-linked insertions were located upstream of the coding sequence (predominantly in 5'UTRs). This preferential targeting of the 5'UTR is in accordance to previous reports (Spradling *et al.* 1995), which reported a tendency for *P* elements to be integrated at the 5'-end of genes. No significant bias for the distribution of landing sites between autosomal and X-linked insertions was found (χ^2 test, $P = 0.3571$).

9.2 Comparison of autosomal and X-linked expression of the *ocnus* construct

The reporter gene expression was measured for all autosomal and all X-linked insertions. In detail, I measured the expression in males and females carrying the *P[wFl-ocn-lacZ]* reporter gene construct. Each insertion was measured with three biological replicates, each with two technical replicates (Appendix C, D). Hense *et al.* (2007) showed that the reporter gene expression was expressed exclusively in testis by staining entire dissected testis and comparing the expression between dissected testis and adult gonadectomized adult male flies. I observed that the expression for the 22 autosomal insertions and the 112 X-linked insertions was significantly greater than zero in males (Student's *t*-test, one sample, $P < 0.0001$). To compare the expression between autosmal and X-linked insertions in males and females, I measured the expression for X-linked insertions in hemizygous males and heterozygous females and for autosomal insertions in heterozygous males and females to rule out any dosage effect. The expression was significantly higher in males than in females (MWW test, $P < 0.0001$, Table 3).

Table 3: Expression for the *P[wFl-ocn-lacZ]* reporter gene construct in males and females. Activity was measured as mean units of β-galactosidase enzymatic activity.

	Average male expression	Standard deviation of male expression	Average female expression	Standard deviation of female expression
Autosomal insertions	8.956	1.653	0.591	0.374
X-linked insertions	2.342	0.330	0.196	0.073

I detected a highly significant difference in expression of X-linked to autosomal insertions in males (MWW test, $P < 0.0001$; Figure 11). I find no evidence for any region along the X chromosome to escape X inactivation in the male germline.

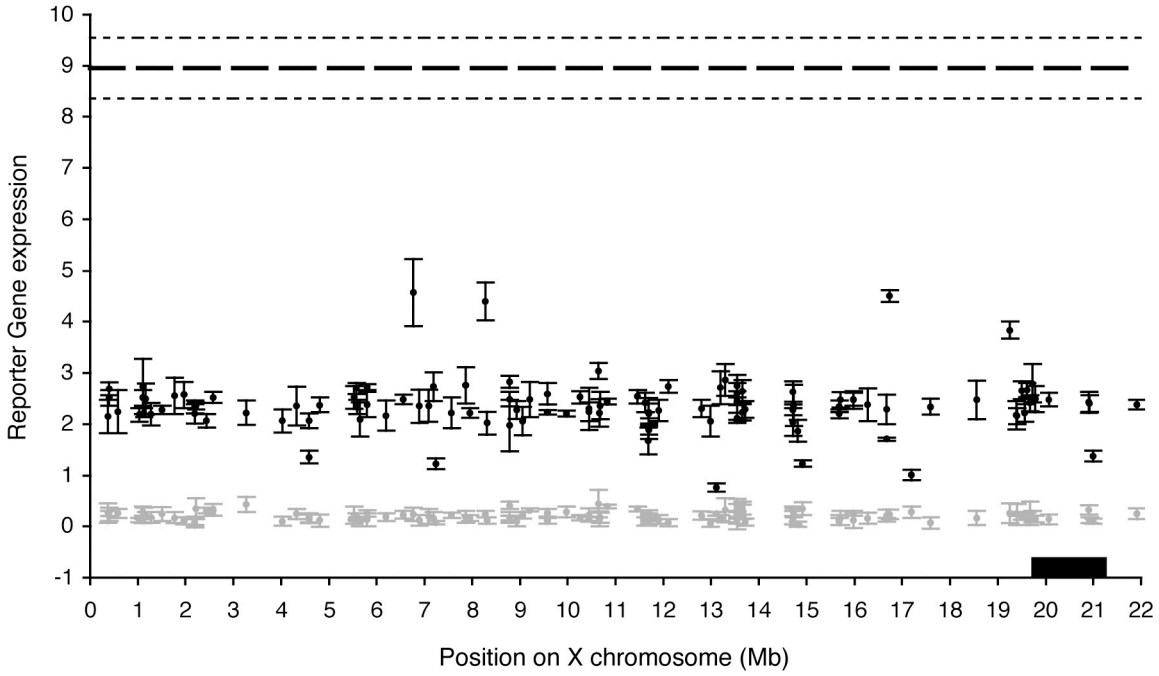

Figure 11: Mean expression (in units of β-galactosidase enzymatic activity) of 112 testis-specific reporter genes inserted on the *D. melanogaster* X chromosome. Black points represent expression in males, while gray points represent expression in females. Error bars indicate the standard deviation. For comparison, the average male expression of 22 autosomal insertions of the same transgene is indicated by a dashed line, with dotted lines indicating the standard deviation. Cytological region 19, which is enriched for newly-evolved and testis-expressed genes, is delineated by a black box on the X-axis.

There was some variation in male expression among transgenes inserted at different locations (Table 2), but no significant difference in expression between X-linked insertions of different landing sites in males was observed (MWW test, $P > 0.09$). However, X-linked transgenes inserted into intergenic regions tended to have a higher expression than those inserted into parts of transcriptional units, including the 5'UTR, coding-exons, or introns (Table 2). The four X-linked transgenes with the highest expression were spread across the X chromosome (at position 6.76 Mb, 8.28 Mb, 16.73 Mb, and 19.25 Mb), with two located in intergenic regions and two located in 5'UTRs. The insertion at 16.73 Mb lies ~500 bp upstream of the gene *CG13004*, which shows male-biased expression according to the SEBIDA database (Gnad and Parsch 2006) and testis enriched expression according to FlyAtlas (Chintapalli *et al.* 2007). However, none of the other three insertions was within 10 Kb of a male-biased or testis-expressed gene. Overall, the observed variation in expression among the X-linked insertions is unlikely to represent variation in X chromosome inactivation, as the coefficient of variation for X-linked insertions (13.2%) was less than that for autosomal insertions (18.5%).

Previous work indicated that there was a good accordance between transgene expression measured as protein abundance (β-galactosidase enzymatic activity) and mRNA abundance measured by quantitative reverse-transcription PCR (qRT-PCR) Hense *et al.* (2007). To confirm this for my transformants, I used qRT-PCR to measure transcript abundance of seven X-linked and seven autosomal transgenes. A significantly positive correlation between protein and mRNA abundance was observed (Figure 12, Appendix E) and there was significantly less transgene mRNA present in flies with X-linked insertions (MWW test, $P = 0.016$), indicating that the enzymatic assays accurately reflect transcript abundance.

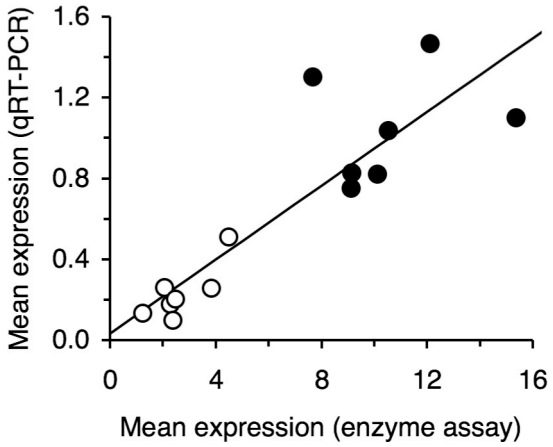

Figure 12: Comparison of expression measured by enzymatic assays and qRT-PCR for seven autosomal (solid circles) and seven X-linked (open circles) transgene insertions. There was a significant correlation between the expressions measured by the two methods (Pearson's $R = 0.859$, $P < 0.001$). The least-squares linear regression line is shown. Values on the X-axis indicate β-galactosidase activity units as defined by Hense *et al.* (2007). Values on the Y-axis indicate the relative threshold cycle difference between the transgene and the control gene, *RpL32*.

9.3 Analysis of male germline X inactivation at cytological band 19

The proposed hotspot for new gene evolution at cytological band 19 lies between nucleotide position 19.8 Mb and 21.2 Mb on the X chromosome (Flybase release 5.30; Tweedie *et al.* (2009)). Four of my transgene insertions (internal reference: 106, 104, 100, 49) fall within this interval, including an insertion at position 20,915,774 that is ~1 Kb away from the 3' end of the gene *Sdic1* (Figure 13). None of these four insertions showed a significantly higher expression than the rest of the X-linked insertions (Student's *t*-test, two-tailed, $P > 0.58$). The conclusion is that this region does not escape male germline X inactivation.

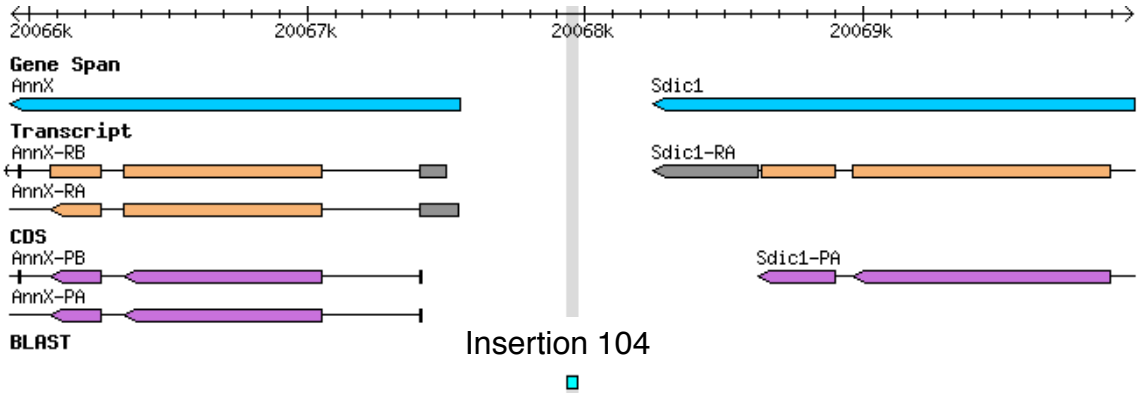

Figure 13: BLAST search of the amplified flanking region of the construct 104 (internal reference). This insertion is located next to the 3'-end of the coding gene *Sdic1*. *Sdic1* encodes a sperm protein and is a candidate for a gene that escapes male germline X inactivation.

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

To functionally test for an increased expression in the male germline associated with escaping the X chromosome, I performed experiments using the upstream regulatory sequences of three X-linked, testis-specific genes: *CG10920*, *CG12681*, and *CG1314*. These genes are located in different regions on the X chromosome and were chosen because they show significantly male- and testis-biased expression (Table 4).

Table 4: Summary of genes used in promoter analysis.

Gene	Cytogenetic map position	Male/female expression[a]	Testis/carcass expression[b]	α[c]	MK-test *P*-value
CG10920	7C	3.76	76.7	0.65	0.010
CG12681	4D	9.15	96.3	0.77	0.049
CG1314	19E	5.20	112.3	0.86	0.001

[a] Ratio of male-to-female expression from SEBIDA database (release 2.0; Gnad and Parsch 2006).

[b] Ratio of testis-to-carcass expression from FlyAtlas database (Chintapalli *et al.* 2007).

[c] Estimated proportion of positively-selected amino acid replacements (Smith and Eyre-Walker 2002).

In addition, for all three genes the McDonald-Kreitman test (McDonald and Kreitman 1991) indicates a significant excess of amino acid replacements between *D. melanogaster* and its sister-species *D. simulans*, which is a hallmark of adaptive evolution (Baines *et al.* 2008). The gene *CG1314* is of particular interest, because it is located at cytological region 19E, a region that is enriched for testis-expressed genes, including several genes that have evolved recently through gene fusion or *de novo* evolution of coding sequences (Nurminsky *et al.* 1998; Boutanaev *et al.* 2002; Levine *et al.* 2006; Chen *et al.* 2007). Thus, it is possible that regulatory sequences in this chromosomal region allow genes to avoid transcriptional silencing in the male germline.

Because functional information about the regulatory sequences of *CG10920*, *CG12681*, or *CG1314* was not available, I identified putative promoter sequences responsible for the testis-expression of the three genes by comparative sequence analysis. Previous studies have shown that testis-specific promoters are often short, conserved sequences located just upstream of the coding sequence (Michiels *et al.* 1989; Yanicostas and Lepesant 1990; Nurminsky *et al.* 1998; Hense *et al.* 2007). I aligned the orthologous upstream sequences from *D. melanogaster*, *D. simulans*, *D. sechellia, D. yakuba*, *D. erecta*, and chose conserved regions of 580 bp (*CG10920*), 765 bp (*CG12681*), and 508 bp (*CG1314*) for further functional analysis.

Putative promoter sequences were fused to the *E. coli lacZ* gene (encoding β-galactosidase) and cloned into the p*P[wFl]* transformation vector (Siegal and Hartl 1996) (Figure 14). Stably

transformed *D. melanogaster* strains were generated by embryo microinjection (Rubin and Spradling 1982; Spradling and Rubin 1982) and subsequent genetic crosses.

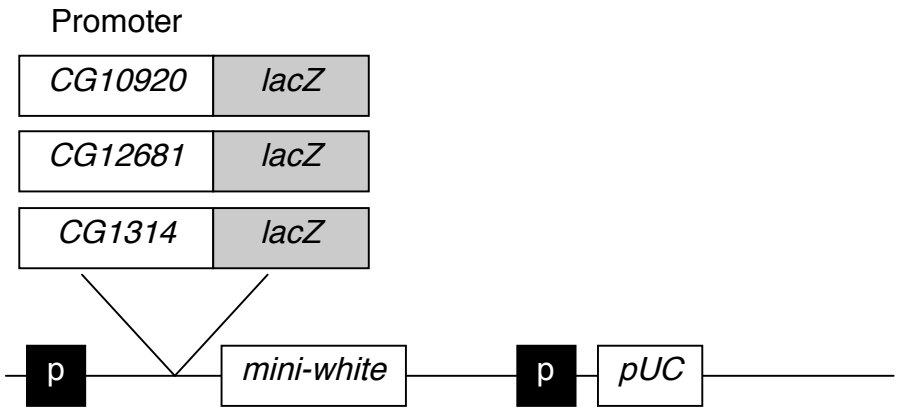

Figure 14: Reporter gene constructs. Promoter sequences of three X-linked, testis-expressed genes were fused to the *E. coli lacZ* reporter gene and independently inserted into the p*P[wFl]* transformation vector (Siegal and Hartl 1996). This vector contains terminal repeat sequences of a *Drosophila* transposable element (*P*) and the *mini-white* gene as a selectable marker (eye color). The portion of the plasmid required for replication in *E. coli* is labeled "pUC".

To control for testis specific expression of the three promoter constructs, I compared the expression in dissected testis to carcass (gonadectomized flies) of one randomly chosen autosomal and X-linked transformed *D. melanogaster* (Table 5) for each construct.

Table 5: Expression (mean units of β-galactosidase enzymatic activity) for one autosomal and one X-linked insertion in testis compared to gonadectomized flies (carcass).

Construct	Autosomal expression carcass	Autosomal expression testis	X-linked expression carcass	X-linked expression testis
CG10920	0.04	11.98	-0.06	2.01
CG12681	0.05	7.30	0.09	1.12
CG1314	0.03	3.95	0.03	1.20

I observed significant higher expression in the testis compared to gonadectomized flies (MWW test, $P < 0.029$). To confirm, I did β-galactosidase staining of entire testis within males (Figure 15). The expression in the testis was highly enriched, especially for autosomal insertions.

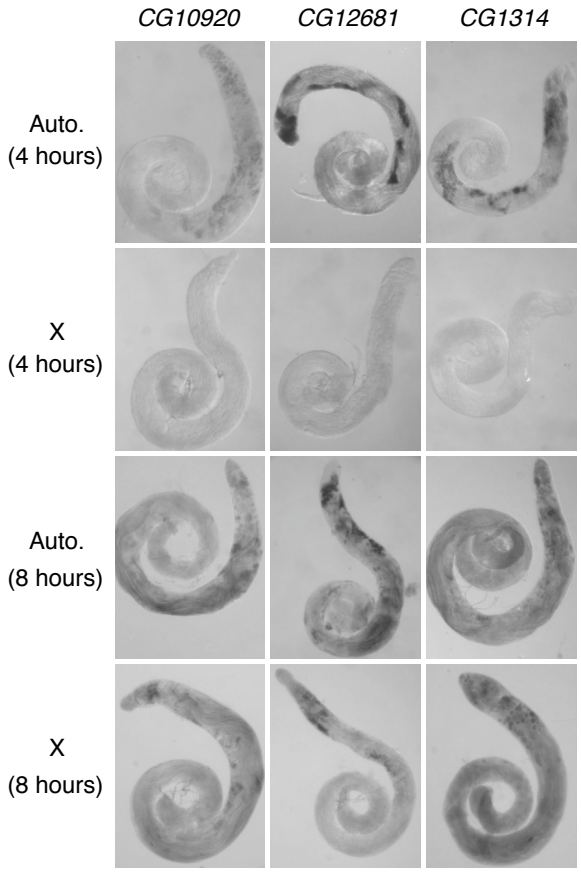

Figure 15: β-galactosidase activity staining in testes. Testes were dissected from males containing autosomal or X-linked insertions of each reporter gene construct and incubated with S-Gal (Sigma-Aldrich) for 4 or 8 hours. Dark areas indicate the presence of reporter gene (β -galactosidase) activity.

9.5 Fine-scale mapping of transgene insertions of three X-linked promoters

In total, I recovered eight, eight, and eight independent autosomal insertions and seven, eight, and nine independent X-linked insertions of the *CG10920*, *CG12681*, and *CG1314* construct, respectively. In order to analyze the local context of the transgene insertions, I performed inverse PCR to map their precise position in the *D. melanogaster* genome (Bellen *et al.* 2004). I was able to map eight autosomal and seven X-linked insertions for the *CG10920* construct, seven autosomal and six X-linked insertions for the *CG12681* construct, and six autosomal insertions and nine X-linked insertions for the *CG1314* construct (Figure 16, Appendix F).

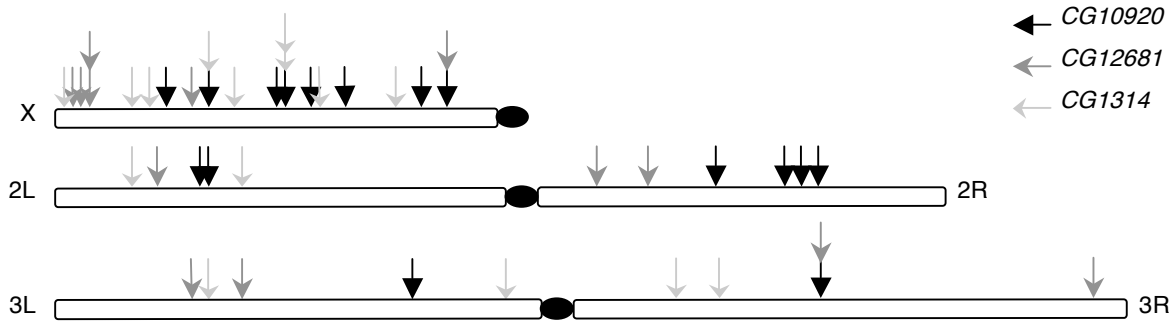

Figure 16: Map of transgene insertion locations. The precise chromosomal location of each insertion was determined by inverse PCR. Each arrow indicates an insertion at a unique site. Multiple arrows at the same position do not indicate insertions at the same site, but insertions that are too close to each other (within 400 kb) to be distinguished on the scale of the figure.

I was able to precisely map 88% of the autosomal insertions and 92% of the X-linked insertions. Further, I analyzed the integration of landing sites into coding or intergenic regions. The landing sites associated with genes were subdivided into insertions associated with the 5'UTR, coding exonic or intronic sequences. A final class includes insertions where the precise location of the construct could not be determined and I was only able to infer autosomal or X-linked inheritance. Of 24 autosomal insertions, five were found in the 5'UTR, seven in coding exonic sequences, one in intronic sequence, eight in intergenic sequences and for three insertions I was only able to infer autosomal linkage by following the inheritance of the *mini-white* gene (red eye color). Similar results were found for the 24 X-linked insertions, one insertion was in the 5'UTR, seven in coding exonic sequences, nine in intronic sequences, five in intergenic sequences and for two I was only able to infer X-linkage by following the inheritance of the *mini-white* gene (red eye color). The distribution of landing sites for X-linked and autosomal insertions (Table 6) showed slightly significant differences (χ^2 test, P = 0.041). However, the differences between autosomal and X-linked insertions could not be explained by a difference in insertion site preference. The expression of the landing site classes was similar in range. Most of the insertions were associated with genes. In detail, 30 out of 48 insertions were associated with genes (5'UTR, coding exonic or intronic sequences).

Table 6: Distribution of independent landing sites for autosomal and X-linked insertions. The expression of each insertion (mean units of β-galactosidase enzymatic activity) was normalized to the average X-linked expression of the corresponding construct for each of the three promoter constructs (*CG10920*, *CG12691*, and *CG1314*).

Location	Autosomal insertions	Autosomal expression	X-linked insertions	X-linked expression
5'UTR	5	3.14	1	0.82
Coding-exon	7	2.54	7	1.06
Intron	1	2.18	9	0.95
Intergenic	8	3.38	5	0.97
Unknown	3	4.00	2	0.92
Total	24	3.11	24	0.98

I observed a lower autosomal expression within introns, but the sample size of one was too small to allow for statistical testing.

9.6 Comparison of X-linked and autosomal reporter gene insertions for three X-linked promoters

For all 48 independent insertions I performed a β-galactosidase assay on male and female flies (Appendix G). The expression of the transgene insertions was measured in three biological replicates, each with two technical replicates. To compare the expression between autosmal and X-linked insertions in males and females, I measured the expression of X-linked insertions in hemizygous males and heterozygous females and for autosomal insertions in heterozygous males and females. For autosomal insertions of the *CG10920* transformants, the average (standard deviation) β-galactosidase activity in males was 6.83 (2.42), while that in females was 0.08 (0.08). For the autosomal *CG12681* transformants, the average β-galactosidase activity in males was 5.20 (1.34), while that in females was 0.14 (0.10). For the autosomal *CG1314* transformants, the average β-galactosidase activity in males was 2.08

(0.29), while that in females was 0.14 (0.09). In all cases, the difference in expression between males and females was significant (MWW test; $P < 1.55*10^{-4}$). I also measured the β-galactosidase activity for X-linked insertions in male and female flies. For the X-linked *CG10920* transformants, the average β-galactosidase activity in males was 2.44 (0.32), while that in females was -0.01 (0.10). For the X-linked *CG12681* transformants, the average β-galactosidase activity in males was 1.35(0.19), while that in females was 0.11 (0.06). For the X-linked *CG1314* transformants, the average β-galactosidase activity in males was 0.72 (0.22), while that in females was 0.05 (0.07). In all cases, the difference in expression between males and females was significant (MWW test, $P < 5.83*10^{-4}$).

Although the X-linked insertions of all three promoters constructs showed expression in testis (Figure 15), their level of expression was significantly lower than that of autosomal insertions (Figure 17–20).

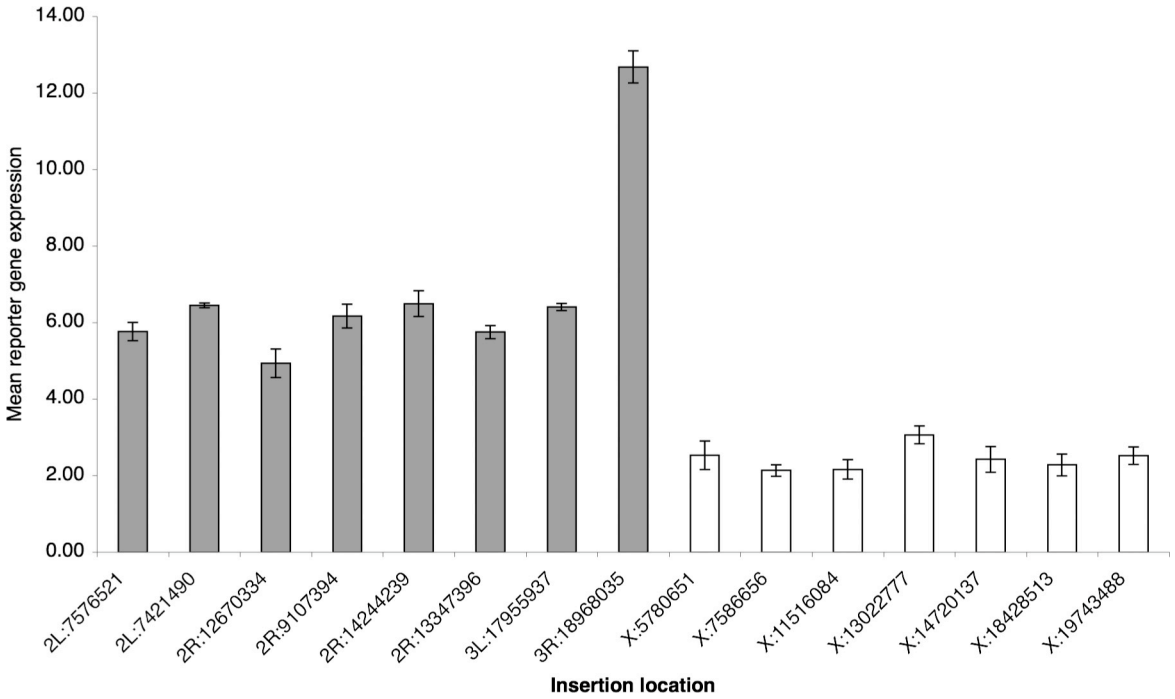

Figure 17: Expression of autosomal and X-linked promoter reporter gene insertions. For the *CG10920* reporter gene construct, the mean β-galactosidase activity of transformants with autosomal (gray bars) and X-linked (open bars) insertions are shown. Each bar represents an independent insertion at a different genomic location. Error bars indicate the standard deviation.

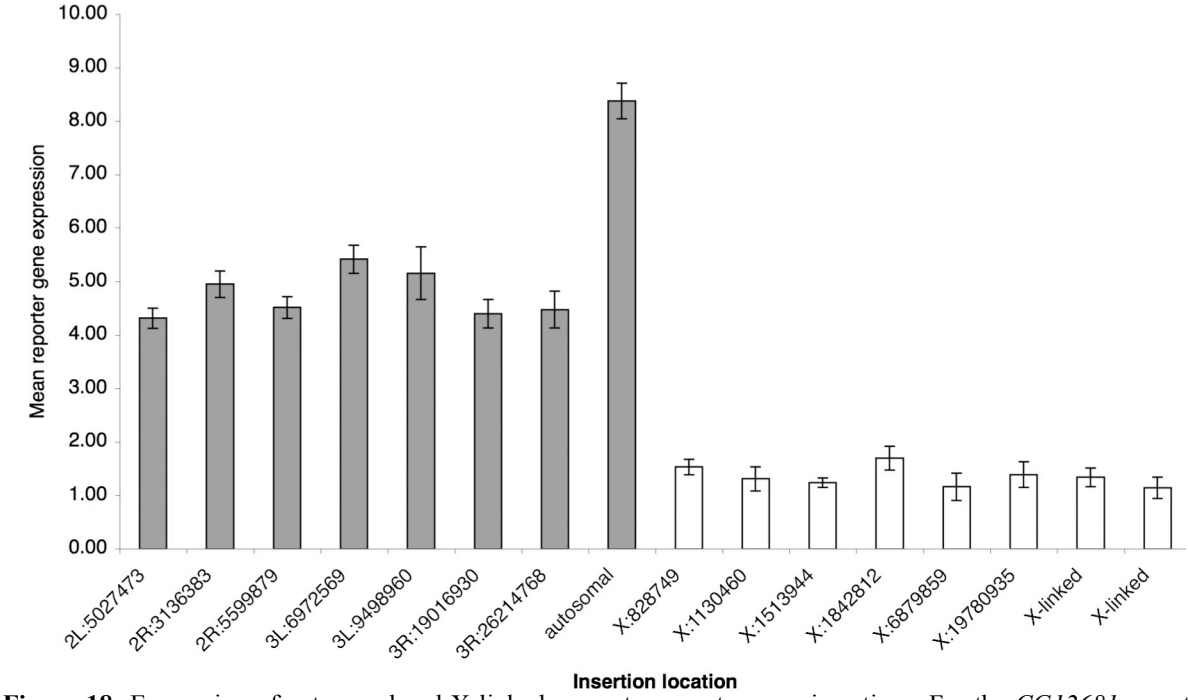

Figure 18: Expression of autosomal and X-linked promoter reporter gene insertions. For the *CG12681* reporter gene construct, the mean β-galactosidase activity of transformants with autosomal (gray bars) and X-linked (open bars) insertions are shown. Each bar represents an independent insertion at a different genomic location. Error bars indicate the standard deviation.

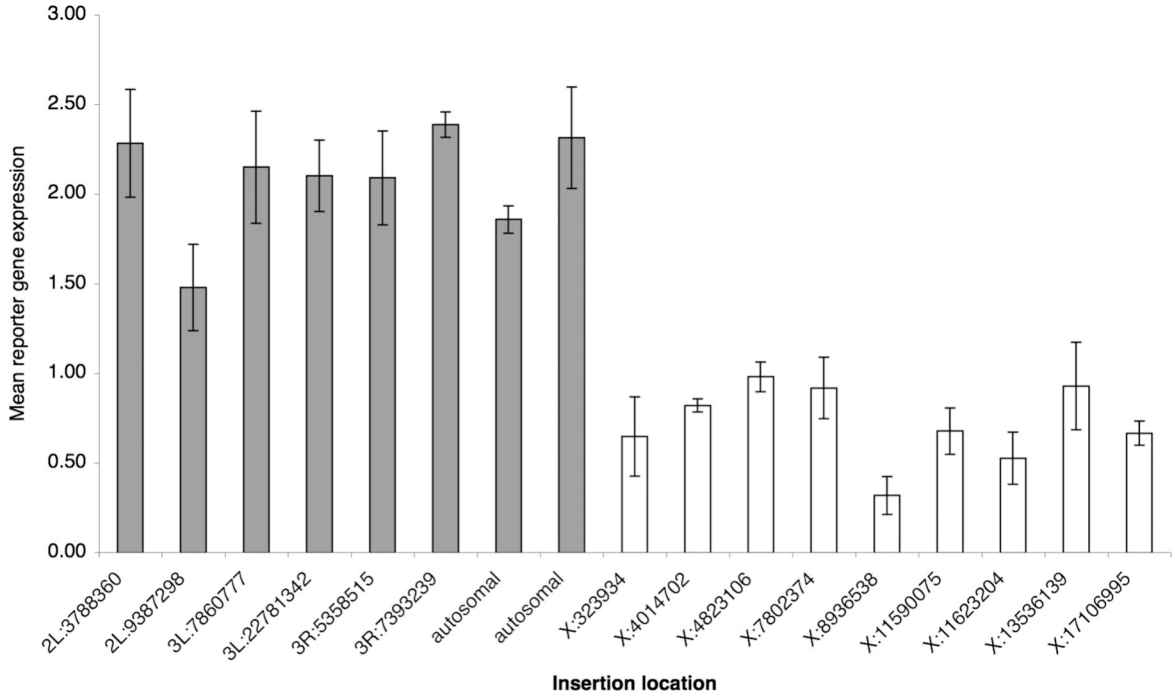

Figure 19: Expression of autosomal and X-linked promoter reporter gene insertions. For the *CG1314* reporter gene construct, the mean β-galactosidase activity of transformants with autosomal (gray bars) and X-linked (open bars) insertions are shown. Each bar represents an independent insertion at a different genomic location. Error bars indicate the standard deviation.

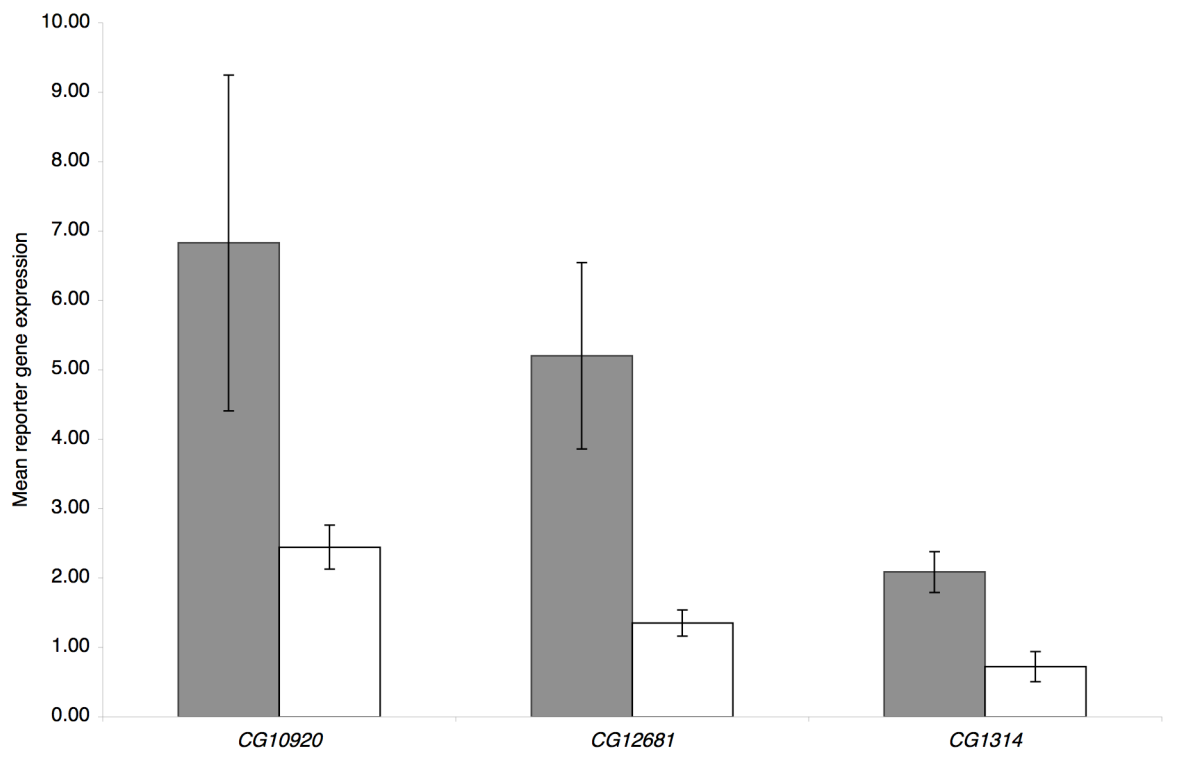

Figure 20: Mean expression of autosomal and X-linked promoter reporter gene insertions. For each reporter gene construct, the mean β-galactosidase activity of transformants with autosomal (gray bars) and X-linked (open bars) insertions are shown. Each bar represents the average expression of independent insertions at different genomic locations from one promoter reporter gene construct, either autosomal or X-linked. In all cases, autosomal expression was significantly greater than X-linked expression (MWW test, $P < 0.001$). Error bars indicate the standard deviation.

The average difference in β-galactosidase enzymatic activity between autosomal and X-linked insertions were 2.8-fold, 3.9-fold, and 2.9 fold for the *CG10920*, *CG12681*, and *CG1314* reporter constructs, respectively.

To confirm these results at the level of transcript abundance, I performed quantitative reverse transcription (qRT)-PCR to estimate relative levels of *lacZ* mRNA. For all three promoter reporter gene constructs, the *lacZ* transcript abundance was significantly higher for autosomal insertions than for X-linked insertions (Appendix H, Figure 21).

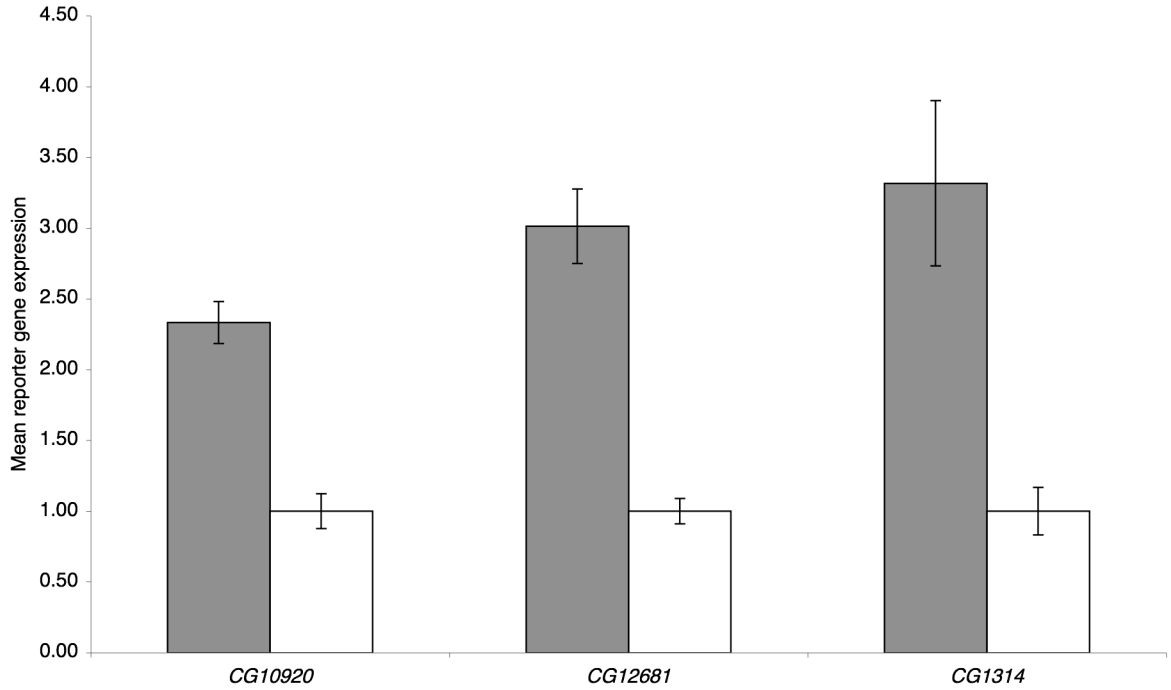

Figure 21: Reporter gene transcript abundance estimated by qRT-PCR. Bars indicate the mean relative *lacZ* transcript abundance of autosomal (gray bars) and X-chromosomal (open bars) transformants of each promoter construct. In all cases, autosomal expression was significantly greater than X-chromosomal expression (MWW test, $P < 0.001$). Error bars indicate the standard deviation.

The average difference in *lacZ* mRNA concentration between autosomal and X-linked insertions were 2.33-fold, 3.01-fold, and 3.32-fold for the *CG10920*, *CG12681*, and *CG1314* reporter constructs, respectively. Thus, the estimates of transcript abundance agree well with the estimates of protein abundance. Furthermore, there was a strong correlation between expression level measured by qRT-PCR and β-galactosiadase activity (*CG10920*: Spearmann's $R = 0.78$, $P = 9.92*10^{-5}$; *CG12681*: $R = 0.82$, $P = 3.97*10^{-7}$; *CG1314*: $R = 0.66$, $P = 0.0024$) (Figure 22).

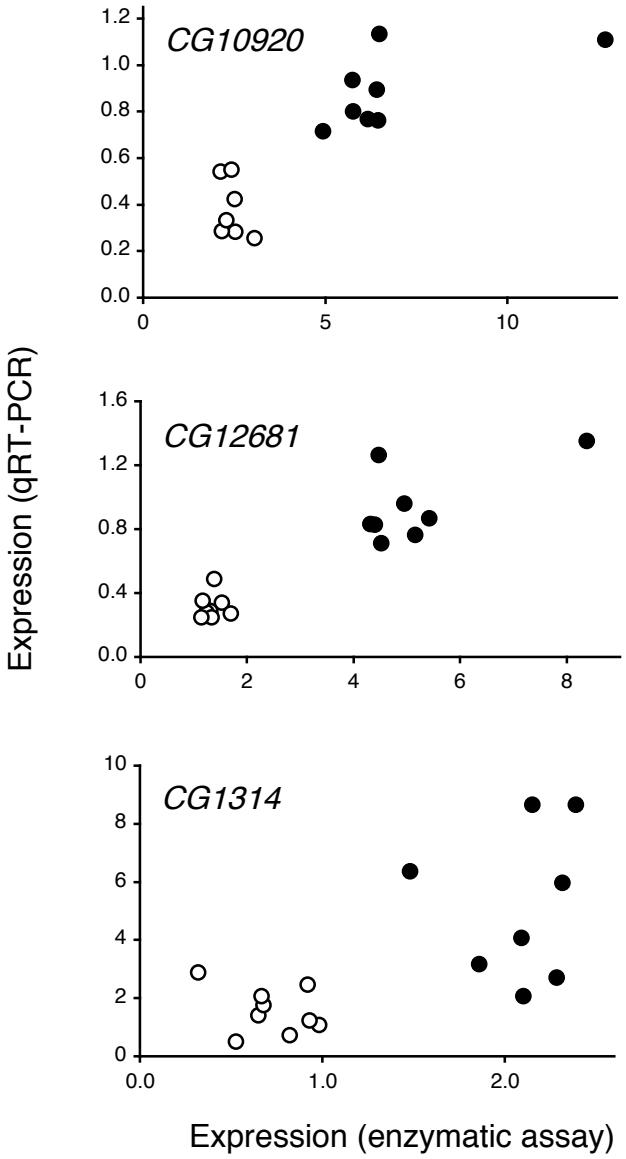

Figure 22: Comparison of reporter gene expression measured at the level of transcript abundance (by qRT-PCR) and protein abundance (by enzymatic assay). X-linked insertions are indicated by open circles, while autosomal insertions are indicated by solid circles. For each of the three promoter constructs (*CG10920*, *CG12681*, and *CG1314*), there was a significant correlation between gene expression levels estimated by the two methods (linear regression, $P < 0.0025$).

In all cases, I found significantly higher expression of transgenes inserted on the autosomes relative to those inserted on the X chromosome. My results are consistent with global transcriptional inactivation of the X chromosome in the male germline and provide direct experimental evidence for an increased expression by escaping the X chromosome.

9.7 Stage specific expression profiling for three X-linked promoters

I also investigated the expression of autosomal and X-linked transgenes during different stages of spermatogenesis by performing qRT-PCR on the dissected apical (mitosis) and proximal (meiosis) regions of testes, which are enriched for mitotic and meiotic cells (Vibranovski *et al.* 2009a). In both stages, there was significantly less expression for X-linked than autosomal transgenes (Figure 23 + 24).

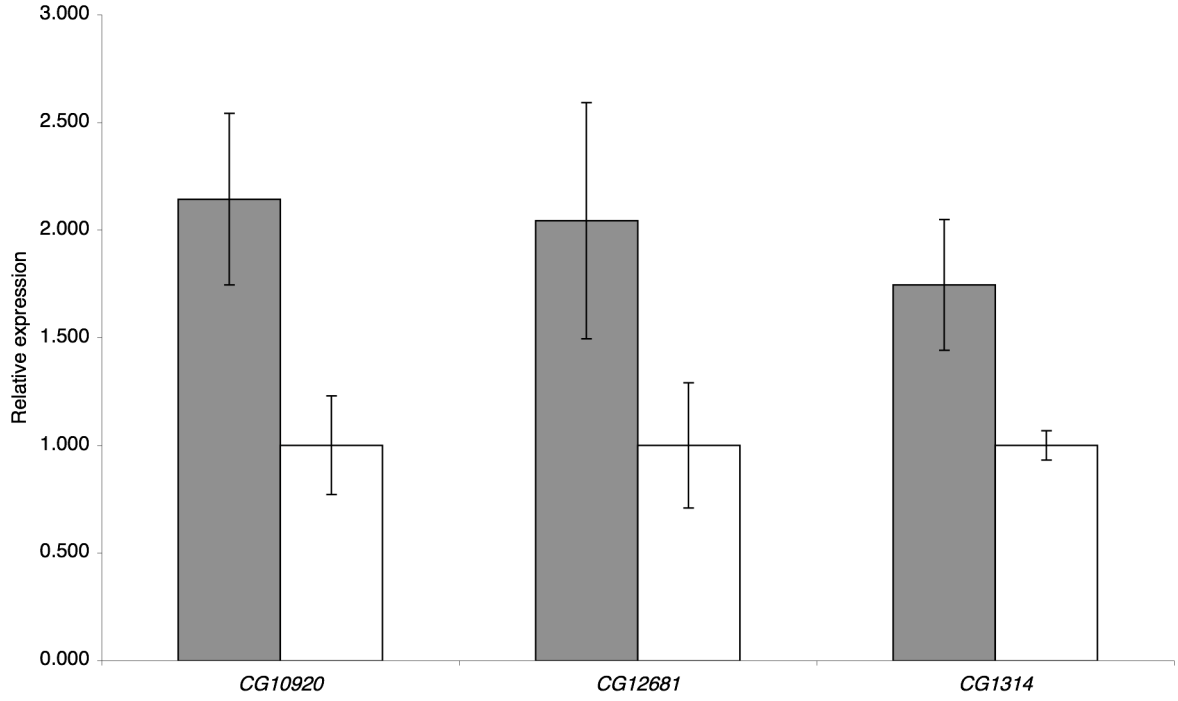

Figure 23: Stage-specific profiling of reporter gene transcript abundance. qRT-PCR was performed on dissected apical (mitosis) region of testes as described in Vibranovski *et al.* (2009a). Bars indicate the mean relative *lacZ* transcript abundance of autosomal (gray bars) and X-chromosomal (open bars) transformants of each promoter construct. For each promoter construct, a single transformed line with expression typical for its class was assayed with two biological replicates, each with two technical replicates. In all cases, autosomal expression was significantly greater than X-chromosomal expression (Student's *t*-test, two-tailed, $P < 0.05$). Error bars indicate the standard deviation.

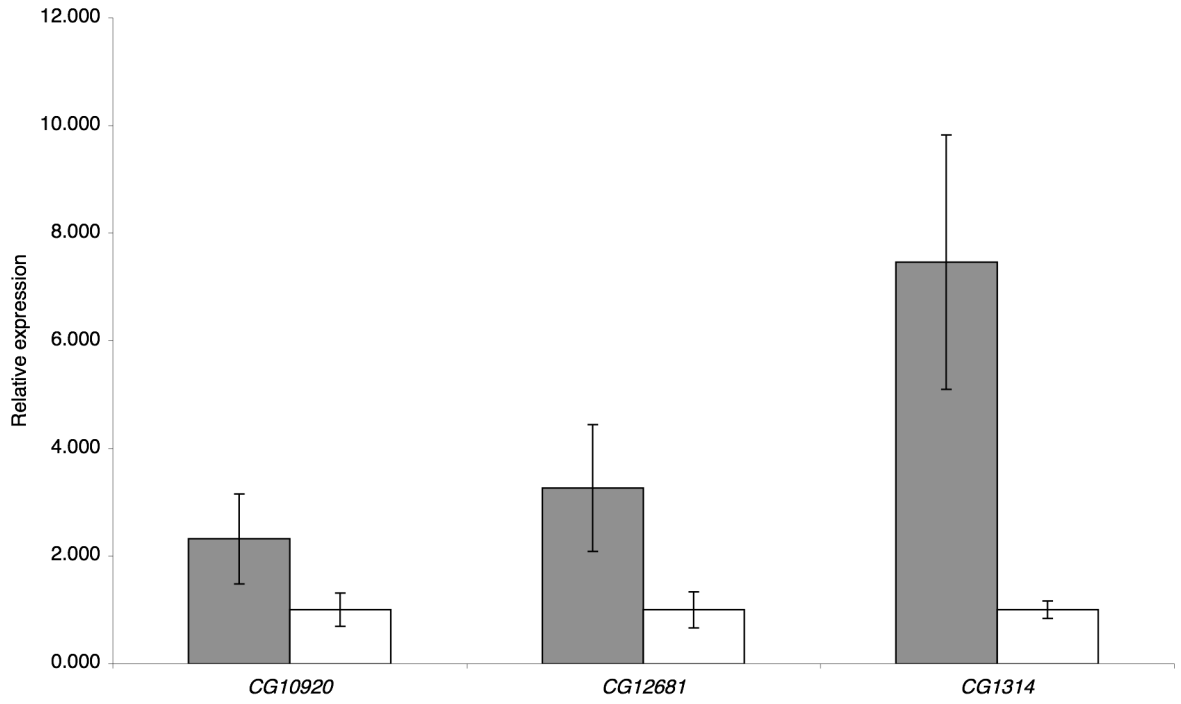

Figure 24: Stage-specific profiling of reporter gene transcript abundance. qRT-PCR was performed on dissected proximal (meiosis) regions of testes as described in Vibranovski *et al.* (2009a). Bars indicate the mean relative *lacZ* transcript abundance of autosomal (gray bars) and X-chromosomal (open bars) transformants of each promoter construct. For each promoter construct, a single transformed line with expression typical for its class was assayed with two biological replicates, each with two technical replicates. In all cases, autosomal expression was significantly greater than X-chromosomal expression (Student's *t*-test, two-tailed, $P < 0.05$). Error bars indicate the standard deviation.

For the *CG10920* and *CG12681* constructs, the ratio of autosomal to X-linked expression was similar in both mitotic and meiosis cells. In contrast, *CG1314* showed a greater enrichment of autosomal expression during meiosis (7.5-fold) than mitosis (1.8-fold). For these reason, MSCI appears to be sufficient to explain my results.

9.8 The expression difference of *CG9509* between European and African populations of *D. melanogaster*

From analyses of gene expression divergence between European and African populations, I obtained a candidate gene (*CG9509*) that is highly overexpressed in the European populations (Hutter *et al.* 2008; Müller *et al.* 2011) and showed sign of positive selection in the putative promoter region of the European population (Saminadin-Peter 2008). To test for functional *cis*-regulatory sequences in the putative promoter region, I amplified the putative promoter region of the European strain E12 and from the African strain Z82. These promoter regions were fused to the *lacZ* reporter gene from *E. coli*, which encodes the β−galactosidase enzyme. The reporter gene constructs were cloned into the p*attB* transformation vector (Bischof *et al.* 2007) and stably transformed *D. melanogaster* strains were generated by microinjection and using the *Φ*C31 transformation system (Bischof *et al.* 2007). In particular, I used the ZH-68E and the ZH-86Fb landing sites to compare the African and European promoters. To confirm the presence of the construct in the *D. melanogaster* genome, I did PCR with primers complementary to the *lacZ* coding region and the genomic flanking region of the landing site. The expression difference between males and females was compared for autosomal insertions in heterozygous males and females. Each enzymatic measurement consisted of three biological replicates, each with two technical replicates. The expression in males and females was significantly higher for the European promoter compared to the African promoter (MWW test, $P < 0.002$) for both landing sites (Table 7, Figure 25). The population difference in expression for the landing site ZH-68E was 2.6-fold in males and 3-fold in females, and for the landing site ZH-86Fb it was 3.5-fold in males and 3.8-fold in females.

Table 7: Male and female expression (β-galactosidase activity) driven by the African or European *CG9509* promoter sequence. The landings sites ZH-68E and ZH-86Fb of the ΦC31 transformation system (Bischof *et al.* 2007) were used.

	ZH-68E		ZH-86Fb	
	average	standard deviation	average	standard deviation
European expression in males	19.04	1.74	20.68	0.77
African expression in males	7.32	0.62	5.91	0.34
European expression in females	17.06	0.41	18.17	1.45
African expression in females	5.55	0.16	4.72	0.81

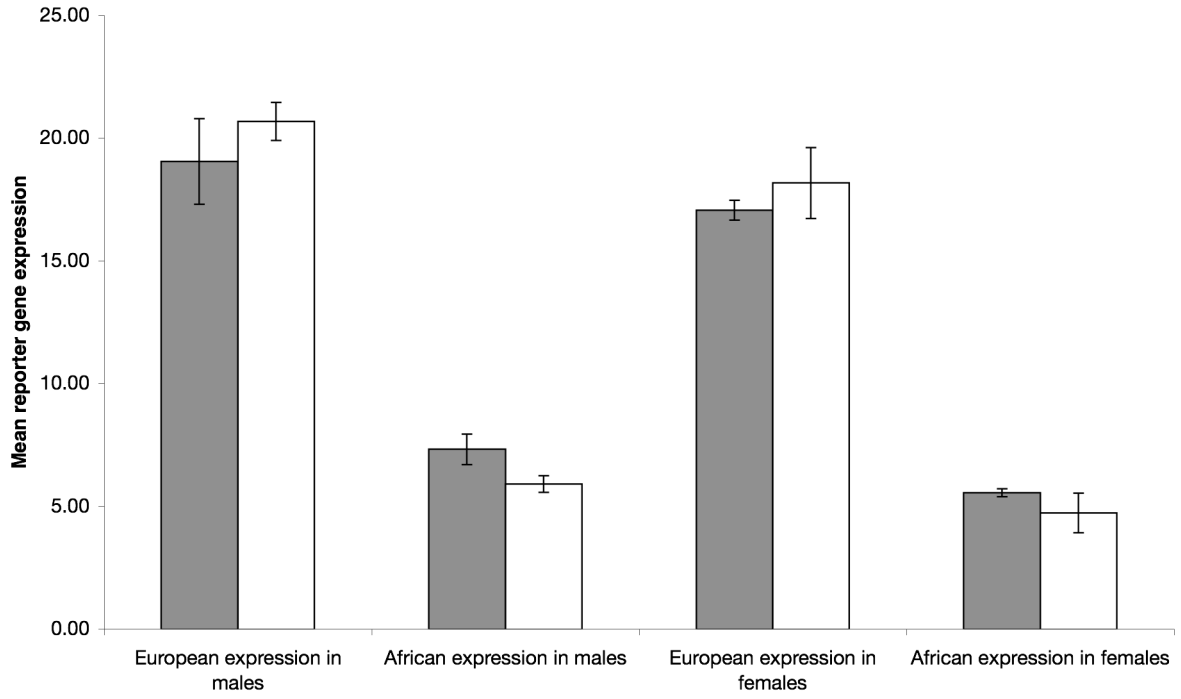

Figure 25: Male and female average expression of the β-galactosidase activity driven by the African or European *CG9509* promoter sequence. The landings sites ZH-68E (gray bars) and ZH-86Fb (open bars) of the ΦC31 transformation system (Bischof *et al.* 2007) were used. Error bars indicate the standard deviation.

I observed a higher expression in males compared to females using the European promoter (ZH-68E: 1.12 fold, $P = 0.13$; ZH-86Fb: 1.14 fold, $P = 0.041$) and the African promoter (ZH-68E: 1.34 fold, $P = 0.002$; ZH-86Fb: 1.25 fold, $P = 0.065$). The difference in expression

between males and females was higher in the African population (1.25- and 1.34-fold) compared to the European population (1.12- and 1.14-fold), which corresponds to the male-biased expression of the *CG9509* gene reported in the SEBIDA database (Gnad and Parsch 2006).

To ensure that the differences in expression I observed at the protein level reflect a difference at the mRNA-level, I performed a qRT-PCR for whole male and female flies carrying either the European promoter reporter gene construct or the African promoter reporter gene construct (Table 8, Figure 26). Each qRT-PCR consisted of two biological replicates, each with two technical replicates. All measurements were performed on heterozygous males and females.

Table 8: Male and female expression of *lacZ* mRNA driven by the African or European *CG9509* promoter sequence. The landings sites ZH-68E and ZH-86Fb of the *Φ*C31 transformation system (Bischof *et al.* 2007) were used.

	ZH-68E		ZH-86Fb	
	average	standard deviation	average	standard deviation
European expression in males	3.91	0.72	2.57	0.50
African expression in males	1.00	0.45	0.68	0.08
European expression in females	0.38	0.09	0.34	0.15
African expression in females	0.09	0.02	0.11	0.02

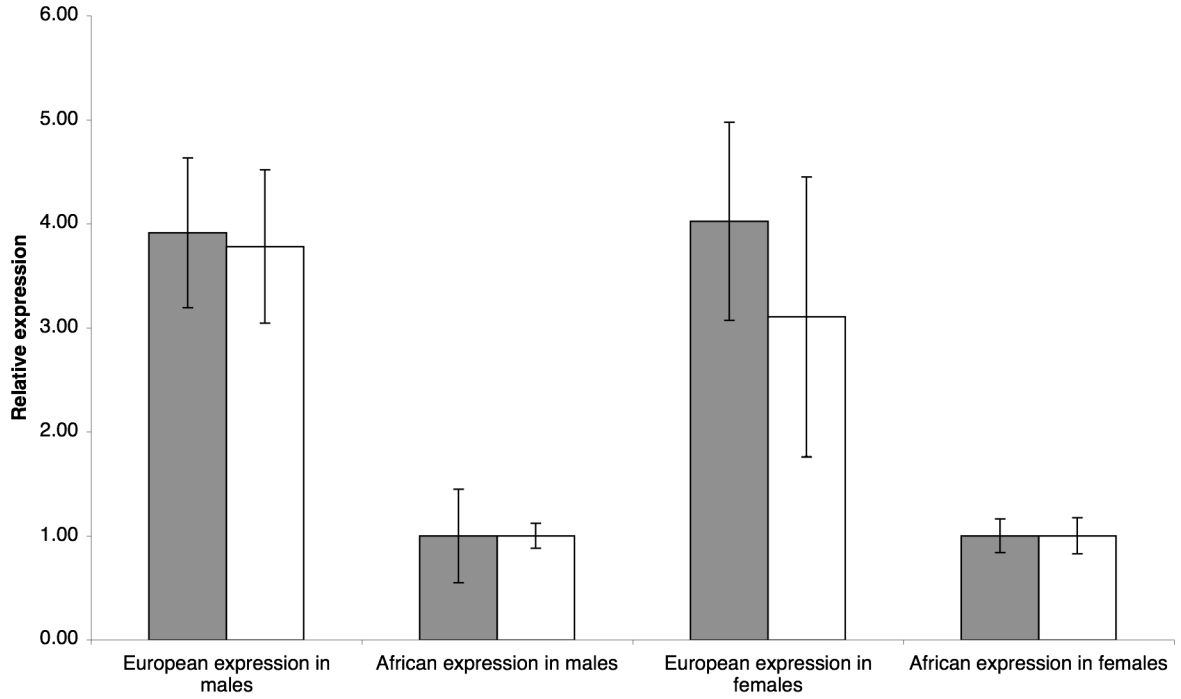

Figure 26: qRT-PCR of *lacZ* mRNA abundance in male and female flies driven by the African or European *CG9509* promoter sequence. The landings sites ZH-68E (gray bars) and ZH-86Fb (open bars) of the ΦC31 transformation system (Bischof *et al.* 2007) were used. To correct for the sex-biased expression of the reference gene *RpL32*, expression of the different landing sites and sexes was normalized to the African expression, which was set to one. Error bars indicate the standard deviation.

I measured a higher expression in males compared to females using the European promoter (ZH-68E: 10.29 fold, $P = 0.001$; ZH-86Fb: 7.54 fold, $P = 0.001$) and the African promoter (ZH-68E: 10.57 fold, $P = 0.001$; ZH-86Fb: 6.19 fold, $P = 0.001$). The higher expression difference between males and females measured by qRT-PCR (European: ~10-fold; African: ~6.5-fold) compared to the enzymatic assay (European: ~1.1-fold; African: ~1.3-fold) is likely due to the sex-biased expression of the reference gene *RpL32*, which showed ~2-4-fold higher expression (SEBIDA; Gnad and Parsch 2006) in female flies. This sex bias has no influence on my results, because I compared the expression between populations and not between the sexes. The expression differences caused by comparing the expression of the *lacZ* gene driven by the European promoter compared to the African promoter are highly significant for both sexes (Student's *t*-test; two-tailed; $P < 0.023$). The population difference in expression for the landing site ZH-68E was 3.9-fold in males and 4.02-fold in females and for the landing site ZH-86Fb 3.78-fold for males and 3.10-fold for females.

The estimates of transcript abundance agree well with the estimates of protein abundance. Furthermore, there was a correlation between expression levels measured by qRT-PCR and β-

galactosiadase activity (males: Spearmann's $R = 0.8$, $P = 0.10$; females: $R = 0.6$, $P = 0.02$) (Figure 27).

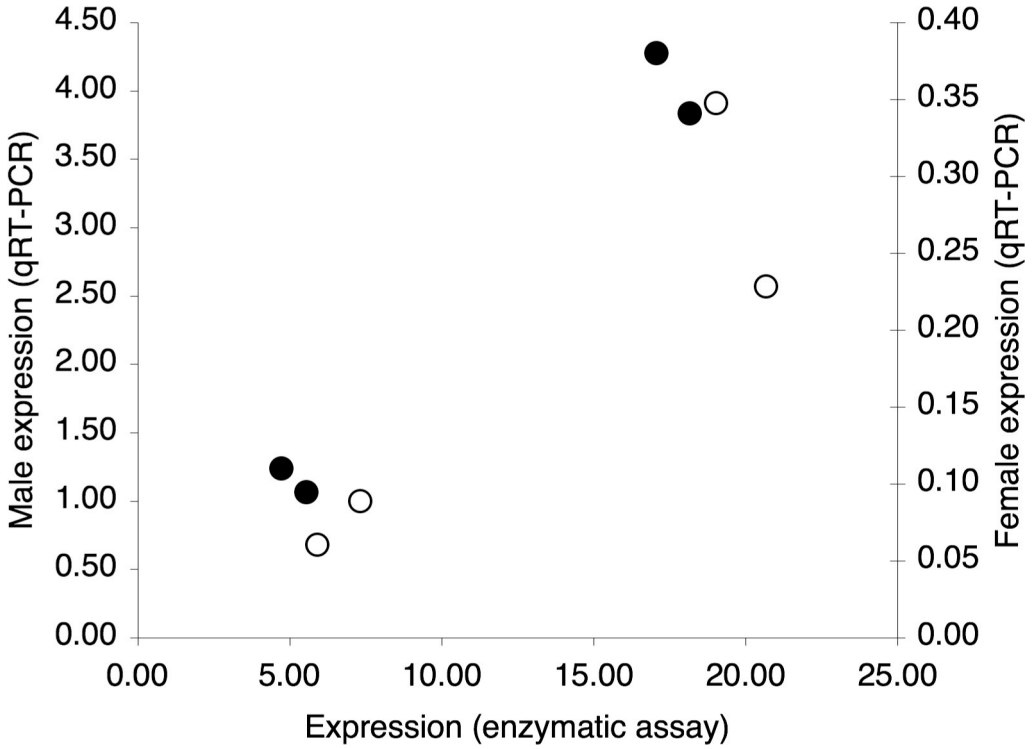

Figure 27: Comparison of reporter gene expression measured at the level of transcript abundance (by qRT-PCR) and protein abundance (by enzymatic assay). Female expression is indicated by open circles, while male expression is indicated by solid circles.

Both, the measurement of expression of the level of protein abundance or mRNA abundance showed a reduced expression for the reporter gene expression driven by the African promoter in comparison to the European promoter. This indicated that differences in the promoter sequence of the European promoter are responsible for the differences in expression between the two populations. Due to the uniform background *yw* flies used for the promoter study no *trans*-regulatory effect could influence these results and the expression differences are caused by *cis*-regulatory elements.

9.9 Expression profiling of the European and African *CG9509* promoter in the malpighian tubule

The previous experiments indicate that *cis*-regulatory elements are responsible for the expression differences between the African and the European populations. The expression was measured in whole flies. However, other expression studies showed that the gene *CG9509* is highly expressed in the malpighian tubule (Chintapalli *et al.* 2007) (Table 9). The expression in the malpighiam tubule is 10-fold higher than in other tissues of adult *Drosophilas*.

Table 9: Expression of the *CG9509* gene in different tissues of adult flies of *D. melanogaster*. The expression was measured by whole transcriptome microarrays (FlyAtlas, Chintapalli *et al.* 2007).

Tissue	mRNA Signal	Present Call
Brain	3 ± 1	1 of 4
Head	5 ± 0	0 of 4
Eye	7 ± 1	1 of 4
Thoracicoabdominal ganglion	6 ± 1	1 of 4
Salivary gland	15 ± 5	1 of 4
Crop	4 ± 0	0 of 4
Midgut	653 ± 54	4 of 4
Tubule	5937 ± 295	4 of 4
Hindgut	359 ± 35	4 of 4
Heart	25 ± 6	4 of 4
Fat body	30 ± 20	4 of 4
Ovary	0 ± 0	0 of 4
Testis	8 ± 1	3 of 4
Male accessory glands	4 ± 0	0 of 4
Virgin spermatheca	14 ± 5	2 of 4
Mated spermatheca	15 ± 2	4 of 4
Adult carcass	16 ± 2	3 of 4

To determine if the difference in population expression observed in whole flies correlates with expression differences in the malpighian tubule, I performed β-galactosidase enzymatic assays on dissected malpighian tubules. The expression assay was performed in heterozygous males and each measurement consisted of two biological and two technical replicates. The expression difference between the European reporter gene construct and the African reporter gene construct in malphigian tubule was 2.25-fold for the ZH-68E landing site and 3.23-fold for the ZH-86Fb landing site (Figure 28). The higher expression in the European population was highly significant (Student's *t*-test, two-tailed, $P < 0.001$).

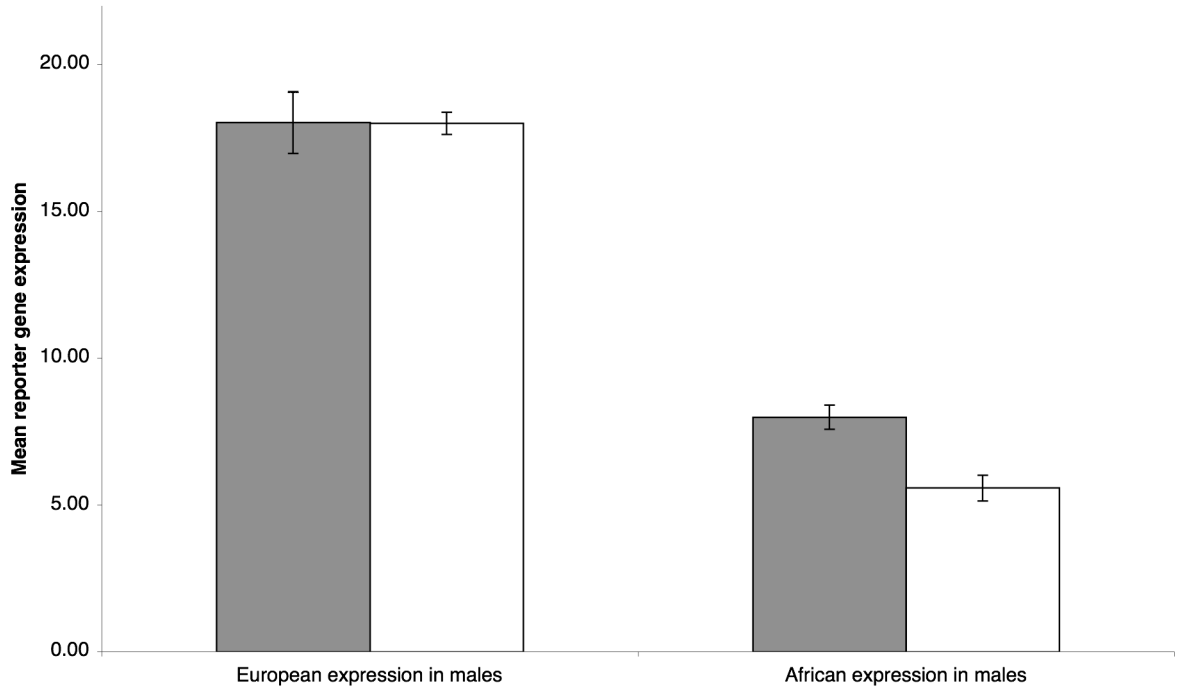

Figure 28: Male and female expression (β-galactosidase activity) driven by the African or European *CG9509* promoter sequence in malpighian tubule. The landings sites ZH-68E (gray bars) and ZH-86Fb (open bars) of the *Φ*C31 transformation system (Bischof *et al.* 2007) were used. Error bars indicate the standard deviation.

The expression differences observed in the malpighian tubule correlate very well with the expression differences observed in whole flies. This indicates that the expression differences measured between the European population and the African population for *CG9509* is result of changes in the *cis*-regulatory sequence of the European promoter and that increases the expression in the malpighian tubule.

10. Discussion

10.1 Global male germline X inactivation

In summary, my results are consistent with global inactivation of the X chromosome in the male germline of *D. melanogaster*. The 112 independent X-chromosomal insertions (*ocn-lacZ* construct) cover the whole euchromatic X chromosome with an average spacing of 194 Kb. None of these insertions showed an expression level that is as high as the 22 independent autosomal insertions. The highest expression achieved by one of the X-chromosomal insertions showed only half of the reporter gene activity of the autosomal average expression. Consistent with this, the new three X-linked promoter reporter constructs (*CG10920-lacZ*-, *CG12681-lacZ*-, and the *CG1314-lacZ*-construct) showed similar expression patterns. The average difference in β-galactosidase enzymatic activity between autosomal and X-linked insertions was 2.8-fold, 3.9-fold, and 2.9-fold for the *CG10920*, *CG12681*, and *CG1314* reporter constructs, respectively. All differences in expression between X-linked and autosomal insertions are highly significant (MWW test, $P < 1*10^{-4}$). The results of the independent 112 *ocn-lacZ* insertions and of the three X-linked promoter reporter gene construct insertions suggest that the male germline X inactivation is a global mechanism affecting the whole X chromosome. My results demonstrate that the X chromosome is an unfavorable environment with respect to expression in male germline. However, many X-linked testis-specific genes are located on the X chromosome and the possibility of *cis*-regulatory sequences, which allow these genes to escape male germline X inactivation, remains.

To test if there is a difference between mRNA abundance and protein abundance, I did qRT-PCR for all four reporter gene constructs. In all four cases the expression measured at the protein-level correlated significantly with the mRNA-level (Spearmann's $R > 0.66$; $P < 0.0024$). This positive correlation indicates that the measurement of the protein-level (β-galactosidase enzymatic activity) reflects accurately the transcript abundance. Both, the

difference measured between autosomal and X-linked insertion at the protein-level and mRNA-level indicate that male germline X inactivation is affecting the whole euchromatic X chromosome.

The global effect of the male germline X inactivation suggests that some major changes in the chromatin structure are down-regulating the expression in the male germline. A similar effect is known for the dosage compensation in *Drosophila*. The dosage compensation complex (DCC) controls the H4 acetylation of the chromatin (Smith *et al.* 2001), which is associated with the up-regulation of male expression on the X chromosome. This acetylation is responsible for the higher expression in hemizygous males of *Drosophila* and this results in an equal expression to homozygous females of *Drosophila*. The DCC is regulating the expression in male flies for the entire X chromosome and a similar process could be responsible for the down-regulation of the X chromosome in the male germline.

My results are consistent with previous reports on the male germline X inactivation. Hense *et al.* (2007) used the same *ocn-lacZ* construct to address experimentally the question of the presence of the male germline X inactivation in *Drosophila*. These authors reported a downregulation of X-linked insertions in comparison to autosomal insertions, which is similar to my results. My work extended the work of (Hense *et al.* 2007), in that I used 107 additional independent insertions of the *ocn-lacZ* construct. Furthermore, I found that the expression downregulation is also present for X-linked promoters driving testis expression (*CG10920-lacZ*, *CG12681-lacZ*, *CG1314-lacZ* construct). The stage specific expression profiling of *Drosophila* spermatogenesis by Vibranovski *et al.* (2009a) reported an underrepresentation of testis-biased genes with higher expression in meiosis in comparison to mitosis on the X chromosome in comparison to the autosomes and an overrepresentation of genes with higher expression during mitosis in comparison to meiosis on the X chromosome in comparison to the autosomes. This stage specific preference for testis-biased genes expressed in mitosis for the X chromosome and the avoidance of testis-biased genes expressed in meiosis for the X chromosome is in accordance with the expectation of the abundance of testis-biased genes expressed late in spermatogenesis (meiosis) on the X chromosome affected by male germline X inactivation. Genes expressed late in spermatogenesis will be down-regulated in expression by the male germline X inactivation. My results agree well with these results. First I observed a down-regulation of the entire X chromosome, which can explain the underrepresentation of testis-biased genes expressed during spermatogenesis. Second, my results indicated that not

only autosomal-linked promoters driving testis expression are affected by male germline X inactivation, when transposed to the X chromosome, but also X-linked promoters driving testis expression are affected by male germline X inactivation.

Overall, my results can explain the chromosomal distribution of male-biased genes in the *Drosophila* genome. The majority of male-biased genes are expressed in reproductive tissues and these genes are significantly under-represented on the X chromosome (Parisi *et al.* 2003; Ranz *et al.* 2003). My results also support the X inactivation hypothesis, which has been proposed to explain the observed excess of X-to-autosome gene movement in *Drosophila* (Betran *et al.* 2002). The hypothesis is discussed in detail later.

10.2 The hotspot for new gene evolution at cytological band 19

It has been proposed that the region around cytological band 19 (19.8 Mb to 21.2 Mb) on the X chromosome is a hotspot for new gene evolution. This region contains and excess of testis-expressed genes (Boutanaev *et al.* 2002), including the newly evolved genes *Sdic*, *CG15323*, and *hydra* (Nurminsky *et al.* 1998; Levine *et al.* 2006; Chen *et al.* 2007). Furthermore, the orthologous region in *D. yakuba* also appears to be a hotspot for *de novo* gene evolution (Begun *et al.* 2007). One explanation for the clustering of testis-biased expressed genes in the cytological band 19 is that this region escapes the male germline X inactivation and allows genes to be expressed at a higher level in the male germline in contrast to the rest of the X chromosome. Four of my transgene insertions fall within this interval and one insertion (internal reference 104) is ~1 Kb away from the 3' end of the gene *Sdic1*. All four transgene insertions showed no higher expression than the average of all X-linked insertions. My results support that escape from X inactivation and increased expression due to this escape are not the reasons for the clustering of testis-biased genes in the cytological band 19. Additionally the global male germline X inactivation I report in this thesis supports these findings.

Further support for the rejection of the escape from X chromosome inactivation of the cytological band 19 came from targeted disruption of three well-defined male-specific gene expression neighbourhoods in the *Drosophila* genome (Meadows *et al.* 2010). One of the

generated inversions disrupts the domain at cytological band 19F (size 190 Kb). By measuring the gene expression between the non disrupted domain and the inverted domain using microarrays and qRT-PCR no significant difference in expression between the genes in the non inverted (wild-type) and the inverted domain were reported. This equal expression for genes in the two domains indicate that no local mechanism is up-regulating the gene expression in the non inverted (wild-typ) domain and no mechanism to escape male germline X inactivation is present for domain up-regulation.

Another possible explanation for the clustering is that some of the genes in this region are expressed in somatic cells of the testis and, thus, are not subject to male germline X inactivation. However, experimental studies of *Sdic* and *hydra* indicate that they are expressed in germline cells (Nurminsky *et al.* 1998; Chen *et al.* 2007). A final possible explanation for the clustering could be that the genes have *cis*-regulatory sequences that allow higher expression despite male germline X inactivation. I cannot reject this explanation with my results, but the insertion next to the 3' end of the *Sdic1* gene showed no higher expression than the average X-linked insertion expression and the three X-linked promoters driving testis expression have no *cis*-regulatory sequence in the amplified promoter region, which drive higher expression in the testis. Especially the *CG1314-lacZ* construct, whose promoter originally was located in the cytological band 19 showed no evidence for higher expression when transposed to other positions on the X chromosome. These findings indicate that local *cis*-regulatory sequences and the corresponding higher expression despite male germline X inactivation are not able to fully overcome the transcriptional down-regulation of the X chromosome in the male germline.

The genes *Sdic1-4* and the gene *hydra* show some uncommon patterns of exon shuffeling and gene duplication. This suggests that the region is maybe a hotspot for chromosomal rearrangements, which facilitates the birth of new genes by relocating and arranging transcriptional units in a new combination and this could be the reason why several newly testis-biased expressed genes are located in this region.

10.3 X-linked promoters driving testis expression

I chose three different X-linked promoters from different positions on the X chromosome. In total, I obtained independent 24 autosomal and 24 X-linked insertions. The distribution of landings sites I mapped showed some deviation from the expectation. In previous reports a preferential targeting of the 5'UTR for *P*-element transformation was reported (Spradling *et al.* 1995). For the three promoter constructs I observed a high number of insertions associated with coding-exonic and intergenic sequences. This deviation from the expectation is due to the relative small number of 24 insertions per targeted chromosome category, either autosome or X chromosome. This effect of preferentially targeting of exonic and intergenic sequences will disappear when the number of insertions is raised, as it is the case for the *ocn-lacZ* construct. The experiment using the *ocn-lacZ* construct showed that when the number of independent insertions is high (112 insertions) there was preferential 5'UTR targeting.

To ensure that the amplified promoter sequences used in my experiments drove testis expression, I performed β-galactosidase staining of entire testis and a measurement of enzymatic activity in dissected testis in comparison to gonadectomized flies. Both tests showed clearly that the amplified promoter sequences were driving exclusively testis expression and were adequate *cis*-regulatory sequences to study X-linked promoters, which drive testis expression to investigate male germline X inactivation. Further support came from different expression atlases, as FlyAtlas (Chintapalli *et al.* 2007) and SEBIDA (Gnad and Parsch 2006), where these genes showed highly male-biased and testis enriched expression.

The three promoter reporter gene constructs showed high expression for autosomal insertions and relatively low expression for X-linked insertions. The average difference in β-galactosidase enzymatic activity between autosomal and X-linked insertions were 2.8-fold, 3.9-fold, and 2.9-fold for the *CG10920*, *CG12681*, and *CG1314* reporter constructs, respectively. When I controlled for transcript abundance using qRT-PCR I obtained similar results. The average difference in *lacZ*-mRNA concentration between autosomal and X-linked insertions were 2.33-fold, 3.01-fold, and 3.32-fold for the *CG10920*, *CG12681*, and *CG1314* reporter construct, respectively. The differences in expression between autosomal and X-linked insertions were highly significant, either tested on the level of protein expression ($P < 3.11*10^{-4}$) or tested on the level of mRNA abundance ($P < 5.8*10^{-4}$). The discrepancy of mRNA abundance and enzymatic activity measurement of the *CG1314* construct showing

relatively low difference expression for the enzymatic test (2.9-fold) and the highest expression difference at the level of transcript abundance (3.32-fold) is likely due to the low absolute expression of this construct. This low absolute expression results in a high coefficient of variation of this construct (enzyme: 0.22, mRNA: 0.5) relative to the other constructs (*CG10920*: enzyme: 0.24, mRNA: 0.25; *CG1268*1: enzyme: 0.2, mRNA: 0.25) and a higher variation in expression, which is indicated by the discrepancy between the mRNA abundance and the protein abundance of the *CG1314* construct. However, I measured a good accordance between mRNA abundance and enzymatic activity, which indicates that the reduced expression for X-linked insertions in comparison to autosomal insertions is present at both the mRNA-level and the protein-level. All three X-linked promoter constructs showed a reduced expression for X-linked insertions. The results suggest that the reduced expression of X-linked insertions in comparison to autosomal insertions is due to male germline X inactivation, which reduce the expression only for X-linked insertions and not for autosomal insertions. To ensure that the observed expression pattern is not affected by gene dosage, I measured all insertions at a heterozygous (autosomal insertions) or hemizygous (X-linked insertions) stage, so that the higher activity of autosomal insertion is not due to the presence of two alleles, which will give higher expression in comparison to only one possible allele for X-linked insertion in male flies.

10.4 *Cis*-regulatory sequences driving testis expression of X-linked genes, despite male germline X inactivation

Despite male germline X inactivation, many genes showing male-biased expression and testis expression are located on the X chromosome. Mechanisms acting on chromatin structure to enable higher expression or enhancer elements causing higher expression to allow chromosomal regions to escape male germline X inactivation were not supported by my results. The results in this thesis showed that the whole X chromosome is affected by male germline X inactivation (Fine scale mapping of male germline X inactivation), and no region could escape X inactivation. Instead, individual genes appear to achieve testis expression through their own *cis*-regulatory sequences. Consistent with this, all three promoter sequences used in my experiments, which were comprised of less than 1 Kb of sequence directly

upstream of the *CG10920*, *CG12681*, and *CG1314* genes, were able to drive levels of testis-specific expression similar to those observed for the native genes (Gnad and Parsch 2006; Chintapalli *et al.* 2007). Since the native *CG1314* gene is located in region 19E, my results provide further evidence that this "gene neighborhood" is not required for proper expression in testis (Meadows *et al.* 2010). For all promoters, reporter gene expression was much higher when inserted on the autosomes than when inserted on the X chromosome, indicating that local *cis*-regulatory elements are not able to achieve higher X-linked expression in comparison to autosomal insertions. The three X-linked promoters used in this study did not share sequence homology with each other or with other known testis-specific regulatory elements, which suggests that they do not have a simple, shared regulatory mechanism. The *CG12681* promoter contains a 20-bp sequence found upstream of the male- and testis-biased gene *CG5732* on chromosome arm 3R (Gnad and Parsch 2006; Chintapalli *et al.* 2007). This region is predicted to contain binding sites for the Even-skipped and Zerknuellt transcription factors (Messeguer *et al.* 2002). However, both of these transcription factors are known to function during early embryogenesis and have no known function in spermatogenesis, nor do they show enriched expression in males and testis (Gnad and Parsch 2006; Chintapalli *et al.* 2007).

Still the question remains, why many male-biased testis expressed genes are located on the X chromosome despite male germline X inactivation. One explanation for this phenomenon could be that these genes are expressed in stages of spermatogenesis that are not affected by male germline X inactivation or that the relatively low expression achieved by the inactivated X-linked genes is sufficient to maintain functionality.

10.5 Stage specific expression profiling of male germline X inactivation

Male germline X inactivation was first discovered in mammals (Lifschytz and Lindsley 1972). In this process, the X chromosome in males is heterochromatized during the first meiotic prophase and the X chromosome is transcriptionally inactivated. In *Drosophila*, male germline X inactivation is also present (Hense *et al.* 2007; Vibranovski *et al.* 2009a). Because mammals and insects diverged hundreds of millions of years ago, it is not known if the male

germline X inactivation is a pleisiomorphic trait or convergent evolution between mammals and insects. To address this question, if male germline X inactivation occurs in *Drosophila*, a microarray analysis of gene expression during different stages of spermatogenesis indicated that there is a significant excess of X-linked genes that are down-regulated during the transition from mitosis to meiosis (Vibranovski *et al.* 2009a). This is consistent with the MSCI present in mammals, however, the average decline in expression between the two stages (~10%) is too small to detect by conventional gene by gene statistical analysis or to account for the observed differences between X-linked and autosomal transgene expression (Meikeljohn unpublished). Furthermore, for the three genes whose promoters were used in the current study (*CG10920*, *CG12681*, and *CG1314*), the stage-specific microarray data indicate that their expression increases during the mitosis-meiosis transition (Vibranovski *et al.* 2009a). In my study I found that X-linked insertions of all three promoter constructs showed significantly less expression than autosomal insertions during both mitotic and meiotic stages of spermatogenesis, with only the *CG1314* construct much stronger down-regulation of X-linked expression during meiosis. For these reasons, MSCI appears to be insufficient explain our results. My data suggest that X-chromosomal gene expression is suppressed in all cells of the *Drosophila* male germline through a mechanism that is independent from the MSCI known to occur in mammals. Meikeljohn (unpublished) found similar results by screening the stage specific expression of the *ocn-lacZ* reporter gene construct. This phenomenon has been termed male germline suppression of the X chromosome (MGSX) and is compatible with our observations, as well as with previous results from experiments using autosomal promoter to drive testis-specific expression of X-linked and autosomal transgenes (Hense *et al.* 2007). Finally these results suggest that the suppression of X-linked expression during spermatogenesis is a case of convergent evolution that occurred in mammals and *Drosophila* independently.

10.6 The excess of X chromosome to autosome gene movement

The distribution of male-biased genes is not random in the *Drosophila* genome. In *Drosophila* an underrepresentation of male-biased genes on the X chromosome has been reported (Parisi *et al.* 2003; Ranz *et al.* 2003). Along with this under-representation of male-biased genes, an excess of X chromosome to autosome movement in comparison to autosome to autosome, and autosome to X chromosome movement was discovered (Vibranovski *et al.* 2009b). One explanation for this phenomenon is male germline X inactivation, which will transcriptionally silence the X chromosome during spermatogenesis. Especially genes expressed during spermatogenesis will be affected and the result of the X inactivation is that male-biased testis specific X-linked genes are not expressed or are expressed at a low level. To avoid this reduction of expression for testis specific genes, these genes escape the X chromosome and move to the autosomes. The new environment of the autosomes, with no expression inactivation, allows the re-located copies to be expressed at a higher level in the male germline.

My results demonstrate that the X chromosome is an unfavorable environment with respect to expression in the male germline. This is in accordance with previous observations that male-biased genes, the majority of which are expressed in reproductive tissues, are significantly under-represented on the X chromosome (Parisi *et al.* 2003; Ranz *et al.* 2003). My results also lend support to the X inactivation hypothesis, which has been proposed to explain the observed excess of X to autosome gene movement in *Drosophila* (Betran *et al.* 2002). This hypothesis posits that genes escaping the X chromosome receive a selective advantage in the form of increased expression in the male germline. Here I show that this is the case for gene expression driven by sequences from three X-linked, testis-expressed genes. In all cases, relocation from the X chromosome to an autosome resulted in an expression increase of ~3-fold in the testis. Although it is difficult to experimentally determine a direct link between an increase in a gene's expression in the testis and in increase in male reproductive fitness, previous findings that testis-expressed genes show exceptionally high rates of adaptive evolution at the protein level (Proschel *et al.* 2006; Baines *et al.* 2008) suggest that positive selection plays an important role in the evolution of genes expressed in the male germline. Similarly, positive selection has been shown to act on testis-expressed retrogenes that have relocated from the X chromosome to an autosome (Betran and Long 2003; Quezada-Diaz *et al.* 2010; Tracy *et al.* 2010). However, not all genes that show male- and testis-expression

escape the X chromosome. These genes could be expressed at low level and not affected by MSCI, because low expression is possible or expressed in different stages of spermatogenesis or in somatic tissues that are not affected by MSCI.

My results support a selective mechanism for the evolutionary redistribution of genes across the genome and provide experimental evidence to explain patterns of inter-chromosomal gene movement observed in *Drosophila* (Vibranovski *et al.* 2009b) and other taxa with herterogametic (XY) males (Emerson *et al.* 2004).

10.7 The *cis*-regulatory sequence of the gene *CG9509* was positively selected in the European population of *D. melanogaster*

The gene *CG9509* showed a significant difference in expression between African and European populations of *D. melanogaster* (Meiklejohn *et al.* 2003; Hutter *et al.* 2008; Muller *et al.* 2011). By sequencing the upstream region of the gene *CG9509* (~1.2 Kb) and analyzing the pattern of polymorphism in and between these populations, it was found that this region showed reduced polymorphism in the European population. Furthermore, two statistical tests applied to the *CG9509* upstream region (CLR test, Kim and Stephan 2002); Sweepfinder, Nielsen *et al.* 2005) showed evidence for positive selection (compared to a standard neutral model) of this region in the European population, also known as a selective sweep. To test the functional basis of the selective sweep in the European population, which may have altered the expression level of *CG9509* the European population, I did an experimental verification of the expression difference by comparing the upstream region of the African population to the upstream region of the European population. The amplified and tested upstream region in both populations consist of 1.2 Kb, which was located between the 3'end of the gene *CG14406* and the 5'end of the *CG9509* gene. By using the entire intergenic region between the two genes, I ruled out that any possible *cis*-regulatory sequence, which controls the expression of the *CG9509* gene is not considered in my approach. I tested the difference in expression with the aid of the ΦC31 transformation system (Bischof *et al.* 2007). This system used pre-defined landings sites and this enables the possibility to compare both promoters at the same genomic location to exclude any influence on expression of different genomic

region by inserting randomly the promoters at different positions in the *Drosophila* genome. The landing sites ZH-68E and ZH-86Fb were used in this approach. The inserted constructs contain the promoter of interest, either European or African, the reporter gene *lacZ* from *E. coli* and a selectable marker the *mini-white* gene (eye color). The *lacZ* gene is a standard reporter gene, which was already used in many studies to investigate promoter dependent expression in *Drosophila* (Hense *et al.* 2007; Kemkemer *et al.* 2011). The differences in expression observed by the enzymatic assay were ~3-fold higher expression for the reporter gene driven by the European promoter in comparison to the African promoter for both landing sites and in both males and females. Similar results were obtained by using qRT-PCR and measuring the mRNA abundance, where the difference in expression was ~3-fold higher expression for the European promoter construct in comparison to the African promoter. Both techniques, either protein abundance (enzymatic assay) or mRNA abundance (qRT-PCR), showed significantly higher expression for the European promoter ($P < 0.02$). Both methods correlate very well by measuring the reporter gene expression (Spearmann's $R > 0.6$; $P < 0.10$). This suggests that the expression differences I observed between the European and the African promoter were due to the different nucleotide sequences of the two population specific promoters. The differences in expression measured with promoter reporter gene constructs reproduce the expression differences measured with microarrays (Hutter *et al.* 2008; Müller *et al.* 2011) or qRT-PCR (Saminadin-Peter 2008; Müller unpublished). In particular, the differences measured with microarrays were 2.31-fold higher expression for the European population in comparison to the African population, the differences measured by qRT-PCR were 2.02-fold for males and 1.68-fold for females and the differences measured with Promoter reporter gene construct were ~3-fold higher expression in European populations. The Promoter reporter gene constructs reproduce very well the differences in expression measured in the natural population, which indicates that the used promoters are able to drive natural expression. These results showed that changes in the promoter region of the European population are responsible for the higher expression of the *CG9509* gene in the European population.

From expression atlases it is known that this gene is highly expressed in the malpighian tubule (Chintapalli *et al.* 2007), showing 10-fold higher expression in the tubule than any other tissue in adult *Drosophila*. To verify that the expression differences I observed in whole flies were due to expression differences in the malpighian tubule, I dissected the malpighian tubule from male flies and performed an enzymatic assay. The expression differences

between the European promoter reporter gene construct and the African promoter reporter gene construct were 2.25-fold for the ZH-68E landing site and 3.23-fold for the ZH-86Fb higher expression in the European population. These results show that the expression differences measured in whole flies are actually caused by expression differences in the malpighian tubule, because the differences in expression measured in the malpighian tubule reproduce the differences measured in whole flies.

The role of *CG9509* in adaptation of the European population is unknown. From expression analysis and comparative computational approaches it is known, that the gene *CG9509* is involved in mesoderm development (Furlong *et al.* 2001), possesses choline dehydrogenase activity, a FAD or FAD2 binding domain and is involved in alcohol metabolic process (Flybase, Tweedie *et al.* 2009). It is possible that the *CG9509* gene is involved in the process of alcohol degradation, which is consistent with its expression in the malpighian tubule, which is a tissue in insects responsible to segregate metabolic endproducts, and necessary for metabolize alcoholic diet, which came along by the colonization of Europe and the increased diet of rotten fruits in Europe compared to Africa. From protein interaction analysis (Biogrid; (Stark *et al.* 2011) it is known that the *CG9509* gene interacts (two hybrid experiments) with two proteins, *CG14216* and *CG4060*. The gene *CG14216* is involved in mRNA processing, possesses a phosphoprotein phosphatase activity and is localizes to the nucleus. The gene *CG4060* has no reported annotation. The interaction of *CG9509* and *CG14216* may be due to the expression of *CG9509* during mesoderm development and the mRNA processing ability of *CG14216.* This could give evidence to the interaction of both proteins involved in mesoderm development and resulting into the development of the malpighian tubule, which is developed from the mesoderm.

The exact cause of the higher expression in the European population has not been identified. With my approach, I showed that variation within the 1.2 Kb upstream regulatory sequence of *CG9509* must be responsible for the expression difference between the populations. Further studies are necessary to identify the specific cause of the expression difference. For example, site-directed mutagenesis could be used to identify the SNP or indel that is responsible for the expression difference.

11. Reference list

Amrein H, Axel R. 1997. Genes expressed in neurons of adult male *Drosophila*. Cell **88**:459–469.

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.

Bauer VL, Aquadro CF. 1997. Rates of DNA sequence evolution are not sex-biased in *Drosophila melanogaster* and *D. simulans*. Mol Biol Evol **14**:1252–1257.

Begun DJ, Lindfors HA, Kern AD, Jones CD. 2007. Evidence for de novo evolution of testis-expressed genes in the *Drosophila yakuba*/*Drosophila erecta* clade. Genetics **176**:1131–1137.

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

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

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

Bischof J, Maeda RK, Hediger M, Karch F, Basler K. 2007. An optimized transgenesis system for *Drosophila* using germ-line-specific *phiC31* integrases. Proc Natl Acad Sci U S A **104**:3312–3317.

Boutanaev AM, Kalmykova AI, Shevelyov YY, Nurminsky DI. 2002. Large clusters of co-expressed genes in the *Drosophila* genome. Nature **420**:666–669.

Bridges CB. 1925. Haploidy in *Drosophila Melanogaster*. Proc Natl Acad Sci U S A **11**:706–710.

Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. 1991. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. Nature **349**:38–44.

Castrillon DH, Gonczy P, Alexander S, Rawson R, Eberhart CG, Viswanathan S, DiNardo S, Wasserman SA. 1993. Toward a molecular genetic analysis of spermatogenesis in *Drosophila melanogaster*: characterization of male-sterile mutants generated by single *P* element mutagenesis. Genetics **135**:489–505.

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

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. Heredity **95**:118–128.

Chen ST, Cheng HC, Barbash DA, Yang HP. 2007. Evolution of *hydra*, a recently evolved testis-expressed gene with nine alternative first exons in *Drosophila melanogaster*. PLoS Genet **3**:e107.

Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat Genet **39**:715–720.

Clark AG, Eisen MB, Smith DR *et al.* 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature **450**:203–218.

Coyne JA. 1992. Genetics and speciation. Nature **355**:511–515.

Emerson JJ, Kaessmann H, Betran E, Long M. 2004. Extensive gene traffic on the mammalian X chromosome. Science **303**:537–540.

Fong Y, Bender L, Wang W, Strome S. 2002. Regulation of the different chromatin states of autosomes and X chromosomes in the germ line of *C. elegans*. Science **296**:2235–2238.

Furlong EE, Andersen EC, Null B, White KP, Scott MP. 2001. Patterns of gene expression during *Drosophila* mesoderm development. Science **293**:1629–1633.

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

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

Haldane JB. 1947. The mutation rate of the gene for *haemophilia*, and its segregation ratios in males and females. Ann Eugen **13**:262–271.

Haldane JBS. 1922. Sex ratio and unisexual sterility in hybrid animals. J. Genet. **12**:101–109.

Handel MA, Park C, Kot M. 1994. Genetic control of sex-chromosome inactivation during male meiosis. Cytogenet Cell Genet **66**:83–88.

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

Hurst LD, Ellegren H. 1998. Sex biases in the mutation rate. Trends Genet **14**:446–452.

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.

Itoh Y, Melamed E, Yang X *et al.* 2007. Dosage compensation is less effective in birds than in mammals. J Biol **6**:2.

Itoh Y, Replogle K, Kim YH, Wade J, Clayton DF, Arnold AP. 2007. Sex bias and dosage compensation in the zebra finch versus chicken genomes: general and specialized patterns among birds. Genome Res **20**:512–518.

Kaiser VB, Ellegren H. 2006. Nonrandom distribution of genes with sex-biased expression in the chicken genome. Evolution **60**:1945–1951.

Kelly WG, Schaner CE, Dernburg AF, Lee MH, Kim SK, Villeneuve AM, Reinke V. 2002. X-chromosome silencing in the germline of *C. elegans*. Development **129**:479–492.

Kemkemer C, Hense W, Parsch J. 2011. Fine-scale analysis of X chromosome inactivation in the male germline of *Drosophila melanogaster*. Mol Biol Evol. (in press; doi: 10.1093/molbev/msq355).

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Paabo S. 2005. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science **309**:1850–1854.

Khil PP, Smirnova NA, Romanienko PJ, Camerini-Otero RD. 2004. The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. Nat Genet **36**:642–646.

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

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. Science **188**:107–116.

Kreitman M, Hudson RR. 1991. Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127**:565–582.

Kuroda MI, Kernan MJ, Kreber R, Ganetzky B, Baker BS. 1991. The maleless protein associates with the X chromosome to regulate dosage compensation in *Drosophila*. Cell **66**:935–947.

Lachaise D, Cariou ML, David JR, Lemeunier F, Tsacas L, Ashburner M. 1988. Historical Biogeography of the *Drosophila-Melanogaster* Species Subgroup. Evolutionary Biology **22**:159–225.

Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ. 2006. Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. Proc Natl Acad Sci U S A **103**:9935–9939.

Li WH, Yi S, Makova K. 2002. Male-driven evolution. Curr Opin Genet Dev **12**:650–656.

Lifschytz E, Lindsley DL. 1972. The role of X-chromosome inactivation during spermatogenesis (*Drosophila*-allocycly-chromosome evolution-male sterility-dosage compensation). Proc Natl Acad Sci U S A **69**:182–186.

Lyon MF. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus L.*). Nature **190**:372–373.

Masly JP, Presgraves DC. 2007. High-resolution genome-wide dissection of the two rules of speciation in *Drosophila*. PLoS Biol **5**:e243.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351**:652–654.

Meadows LA, Chan YS, Roote J, Russell S. Neighbourhood continuity is not required for correct testis gene expression in *Drosophila*. PLoS Biol **8**:e1000552.

Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. Proc Natl Acad Sci U S A **100**:9894–9899.

Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL. 1997. *roX1* RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. Cell **88**:445–457.

Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM. 2002. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics **18**:333–334.

Michiels F, Gasch A, Kaltschmidt B, Renkawitz-Pohl R. 1989. A 14 bp promoter element directs the testis specificity of the *Drosophila beta 2 tubulin* gene. EMBO J **8**:1559–1565.

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

Nei M. 1969. Linkage modifications and sex difference in recombination. Genetics **63**:681–699.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. 2005. Genomic scans for selective sweeps using SNP data. Genome Res **15**:1566–1575.

Nurminsky DI, Nurminskaya MV, De Aguiar D, Hartl DL. 1998. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. Nature **396**:572–575.

Orr HA, Betancourt AJ. 2001. Haldane's sieve and adaptation from the standing genetic variation. Genetics **157**:875–884.

Palmer MJ, Mergner VA, Richman R, Manning JE, Kuroda MI, Lucchesi JC. 1993. The *male-specific lethal-one* (*msl-1*) gene of *Drosophila melanogaster* encodes a novel protein that associates with the X chromosome in males. Genetics **134**:545–557.

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

Presgraves DC. 2002. Patterns of postzygotic isolation in Lepidoptera. Evolution **56**:1168–1183.

Price TD, Bouvier MM. 2002. The evolution of F1 postzygotic incompatibilities in birds. Evolution **56**:2083–2089.

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

Quezada-Diaz JE, Muliyil T, Rio J, Betran E. 2010. *Drcd-1* related: a positively selected spermatogenesis retrogene in *Drosophila*. Genetica **138**:925–937.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science **300**:1742–1745.

Reinke V, Gil IS, Ward S, Kazmer K. 2004. Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*. Development **131**:311–323.

Rice WR. 1984. Sex Chromosome and the Evolution of Sexual Dimorphism. Evolution **38**:735–742.

Richler C, Soreq H, Wahrman J. 1992. X inactivation in mammalian testis is correlated with inactive X-specific transcription. Nat Genet **2**:192–195.

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

Rubin GM, Spradling AC. 1982. Genetic transformation of *Drosophila* with transposable element vectors. Science **218**:348–353.

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

Sandler E. NE. 1957. Meiotic drive as an evolutionary force. American Naturalist **91**:105–110.

Shiu PK, Raju NB, Zickler D, Metzenberg RL. 2001. Meiotic silencing by unpaired DNA. Cell **107**:905–916.

Siegal ML, Hartl DL. 1996. Transgene Coplacement and high efficiency site-specific recombination with the *Cre/loxP* system in *Drosophila*. Genetics **144**:715–726.

Smith ER, Allis CD, Lucchesi JC. 2001. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. J Biol Chem **276**:31483–31486.

Smith NG, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. Nature **415**:1022–1024.

Spradling AC, Rubin GM. 1982. Transposition of cloned *P* elements into *Drosophila* germ line chromosomes. Science **218**:341–347.

Spradling AC, Stern DM, Kiss I, Roote J, Laverty T, Rubin GM. 1995. Gene disruptions using *P* transposable elements: an integral component of the *Drosophila* genome project. Proc Natl Acad Sci U S A **92**:10824–10830.

Stark C, Breitkreutz BJ, Chatr-Aryamontri A *et al.* 2011. The BioGRID Interaction Database: 2011 update. Nucleic Acids Res **39**:D698–704.

Steinemann M, Steinemann S. 2000. Common mechanisms of Y chromosome evolution. Genetica **109**:105–111.

Steinemann S, Steinemann M. 2001. Biased distribution of repetitive elements: a landmark for neo-Y chromosome evolution in *Drosophila miranda*. Cytogenet Cell Genet **93**:228–233.

Su AI, Wiltshire T, Batalov S *et al.* 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A **101**:6062–6067.

Torgerson DG, Singh RS. 2003. Sex-linked mammalian sperm proteins evolve faster than autosomal ones. Mol Biol Evol **20**:1705–1709.

Tracy C, Rio J, Motiwale M, Christensen SM, Betran E. 2010. Convergently recruited nuclear transport retrogenes are male biased in expression and evolving under positive selection in *Drosophila*. Genetics **184**:1067–1076.

Turner JM. 2007. Meiotic sex chromosome inactivation. Development **134**:1823–1831.

Tweedie S, Ashburner M, Falls K *et al.* 2009. FlyBase: enhancing *Drosophila* Gene Ontology annotations. Nucleic Acids Res **37**:D555–559.

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

Vibranovski MD, Zhang Y, Long M. 2009b. General gene movement off the X chromosome in the *Drosophila* genus. Genome Res **19**:897–903.

Vicoso B, Charlesworth B. 2009. Effective population size and the faster-X effect: an extended model. Evolution **63**:2413–2426.

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

Wang X, Zhang J. 2004. Rapid evolution of mammalian X-linked testis-expressed homeobox genes. Genetics **167**:879-888.

Wittkopp PJ, Haerum BK, Clark AG. 2008. Regulatory changes underlying expression differences within and between *Drosophila* species. Nat Genet **40**:346–350.

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

Yanicostas C, Lepesant JA. 1990. Transcriptional and translational *cis*-regulatory sequences of the spermatocyte-specific *Drosophila janusB* gene are located in the 3' exonic region of the overlapping *janusA* gene. Mol Gen Genet **224**:450–458.

12. Appendix

Appendix A: Chromosomal locations of X-linked transgene insertions.

Internal reference	Chromosome	Cytological band	Mapped position (v5.30)	Landing site class	Affected gene	Proximal gene within 10Kb	Distal gene within 10Kb
60	X	1B4	371549	Exon	CG13373		
58	X	1B5	391321	Exon	CG4122		
7	X	1B6	392782	Intron	CG4122		
82	X	1C2	580780	Intergenic			CG5227
59	X	1E3	1028402	5'UTR	CG3655		
129	X	1E4	1103391	Intergenic		CG14624	CG11382
77	X	1E4	1103702	Intergenic		CG14624	CG11382
23	X	1E5	1129003	5'UTR	CG3638		
120	X	1F1	1170568	Intergenic		CG11405	
22	X	2A1	1275081	5'UTR	CG32813		
78	X	2B4	1513944	5'UTR	CG11491		
29	X	2B13	1767523	5'UTR	Pgam5		
127	X	2C10	1967570	5'UTR	CG4061		
25	X	2F5	2187197	Intergenic		CG2865	
51	X	2F5	2187547	Intergenic		CG2865	
72	X	2F5	2211614	Intergenic			CG34052
73	X	3A4	2439975	Intergenic		CG33950	
128	X	3B1	2579132	Intergenic			CG2647
48	X	3D2	3266905	Intergenic		CG10798	
97	X	4B1	4025208	Intergenic			CG32775
42	X	4C3	4322739	5'UTR	CG3578		
8	X	4C13	4579832	5'UTR	CG2984		
35	X	4C13	4582188	5'UTR	CG6998		
81	X	4D5	4803582	5'UTR	CG32772		
87	X	5A9	5529251	Intron	CG42492		
94	X	5A12	5573943	5'UTR	CG3171		
85	X	5A13	5584547	5'UTR	CG12410		
9	X	5B8	5650466	Intron	CG15771		
28	X	5C7	5795683	5'UTR	CG4027		
24	X	5C7	5796196	5'UTR	CG4027		
1	X	5E7	6197970	5'UTR	CG3823		

39	X	6C4	6556306	5'UTR	CG3977		
26	X	6E1	6760736	Intergenic			CG33692
76	X	6E4	6892543	Intron	CG2977		
18	X	7A3	7089164	5'UTR	CG9650		
53	X	7B1	7185793	Intergenic		CG1659	
wol12X	X	7B1	7231447	5'UTR	CG1435		
115	X	7B6	7574392	Intergenic		CG11387	CHES 1
36	X	7D1	7863357	5'UTR	CG32858		
90	X	7D5	7950815	5'UTR	CG2252		
13	X	7E5	8280388	5'UTR	CG1387		
12	X	7E7	8305832	5'UTR	CG18009		
43	X	8B6	8787956	5'UTR	CG10701		
96	X	8B6	8788272	5'UTR	CG10701		
5	X	8B6	8788760	5'UTR	CG10701		
6	X	8C4	8936482	Intron	CG42388		
88	X	8C14	9050450	5'UTR	CG8989		
61	X	8D6	9200323	Intergenic		CG1689	
50	X	8F9	9580425	5'UTR	CG15319		
62	X	8F9	9580484	5'UTR	CG15319		
41	X	9B1	9966318	Exon	CG32685		
114	X	9B11	10259107	Intron	CG2221		
122	X	9D3	10440811	5'UTR	CG34414		
116	X	9D3	10441711	5'UTR	CG34414		
20	X	9E2	10638737	5'UTR	CG32676		
52	X	9E7	10662785	5'UTR	CG1826		
111	X	9F1	10677382	5'UTR	CG1683		
40	X	9F12	10823647	Exon	CG2145		
109	X	10C5	11454011	Intergenic		CG1572	CG11709
84	X	10D8	11622650	5'UTR	inaF-D		
68	X	10E3	11687281	5'UTR	CG15224		
66	X	10E3	11687683	5'UTR	CG15224		
80	X	10E3	11687934	5'UTR	CG15224		
108	X	10E3	11699346	5'UTR	CG4147		
46	X	10E3	11699401	5'UTR	CG4147		
63	X	11A1	11901124	5'UTR	CG1806		
11	X	11A6	12097826	Intron	CG42338		
125	X	11D1	12796913	5'UTR	CG4407		
65	X	11D10	12985294	5'UTR	CG12244		
wol13X	X	11E3	13101216	5'UTR	CG1903		
54	X	11E9	13195229	Intergenic		CG1622	
110	X	11F1	13291688	Intron	CG1673		

16	X	12A9	13534378	5'UTR	CG11172		
101	X	12A9	13534847	5'UTR	CG11172		
93	X	12A9	13534954	5'UTR	CG11172		
10	X	12A9	13534954	5'UTR	CG11172		
126	X	12A9	13535895	5'UTR	CG11172		
79	X	12C1	13656667	Intergenic		CG11129	CG11111
103	X	12C1	13656798	5'UTR	CG11111		
32	X	12C6	13716345	5'UTR	CG10997		
27	X	12F4	14717999	5'UTR	CG9533		
34	X	12F4	14719839	5'UTR	CG9533		
37	X	12F5	14720092	5'UTR	CG9533		
95	X	12F5	14726724	5'UTR	CG14411		
21	X	13A1	14817805	Intron	CG32593		
71	X	13A5	14917818	Intron	CG32592		
118	X	13E18	15679019	Exon	PafAhα		
19	X	13E18	15682937	5'UTR	CG8497		
112	X	13F1	15705777	5'UTR	CG8544		
130	X	14A8	15980131	5'UTR	CG9214		
30	X	14A8	15985161	5'UTR	CG9214		
86	X	14C2	16279793	5'UTR	CG4239		
wol20X	X	15A7	16677891	Intergenic	CG9623		CG12220
3	X	15A7	16677901	5'UTR	CG4742		
45	X	15A11	16730463	Intergenic		CG13004	
wol19X	X	16A1	17197389	Intron	CG5445		
55	X	16C1	17592835	Intergenic		CG32556	CG8188
31	X	17D1	18559749	Intergenic		CG6696	
15	X	18C3	19247730	5'UTR	CG12199		
17	X	18C8	19392349	5'UTR	CG3400		
4	X	18C8	19399583	5'UTR	CG3400		
70	X	18D3	19498575	Intergenic		CG14220	
89	X	18D13	19561872	Intron	CG12529		
57	X	18E3	19607504	Intergenic	CG14233		
83	X	18F2	19677223	Intergenic	CG12701		
33	X	18F2	19717282	Intergenic			CG11942
106	X	18F4	19780935	Exon	CG11937		
104	X	19C1	20067935	Intergenic		CG9579	CG9580
100	X	19E7	20915774	Intergenic		Mgst1	CG1753
49	X	19E7	20925189	5'UTR	CG32513		
wol23X	X	19F1	20994197	Intergenic		CG15445	CG34120
64	X	20C1	21917264	5'UTR	CG17600		

Appendix B: Chromosomal locations of autosomal transgene insertions.

Internal reference	Chromosome	Cytological band	Mapped position (v5.30)	Landing site class	Affected gene	Proximal gene within 10Kb	Distal gene within 10Kb
control 3	2L	25C6	5108428	Intergenic			
control 2	2L	26D9	6498770	Econ	CG9550		
wol4	2L	27F4	7423613	Intergenic		CG5229	CG5261
control 4	2L	28D3	7984133	5'UTR	CG7231		
wol7	2R	42C6	2603250	5'UTR	CG3409		
control 8	2R	50B3	9465619	Intergenic		CG13335	CG6191
wol9	2R	56E1	15518667	5'UTR	CG9218		
control 11	3L	61C9	746383	Intergenic			CG1007
wol11	3L	61C9	749342	Intergenic			CG1007
wol6	3L	66C12	8414592	Intergenic		CG32354	
wol18	3L	70F4	14751002	5'UTR	CG42507		
control 9	3L	75E2	18839391	5'UTR	CG3979		
wol16	3L	79A2	21872686	Intergenic		CG7437	
wol2	3R	82E4	790870	Intergenic			
wol1	3R	84B1	279214	5'UTR	CG31522		
wol14	3R	85F10	5920571	Intergenic			
control 6	3R	86E18	7589977	5'UTR	CG17342		
wol3	3R	89E11	12881438	5'UTR	CG5201		
wol15	3R	91D4	14743978	5'UTR	Xrp1		
wol17	3R	91F4	14983880	Intron	CG6713 & CG11779		
wol10	Autosome						
wol8	Autosome						

Appendix C: Expression (mean units of β-galactosidase enzymatic activity) of X-linked insertions. Every insertion was measured with three biological replicates and two technical replicates.

Internal refereence	Chromosome	Mapped position (v5.30)	Average male expression	Standard deviation of male expres.	Average female expression	Standard deviation of female expres.
60	X	371549	2.151	0.322	0.264	0.190
58	X	391321	2.691	0.129	0.204	0.094
7	X	392782	2.512	0.157	0.284	0.084
82	X	580780	2.243	0.419	0.253	0.090
59	X	1028402	2.179	0.134	0.124	0.044
129	X	1103391	2.512	0.151	0.133	0.075
77	X	1103702	2.741	0.542	0.279	0.121
23	X	1129003	2.223	0.085	0.257	0.116
120	X	1170568	2.492	0.306	0.218	0.057
22	X	1275081	2.192	0.218	0.173	0.107
78	X	1513944	2.275	0.077	0.247	0.137
29	X	1767523	2.551	0.357	0.168	0.108
127	X	1967570	2.572	0.260	0.108	0.072
25	X	2187197	2.200	0.187	0.088	0.059
51	X	2187547	2.328	0.128	0.086	0.105
72	X	2211614	2.351	0.069	0.354	0.204
73	X	2439975	2.064	0.134	0.315	0.053
128	X	2579132	2.522	0.106	0.324	0.112
48	X	3266905	2.220	0.235	0.427	0.149
97	X	4025208	2.062	0.221	0.098	0.085
42	X	4322739	2.356	0.380	0.245	0.098
8	X	4579832	2.063	0.148	0.146	0.088
35	X	4582188	1.357	0.118	0.173	0.086
81	X	4803582	2.372	0.147	0.114	0.120
87	X	5529251	2.525	0.227	0.219	0.172
94	X	5573943	2.698	0.111	0.197	0.064
85	X	5584547	2.329	0.138	0.142	0.039
9	X	5650466	2.093	0.332	0.091	0.086
28	X	5795683	2.380	0.249	0.252	0.067
24	X	5796196	2.722	0.059	0.132	0.126
1	X	6197970	2.163	0.292	0.183	0.078
39	X	6556306	2.484	0.091	0.225	0.077
26	X	6760736	4.569	0.655	0.239	0.133
76	X	6892543	2.352	0.330	0.120	0.100
18	X	7089164	2.358	0.319	0.195	0.144

53	X	7185793	2.733	0.280	0.156	0.123
wol12X	X	7231447	1.228	0.101	0.070	0.037
115	X	7574392	2.218	0.298	0.219	0.054
36	X	7863357	2.757	0.359	0.183	0.114
90	X	7950815	2.215	0.090	0.165	0.061
13	X	8280388	4.397	0.368	0.239	0.069
12	X	8305832	2.018	0.226	0.113	0.068
43	X	8787956	1.974	0.502	0.164	0.049
96	X	8788272	2.487	0.141	0.174	0.104
5	X	8788760	2.828	0.118	0.410	0.076
6	X	8936482	2.278	0.171	0.096	0.096
88	X	9050450	2.060	0.277	0.220	0.062
61	X	9200323	2.479	0.345	0.318	0.030
50	X	9580425	2.589	0.211	0.120	0.072
62	X	9580484	2.230	0.049	0.254	0.086
41	X	9966318	2.201	0.058	0.282	0.109
114	X	10259107	2.524	0.124	0.168	0.054
122	X	10440811	2.239	0.182	0.148	0.040
116	X	10441711	2.298	0.413	0.226	0.125
20	X	10638737	3.041	0.152	0.443	0.272
52	X	10662785	2.219	0.263	0.145	0.144
111	X	10677382	2.357	0.272	0.171	0.095
40	X	10823647	2.432	0.058	0.395	0.031
109	X	11454011	2.538	0.134	0.346	0.049
84	X	11622650	2.417	0.191	0.096	0.081
68	X	11687281	1.673	0.260	0.098	0.080
66	X	11687683	2.205	0.191	0.185	0.075
80	X	11687934	1.912	0.207	0.157	0.102
108	X	11699346	1.882	0.087	0.265	0.053
46	X	11699401	2.228	0.273	0.169	0.020
63	X	11901124	2.262	0.212	0.134	0.086
11	X	12097826	2.732	0.128	0.075	0.068
125	X	12796913	2.301	0.166	0.216	0.081
65	X	12985294	2.055	0.296	0.063	0.074
wol13X	X	13101216	0.768	0.079	0.247	0.037
54	X	13195229	2.712	0.326	0.167	0.056
110	X	13291688	2.863	0.311	0.324	0.237
16	X	13534378	2.743	0.227	0.366	0.178
101	X	13534847	2.440	0.363	0.232	0.076
93	X	13534954	2.506	0.137	0.118	0.066
10	X	13534954	2.117	0.097	0.198	0.247

126	X	13535895	2.593	0.210	0.216	0.064
79	X	13656667	2.230	0.161	0.357	0.125
103	X	13656798	2.641	0.218	0.288	0.145
32	X	13716345	2.287	0.162	0.084	0.070
27	X	14717999	2.276	0.103	0.118	0.081
34	X	14719839	2.035	0.272	0.171	0.132
37	X	14720092	2.638	0.206	0.344	0.047
95	X	14726724	2.363	0.405	0.267	0.095
21	X	14817805	1.861	0.213	0.045	0.051
71	X	14917818	1.230	0.061	0.351	0.130
118	X	15679019	2.202	0.084	0.100	0.070
19	X	15682937	2.345	0.113	0.122	0.098
112	X	15705777	2.467	0.152	0.169	0.140
130	X	15980131	2.479	0.126	0.253	0.047
30	X	15985161	2.471	0.176	0.121	0.151
86	X	16279793	2.377	0.325	0.156	0.107
wol20X	X	16677891	1.708	0.028	0.133	0.033
3	X	16677901	2.288	0.288	0.210	0.116
45	X	16730463	4.500	0.110	0.233	0.097
wol19X	X	17197389	1.008	0.107	0.283	0.114
55	X	17592835	2.337	0.153	0.069	0.107
31	X	18559749	2.470	0.378	0.163	0.137
15	X	19247730	3.836	0.165	0.260	0.197
17	X	19392349	2.173	0.276	0.249	0.034
4	X	19399583	2.157	0.157	0.233	0.211
70	X	19498575	2.660	0.180	0.201	0.073
89	X	19561872	2.218	0.177	0.240	0.043
57	X	19607504	2.669	0.141	0.150	0.124
83	X	19677223	2.411	0.154	0.258	0.234
33	X	19717282	2.799	0.378	0.192	0.110
106	X	19780935	2.490	0.257	0.146	0.066
104	X	20067935	2.476	0.128	0.140	0.099
100	X	20915774	2.423	0.207	0.327	0.095
49	X	20925189	2.402	0.160	0.118	0.062
wol23X	X	20994197	1.374	0.104	0.103	0.051
64	X	21917264	2.377	0.090	0.244	0.106

Appendix D: Expression (mean units of β-galactosidase enzymatic activity) of autosomal insertions. Every insertion was measured with three biological replicates and two technical replicates.

Internal refereence	Chromosome	Mapped position (v5.30)	Average male expression	Standard deviation of male expres.	Average female expression	Standard deviation of female expres.
control 3	2L	5108428	8.166	0.370	0.352	0.057
control 2	2L	6498770	9.153	0.393	0.561	0.745
wol4	2L	7423613	7.035	1.329	1.644	2.647
control 4	2L	7984133	9.132	0.230	0.371	0.152
wol7	2R	2603250	10.557	2.620	0.758	0.593
control 8	2R	9465619	10.545	0.409	0.274	0.089
wol9	2R	15518667	6.238	2.164	1.086	1.013
control 11	3L	746383	10.127	0.400	0.373	0.209
wol11	3L	749342	7.103	2.415	0.497	1.225
wol6	3L	8414592	7.009	1.701	1.416	2.257
wol18	3L	14751002	5.958	4.580	0.001	0.320
control 9	3L	18839391	7.677	0.346	0.459	0.135
wol16	3L	21872686	8.548	1.296	0.133	0.630
wol2	3R	790870	7.447	1.656	2.002	4.021
wol1	3R	279214	15.363	3.909	0.743	1.642
wol14	3R	5920571	9.249	2.576	0.391	1.710
control 6	3R	7589977	12.125	0.382	0.209	0.196
wol3	3R	12881438	10.142	2.301	0.600	1.223
wol15	3R	14743978	10.770	4.434	0.489	0.559
wol17	3R	14983880	9.539	2.192	0.338	0.641
wol10	Autosome		7.542	1.233	0.156	2.067
wol8	Autosome		7.605	2.197	0.150	1.481

Appendix E: Comparison of X-linked and autosomal gene expression for protein abundance and mRNA abundance.

Internal refereence	Chromosome	Cytological band	Mapped position (v5.30)	Enzymatic assay expression in males (average)	Enzymatic assay expression in males (standard deviation)	qRT-PCR expression in males (average)	qRT-PCR expression in males (standard error)
71	X	13A5	14917818	1.230	0.061	0.134	0.011
97	X	4B1	4025208	2.062	0.221	0.259	0.051
3	X	15A7	16677901	2.288	0.288	0.176	0.012
64	X	20C1	21917264	2.377	0.090	0.099	0.027
104	X	19C1	20067935	2.476	0.128	0.202	0.020
15	X	18C3	19247730	3.836	0.165	0.258	0.015
45	X	15A11	16730463	4.500	0.110	0.511	0.046
control 9	3L	75E2	18839391	7.677	0.346	1.302	0.164
control 4	2L	28D3	7984133	9.132	0.230	0.752	0.057
control 2	2L	26D9	6498770	9.153	0.393	0.827	0.024
control 11	3L	61C9	746383	10.127	0.400	0.821	0.056
control 8	2R	50B3	9465619	10.545	0.409	1.037	0.148
control 6	3R	86E18	7589977	12.125	0.382	1.466	0.147
wol1	3R	84B1	279214	15.363	3.909	1.098	0.223

Appendix F: Chromosomal locations of autosomal and X-linked transgene insertions of the *CG10920*, *CG12681*, and *CG1314* construct.

Construct	Internal reference	Chromosome	Cytological band	Mapped position (v5.30)	Landing site class	Affected gene	Proximal gene within 10Kb	Distal gene within 10Kb
CG10920	A2	2L	28B1	7576521	5'UTR	CG34374		
CG10920	A10	2L	27F3	7421490	Exon	CG5229		
CG10920	A1	2R	53D8	12670334	Exon	CG15920		
CG10920	A6	2R	49F10	9107394	Exon	CG4646		
CG10920	A7	2R	55C4	14244239	Exon	CG5580		
CG10920	A8	2R	54B16	13347396	Intron	CG14478		
CG10920	A13	3L	75B1	17955937	Exon	CG8127 & CG32193		
CG10920	A3	3R	94E1	18968035	Intergenic		CG4637	
CG10920	X7	X	5C6	5780651	Intergenic		CG16721	

CG10920	X11	X	7B6	7586656	Intron	CG12690		
CG10920	X5	X	10D1	11516084	5' UTR	CG1817		
CG10920	X8	X	11E1	13022777	Exon	CG32638		
CG10920	X6	X	12F5	14720137	Intergenic		CG9533	CG14408
CG10920	X4	X	17C2	18428513	Intergenic			CG6500
CG10920	X3	X	18F3	19743488	Intron	CG11940		
CG12681	A15	2L	25C1	5027473	Intergenic		CG16858	CG4145
CG12681	A09	2R	43A2	3136383	Intergenic		CG1851	CG11086
CG12681	A04	2R	46B1	5599879	5'UTR	CG1772		
CG12681	A01	3L	65D5	6972569	5'UTR	CG10060		
CG12681	A17	3L	67B10	9498960	Intergenic		CG3424	CG3408
CG12681	A13	3R	94E5	19016930	5'UTR	CG17894		
CG12681	A05	3R	99F2	26214768	Exon	CG1469		
CG12681	A10	Autosome						
CG12681	X03	X	1D2	828749	Exon	CG32815		
CG12681	X05	X	1E5	1130460	Intron	CG3638		
CG12681	X10	X	2B4	1513944	Intron	CG11491		
CG12681	X01	X	2B17	1842812	Intron	CG3600		
CG12681	X11	X	6E4	6879859	Intergenic			CG14430
CG12681	X02	X	18F4	19780935	Intron	CG32529 & CG11937		
CG12681	X06	X						
CG12681	X07	X						
CG1314	A13	2L	24C4	3788360	Intergenic		CG31958	
CG1314	A01	2L	30B1	9387298	Intergenic			CG3752
CG1314	A15	3L	66A17	7860777	Intergenic		CG12151	CG32364
CG1314	A14	3L	80A2	22781342	5'UTR	CG14448		
CG1314	A02	3R	85D22	5358515	Exon	CG9379		
CG1314	A08	3R	86E10	7393239	Intergenic		CG6783	CG14709
CG1314	A11	Autosome						
CG1314	A12	Autosome						
CG1314	X08	X	1B2	323934	Intron	CG32816		
CG1314	X10	X	4B1	4014702	Exon	CG4857		
CG1314	X12	X	4D6	4823106	Exon	CG4068		
CG1314	X06	X	7C2	7802374	Intergenic		CG10946	CG1444
CG1314	X04	X	8C4	8936538	Intron	CG42388 & CG10962		
CG1314	X09	X	10B5	11590075	Exon	CG1830		
CG1314	X03	X	10D8	11623204	Exon	inaF cluster		
CG1314	X02	X	12A9	13536139	Intron	CG11172		
CG1314	X01	X	15F3	17106995	Exon	CG18258		

Appendix G: Expression (mean units of β-galactosidase enzymatic activity) of autosomal and X-linked insertions. Every insertion was measured with three biological replicates and two technical replicates.

Construct	Internal refereence	Chrom osome	Mapped position (v5.30)	Average male expression	Standard deviation of male expres.	Average female expression	Standard deviation of female expres.
CG10920	A2	2L	7576521	5.76	0.24	-0.01	0.06
CG10920	A10	2L	7421490	6.45	0.06	-0.01	0.05
CG10920	A1	2R	12670334	4.93	0.37	-0.05	0.10
CG10920	A6	2R	9107394	6.16	0.31	0.12	0.06
CG10920	A7	2R	14244239	6.49	0.34	0.05	0.14
CG10920	A8	2R	13347396	5.75	0.17	0.01	0.14
CG10920	A13	3L	17955937	6.40	0.10	0.15	0.24
CG10920	A3	3R	18968035	12.68	0.42	0.15	0.12
CG10920	X7	X	5780651	2.53	0.37	-0.03	0.08
CG10920	X11	X	7586656	2.13	0.15	0.05	0.14
CG10920	X5	X	11516084	2.16	0.26	-0.04	0.12
CG10920	X8	X	13022777	3.06	0.23	-0.02	0.12
CG10920	X6	X	14720137	2.42	0.34	-0.20	0.18
CG10920	X4	X	18428513	2.28	0.29	0.06	0.08
CG10920	X3	X	19743488	2.52	0.23	0.08	0.05
CG12681	A15	2L	5027473	4.31	0.19	0.19	0.10
CG12681	A09	2R	3136383	4.95	0.25	0.23	0.12
CG12681	A04	2R	5599879	4.51	0.20	0.00	0.06
CG12681	A01	3L	6972569	5.42	0.26	0.08	0.23
CG12681	A17	3L	9498960	5.15	0.49	0.08	0.22
CG12681	A13	3R	19016930	4.40	0.27	0.08	0.09
CG12681	A05	3R	26214768	4.47	0.35	0.20	0.04
CG12681	A10	Autosome		8.38	0.33	0.29	0.19
CG12681	X03	X	828749	1.53	0.15	0.11	0.07
CG12681	X05	X	1130460	1.31	0.23	0.11	0.13
CG12681	X10	X	1513944	1.24	0.09	0.17	0.11
CG12681	X01	X	1842812	1.69	0.22	0.06	0.12
CG12681	X11	X	6879859	1.16	0.26	0.01	0.15
CG12681	X02	X	19780935	1.39	0.24	0.10	0.13
CG12681	X06	X		1.34	0.17	0.21	0.10
CG12681	X07	X		1.14	0.20	0.11	0.09
CG1314	A13	2L	3788360	2.28	0.30	0.13	0.07
CG1314	A01	2L	9387298	1.48	0.24	-0.02	0.15
CG1314	A15	3L	7860777	2.15	0.31	0.10	0.07

CG1314	A14	3L	22781342	2.10	0.20	0.17	0.12
CG1314	A02	3R	5358515	2.09	0.26	0.09	0.15
CG1314	A08	3R	7393239	2.39	0.07	0.26	0.22
CG1314	A11		Autosome	1.86	0.08	0.26	0.18
CG1314	A12		Autosome	2.32	0.28	0.09	0.08
CG1314	X08	X	323934	0.65	0.22	0.04	0.14
CG1314	X10	X	4014702	0.82	0.04	0.04	0.08
CG1314	X12	X	4823106	0.98	0.08	-0.06	0.11
CG1314	X06	X	7802374	0.92	0.17	0.09	0.18
CG1314	X04	X	8936538	0.32	0.11	0.12	0.10
CG1314	X09	X	11590075	0.68	0.13	0.14	0.07
CG1314	X03	X	11623204	0.53	0.15	0.05	0.04
CG1314	X02	X	13536139	0.93	0.24	0.02	0.14
CG1314	X01	X	17106995	0.67	0.07	-0.04	0.10

Appendix H: Comparison of X-linked and autosomal gene expression for protein abundance and mRNA abundance of the *CG10920, CG12681, and CG1314*.

Construct	Internal refereence	Chromo some	Mapped position (v5.30)	Average male expression	Standard deviation male expres.	qRT-PCR expression males	qRT-PCR expression males standard deviation
CG10920	A2	2L	7576521	5.76	0.24	0.801	0.067
CG10920	A10	2L	7421490	6.45	0.06	0.763	0.133
CG10920	A1	2R	12670334	4.93	0.37	0.716	0.166
CG10920	A6	2R	9107394	6.16	0.31	0.768	0.248
CG10920	A7	2R	14244239	6.49	0.34	1.135	0.131
CG10920	A8	2R	13347396	5.75	0.17	0.936	0.225
CG10920	A13	3L	17955937	6.40	0.10	0.896	0.44
CG10920	A3	3R	18968035	12.68	0.42	1.11	0.225
CG10920	X7	X	5780651	2.53	0.37	0.283	0.059
CG10920	X11	X	7586656	2.13	0.15	0.542	0.152
CG10920	X5	X	11516084	2.16	0.26	0.286	0.043
CG10920	X8	X	13022777	3.06	0.23	0.256	0.025
CG10920	X6	X	14720137	2.42	0.34	0.55	0.119
CG10920	X4	X	18428513	2.28	0.29	0.332	0.174
CG12681	A15	2L	5027473	4.31	0.19	0.831	0.137
CG12681	A09	2R	3136383	4.95	0.25	0.962	0.132
CG12681	A04	2R	5599879	4.51	0.20	0.714	0.054
CG12681	A01	3L	6972569	5.42	0.26	0.87	0.152
CG12681	A17	3L	9498960	5.15	0.49	0.766	0.298

CG12681	A13	3R	19016930	4.40	0.27	0.828	0.127
CG12681	A05	3R	26214768	4.47	0.35	1.265	0.125
CG12681	A10		Autosome	8.38	0.33	1.354	0.154
CG12681	X03	X	828749	1.53	0.15	0.342	0.046
CG12681	X05	X	1130460	1.31	0.23	0.287	0.037
CG12681	X10	X	1513944	1.24	0.09	0.282	0.063
CG12681	X01	X	1842812	1.69	0.22	0.273	0.023
CG12681	X11	X	6879859	1.16	0.26	0.351	0.033
CG12681	X02	X	19780935	1.39	0.24	0.487	0.122
CG12681	X06		X	1.34	0.17	0.249	0.07
CG12681	X07		X	1.14	0.20	0.248	0.052
CG1314	A13	2L	3788360	2.28	0.30	2.711	0.94
CG1314	A01	2L	9387298	1.48	0.24	6.359	3.386
CG1314	A15	3L	7860777	2.15	0.31	8.663	2.232
CG1314	A14	3L	22781342	2.10	0.20	2.077	1.059
CG1314	A02	3R	5358515	2.09	0.26	4.083	0.556
CG1314	A08	3R	7393239	2.39	0.07	8.663	2.232
CG1314	A11		Autosome	1.86	0.08	3.169	1.207
CG1314	A12		Autosome	2.32	0.28	5.977	3.076
CG1314	X08	X	323934	0.65	0.22	1.419	0.158
CG1314	X10	X	4014702	0.82	0.04	0.72	0.12
CG1314	X12	X	4823106	0.98	0.08	1.071	0.389
CG1314	X06	X	7802374	0.92	0.17	2.471	1.141
CG1314	X04	X	8936538	0.32	0.11	2.888	0.549
CG1314	X09	X	11590075	0.68	0.13	1.764	0.312
CG1314	X03	X	11623204	0.53	0.15	0.504	0.312
CG1314	X02	X	13536139	0.93	0.24	1.236	0.204
CG1314	X01	X	17106995	0.67	0.07	2.072	0.469

13. Curriculum vitae

Name:	Claus Kemkemer
Address:	Greinerberg 9, 81371 München
Date of birth:	09.12.1977
Place of birth:	Neu-Ulm
Nationality:	German
E-mail:	ck_dionysos@yahoo.de

Dissertation:

01.07.2007–01.04.2011	Ludwig-Maximilians University Munich, Department: Evolutionary biology, AG Parsch; Topic: Functional analysis of X-chromosomal gene expression in *Drosophila melanogaster*

Study:

01.09.1997–31.08.2002	University of Applied Science Mannheim, Study of biotechnology
01.09.1998–28.02.1999	Internship at the German Cancer Research Centre in Heidelberg, Department: Interaction of carcinogenic agents with biological macromolecules
01.03.2000–31.08.2000	Internship at the Pharma Research Centre of the Bayer AG in Wuppertal, Department: Pharma-Research Antiinfektiva II
01.10.2001–31.08.2002	Diplom thesis at the Frauenhofer-Institute of Interfacial Engineering and Biotechnology, Topic: Purification of the Dihydroliponamid Dehydrogenase from *Scyliorhinus canicula*
26.09.2002	Degree: Graduate Engineer of (FH) Biotechnology
01.10.2002–31.03.2007	University Ulm, Study of Biology

01.07.2006–28.02.2007	Diplom thesis at the University of Ulm, Institute of Human Genetics, Topic: A comparative expression analysis of different tissues from chicken and mouse with the aid of the array technique
01.02.2007	Degree: Diplom in Biology
01.03.2007–30.06.2007	Employed at the University of Ulm, Department: Human Genetics

Conferences:

1. European Society for Evolutionary Biology (ESEB), August 2009, Turin (Poster)
2. Conference of the German Society of Human Genetics, March 2007, Bonn (Poster)
3. Genome Informatics, September 2006, Hinxton (Poster)
4. Society of Molecular Biology and Evolution, Juli 2010, Lyon (Talk)

Publication list:

1. Kemkemer C, Kohn M, Kehrer-Sawatzki H, Fundele RH, Hameister H. Enrichment of brain-related genes on the mammalian X chromosome is ancient and predates the divergence of synapsid and sauropsid lineages. Chromosome Res. 2009;17(6):811-20.

2. Kemkemer C, Kohn M, Cooper DN, Froenicke L, Högel J, Hameister H,Kehrer-Sawatzki H. Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. BMC Evol Biol. 2009 Apr 24;9:84.

3. Kemkemer C, Kohn M, Kehrer-Sawatzki H, Minich P, Högel J, Froenicke L, Hameister H. Reconstruction of the ancestral ferungulate karyotype by electronic chromosome painting (Epainting). Chromosome Res. 2006;14(8):899-907. Epub 2007Jan 19.

4. Kemkemer C, Hense W, Parsch J. Fine-scale analysis of X chromosome inactivation in the male germline of Drosophila melanogaster. Mol Biol Evol. (in press; doi: 10.1093/molbev/msq355).

14. Acknowledgements

I would like to thank Prof. John Parsch for giving me the opportunity to work in his lab. His advices and his supervision helped me in extending my scientific knowledge and my method of scientific working. He promoted my scientific career in many ways. Aside of the scientific relationship with John Parsch I also want to thank him for his colleagueship and the time we spend outside the university, such as the visits of the “Oktoberfest” and visits of several “Biergarten” around Munich. The members of my committee, in particular Prof. Susanne Renner, Prof. John Baines, Prof. Wolfgang Stephan, Prof. Dirk Metzler, Prof. Wilfried Gabriel, and Prof. Thomas Lahaye, I want to thank for the scientific support during my PhD and the assessment of this thesis.

The environment offered by the evolutionary biology group in Munich was very encouraging and stimulating. The cooperation with Winfried Hense, Pavlos Pavlidis and Sarah Saminiadin-Peter during my PhD was a big support and help for my projects. For these colleagueship I want to thank these three people and for the discussion and friendship.

I’m thankful for the scientific support/work of Hedwig Gebhart, Carmen Iannitti and Yvonne Cämmerer, which they contributed to this thesis. An extra thank to Hedwig Gebhart for the enormous help for mastering and mapping over 100 insertions together.

Many thanks for the help with administrative things to Kathrin Kümpfbeck and Ingrid Kroiß.

I want to thank Anja Hörger, Iris Fischer, Simone Lange, Hildegard Lainer, Stefan Laurant, Aurelien Tellier, Robert Piskol, Gisela Brinkmann, and Annica Vrljic for giving me a great time and for spending the entire PhD together. The friendship, which I have to all of them, is going beyond the PhD time. We spend many times together and I enjoyed every second.

Furthermore, I’m thankful for the friendship of Stephan Hutter, Martin Hutzenthaler, Ricardo Wilches, Pablo Duchen, Susanne Voigt, Daniel Zivkovic, Mamadou Mboub, Francesco Paparazzo, Rayna Stamboliyska, Annegret Werzner, Ana Catalan, Erin Foley and Meike Wittmann. All of them were a great help, either for the support or the nice conversation.

Big thanks go to Miriam Linnenbrink, Sonja Grath, Lisha Naduvilezhath and Lena Müller for being patient and helpful during my endless visits and questions. Special thanks to Lena Müller for helping me correcting my English and reading many of my proposals and abstracts.

Last but not least I want to thanks my parents for endless support not only during my PhD, but also for the support they gave me my entire life. Furthermore, I'm thankful for my relatives, which support me in many ways.