Age-dependent chromosomal distribution of male-biased genes in *Drosophila*

Yong E. Zhang,[1] Maria D. Vibranovski,[1] Benjamin H. Krinsky,[2] and Manyuan Long[1,2,3]

[1]*Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA;* [2]*Committee on Evolutionary Biology, University of Chicago, Chicago, Illinois 60637, USA*

We investigated the correlation between the chromosomal location and age distribution of new male-biased genes formed by duplications via DNA intermediates (DNA-level) or by de novo origination in *Drosophila*. Our genome-wide analysis revealed an excess of young X-linked male-biased genes. The proportion of X-linked male-biased genes then diminishes through time, leading to an autosomal excess of male-biased genes. The switch between X-linked and autosomal enrichment of male-biased genes was also present in the distribution of both protein-coding genes on the *D. pseudoobscura* neo-X chromosome and microRNA genes of *D. melanogaster*. These observations revealed that the evolution of male-biased genes is more complicated than the previously detected one-step X→A gene traffic and the enrichment of the male-biased genes on autosomes. The pattern we detected suggests that the interaction of various evolutionary forces such as the meiotic sex chromosome inactivation (MSCI), faster-X effect, and sexual antagonism in the male germline might have shaped the chromosomal distribution of male-biased genes on different evolutionary time scales.

[Supplemental material is available online at http://www.genome.org.]

It has been observed that male-biased genes in *Drosophila* are overrepresented on autosomes (Parisi et al. 2003; Ranz et al. 2003). Consistent with this result, a dynamic process that can explain the nonrandom autosomal distribution has also been observed, in which autosomal new genes with X-linked parental genes are often male-biased. Specifically, a significant excess of autosomal testis-expressed retrogenes were identified as RNA-duplicates of X-linked parental genes (Betran et al. 2002). Recently, similar X→A gene traffic was observed in the DNA-level duplication and relocation data set of the *Drosophila* genus (Vibranovski et al. 2009b), and was further confirmed for DNA-level duplications in the *D. pseudoobscura* neo-X chromosome (Meisel et al. 2009). In addition, selective extinction of neo-X linked male-biased genes also occurred in *D. pseudoobscura* (Sturgill et al. 2007). These three lines of genome-wide investigation support a common pattern of out-of-X traffic for male-biased genes, resulting in an enrichment of these genes on autosomes in the long term.

Various hypotheses have been proposed to explain the chromosomal distribution of sex-biased genes. The sexual antagonism hypothesis (Rice 1984) predicts an increase of X-linked sexually antagonistic genes in a population over a wide range of parameters (e.g., recessive mutations advantageous for males and disadvantageous for females). With the assumption of a modifier gene that may restrict expression to the advantageous sex, X-linked antagonistic genes would be more likely to be involved in the evolution of sexually dimorphic traits (Rice 1984; Vicoso and Charlesworth 2006). In autosomal inheritance, when the advantageous effect on one sex is greater than the disadvantageous effect on the other sex, sexually antagonistic genes can spread in the population (Rice 1984). Although sexual antagonism has often been used to explain the evolution of sex-biased genes (Betran et al. 2002; Parisi et al. 2003; Ranz et al. 2003; Meisel et al. 2009), recent reports questioned the association between sex-biased expression and sexual antagonism

(Vicoso and Charlesworth 2009; Innocenti and Morrow 2010). On the other hand, meiotic sex chromosome inactivation (MSCI) in spermatogenesis predicts X demasculinization via selection favoring autosomal male-biased genes, if the X chromosome is inactive during meiosis (Lifschytz and Lindsley 1972; Betran et al. 2002; Hense et al. 2007; Vibranovski et al. 2009a, 2010). Thus, if sex-biased expression were a proxy for sexual antagonism, sexually antagonistic selection is compatible with a paucity of X-linked somatic male-biased genes (Parisi et al. 2003). Additionally, the existence of MSCI-based selection is supported by the recent finding that autosomal genes retroposed from the X chromosome frequently demonstrate complementary expression in meiosis compared to their X-linked parental genes (Vibranovski et al. 2009a).

However, a number of X-linked evolutionarily young genes have been identified recently (Arguello et al. 2006; Levine et al. 2006; Chen et al. 2007), all of which are male-biased. In order to understand whether these cases represent a general pattern, we examined the chromosomal distribution of male-biased genes of different evolutionary ages. Unexpectedly, we observed that the X chromosome has undergone an initial enrichment of young male-biased genes through intrachromosomal origination. However, as gene age increases, this excess gradually diminishes and is finally reversed, resulting in an overrepresentation of male-biased genes on the autosomes. This dynamic suggests a significant impact of the evolutionary time scale on the different mechanisms responsible for the evolution of male-biased genes. Here, we discuss how different evolutionary forces may impact the chromosomal distribution of male-biased genes with different ages.

Results

The X chromosome shows a paucity of old male-biased genes but an excess of young male-biased genes

In order to date the origination of *D. melanogaster* protein-coding genes, we determined the presence or absence of their orthologs across all other 11 *Drosophila* species with reference to *D. melanogaster*–centric syntenic alignments (Kuhn et al. 2007).

We inferred the origination time by following the parsimony principle and assigned genes to their corresponding phylogenetic branches (see Methods). We identified 947 genes (representing 7% of all genes in the genome) within the six branches leading to *D. melanogaster* (Fig. 1; Supplemental Table S1). Therefore, the *Sophophora* subgenus displays a gene origination rate of 15 genes per million years (Myr) after its divergence with the *Drosophila* subgenus. For the following analyses described in this work, unless otherwise specified, these 947 genes are referred to as young genes and the remaining ones as old genes.

Based on exon–intron structure and sequence similarity, young genes were classified as follows: 102 (11%) retrogenes and 845 nonretroposed genes, including 741 (78%) DNA-level duplicates, and 104 (11%) de novo genes (Supplemental Table S2A). We defined de novo genes as those without ancestral coding genes (see Methods). The majority of DNA-level duplicates (578 or 78%) arose via intrachromosomal events, while the remaining 163 duplicates (22%) represent interchromosomal events. Intrachromosomal events were further divided into tandem (49%) and dispersed (51%) duplications (X chromosome: 40%, 60%; autosomes: 51%, 49%, respectively). Our analyses of the 845 nonretroposed young genes (DNA-level duplicates and de novo genes) revealed interesting patterns distinct from those of previously analyzed retroposed genes (Betran et al. 2002; Bai et al. 2007).

For 38 out of 163 interchromosomal DNA-level duplicates with unambiguous chromosomal movement directionality and high sequence identity (Supplemental Table S2B), we detected a significant excess of duplicated genes that moved from the X chromosome to an autosome (X→A traffic, Table 1). Moreover, according to FlyAtlas testis and ovary microarray data (Chintapalli et al. 2007), 50% (seven out of 14) of X-derived autosomal genes showed male-biased expression, which is significantly higher than the genomic background (24%, One-sided Fisher's exact test, $P = 0.03$). This X→A excess in the *D. melanogaster* lineage is consistent with our previous observation for the entire genus (Vibranovski et al. 2009b), based on a DNA-level relocation database (Bhutkar et al. 2007). Moreover, this pattern also held when we used the whole data set of 163 genes (Supplemental Table S3).

Interestingly, reanalysis of FlyAtlas (Chintapalli et al. 2007) expression data (see Methods) led us to observe a remarkable difference between the location of old (branch 0) male-biased genes and that of recently evolved genes (branches 5 and 6). Old male-biased genes were underrepresented on the X chromosome (15% vs. 23%~28%), consistent with previous findings (Parisi et al. 2003; Sturgill et al. 2007) (Fig. 2). However, recently evolved genes showed different patterns. First, consistent with previous reports (Arguello et al. 2006; Levine et al. 2006; Begun et al. 2007b; Chen et al. 2007; Meisel et al. 2009), we observed that the proportion of recently evolved genes with male-biased expression (32 out of 52 or 61%) was significantly higher than the proportion of old genes with male-biased expression (2592 out of 11,397 or 23%) (Fig. 2; χ^2 test: degrees of freedom [df] = 1, $\chi^2 = 42$, $P = 9 \times 10^{-11}$). Second, a larger proportion of X-linked recently evolved genes showed male-biased expression as compared to X-linked old genes (Fig. 2; χ^2 test: df = 1, $\chi^2 = 73$, $P \leq 2.2 \times 10^{-16}$). More surprisingly, among recently evolved genes, there was a significant excess of male-biased genes on the X chromosome (Fig. 2B; χ^2 test: df = 1, $\chi^2 = 5$, $P = 0.03$). In other words, male-biased young genes showed the opposite chromosomal distribution as compared to male-biased old genes.

The enrichment of X-linked male-biased genes does not appear to be an exclusive feature of recently evolved genes (branches 5 and 6). Our observation still held (Supplemental Fig. S1A) when we analyzed all *D. melanogaster* young genes as a whole, i.e., those that originated after the split of the *Sophophora* subgenus (branches 1 to 6). Moreover, the results were the same when we analyzed several subsets of young genes under more stringent criteria. In the first subset, we excluded young genes that originated on branches 1 and 6 to control for possible unreliable branch assignment. More explicitly, genes present only in the *Sophophora* subgenus (assigned to branch 1) might have experienced gene loss prior to the *D. grimshawi* and *D. virilis* divergence instead of gene gain. *D. melanogaster* genes assigned to branch 6, on the other hand, could be polymorphisms (Zhou et al. 2008). In the second subset, to avoid possible incorrect gene annotation, we excluded young genes lacking peptide evidence (Brunner et al. 2007), because they were more likely to be pseudogenes. In the third subset, in order to obtain a conservative call of male-biased expression, we used male-specific genes,

Table 1. Interchromosomal DNA-level gene duplications

	A→X[a]	X→A[b]	A→A[c]
Observed	8 (3,5)[d]	16 (7,7)[e]	14 (1,12)
Expected[f] (excess)	7.6 (5%)	6.1 (163%)	24.3 (−42%)

χ^2 test: $\chi^2 = 20.6$, df = 2, $P = 3 \times 10^{-5}$.
[a]Autosome to X chromosome duplicates.
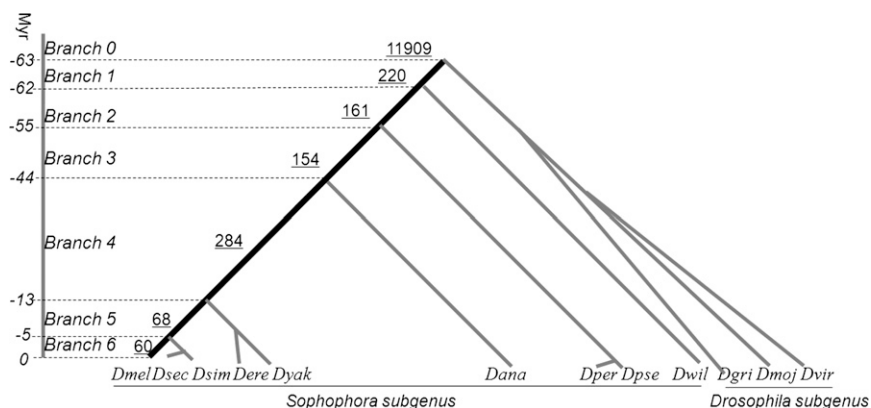[b]X chromosome to autosome duplicates.
[c]Interautosomal duplicates.
[d]Across genes with unique probes, we show the number of male-biased and non-male-biased gene numbers in parentheses.
[e]Two genes lack specific probes.
[f]We calculated expectation numbers according to Vibranovski et al. (2009b).

Figure 1. Age distribution of *D. melanogaster* protein-coding genes. Divergence times between species are given in millions of years according to Tamura et al. (2004). The bold line indicates the branch leading toward *D. melanogaster*. Evolutionary time and corresponding branch assignment are shown in italics. For example, branch 6 represents the *D. melanogaster*–specific branch. Numbers of gene origination events assigned to each branch are underlined (*Dmel*, *D. melanogaster*; *Dsim*, *D. simulans*; *Dsec*, *D. sechellia*; *Dyac*, *D. yakuba*; *Dere*, *D. erecta*; *Dana*, *D. ananassae*; *Dper*, *D. persimilis*; *Dpse*, *D. pseudoobscura*; *Dwil*, *D. willistoni*; *Dmoj*, *D. mojavensis*; *Dvir*, *D. virilis*; and *Dgri*, *D. grimshawi*).
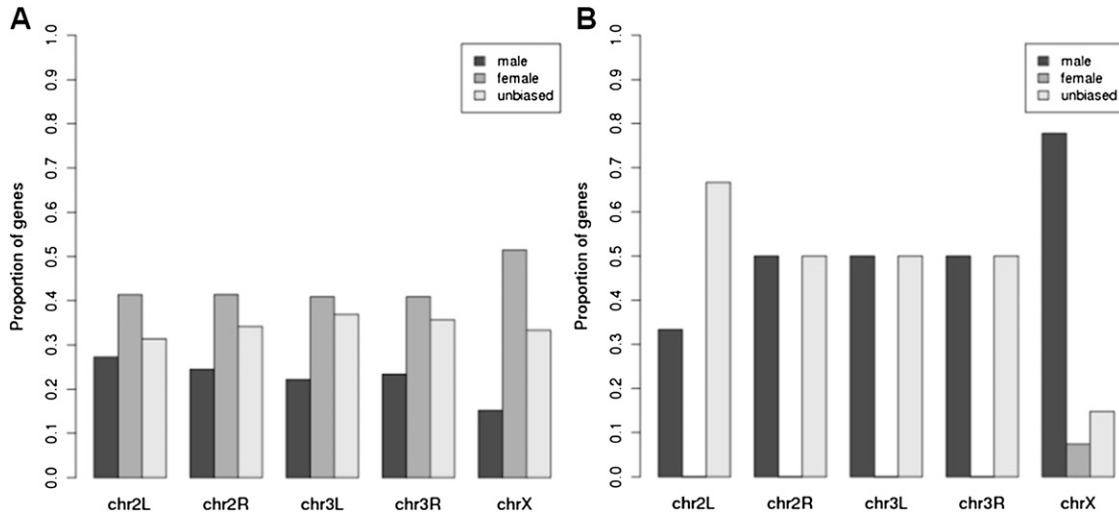
Figure 2. Chromosomal distributions of male-biased, female-biased, and unbiased *D. melanogaster* genes. (*A*) Old genes that originated before the *Sophophora* and *Drosophila* subgenus divergence (Fig. 1, branch 0); (*B*) recently evolved genes that originated after the *D. yakuba*/*D. erecta* split from the *D. melanogaster* subgroup (Fig. 1, branches 5 and 6).

i.e., those showing expression in males but not in females. Our observation was still the same for all three different subsets of genes (Supplemental Fig. S1B–D).

The enrichment of young (branch 1 to 6) male-biased genes on the X chromosome stands in contrast to the paucity of X-linked old male-biased genes. This pattern also indicates a decline in the proportion of male-biased X linkage as genes age, although all young genes as a whole are abundant on the X chromosome (Supplemental Fig. S1A). We investigated this intriguing phenomenon by analyzing the within-chromosomal proportions of male-biased genes that originated at different evolutionary times (Fig. 3). We observed that the proportion of X-linked male-biased genes gradually declined as genes age (Fig. 3, black line; Spearman $\rho = 0.99$, $P = 3 \times 10^{-4}$). Autosomal genes, in contrast, did not show a significant correlation (Fig. 3, gray line; Spearman $\rho = 0.71$, $P = 0.14$).

Furthermore, we determined how the genomic distribution of male-biased genes between autosomes and the X chromosome changed with evolutionary time (Fig. 4). Recently evolved genes, such as *D. melanogaster* specific (<6 Myr) or *D. melanogaster* and *D. simulans* lineage specific (<13 Myr) genes, were significantly enriched on the X chromosome (66% of all male-biased genes with two branches combined). By contrast, branches older than 55 Myr had a decreased proportion of male-biased X-linked genes ($\sim$12%), which is lower than the genomic background of all X-linked genes (16%).

Given such a high abundance of young male-biased genes, we asked whether their parental genes are also male-biased. We found that fewer parental genes of X-linked male-biased duplicates were also male-biased (20%, 2/10) compared to the parental genes of autosomal young male-biased duplicates (32%, 12/37). These data, despite the small sample sizes and being statistically not significant, may suggest that compared to autosomal young genes, X-linked young genes more often evolved novel male-biased expression. However, as the majority of young genes are the result of intrachromosomal duplication events, the pattern might also reflect the fact that X-linked old genes are less likely to be male-biased.

Notably, female-biased genes were scarce among the recently evolved genes (branches 5 and 6), but abundant in the old gene group (43% vs. 3%, Fisher's exact test, $P \sim 1.6 \times 10^{-11}$) (Fig. 2). For all young genes identified from branches 1 to 6, there were 62 (9%)

female-biased genes compared to 301 (42%) male-biased genes and 353 (49%) unbiased genes. This pattern reflects a slow pace of female-biased gene evolution, but that a high proportion of these genes can accumulate over a long evolutionary time (Ranz et al. 2003).

Positive selection detected in the evolution of X-linked young male-biased genes

Next, we investigated the evolutionary forces acting on young X-linked genes. Previously, a genome-wide analysis regardless of
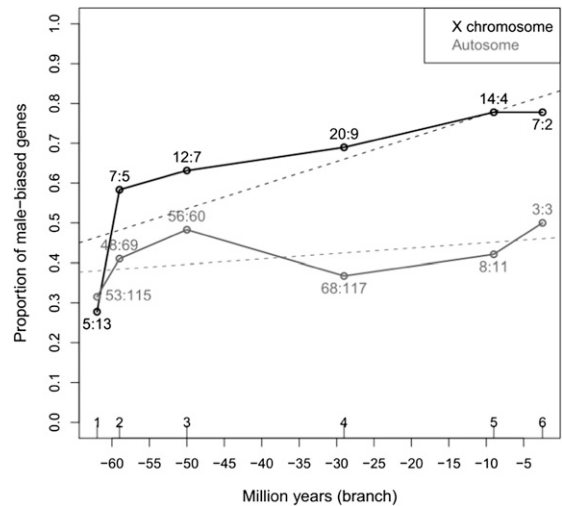
Figure 3. The proportions of male-biased genes in all evolutionary periods. The *y*-axis presents proportions of male-biased genes for a given chromosome category (X chromosome or autosome), whereas the *x*-axis indicates evolutionary time scale (branch or Myr). Here, time in millions of years is calculated as the middle point for each branch. For instance, *D. melanogaster*–specific genes assigned to branch 6 are shown at −2.5 Myr, representing the origination time average for an interval ranging from −5 to 0 Myr. Linear regressions are given as dashed lines. For each data point, numbers of male-biased and non-male-biased genes, respectively, are shown. The data is based on 715 out of the 843 young genes that have unique Affymetrix probes.
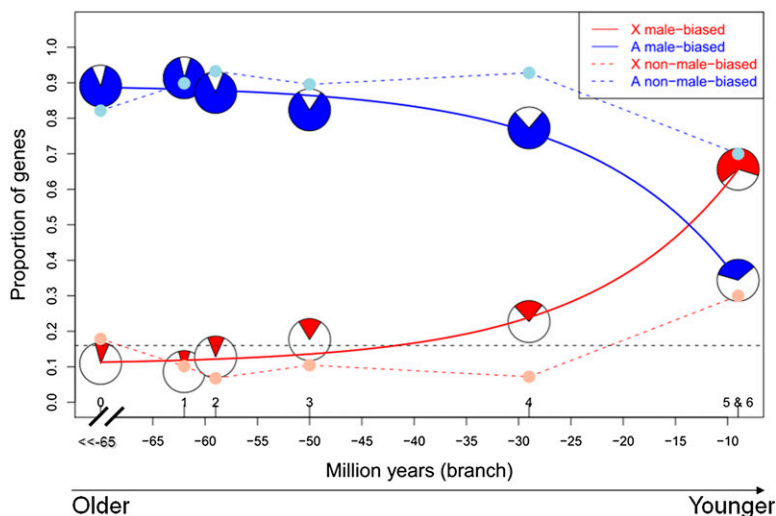
Figure 4. Proportions of male-biased and non-male-biased (female-biased and unbiased combined) genes originating in different evolutionary periods. For male-biased genes, we calculated the proportion as the number of male-biased genes in a given chromosome out of the whole genome. Analogously, we calculated non-male-biased gene proportions. Pie plots mark chromosomal proportions of male-biased genes. We pooled branches 5 and 6, considering individual small sample size. The black dashed line indicates the genomic proportion of all X-linked genes, 0.16. We used the exponential decay formula, $f(t) = N[e^{rt}(1 - d) + d]$, to model the origination process of male-biased genes, which converged as: $N = 1.14$, $r = 0.07$, and $d = 0.09$.

number of X-linked young male-biased genes is small, the likelihood ratio test still showed that this group was subject to stronger positive selection as compared to other groups with at least a marginal significance ($P \sim 0.1$; Supplemental Table S5). Thus, together with the previous pairwise comparison between parent and offspring genes (Fig. 5), these analyses revealed that adaptive selection drives the evolution of young male-biased genes on the X chromosome.

It has been reported that the initial manifestations of new gene emergence, namely polymorphic duplicates, occur at a lower frequency on the X chromosome, thus indicating that these duplicates are subject to stronger purifying selection (Emerson et al. 2008). Therefore, the excessive fixation of X-linked duplicates might not occur via neutral processes. Positive selection could have facilitated the fixation of X-linked young genes in addition to driving their subsequent sequence evolution.

gene age found that X-linked duplicates are more diverged as measured by the ratio of the nonsynonymous substitution rate to the synonymous substitution rate (d_N/d_S) (Thornton and Long 2002). Consistently, we observed a higher d_N/d_S among X-linked young genes compared to autosomal young genes (Wilcoxon rank sum test, $P = 0.005$) (Fig. 5). Moreover, X-linked male-biased genes showed a higher rate of d_N/d_S than male-biased genes located on autosomes (Wilcoxon rank sum test, $P = 0.009$). Such a pattern suggests that X-linked male-biased genes have evolved either under stronger positive selection or under relaxation of functional constraint.

In order to distinguish between these two possibilities, we also compared fixed and polymorphic substitutions between and within species using the McDonald-Kreitman (MK) test (McDonald and Kreitman 1991). Using population genomic sequences from *D. melanogaster* (7 Mb from 2L and X in 50 strains; http://www.dpgp.org/1K/) and from *D. simulans* (whole genome from six strains) (Begun et al. 2007a), we identified synonymous and replacement substitutions in the young genes (Supplemental Table S4). Based on a maximum likelihood framework (Bierne and Eyre-Walker 2004), we conducted a multi-locus MK test to estimate the proportion of substitutions fixed by adaptive mutation, α. Consistent with the previous genome-wide random sampling of 45 X-linked genes and 91 autosomal genes (Baines et al. 2008), X-linked young male-biased genes had a higher proportion of adaptive substitutions compared to the neutral expectation ($\alpha = 0.4 \sim 0.7$, Supplemental Table S5). More importantly, although the

The age-dependent chromosomal distribution of male-biased neo-X linked genes and male-biased microRNA genes

Next, we investigated whether the enrichment of X-linked male-biased expression is limited to young *D. melanogaster* protein coding genes. Two independent analyses using *D. pseudoobscura* coding genes and *D. melanogaster* noncoding microRNA (miRNAs) genes demonstrated that the initial enrichment of new male-biased genes on the X chromosome appears to be a general process in *Drosophila* evolution.

First, in *D. pseudoobscura* we analyzed the distribution of new genes linked to the ancestral X (XL) and to the neo-X chromosomes (XR, derived from the ancestral autosomal Muller element D) (Richards et al. 2005). Based on genomic alignments between
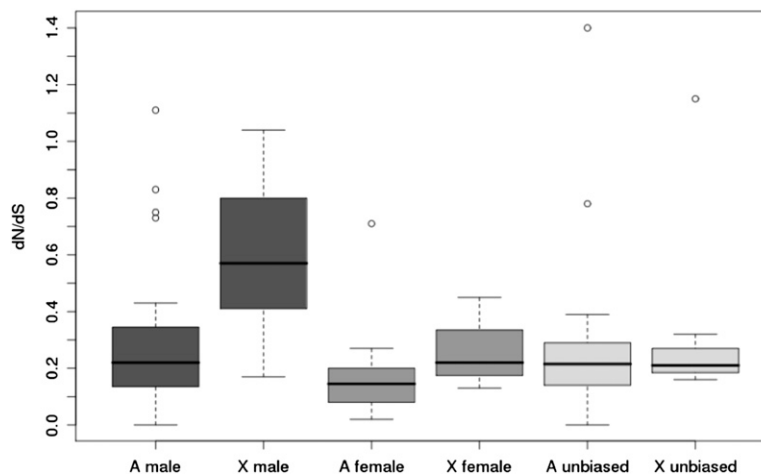
Figure 5. d_N/d_S distribution across different expression groups. Distinct shades indicate the three groups compared, i.e., male-biased (dark gray), female-biased (medium gray), and unbiased (light gray).

D. pseudoobscura and *D. melanogaster* provided by the UCSC Genome Browser and orthologous gene annotation in FlyBase, we built a conservative data set (Supplemental Table S6) of new genes that originated either in the *D. pseudoobscura* lineage (*D. pseudoobscura*-specific) or within the obscura branch (shared by *D. pseudoobscura* and *D. persimilis*). These genes represented a completely nonoverlapping data set from previously analyzed *D. melanogaster* young genes. Compared to old genes shared by the *Sophophora* and *Drosophila* subgenus, we detected an excess of new genes linked to the ancestral XL when both the *D. pseudoobscura* and obscura branches were considered together (27% vs. 16%, χ^2 test: df = 1, $\chi^2 = 124$, $P \leq 2.2 \times 10^{-16}$). In addition, compared to the obscura branch, XR showed a significant excess of *D. pseudoobscura* specific new genes (20% vs. 14%, χ^2 test: df = 1, $\chi^2 = 9$, $P = 0.002$). Since XR fused to the ancestral X chromosome before the divergence of *D. pseudoobscura* and *D. persimilis* less than 1 Myr ago (Hey and Nielsen 2004), our results demonstrate incipient and recent new gene enrichment on the X chromosome.

Furthermore, we identified 102 new male-biased genes and 1778 old male-biased genes using expression data from a male vs. female whole-body comparison (Sturgill et al. 2007). A larger proportion of XL- and XR-linked genes that originated within the *D. pseudoobscura* and obscura branches showed male-biased expression compared to genes shared by the whole *Drosophila* genus (24% vs. 10%, χ^2 test: df = 1, $\chi^2 = 4$, $P = 0.05$), indicating that new male-biased genes also tend to be enriched in *D. pseudoobscura* X chromosomes. We also confirmed that male-biased genes are underrepresented on the neo-X chromosome as a whole compared to autosomes (11% vs. 19%, χ^2 test df = 1, $\chi^2 = 64$, $P = 2 \times 10^{-15}$) (Sturgill et al. 2007). Thus, both the selective extinction of old male-biased genes (Sturgill et al. 2007) and the creation of new ones could have occurred on the neo-X chromosome.

The second independent line of evidence supporting the generality of new X-linked male-biased gene enrichment comes from miRNA analysis. Using recent sequencing-based small RNA expression data (Ruby et al. 2007; Czech et al. 2008) and UCSC genome alignment data, we identified 29 new miRNA genes that originated after the divergence of *D. melanogaster* and *D. pseudoobscura* (Supplemental Table S7). We observed that the X chromosome was enriched with new male-biased miRNA genes, compared to old X-linked miRNA genes (Fisher's exact test, $P = 0.03$) or compared to new autosomal miRNA genes (Fisher's exact test, $P = 0.04$) (Table 2). The result was similar when we used more stringent criteria, i.e., sex specific expression (Supplemental Table S8).

Age-dependent pattern of male-biased genes is not dominated by de novo genes

Our analyses above were conducted for both DNA-level duplicates and de novo genes. It is known that de novo protein-coding genes

Table 2. Sex bias of miRNAs

Chromosome	Young miRNAs[a]		Old miRNAs[a]		Fisher's exact test *P*-value[b]
	Male-biased	Non-male-biased	Male-biased	Non-male-biased	
X	13	1	7	6	0.03
Autosome	8	7	40	65	0.27

[a]"Old" and "Young" stand for miRNAs that originated before or after the divergence of *D. melanogaster* and *D. pseudoobscura*, respectively.
[b]For contingency table tests comparing male-biased and non-male-biased gene proportions between new and old miRNAs.

are often X-linked and male-biased (Levine et al. 2006; Begun et al. 2007b; Chen et al. 2007). This raised a question as to whether or not the age distribution of male-biased genes described above is predominantly affected by de novo loci. Thus, we repeated the analyses using only DNA-level duplicates. The result remained similar to what was observed in the whole young gene data set (e.g., Supplemental Figs. S2–S4). This indicated that the patterns were not exclusively generated by de novo genes, but were also typical of DNA-level duplicates.

Discussion

Chromosomal distribution and evolution of male-biased genes

We observed that new male-biased genes are overrepresented on the X chromosome, but old ones are overrepresented on the autosomes. This pattern suggests a two-step process of male-biased gene evolution. Specifically, new genes arose via DNA-mediated mechanisms in the X chromosome and performed male-related functions. As these genes aged, they were transposed into autosomes, were lost, or evolved to become non-sex-biased genes. Thus, the autosomes can accumulate male-biased genes either by accepting new genes via transposition from the X, or evolving new male-biased genes. Although X-derived autosomal duplicates are often male-biased (Table 1), the fact that we did not observe an excess of new, male-biased genes on the autosomes might be explained by the preponderance of intrachromosomal duplication events. Since between-chromosome duplications are much more rare than within-chromosome duplications, one may expect that the process of demasculinization by the interchromosomal copy mechanism is slow and only a small proportion of the X-linked male-biased genes can be copied into the autosomes. Thus, autosomes would not accumulate a large proportion of young male-biased genes. However, when this process occurs over an adequately long period of time, the autosomes become enriched with older male-biased genes, as has been previously observed (Parisi et al. 2003; Ranz et al. 2003).

This suggested model is not ready for a direct test, because it requires information about the ancestral expression states for the parental genes of autosomal male-biased genes, which often changed in the course of their evolution (Zhang et al. 2007). Moreover, the parental copies could have degenerated or evolved complementary functions (Dai et al. 2006; Potrzebowski et al. 2008). For such an analysis, having genome-wide data from multiple closely related species is essential. However, our genome-wide scan identified seven cases of pseudogenization and relocation (Supplemental Table S9). For example, an X-linked gene, *kek4* (*CG9431*), was relocated to chromosome 2L in the common ancestor of the *melanogaster* group and a male-biased gene, *CG14810*, encoded by the X chromosome of *D. melanogaster* and *D. simulans* degenerated in *D. sechellia* through one indel mutation. Furthermore, many old male-biased autosomal genes were copied from X-linked parental genes by RNA-based and DNA-level duplication (Betran et al. 2002; Bai et al. 2007; Vibranovski et al. 2009b).

It should be pointed out that the relative importance of all three mechanisms, namely transposition, extinction, and change of expression bias, is not well understood. In the case of the *D. pseudoobscura* neo-X chromosome, the former two mechanisms are important while the last one is not (Sturgill et al. 2007). However, it is unclear whether or not this conclusion is also applicable to the ancestral X chromosome.

Multiple mechanisms may contribute to evolution of sex-biased genes

A number of possible scenarios may explain why young male-biased genes are enriched on the X chromosome, but old genes are not. Sexual antagonism could contribute to the X-enrichment of young male-biased genes, assuming that copy number duplicates or new de novo alleles are recessive, advantageous for males, and disadvantageous for females (Rice 1984). The fixation of X-linked sexually antagonistic alleles that are beneficial to males has also been proposed based on some specific assumptions, e.g., the even distribution of dominance effects and fitness effects on males and females (Patten and Haig 2009). Alternatively, the emergence of young X-linked male-biased genes could also be explained by the faster-X hypothesis (Charlesworth et al. 1987). In this model, recessive male-advantageous X-linked mutations are more often fixed, and they do not necessarily have an adverse effect on female fitness.

Regarding the second step in the evolution of male-biased genes, namely X→A transposition, sexual antagonism favorable for autosomal fixation (Vicoso and Charlesworth 2006) and/or MSCI (Lifschytz and Lindsley 1972; Betran et al. 2002) may play a role in this process. On the other hand, the within-chromosomal duplication rate is higher than the between-chromosomal duplication rate (Emerson et al. 2008), which may contribute to the slow pace of X→A transposition.

A slight excess of X-linked female-biased genes was also detected (Fig. 2). Although most of them are old, a few recently arose on the X chromosome over 4 to 6 Myr in the common ancestor of the *D. melanogaster* and *D. simulans* clade (branch 5). This can be interpreted in the context of the dominance model of the sexual antagonism hypothesis. In this case, a dominant, X-linked gene that is favorable to females but disadvantageous for males can become fixed. The slow accumulation of female-biased genes in the X reflects an overall low rate of female gene origination, either due to a small dominance effect (the degree of dominance $h{\to}1/2$), or a minor disadvantageous effect on males (the ratio of fitness effects of male relative to female $k{\to}0$) along with a favorable effect on females (Vicoso and Charlesworth 2006, Equation 10).

Although sexual antagonism has often been thought to play an important role in the evolution of sex-biased genes (Betran et al. 2002; Parisi et al. 2003; Ranz et al. 2003; Meisel et al. 2009), some recent reports have shown that the association between sex-biased expression and sexual antagonism might not be significant (Connallon and Knowles 2005; Vicoso and Charlesworth 2009; Innocenti and Morrow 2010). Specifically, in Rice's (1984) model, sexually antagonistic genes become sex-biased when their expression in the harmed sex is decreased or silenced. Recent analyses (Connallon and Knowles 2005; Vicoso and Charlesworth 2009) suggest that sex-biased expression evolves by an increase in expression in one sex, rather than a decrease in expression in the opposite sex. Furthermore, Innocenti and Morrow (2010) found that there was little overlap between sex-biased genes and sexually antagonistic genes. These analyses suggest that the sexual antagonism, although consistent with the observed age-dependent chromosomal evolution of male-biased genes, needs to be further examined in the evolution of male-biased genes.

Young male-biased genes show faster protein evolution

Origination of new genes can be divided into two stages, namely the fixation of new genes, followed by subsequent sequence evolution. While the discussion above suggests a possible role of natural selection in the fixation of new genes, we also found that sequence changes of new genes were driven by natural selection. The comparison (Fig. 5) of sequence evolution between the X-linked and autosomal genes showed that unbiased genes are similar in terms of their protein sequence evolution, suggesting that similar forces were operating on the two types of chromosomes for non-male biased genes. However, remarkably, the comparison of male-biased genes on the X chromosome and autosomes revealed significantly elevated rates of protein sequence evolution for X-linked male-biased genes, suggesting that positive selection has acted on newly originated male-biased genes. Furthermore, the test based on between-species and within-species variation detected adaptive evolution in a large proportion of these new male-biased genes (Supplemental Table S5).

The faster-X hypothesis predicts high fixation rates for recessive advantageous X-linked mutations (Charlesworth et al. 1987). Consistent with previous reports mainly based on old genes (Thornton and Long 2002; Vicoso and Charlesworth 2006; Ellegren and Parsch 2007; Baines et al. 2008), our observations also revealed elevated substitution rates in X-linked young male-biased genes. Furthermore, we observed that significantly more protein sites of the young X-linked male-biased genes in this study were under positive selection than those of the old male-biased genes (Supplemental Table S5). These two lines of evidence support the faster-X hypothesis in the evolution of male-biased genes. However, young X-linked male-biased genes represent a small proportion (4%) of all X-linked genes and the X chromosome is overall demasculinized, which might explain why the faster-X effect was not observed in large samples (Thornton et al. 2006).

DNA-level out-of-X gene traffic

We confirmed DNA-level out-of-X gene traffic (Vibranovski et al. 2009b) for the *D. melanogaster* lineage (Table 1). However, it was reported recently that DNA-based duplication was randomly distributed among chromosomes in many *Drosophila* species, with the exception of *D. pseudoobscura* (Meisel et al. 2009). This apparent contradiction may be due to the fact that the different sample sizes in the two analyses, with the former test using a larger data set, have different statistical power. Consistently, the Meisel et al. (2009) study provided independent evidence for X→A traffic of retrogenes in the *Drosophila* genus and DNA-level duplications in *D. pseudoobscura*.

Evolution of sex-biased genes in other animals

It is tempting to compare our findings in *Drosophila* to analyses of other organisms. In mammals, the X chromosome appears to contain either an excess or scarcity of male-biased genes, depending on the stages of spermatogenesis (Khil et al. 2004; Potrzebowski et al. 2008). However, regarding the whole testis, the X chromosome is underrepresented (Zhang et al. 2010). Gene movement between the X chromosome and autosomes appears to be bidirectional, but with an excess of X→A male-biased genes (Emerson et al. 2004). As revealed in this study, the *Drosophila* new male-biased genes generated at the DNA level, consistent with previous analyses of both RNA- and DNA-level duplication (Betran et al. 2002; Vibranovski et al. 2009b), also showed a similar X→A process, but without the recruitment process into the X. In chicken, it was recently observed that female-detrimental genes are overrepresented on the Z chromosome (Mank and Ellegren 2009). These genes possibly contain some mutations occurring in coding regions, which are beneficial for homogametic males (Mank and

Ellegren 2009). It could be a general case that sex chromosomes often fix mutations beneficial for the homogametic sex (male in the ZW system and female in the XY system). From this perspective, *Drosophila* appears to be similar in that the X chromosome has fixed an excess of old female-biased (supposedly female-beneficial) genes (Fig. 2). Interestingly, male-biased genes in the mammalian and *Drosophila* X- and Z-linked genes in birds all showed faster-X/Z effects although driven by different forces of selection and genetic drift, respectively (Vicoso and Charlesworth 2006; Mank et al. 2007, 2010).

Methods

For detailed methods, see Supplemental material.

General pipeline

We downloaded FlyBase release V5.3 (http://hgdownload.cse. ucsc.edu/goldenPath/dm3/database/flyBaseGene.txt.gz) from the UCSC Genome Browser. We used MySQL V5.0.45 to organize the data, BioPerl (Stajich et al. 2002) and BioEnsembl (Stabenau et al. 2004) to fold the pipeline, and R V2.7.1 (http://www.R-project.org) to perform several statistical tests. Notably, for two-by-two contingency table tests, we used the χ^2 test generally; however, when one cell had not more than five samples, Fisher's exact test was used. Through the analyses of *D. melanogaster* genes, we generated three gene groups of different ages: (1) those that arose along the *D. melanogaster* branch after the most recent common ancestor (MRCA) of the genus; (2) those that arose after the MRCA of *D. melanogaster* and *D. yakuba*; (3) those that were present in the MRCA of the genus. We defined them as young genes, recently evolved genes, and old genes, respectively. For a few cases, we explicitly added branch information. For general discussion of gene origination (without specific age data), we simply referred to them as new genes.

Protein-coding genes' origination dating

For a *D. melanogaster* gene, we deduced its origination time by analyzing its ortholog distribution in different species. Briefly, we extended a published synteny-based strategy (Zhou et al. 2008) by integrating syntenic genomic alignment generated by the UCSC Genome Browser (Kuhn et al. 2007). If a candidate gene overlapped with reciprocal best genomic mapping, we assigned "gene presence" in the reference genome. Given presence and absence information across 12 species, we assigned genes onto phylogenetic branches by following a parsimony rule.

Gene origination mechanism classification

We made a new improvement in that we parsed out self genome alignment (chainSelf) information (Kuhn et al. 2007) to infer the most likely parent–child gene mappings. Then, we generally followed the strategy of previous studies (Levine et al. 2006; Bai et al. 2007; Zhou et al. 2008) and classified all new genes into three categories: DNA-based duplication; retroposition; and, de novo gene origination. We followed the *D. melanogaster* karyotype and treated the two major autosomes as single autosomes each while defining interchromosomal and intrachromosomal duplication events. However, 2L–2R and 3L–3R fusion actually occurred after the split of *D. melanogaster* and *D. pseudoobscura* (http://flybase. org/maps/chromosomes/synteny_table.html). After taking this factor into consideration, our related analyses were not significantly altered. First, the majority (74%) of DNA-level duplicates were still intrachromosomal duplicates, which did not change

much compared to the previous proportion (78%). Second, we had relatively more A→A cases in Table 1 (14 cases increased to 18). However, the excess of X→A still holds (χ^2 test: $\chi^2 = 15.8$, df = 2, $P = 4 \times 10^{-4}$).

Microarray data analysis

We reanalyzed the microarray data sets of FlyAtlas (Chintapalli et al. 2007) as the basic source to identify sex-biased genes. These data sets cover 12,277 genes with unique probes. We performed all the analyses based on Bioconductor software (Gentleman et al. 2004). We used the GCRMA package (v2.12.1) to adjust the background intensity and to normalize and summarize the expression value, and the MAS5 function of the Affy package (v1.18.2) to call the presence or absence of each gene. After that, we implemented the linear models of the *limma* package (v2.14.6) to assess whether genes show differential expression between testis and ovary (Smyth 2004). We defined genes with a false discovery rate (FDR) controlled p or q of 0.05 as a threshold for candidate sex-biased genes. Genes with significantly higher expression in testis, which were also called as "present" in testis across all four replicates by MAS5, were defined as male-biased genes since the previous study (Parisi et al. 2003) showed that the proportion of germline male-biased genes is much higher than that of somatic male-biased genes (20% vs. 2%). We identified 4958 female-biased genes, 3007 male-biased genes, and 4312 unbiased genes.

Evolutionary analysis

We calculated d_N and d_S based on codon-level alignments between parental genes and children genes. In order to perform the MK test, we used both 7 Mb of resequencing data from 50 *D. melanogaster* strains released by DPGP and whole-genome population data based on the syntenic alignment of six strains of *D. simulans* (Begun et al. 2007a). We filtered a low quality alignment region before counting substitutions.

miRNA origination dating

We profiled miRNAs by mapping 14 small RNA deep-sequencing samples (Ruby et al. 2007; Czech et al. 2008) onto the genome. Based on UCSC genomic alignment, we defined the newly emerged or evolved miRNA as miRNA absent both in *D. pseudoobscura* and *D. persimilis*. The putative new miRNA either had to have no orthologous locus in these species, or the seed region included mismatches in both *D. ananassae* and *D. pseudoobscura* or *D. persimilis*.

Acknowledgments

References

Arguello JR, Chen Y, Yang S, Wang W, Long M. 2006. Origination of an X-linked testes chimeric gene by illegitimate recombination in *Drosophila*. *PLoS Genet* **2:** e77. doi: 10.1371/journal.pgen.0020077.

Bai Y, Casola C, Feschotte C, Betran E. 2007. Comparative genomics reveals a constant rate of origination and convergent acquisition of functional retrogenes in *Drosophila*. *Genome Biol* **8:** R11. doi: 10.1186/gb-2007-8-1-r11.

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.

Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW, Nista PM, Jones CD, Kern AD, Dewey CN, et al. 2007a. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* **5:** e310.

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

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

Bhutkar A, Russo SM, Smith TF, Gelbart WM. 2007. Genome-scale analysis of positionally relocated genes. *Genome Res* **17:** 1880–1887.

Bierne N, Eyre-Walker A. 2004. The genomic rate of adaptive amino acid substitution in *Drosophila*. *Mol Biol Evol* **21:** 1350–1360.

Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, Deutsch EW, Panse C, de Lichtenberg U, Rinner O, et al. 2007. A high-quality catalog of the *Drosophila melanogaster* proteome. *Nat Biotechnol* **25:** 576–583.

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

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. doi: 10.1371/journal.pgen.0030107.

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

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

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **453:** 798.

Dai H, Yoshimatsu TF, Long M. 2006. Retrogene movement within- and between-chromosomes in the evolution of *Drosophila* genomes. *Gene* **385:** 96–102.

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

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

Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M. 2008. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. *Science* **320:** 1629–1631.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al. 2004. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* **5:** R80. doi: 10.1186/gb-2004-5-10-r80.

Hense W, Baines JF, Parsch J. 2007. X chromosome inactivation during *Drosophila* spermatogenesis. *PLoS Biol* **5:** e273. doi: 10.1371/journal.pbio.0050273.

Hey J, Nielsen R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **167:** 747–760.

Innocenti P, Morrow EH. 2010. The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biol* **8:** e1000335. doi: 10.1371/journal.pbio.1000335.

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.

Kuhn RM, Karolchik D, Zweig AS, Trumbower H, Thomas DJ, Thakkapallayil A, Sugnet CW, Stanke M, Smith KE, Siepel A, et al. 2007. The UCSC genome browser database: Update 2007. *Nucleic Acids Res* **35:** D668–D673.

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* **103:** 9935–9939.

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

Mank J, Ellegren H. 2009. Sex-linkage of sexually antagonistic genes is predicted by female, but not male, effects in birds. *Evolution* **63:** 1464–1472.

Mank JE, Axelsson E, Ellegren H. 2007. Fast-X on the Z: Rapid evolution of sex-linked genes in birds. *Genome Res* **17:** 618–624.

Mank JE, Nam K, Ellegren H. 2010. Faster-Z evolution is predominantly due to genetic drift. *Mol Biol Evol* **27:** 661–670.

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

Meisel RP, Han MV, Hahn MW. 2009. A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. *Genome Biol Evol* **1:** 176–188.

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.

Patten MM, Haig D. 2009. Maintenance or loss of genetic variation under sexual and parental antagonism at a sex-linked locus. *Evolution* **63:** 2888–2895.

Potrzebowski L, Vinckenbosch N, Marques AC, Chalmel F, Jegou B, Kaessmann H. 2008. Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biol* **6:** e80.

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

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

Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP. 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: Chromosomal, gene, and cis-element evolution. *Genome Res* **15:** 1–18.

Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res* **17:** 1850–1864.

Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3:** Article3. doi: 10.2202/1544-6115.1027.

Stabenau A, McVicker G, Melsopp C, Proctor G, Clamp M, Birney E. 2004. The EnsemblCore software libraries. *Genome Res* **14:** 929–933.

Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H, et al. 2002. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res* **12:** 1611–1618.

Sturgill D, Zhang Y, Parisi M, Oliver B. 2007. Demasculinization of X chromosomes in the *Drosophila* genus. *Nature* **450:** 238.

Tamura K, Subramanian S, Kumar S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* **21:** 36–44.

Thornton K, Long M. 2002. Rapid divergence of gene duplicates on the *Drosophila melanogaster* X chromosome. *Mol Biol Evol* **19:** 918–925.

Thornton K, Bachtrog D, Andolfatto P. 2006. X chromosomes and autosomes evolve at similar rates in *Drosophila*: No evidence for faster-X protein evolution. *Genome Res* **16:** 498–504.

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. doi: 10.1371/journal.pgen.1000731.

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

Vibranovski MD, Chalopin DS, Lopes HF, Long M, Karr TL. 2010. Direct evidence for postmeiotic transcription during *Drosophila melanogaster* spermatogenesis. *Genetics* **186:** 431–433.

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

Vicoso B, Charlesworth B. 2009. The deficit of male-biased genes on the *D. melanogaster* X chromosome is expression-dependent: A consequence of dosage compensation? *J Mol Evol* **68:** 576–583.

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

Zhang YE, Vibranovski MD, Landback P, Marais G, Long M. 2010. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on X chromosome. *PLoS Biol* **8:** e1000494. doi: 10.1371/journal.pbio.1000494.

Zhou Q, Zhang G, Zhang Y, Xu S, Zhao R, Zhan Z, Li X, Ding Y, Yang S, Wang W. 2008. On the origin of new genes in *Drosophila*. *Genome Res* **18:** 1446–1455.