Genes relocated between *Drosophila* chromosomes evolve under selection pressures different from canonical dispersed duplicates

Margaret L. I. Hart[*†], Ban L. Vu[*‡], Quinten Bolden[*§], Keith T. Chen[*,**], Casey L. Oakes[*,††], Lejla Zoronjic[*,‡‡], and Richard P. Meisel[*,1]

∗Department of Biology and Biochemistry, University of Houston, TX 77204

†Current affiliation: Baylor College of Medicine, Houston, TX 77030

‡Current affiliation: College of Pharmacy, University of Houston, TX 77204

§Current affiliation: University of Texas Health Science Center at Houston School Of Nursing, Houston, TX 77030

∗∗Current affiliation: School of Graduate Studies, Rutgers University, New Brunswick, NJ 08901

††Current affiliation: Houston Department of Health and Human Services, Houston, TX 77054

‡‡Current affiliation: University of Texas School of Dentistry, Houston, TX 77054

Running title: *Drosophila* relocated genes

Keywords: gene relocation, gene duplication, gene expression, RNAi, selective constraints

1. Corresponding author:

Richard P. Meisel

3455 Cullen Blvd., Suite 342

Houston, TX

77204-5001

713-743-3607

rpmeisel@uh.edu

Abstract

Gene duplication creates a second copy of a gene either in tandem to the ancestral locus or dispersed to another chromosomal location. When the ancestral copy of a dispersed duplicate is lost from the genome, it creates the appearance that the gene was "relocated" from the ancestral locus to the derived location. Gene relocation may be as common as canonical dispersed duplications in which both the ancestral and derived copies are retained, but relocated genes have received far less attention than canonical duplicates. To test this hypothesis, we combined comparative genomics, population genetics, gene expression, and functional analyses to assess the selection pressures acting on relocated, duplicated, and non-relocated single-copy genes in *Drosophila* genomes. We find that relocated genes evolve faster than single-copy non-relocated genes, and there is no evidence that this faster evolution is driven by positive selection. In addition, relocated genes are less essential for viability and male fertility than single-copy non-relocated genes, suggesting that relocated genes evolve fast because of relaxed selective constraints. The derived copies of canonical dispersed duplicated genes evolve even faster than relocated genes, and there is no evidence that this faster evolution is driven by positive selection. The derived copies of dispersed duplicates are also more narrowly expressed than relocated genes, which suggests that the evolution of canonical duplicates is less selectively constrained than relocated genes. We therefore conclude that relocated genes are under more selective constraints than canonical duplicates, but are not as conserved as single-copy non-relocated genes.

Introduction

Duplicated genes are important contributors to molecular evolution (Ohno, 1970; Conant and Wolfe, 2008; Dittmar and Liberles, 2010; Innan and Kondrashov, 2010). A gene duplication event creates a second (derived) copy of a gene via one of many molecular mechanisms, including non-allelic recombination and reverse transcription of mRNA (Zhang, 2003; Kaessmann *et al.*, 2009; Marques-Bonet *et al.*, 2009). The derived copy can acquire novel functions and/or the ancestral and derived loci can each evolve a subset of functions present prior to duplication (Spofford, 1969; Hughes, 1994; Force *et al.*, 1999; Lynch and Force, 2000). When functions are partitioned between the paralogous copies, gene duplication can resolve pleiotropic conflicts present in the single-copy ancestor (Hittinger and Carroll, 2007; Des Marais and Rausher, 2008; Connallon and Clark, 2011; Gallach and Betrán, 2011; Abascal *et al.*, 2013).

Derived copies of duplicated genes in animal genomes are often testis-expressed (Vinckenbosch *et al.*, 2006; Meisel *et al.*, 2009, 2010; Baker *et al.*, 2012). It has therefore been hypothesized that gene duplication frequently resolves pleiotropic conflicts involving male-specific functions (Gallach and Betrán, 2011). Testis-expressed derived copies often harbor population genetic signatures of adaptive evolution, suggesting that positive selection fixes mutations that improve testis-specific functions once pleiotropic constraints are relaxed by duplication (Betrán and Long, 2003; Torgerson and Singh, 2004; Betrán *et al.*, 2006; Rosso *et al.*, 2008; Meisel *et al.*, 2010; Quezada-Diaz *et al.*, 2010; Tracy *et al.*, 2010).

Gene duplication can give rise to a derived copy located in tandem to the ancestral copy, non-adjacent on the same chromosome, or on a different chromosome. The ancestral copy of a dispersed duplicate can be lost from the genome, creating the appearance that the gene was "relocated" to the derived locus (**Fig 1**). Comparative genomic analyses in animals and plants have revealed that gene relocation occurs frequently, and relocated genes may be as common as canonical dispersed duplications (Bhutkar *et al.*, 2007; Meisel *et al.*, 2009;

Wicker *et al.*, 2010; Han and Hahn, 2012; Ciomborowska *et al.*, 2013). Furthermore, gene relocation can promote reproductive isolation between species because some F_2 hybrids lack the relocated gene (Masly *et al.*, 2006; Bikard *et al.*, 2009; Moyle *et al.*, 2010).

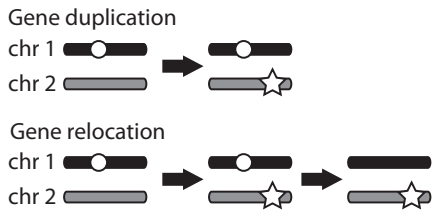


Figure 1: Gene duplication and relocation. In the ancestral arrangement, a gene (white circle) is located on chromosome 1. After gene duplication, the derived copy (star) is located on chromosome 2. In the case of gene relocation, the copy at the ancestral locus is subsequently lost.

Despite the prevalence and evolutionary importance of gene relocation, the selection pressures acting on relocated genes have received considerably less attention than the evolutionary dynamics of canonical duplicated genes. The analyses that have been performed identified some important differences between relocated genes and canonical dispersed duplicates. In contrast to the derived copies of dispersed duplicates, which are usually narrowly expressed in reproductive tissues, *Drosophila* and human relocated genes tend to be broadly expressed across many tissues (Meisel *et al.*, 2009; Ciomborowska *et al.*, 2013). In addition, mammalian relocated genes appear to evolve under strong purifying selection (Ciomborowska *et al.*, 2013), while the derived copies of dispersed duplicates tend to experience positive selection or relaxed constraints (Kondrashov *et al.*, 2002; Conant and Wagner, 2003; Han *et al.*, 2009; Han and Hahn, 2012). Furthermore, unlike duplicated genes, relocated genes probably do not resolve pleiotropic conflicts because a second copy is not retained. However, direct comparisons between relocated genes and dispersed duplicates are rarely performed to test the hypothesis that relocated genes evolve under different selection pressures than dispersed duplicates (Meisel *et al.*, 2009; Ciomborowska *et al.*, 2013). To address this shortcoming, we combined population genetic, functional genomic, and experimental approaches to test the hypothesis that *Drosophila* relocated genes are under greater pleiotropic constraints than

canonical dispersed duplicates, which leads to stronger purifying selection on relocated genes.

Materials and Methods

Identifying duplicated and relocated genes

Drosophila genomes have six chromosome arms, known as Muller elements A–F (Muller, 1940; Schaeffer *et al.*, 2008). We analyzed previously annotated inter-chromosome arm duplicated and relocated genes that occurred along the lineages leading to *Drosophila melanogaster* and *Drosophila pseudoobscura* (Hahn *et al.*, 2007; Meisel *et al.*, 2009), ignoring duplication and relocation events involving the minute element F. The lineage-specific dispersed duplicates were identified by examining phylogenetic reconstructions of gene families from the *D. melanogaster*, *D. pseudoobscura*, *Drosophila willistoni*, *Drosophila virilis*, and *Drosophila grimshawi* genomes. We selected paralogs in the *D. melanogaster* and *D. pseudoobscura* genomes that have common ancestors after the divergence of the lineages leading to the two species. The ancestral copy of a duplicated gene in one species' genome was inferred to be the copy found on the same chromosome arm as in the other four species, and the derived copy is the one on a different chromosome arm. Relocated genes are present in a single copy in *D. melanogaster* or *D. pseudoobscura*, with single-copy orthologs on a different chromosome arm in the other four species. As a control, we also analyzed single-copy non-relocated genes that are retained as 1:1:1:1:1 orthologs on the same chromosome arm across all five species (Meisel *et al.*, 2009). We excluded genes on element F from our control set.

Divergence and selection

We obtained estimates of polymorphism and divergence for relocated genes, non-relocated single-copy genes, and the ancestral and derived copies of dispersed duplicated genes in the *D. melanogaster* genome from published datasets. Divergence was calculated between *D. melanogaster* and *Drosophila simulans*. Importantly, all duplications happened before the divergence of the *D. melanogaster* and *D. simulans* lineages, so that our estimates of

divergence are specific to either the ancestral or derived copy. First, we obtained estimates of nucleotide sequence divergence along the *D. melanogaster* lineage after the split with *D. simulans* for all 1:1 orthologous genes between these two closely related species (Hu *et al.*, 2013). In the results presented here, we analyzed substitutions per site for 0-fold and 4-degenerate sites within protein coding regions. We also obtained estimates of the ratio of non-synonymous to synonymous substitutions per site (d_N/d_S) from published analyses comparing *D. melanogaster* and *D. simulans* (Stanley and Kulathinal, 2016).

We obtained the amount non-synonymous (P_N) and synonymous (P_S) polymorphic sites within *D. melanogaster* genes from the *Drosophila* Genetic Reference Panel (DGRP; Mackay *et al.*, 2012; Ràmia *et al.*, 2012). We only included polymorphic sites with a minor allele frequency >5% to minimize the inclusion of segregating deleterious alleles (Fay *et al.*, 2001). We also obtained the number of non-synonymous (D_N) and synonymous (D_S) substitutions between *D. melanogaster* and *D. simulans* from the DGRP data. We analyzed the polymorphism and divergence data for single-copy non-relocated genes, the ancestral and derived copies of dispersed duplicates, and relocated genes within the framework of McDonald and Kreitman (1991). First, we used a χ^2 test of independence to identify genes with an excess or deficiency of non-synonymous substitutions. We assigned genes as evolving under positive selection if they have a significant excess of non-synonymous substitutions, and we assigned genes as evolving under strong negative selection if they have a significant deficiency of non-synonymous substitutions. Second, we calculated $\overline{\alpha}$, the fraction of non-synonymous substitutions fixed by selection (Smith and Eyre-Walker, 2002), for each group of genes:

$$\overline{\alpha} = 1 - \frac{\overline{D_S}}{\overline{D_N}} \overline{\left(\frac{P_N}{P_S} \right)}. \tag{1}$$

We calculated $\overline{\alpha}$ separately for single-copy non-relocated genes, ancestral copies of dispersed duplicates, derived copies, and relocated genes. We performed 1,000 bootstrapped replicate analyses to calculate a confidence interval (CI) for each $\overline{\alpha}$ estimate.

Gene expression profiles

We analyzed available microarray data to assess the expression across adult tissues of *D. melanogaster* relocated genes, duplicated genes, and single-copy genes. Expression measurements were taken from FlyAtlas, which includes 11 non-redundant adult non-sex-specific tissue samples (brain, crop, midgut, hindgut, Malpighian tubule, thoracicoabdominal ganglion, salivary gland, fat body, eye, heart, and trachea), two male-specific organs (testis and accessory gland), and two female-specific organs (ovary and spermatheca) (Chintapalli *et al.*, 2007). Expression levels for spermatheca were averaged between mated and unmated females (Meisel, 2009). We used τ as a measure of expression breadth for each gene:

$$\tau = \frac{\sum_{i=1}^{N} 1 - \frac{log_{10} S_i}{log_{10} S_{max}}}{N-1}, \tag{2}$$

where N is the number of tissues (15), S_i is the expression level in tissue i, and S_{max} is the maximum expression of that gene across all tissues (Yanai *et al.*, 2005; Larracuente *et al.*, 2008; Meisel *et al.*, 2009). All S_i<1 were set to 1 for this analysis. Values of τ range from 0 to 1, with higher values corresponding to more tissue-specific expression. We also analyzed microarray data from *D. melanogaster* testis and RNA-seq data from *D. pseudoobscura* testis to infer the expression levels of relocated, duplicated, and single-copy genes (Chintapalli *et al.*, 2007; Meisel *et al.*, 2010). Finally, we analyzed sex-specific microarray data from *D. melanogaster* and *D. pseudoobscura* heads and whole flies to calculate the relative expression of genes in males and females ($\log_2 \frac{M}{F}$), or "sex-biased" expression (Meisel *et al.*, 2012).

Viability and fertility effects of knockdown

To assess if relocated and non-relocated single-copy *D. melanogaster* genes are essential for viability and male fertility, we used Gal4-UAS inducible RNA interference (RNAi) to knockdown the expression of relocated and non-relocated genes. Flies carrying an inducible

construct containing a hairpin sequence that silences the expression of a target gene via RNAi (UAS-RNAi) were obtained from the Vienna Drosophila Resource Center (VDRC; Dietzl *et al.*, 2007). Knockdown was performed using two different sets of RNAi lines. The first set, known as "GD" lines, were produced by random integration into the *D. melanogaster* genome of a P-element construct carrying a pUAST vector with 10 copies of the UAS and a 300–400bp inverted repeat targeting the gene of interest. The second set, known as "KK" lines, also carry 10 copies of UAS and a long inverted repeat, but they were inserted into specific sites in the genome using ϕC31 targeted integration (Groth *et al.*, 2004; Bateman *et al.*, 2006). Expression of the RNAi construct in some of the KK lines can be lethal because of mis-expression of the developmental gene *tiptop* (Green *et al.*, 2014; Vissers *et al.*, 2016). We therefore performed analyses of both all lines and only GD lines to confirm that our results are not driven by misregulation of *tiptop*.

To assess the effect of knockdown on viability, individual UAS-RNAi males were crossed to individual females carrying a Gal4 driver construct that is ubiquitously expressed under the *tubulin 1*α promoter ($P\{$*tubP-Gal4*$\}$). $P\{$*tubP-Gal4*$\}$ is expressed in many tissues and throughout development (Lee and Luo, 1999), which causes constitutive knockdown of the target gene when combined in the same genotype with the RNAi construct. In addition, $P\{$*tubP-Gal4*$\}$ is balanced over the TM3 chromosome, which carries the dominant *Stubble* (*Sb*) allele, allowing us to differentiate between knockdown and non-knockdown (control) siblings within each cross. We assessed the viability of the knockdown flies by comparing the counts of knockdown progeny with their control siblings. We also performed control crosses in which the UAS-RNAi male is replaced with a male from the progenitor stock from which the RNAi lines were derived—GD lines were created by transforming w^{1118} flies (VDRC line 60000), and KK lines were created by transforming $y,w^{1118};P\{\text{attP},y^{+},w^{3'}\}$ flies (VDRC line 60100). These control males do not carry a UAS-RNAi construct.

We used linear models to assess the effect of RNAi knockdown on viability. For each RNAi line, we modeled the number of progeny recovered (N_{ij}) with the phenotype associated with

either knockdown (wild-type bristles) or non-knockdown (*Sb*) from crosses involving a fly either carrying the UAS-RNAi construct or from the progenitor non-RNAi strain:

$$N_{ij} \sim L_i + P_j + L_i \times P_j + b_k, \tag{3}$$

where L_i is a fixed effect indicating the line used in the cross (either carrying the UAS-RNAi construct or control), P_j is a fixed effect indicating the phenotype of the progeny (either knockdown or *Sb* control), and b_k is a random effect indicating the replicate block in which the viability assay was performed. The effect of knockdown on viability was estimated as the interaction between line and progeny phenotype ($L_i \times P_j$), when the line carries the UAS-RNAi construct and the phenotype is wild-type bristles (knockdown). If the interaction between line and phenotype ($L_i \times P_j$) has a significant effect on the number of progeny recovered from the cross (N_{ij}), then there is an effect of knockdown on viability. To test for significance of the interaction term, we used a drop in deviance test to compare the fit of the full model with a model excluding the interaction term.

To assess the effects of RNAi knockdown on male fertility, we crossed UAS-RNAi males with females carrying a Gal4 driver construct that is constitutively expressed under the *bag of marbles* (*bam*) promoter (*P{bam-Gal4-VP16}*) to create male progeny in which the target gene is knocked down the germline (Sartain *et al.*, 2011). The *bam* promoter drives expression in germ cells after differentiation from the stem cells (Chen and McKearin, 2003). We assessed the fertility of the knockdown male progeny by crossing individual males to single Orergon R (OreR) or Canton S (CanS) virgin females and counting the number of adult progeny that emerge from pupae. We also assessed the fertility of control males that were created by crossing *bam-Gal4* females with males from the progenitor strains that do not carry the UAS-RNAi constructs. We only considered the results of matings in which we observed copulation between the male and the CanS/OreR female to ensure that fertility was not confounded by behavioral effects that interfere with mating success.

For each RNAi line, we modeled the number of progeny recovered (N_{ij}) from matings involving either a male carrying the UAS-RNAi construct or a control male carrying a chromosome from the progenitor line:

$$N_{ij} \sim L_i + T_j + b_k, \tag{4}$$

where L_i is a fixed effect indicating the male genotype used in the mating (either carrying the UAS-RNAi construct or control), T_j is a fixed effect indicating the genotype of the female used to assess fertility (either CanS or OreR), and b_k is a random effect indicating the replicate block in which the fertility assay was performed. The effect of knockdown on fertility is the male genotype term (L_i), when the male carries the UAS-RNAi construct. If male genotype (L_i) has a significant effect on the number of progeny (N_{ij}), then there is an effect of germline knockdown on male fertility. To test for a significant effect of male genotype, we used a drop in deviance test to compare the fit of the full model with a model excluding the male genotype (L_i). For some of the genes, only one female genotype was used for the assay in a single block, and we therefore could not use the drop in deviance to assess the effect of knockdown on fertility. In those cases, we assessed the effect of the male genotype using a single factor ANOVA (equivalent to a Student's T-test): $N_i \sim L_i$. All analyses were performed in the R statistical programming environment (R Core Team, 2015).

Data availability

All divergence data, gene expression data, and results from RNAi experiments are available as supplemental files. File S1 contains a description of all supplemental data.

Results

Dispersed duplicates and relocated genes evolve fast because of relaxed selective constraints

We tested if the protein coding sequences of genes that were duplicated or relocated to other chromosome arms along the *D. melanogaster* lineage evolve at different rates than single-copy non-relocated genes. The ancestral copies of dispersed duplicated genes, derived copies, and relocated genes all evolve faster at 0-fold degenerate (amino acid changing) sites than single-copy non-relocated genes (**Fig 2A**). Accelerated amino acid sequence evolution can be driven by positive selection, relaxed constraints, or higher mutation rates. The derived copies of duplicated genes evolve faster than single-copy genes at 4-fold degenerate (silent) sites (**Fig 2B**), suggesting that higher mutation rates could explain the faster evolution of derived copies at 0-fold degenerate sites. However, d_N/d_S is significantly elevated in the ancestral copies, derived copies, and relocated genes relative to single-copy non-relocated genes (**Fig 2C**). We therefore conclude that mutational bias cannot entirely explain the faster amino acid sequence evolution of duplicated and relocated genes.

Other analyses have found that the derived copies of duplicated genes evolve faster and experience more positive selection than the ancestral copies (Kondrashov *et al.*, 2002; Conant and Wagner, 2003; Han *et al.*, 2009; Han and Hahn, 2012). We fail to detect significant differences between ancestral and derived copies of dispersed duplicated genes in divergence at 0-fold degenerate sites ($P = 0.373$), divergence at 4-fold degenerate sites ($P = 0.553$), or d_N/d_S ($P = 0.208$; all P values from Mann-Whitney U tests). We have small sample sizes of ancestral and derived duplicates (14–30 depending on the divergence estimate), which likely limits our power to detect significant differences in evolutionary rates. There are substantially more relocated genes with divergence estimates (34–60), and we detect significantly elevated

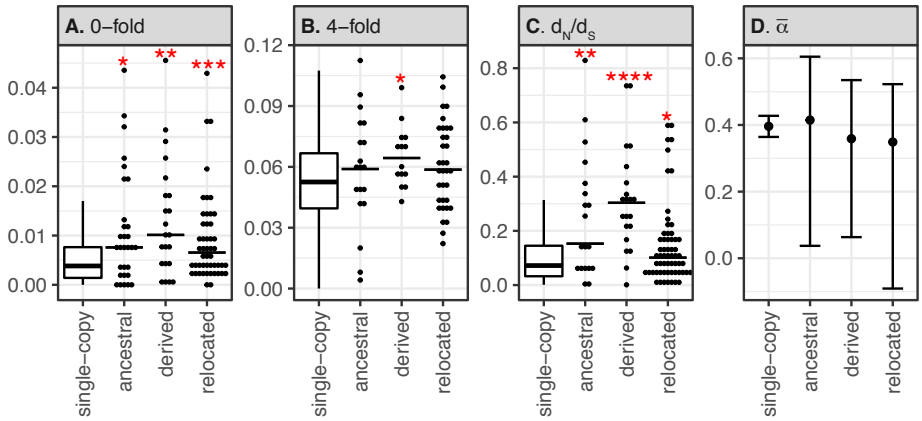


Figure 2: Divergence and polymorphism-divergence statistics for single-copy non-relocated genes, the ancestral copies of duplicated genes, the derived copies of duplicated genes, and relocated genes are plotted. Divergence estimates are between *D. melanogaster* and *D. simulans* at (A) 0-fold degenerate sites, (B) 4-fold degenerate sites, and (C) d_N/d_S. The distribution of divergence values for single-copy genes is represented by a boxplot, while individual divergence values are shown for each of the other genes as a point (with the median indicated by a horizontal line). Significant differences in divergence when comparing single-copy genes with either ancestral copies, derived copies, or relocated genes are shown by red asterisks (*$P < 0.05$, **$P < 0.005$, ***$P < 0.0005$, and ****$P < 0.00005$ in a Mann-Whitney U test). (D) Point estimates of $\overline{\alpha}$ are plotted along with the 95% CI.

d_N/d_S in the derived copies of duplicated genes relative to relocated genes ($P = 7.5 \times 10^{-5}$ in a Mann-Whitney U test). These results demonstrate that the protein coding sequences of relocated genes, similar to both the ancestral and derived copies of dispersed duplicated genes, evolve faster than single-copy non-relocated genes. However, the derived copies of duplicated genes evolve faster than relocated genes.

	no selection	positive selection	negative selection
single-copy	3459	539	161
ancestral	19	1	0
derived	18	2	1
relocated	28	4	1

Table 1: Counts of *D. melanogaster* single-copy non-relocated genes, ancestral copies of duplicated genes, derived copies, and relocated genes with no evidence for strong selection, evidence for positive selection, and evidence for negative selection.

To distinguish between relaxed selective constraints (decreased purifying selection) and increased adaptive substitutions (positive selection) driving the rapid evolution of *D. melanogaster* duplicated and relocated genes, we analyzed polymorphism and divergence data. If accelerated evolutionary divergence is driven by positive selection, we expect the ratio of non-synonymous to synonymous substitutions to be greater than non-synonymous to synonymous polymorphisms (McDonald and Kreitman, 1991). Only a handful of duplicated and relocated genes have an excess of non-synonymous substitutions (**Table 1**). In addition, the proportion of ancestral copies, derived copies, or relocated genes with evidence for positive selection is not greater than the proportion of single-copy non-relocated genes with evidence for positive selection (**Table 1**). Furthermore, the fraction of non-synonymous substitutions fixed by selection ($\overline{\alpha}$) for ancestral copies, derived copies, and relocated genes all fall within or below the 95% CI for single-copy non-relocated genes (**Fig 2D**). There is also no evidence for a difference in $\overline{\alpha}$ between derived copies of duplicated genes and either ancestral copies or relocated genes (**Fig 2D**). In summary, our results provide no evidence that dispersed duplicated genes (either ancestral or derived copies) or relocated genes experience disproportionate positive selection relative to single-copy non-relocated genes. We therefore conclude that the accelerated evolution of dispersed duplicates and relocated genes is driven by relaxed selective constraints.

Relocated genes are not narrowly expressed but are highly expressed in testis and have male-biased expression

The derived copies of *D. melanogaster* dispersed duplicated genes tend to be narrowly expressed in male reproductive tissues (Meisel *et al.*, 2009). We used microarray measurements from 15 adult *D. melanogaster* tissues to compare expression breadth (τ) for single-copy non-relocated genes, ancestral copies of dispersed duplicates, derived copies, and relocated genes (Yanai *et al.*, 2005; Larracuente *et al.*, 2008; Meisel *et al.*, 2009). Broadly expressed genes have low values of τ (close to 0), whereas narrowly expressed genes have high values of τ (close

to 1). We confirmed that the derived copies of *D. melanogaster* duplicated genes are more narrowly expressed than single-copy genes (**Fig 3A**). In addition, we found that relocated genes do not significantly differ in their expression breadth from single-copy non-relocated genes (**Fig 3A**) or the ancestral copies of duplicated genes ($P = 0.900$ in a Mann-Whitney U test).

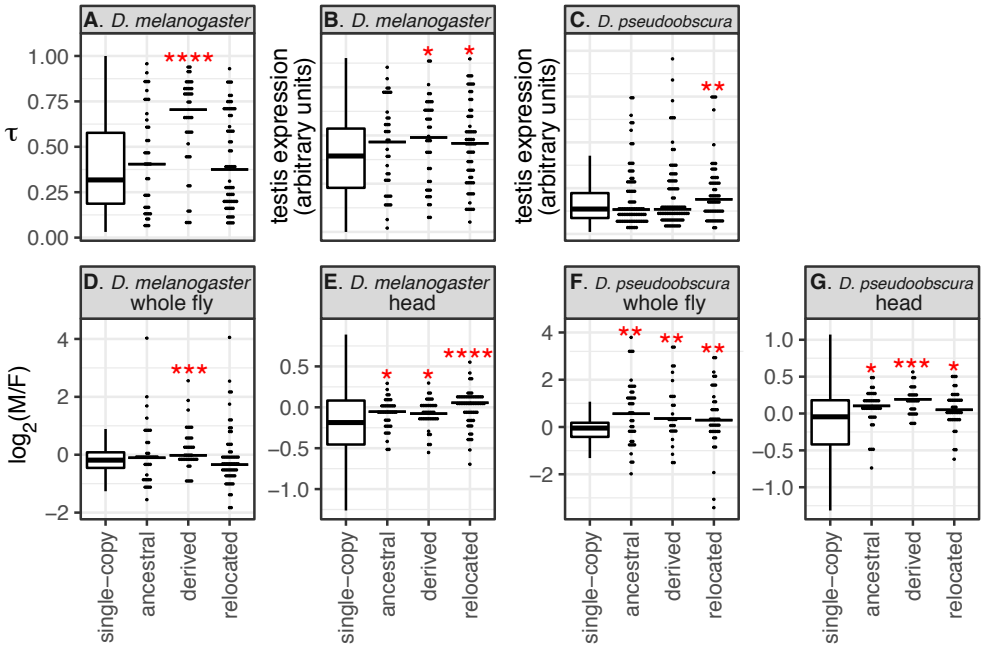


Figure 3: Expression of single-copy non-relocated genes, the ancestral copies of dispersed duplicated genes, the derived copies of dispersed duplicated genes, and relocated genes are plotted. (A) Distributions of τ for *D. melanogaster* genes are plotted. Distributions of testis expression for (B) *D. melanogaster* and (C) *D. pseudoobscura* genes are plotted. Distributions of $\log_2 \frac{M}{F}$ in (D) *D. melanogaster* whole fly, (E) *D. melanogaster* head, (F) *D. pseudoobscura* whole fly, and (G) *D. pseudoobscura* head are plotted. The distribution of $\log_2 \frac{M}{F}$ for single-copy genes is represented by a boxplot, while individual values are shown for each of the other genes as a point (with the median indicated by a horizontal line). Significant differences in τ when comparing single-copy genes with either ancestral copies, derived copies, or relocated genes are shown by red asterisks (*$P < 0.05$, **$P < 0.005$, ***$P < 0.0005$, and ****$P < 0.00005$ in a Mann-Whitney U test).

The derived copies of duplicated genes in animal genomes are often testis-expressed (Vinckenbosch *et al.*, 2006; Meisel *et al.*, 2009, 2010; Baker *et al.*, 2012). We indeed find that the derived copies of *D. melanogaster* duplicates in our dataset are more highly expressed in testis than single-copy non-relocated genes (**Fig 3B**) and the ancestral copies

of duplicated genes ($P = 2.8 \times 10^{-3}$ in a Mann-Whitney U test). In addition, *D. melanogaster* relocated genes are also more highly expressed in testis than non-relocated genes (**Fig 3B**). However, the derived copies of *D. pseudoobscura* duplicated genes are not more highly expressed in testis than either the ancestral copies or single-copy genes (**Fig 3C**). *D. pseudoobscura* relocated genes, on the other hand, are more highly expressed in testis than non-relocated single-copy genes (**Fig 3C**). We therefore conclude that *Drosophila* relocated genes are highly expressed in testis, but the testis-expression of the derived copies of duplicated genes is species-dependent.

Testis expression is the primary driver of male-biased gene expression in *Drosophila* (Parisi *et al.*, 2003). In addition, male-biased and testis expression are among the best predictors of evolutionary rates of protein coding genes (Meisel, 2011). Because duplicated and relocated genes evolve fast (**Fig 2A–C**) and are testis expressed (**Fig 3B–C**), we assessed whether duplicated and relocated genes also have male-biased expression. The derived copies of *D. melanogaster* and *D. pseudoobscura* duplicated genes do indeed have more male-biased expression (increased $\log_2 \frac{M}{F}$) relative to single-copy non-relocated genes when expression is measured in either whole fly or head (**Fig 3D–G**). In addition, *D. melanogaster* relocated genes and the ancestral copies of *D. melanogaster* duplicates have more male-biased expression than single-copy non-relocated genes in head, but not in whole fly (**Fig 3D–E**). *D. pseudoobscura* relocated genes and the ancestral copies of *D. pseudoobscura* duplicates also have more male-biased expression than single-copy non-relocated genes in both whole fly and head (**Fig 3F–G**). *Drosophila* relocated genes therefore are broadly expressed across many tissues, are highly expressed in male-limited tissues, and have elevated expression relative to females in non-sex-limited tissues (**Fig 3**). However, unlike the derived copies of duplicated genes, relocated genes do not have limited expression in male-specific tissues.

To assess how the expression profiles of relocated genes affect their rates of evolution, we calculated Spearman's non-parametric rank order correlation (ρ) between each of our divergence estimates and expression metrics for *D. melanogaster* single-copy genes, ancestral

copies of dispersed duplicates, derived copies, and relocated genes. Consistent with previous results (Meisel, 2011), faster evolution of single-copy non-relocated genes is associated with more male-biased expression in whole fly (higher $\log_2 \frac{M}{F}$), narrower expression (greater τ), and higher testis expression (**Supplementary Fig S1**). Faster evolution of relocated genes is also positively correlated with narrower expression (**Supplementary Fig S1**), even though relocated genes are not narrowly expressed (**Fig 3A**). In contrast, testis expression levels are not positively correlated with evolutionary rate for relocated genes (**Supplementary Fig S1**), even though relocated genes evolve fast (**Fig 2A-C**) and are highly expressed in testis (**Fig 3B**). We observe similar results for the derived copies of dispersed duplicated genes (**Supplementary Fig S1**). We therefore conclude that higher testis expression could explain the faster evolution of relocated genes and dispersed duplicates when compared to non-relocated genes, but expression breadth is the best predictor of evolutionary rates within relocated genes and dispersed duplicates.

Relocated genes are not disproportionately essential for viability

The broad expression of relocated genes suggests that they may be essential for viability. To test this hypothesis, we compared the effects of RNAi knockdown of relocated genes to knockdown of single-copy non-relocated genes in *D. melanogaster*. Within both relocated and non-relocated genes, there is not a significant correlation between τ and the effect of knockdown on viability (**Supplementary Fig S2**). This suggests expression breadth is not a reliable proxy for gene essentiality, which could explain why relocated genes evolve fast because of relaxed constraints even though they are broadly expressed.

We find some evidence that relocated genes are less essential than non-relocated genes. First, there is not a significant difference in the fraction of RNAi lines that decrease viability between relocated and non-relocated genes ($P = 0.83$ in Fisher's exact test): approximately half (28/55) of the RNAi lines targeting relocated genes induced a decrease in viability (**Fig 4A**), and a similar fraction (18/33) of RNAi lines targeting non-relocated single-copy

genes decreased viability (**Fig 4B**). In addition, we used 1–3 RNAi lines per each gene, and we considered a gene to have a viability effect if at least one RNAi line targeting the gene induced a decrease in viability. Two-thirds of relocated genes have a viability effect (24/36), which is approximately the same fraction (12/19) as single-copy non-relocated genes ($P = 1$ in Fisher's exact test). However, the knockdown effect on viability is greater for RNAi lines targeting non-relocated genes ($P = 0.027$ in a Mann-Whitney U test).

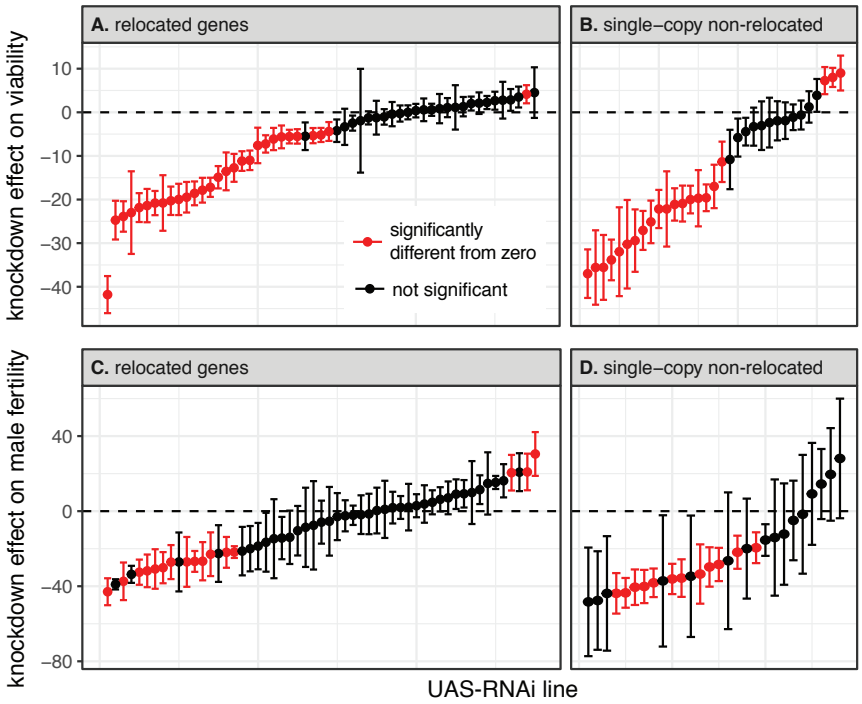


Figure 4: The effects of RNAi knockdown on viability and fertility are plotted. Knockdown was performed using (A–B) ubiquitous expression of Gal4 to assess viability and (C–D) germline expression of Gal4 to assess male fertility. RNAi targeted (A & C) relocated genes or (B & D) single-copy non-relocated genes. Dots indicate the mean effect of knockdown across replicates, and the vertical bars show the standard error. Each point is a RNAi line, and those colored red are significantly different from zero.

We did not perform RNAi knockdown experiments on duplicated genes, but another experiment assessed if ubiquitous knockdown of the derived copies of *D. melanogaster* duplicated genes induces lethality (Chen *et al.*, 2010). We compared our analysis of relocated and non-relocated genes to the inter-chromosome-arm duplications from the published RNAi experiment (**Table 2**). We considered knockdown of relocated and non-relocated genes to

induce lethality if there were no knockdown progeny recovered in at least 90% of replicate experiments we performed with at least one RNAi line. This is a similar criterion as employed by Chen *et al.* (2010) for derived copies of duplicated genes. There is not a significant difference in the frequency of RNAi-induced lethality between relocated and non-relocated genes ($P = 1$ in Fisher's exact test), nor is there a difference between derived copies of dispersed duplicates and single-copy non-relocated genes ($P = 1$ in Fisher's exact test). Our results therefore suggest that relocated genes, the derived copies of duplicated genes, and single-copy non-relocated genes are all equally likely to be essential for viability (**Table 2**). However, ubiquitous knockdown of non-relocated genes has a significantly greater effect on viability than knockdown of relocated genes (**Fig 4A–B**).

	lethal	non-lethal	perc lethal
Relocated	13	23	36.1%
Derived dups	21	39	35.0%
Single-copy	7	12	36.8%

Table 2: Counts of relocated genes, derived copies of dispersed duplicates, and single-copy non-relocated genes that are lethal or non-lethal to knockdown. The criterion for lethality is no knockdown progeny recovered in 90% of replicate experiments.

The genes included in the viability assays are a subset of all single-copy non-relocated, duplicated, and relocated genes in the *D. melanogaster* genome. We tested if they are representative of the patterns of divergence, selection, and expression we observe in the full set of single-copy, duplicated, and relocated genes. We confirmed that the derived copies of duplicated genes that were included in the viability assays do indeed evolve faster than the single-copy non-relocated genes included in the viability assays, and there is no evidence for more positive selection on duplicated or relocated genes included in the viability assays (**Supplementary Fig S3**). In addition, the derived copies of duplicated genes in our viability assays are narrowly expressed, and they have more male-biased expression (greater $\log_2 \frac{M}{F}$) in whole fly than single-copy genes (**Supplementary Fig S3**). In contrast, derived

copies of duplicated genes included in the viability assay are expressed at a lower level in testis than the single-copy genes (**Supplementary Fig S3**), which is the opposite pattern observed in the full dataset of duplicated and single-copy genes (**Fig 2**). However, testis expression is not likely to be an important factor in viability assays. Because the evolutionary patterns and most expression profiles of the genes included in the viability assays are similar to the complete set of genes, we conclude that the genes included in the viability assays are representative of single-copy, duplicated, and relocated genes in the *D. melanogaster* genome.

We and others used two types of RNAi lines in the knockdown experiments described above (GD and KK). The KK lines can have developmental effects independent of RNAi knockdown of the target gene (Green *et al.*, 2014; Vissers *et al.*, 2016). We therefore repeated our analysis considering only GD lines (**Supplementary Fig S4**). RNAi knockdown using the GD lines decreased viability for a smaller proportion of lines targeting relocated genes (8/30) than for the KK lines (20/25; $P = 10^{-4}$ in Fisher's exact test). This is consistent with the KK lines having lethality effects independent of knockdown of the target genes. Over half of the GD lines targeting single-copy non-relocated genes decreased viability (10/19), which is marginally greater than the fraction of GD lines targeting relocated genes ($P = 0.078$ in Fisher's exact test). There also was a significantly greater effect on viability of knocking down non-relocated genes with GD lines compared to relocated genes ($P = 9 \times 10^{-5}$ in a Mann-Whitney U test). In addition, knockdown of a larger frequency non-relocated genes caused a significant decrease in viability (9/15) than knockdown of relocated genes (6/24) when using at least one GD line ($P = 0.044$ in Fisher's exact test). However, there is not a significant difference in the frequency of RNAi-induced lethality between relocated and non-relocated genes when using only GD lines ($P = 0.69$ in Fisher's exact test; **Supplementary Table S1**). There is also not a significant difference in lethality using GD lines between derived copies of dispersed duplicates and single-copy non-relocated genes ($P = 0.43$ in Fisher's exact test; **Supplementary Table S1**). These results therefore demonstrate that

knockdown of relocated genes has less of an effect on viability than knockdown of single-copy non-relocated genes, but there is not a difference in the frequency of essential relocated genes, derived duplicates, and single-copy non-relocated genes.

Relocated genes are not disproportionately essential for male fertility

Relocated genes are highly expressed in testis (**Fig 3B–C**), suggesting that their products may perform essential roles in spermatogenesis. To test this hypothesis, we assessed the fertility of *D. melanogaster* males in which relocated genes were knocked down in the male germline using RNAi. We compared our results to germline knockdown of single-copy non-relocated genes. The decrease in male fertility induced by germline knockdown is positively correlated with testis expression level for both relocated and non-relocated genes (**Supplementary Fig S5**). Therefore, testis expression is a good proxy for the degree to which a gene is required for male fertility. No other expression measures are correlated with germline knockdown effects (**Supplementary Fig S5**).

Surprisingly, despite their higher testis expression, we find evidence that relocated genes are less essential for male fertility than non-relocated genes. Germline knockdown induced a significant decrease in male fertility in less than a quarter (13/55) of RNAi lines targeting relocated genes (**Fig 4C**). In contrast, a larger fraction (12/28) of RNAi lines targeting non-relocated single-copy genes induced a decrease in male fertility ($P = 0.082$ in Fisher's exact test). In addition, the knockdown effect on male fertility is greater for RNAi lines targeting non-relocated genes than relocated genes ($P = 0.0011$ in a Mann-Whitney U test). We considered a gene to have a fertility effect if at least one RNAi line targeting the gene induced a decrease in male fertility. Less than a third of relocated genes have a fertility effect (9/32), and 9/17 single-copy non-relocated genes have a fertility effect ($P = 0.122$ in Fisher's exact test).

We additionally tested if the genes included in the fertility assay are a representative subset of all non-relocated and relocated genes in the *D. melanogaster* genome. The re-

located genes in our fertility assay evolve faster at 4-fold degenerate sites than the non-relocated genes, and the trend is in that direction at 0-fold degenerate sites and d_N/d_S (**Supplementary Fig S6**). However, the relocated genes in our fertility assay are not more highly expressed in testis than the non-relocated genes (**Supplementary Fig S6**). On the other hand, the relocated genes in our experiments do have more male-biased expression (higher $\log_2 \frac{M}{F}$) than the non-relocated genes (**Supplementary Fig S6**), suggesting that the relocated genes are up-regulated in male relative to female reproductive tissues (Parisi *et al.*, 2003). It is therefore possible that the reduced knockdown effect on male-fertility that we observe amongst relocated genes could be a result of selecting relocated genes with lower testis expression or non-relocated genes with higher testis expression.

We also repeated our analysis of the effect of RNAi knockdown on male fertility considering only GD lines (**Supplementary Fig S7**). We found significant decreases in male fertility in 9/34 GD lines targeting relocated genes, which is not significantly different from the fraction of KK lines (4/21) that significantly decrease male fertility ($P = 0.75$ in Fisher's exact test). This suggests KK lines do not have RNAi-independent effects on male germline function. There is not a significant difference in the fraction of GD lines that decrease male fertility targeting relocated genes (9/34) compared to those targeting non-relocated genes (8/16 ; $P = 0.121$ in Fisher's exact test). However, using GD lines, there is a significantly greater effect on male fertility when non-relocated genes are knocked down compared to relocated genes ($P = 2.6 \times 10^{-4}$ in a Mann-Whitney U test). In contrast, there is not a significant difference in the fraction of relocated genes with a fertility effect (7/24) than non-relocated single-copy genes (7/13) when using only GD lines ($P = 0.171$ in Fisher's exact test). These results suggest that, despite the elevated testis expression of relocated genes, knockdown of non-relocated genes has a greater effect on male fertility than silencing relocated genes. Relocated genes are also not more likely to be essential for male fertility than single-copy non-relocated genes. Unfortunately, there are not available data on the effect of germline knockdown of duplicated genes for comparison.

Discussion

Relocated genes are as common as canonical inter-chromosome duplicated genes in eukaryotic genomes, and relocated genes were thought to evolve under different selection pressures than the derived copies of dispersed duplicated genes (Bhutkar *et al.*, 2007; Meisel *et al.*, 2009; Wicker *et al.*, 2010; Han and Hahn, 2012; Ciomborowska *et al.*, 2013). To test this hypothesis, we compared the evolutionary rates of relocated genes with the ancestral copies of dispersed duplicates, the derived copies, and single-copy non-relocated genes. We found that relocated genes and derived copies of dispersed duplicates both evolve faster than non-relocated genes, but there is no evidence that this faster evolution is driven by positive selection (**Fig 2** & **Table 1**). In addition, the derived copies of dispersed duplicates evolve even faster than relocated genes. We therefore conclude that relocated genes evolve fast because they are under relaxed constraints, and the derived copies of dispersed duplicates evolve even faster because they have fewer selective constraints than relocated genes. This is consistent with relocated genes and dispersed duplicates experiencing different selection pressures.

Relocated genes are broadly expressed, while the derived copies of dispersed duplicates are narrowly expressed (**Fig 3**). Broad expression and high expression levels are associated with slower evolution of *Drosophila* genes (Larracuente *et al.*, 2008; Meisel, 2011). It is therefore not surprising that the derived copies of dispersed duplicates evolve faster than relocated genes (**Fig 2**). Our results suggest that if expression breadth/level are directly associated with evolutionary rate, it is because lower and narrower expression allow for more relaxed constraints (**Table 1**). The broad expression and high testis expression of relocated genes also suggest that they may be essential for viability and male fertility (**Fig 3**). However, our RNAi experiments revealed that the effect of knocking down relocated genes on both viability and male fertility is less than the effect of knocking down single-copy non-relocated genes (**Fig 4**). This suggests that relocated genes are less essential than single-

copy non-relocated genes. This is consistent with our population genetic data that suggest relocated genes evolve faster than single-copy non-relocated genes because relocated genes are under relaxed selective constraints (**Fig 2** & **Table 1**). These results also demonstrate that functional analyses that complement expression measurements are necessary to identify subtle differences in selective constraints acting on different classes of genes.

An excess of genes has been relocated from the X chromosome to the autosomes across the *Drosophila* genus (Meisel *et al.*, 2009; Vibranovski *et al.*, 2009b). Three hypotheses could explain this phenomenon. First, the female-biased transmission of the X chromosome may favor X-linked female-beneficial mutations and prevent the fixation of male-beneficial mutations on the X (Rice, 1984). This sexually antagonistic selection could favor the X-to-autosome relocation of genes that perform male-beneficial functions (Wu and Xu, 2003). Second, expression of the X chromosome is down-regulated in spermatogenesis (Vibranovski *et al.*, 2009a; Meiklejohn *et al.*, 2011), which could favor the X-to-autosome relocation of genes that have beneficial effects when highly expressed in spermatogenesis (Betrán *et al.*, 2002; Emerson *et al.*, 2004; Meisel *et al.*, 2009). Third, there may be a mutational bias in favor of X-to-autosome duplications (Metta and Schlotterer, 2010; Díaz-Castillo and Ranz, 2012), but this is not supported by copy number polymorphisms (Schrider *et al.*, 2011). We find no evidence that knockdown of relocated genes disproportionately affects male-specific fitness—knockdown of relocated genes does not affect male-specific viability (results not shown) nor does knockdown of relocated genes disproportionately affect male fertility (**Fig 4**). However, relocated genes are more highly expressed in testis than non-relocated single-copy genes (**Fig 3**), which holds even if we only consider autosomal genes ($P < 0.05$ for both *D. melanogaster* and *D. pseudoobscura*). We therefore conclude that our results support the hypothesis that the X-to-autosome relocation bias is driven by selection in favor of higher testis expression on the autosomes.

There are some technical limitations of our experiments that may limit our ability to detect male-specific functions of relocated genes. We used ubiquitous and germline knockdown

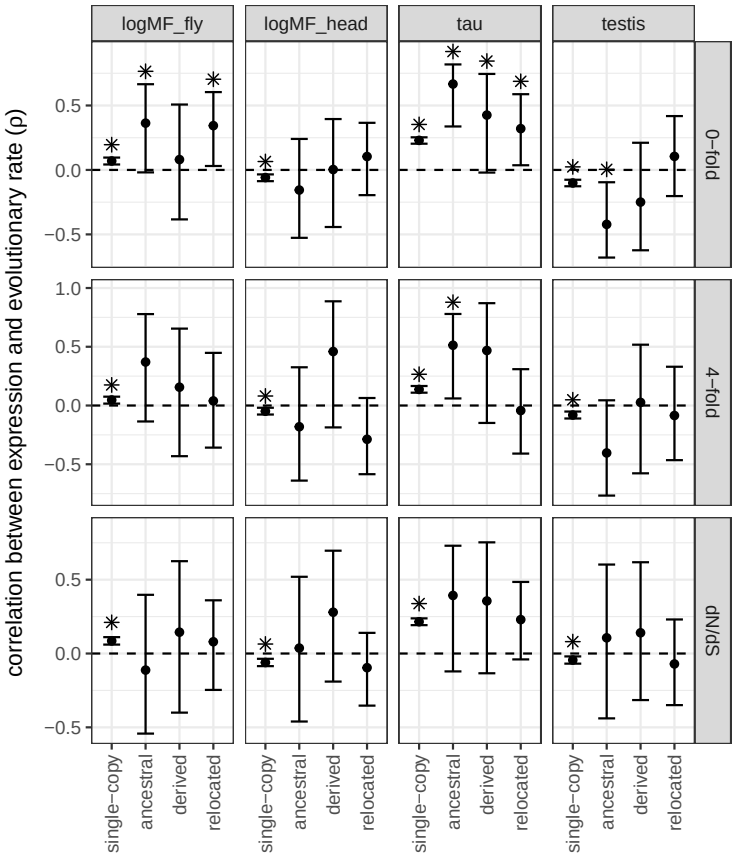
to assess if relocated and single-copy non-relocated genes are essential for viability and male fertility. We chose to assay the effect of germline knockdown because relocated genes are highly expressed in testis (**Fig 3**), and the derived copies of duplicated genes are hypothesized to be specialized for germline functions (Marques *et al.*, 2005; Vinckenbosch *et al.*, 2006; Potrzebowski *et al.*, 2008; Meisel *et al.*, 2010; Tracy *et al.*, 2010). We demonstrated that, even though they are highly expressed in testis, relocated genes are not disproportionately essential for spermatogenesis (**Fig 4**). However, our results may be biased by the Gal4 driver that we selected (*bam*), which is expressed early in spermatogenesis immediately after differentiation from the stem cell niche (Chen and McKearin, 2003). Knockdown in later stages of spermatogenesis or in somatic testis tissue may reveal testis-biased functions for relocated genes. In addition, germline knockdown of a panel of duplicated genes would allow for a comparison of male-specific functions between relocated genes and derived copies of dispersed duplicates. This may reveal additional insights into the causes of differences in the evolutionary rates of duplicated and relocated genes (**Fig 2**).

In conclusion, we demonstrated that *Drosophila* relocated genes and the derived copies of dispersed duplicates both evolve fast, and this rapid evolution is likely the result of relaxed selective constraints (**Fig 2** & **Table 1**). This differs from mammals, where relocated genes are under stronger purifying selection than other genes in the genome (Ciomborowska *et al.*, 2013). *Drosophila* relocated genes are also less essential for viability and male fertility than single-copy non-relocated genes (**Fig 4**), which is consistent with relocated genes evolving under relaxed constraints. In addition, the derived copies of dispersed duplicates appear to be under even more relaxed constraints than relocated genes, which allows them to evolve even faster. The narrow expression profiles of derived copies of *Drosophila* duplicated genes are also consistent with relaxed selective constraints (**Fig 3**), which agrees with previous work in humans (Ciomborowska *et al.*, 2013). However, additional work is necessary to determine the causes of taxon-specific selection pressures acting on relocated genes.

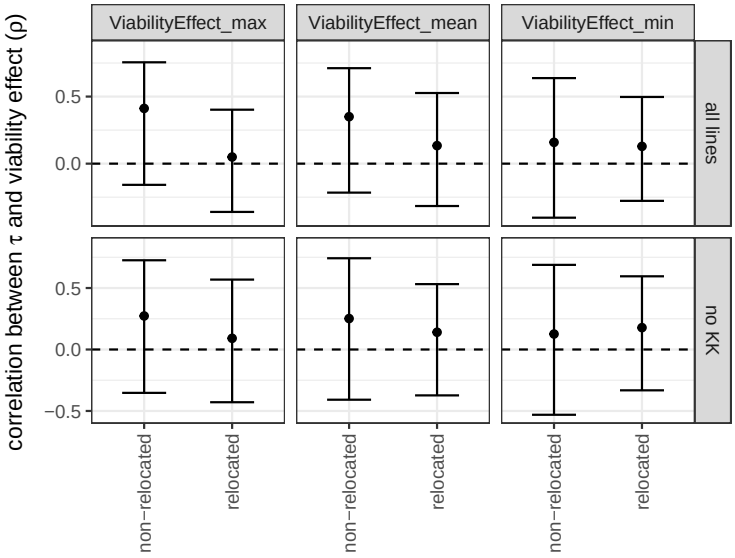
Acknowledgements

We thank members of the Meisel lab at the University of Houston and Andy Clark's lab at Cornell University for assistance with the RNAi experiments. Mariana Wolfner kindly supplied the *bam-Gal4* line, which was originally produced in Margaret Fuller's laboratory. Erin Kelleher provided valuable feedback on the preparation of this manuscript. This work was supported by start up funds from the University of Houston to RPM and a University of Houston Summer Undergraduate Research Fellowship to LZ.

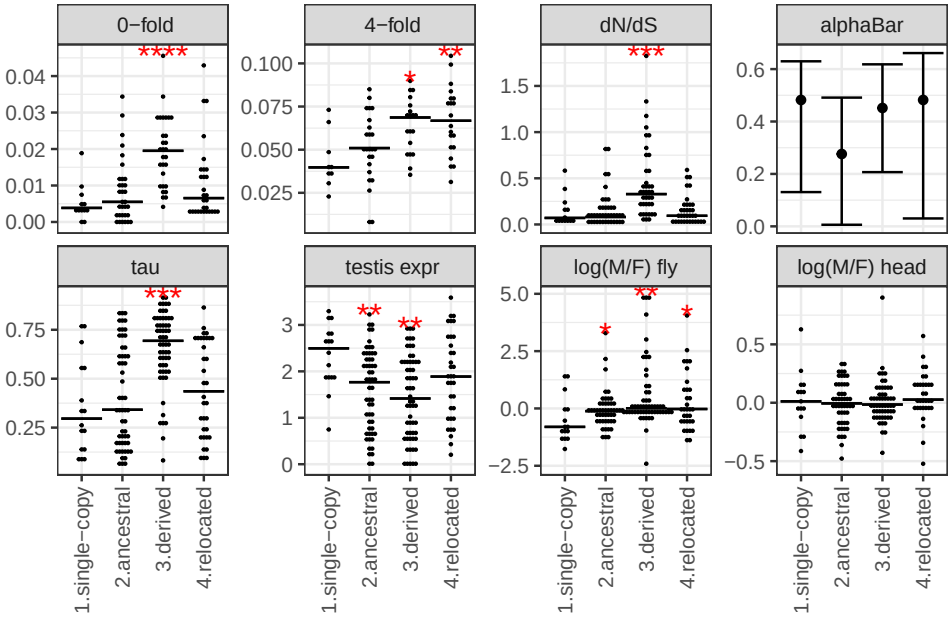
Supplemental Figures and Tables



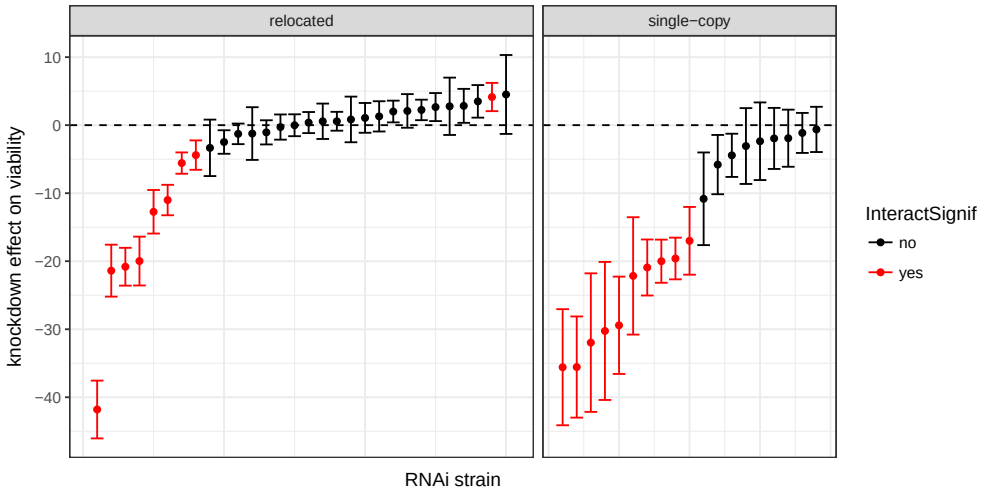
Supplemental Fig S1: Spearman's non-parametric rank order correlation (ρ) between all pairwise combinations of expression metrics (columns) and evolutionary divergence (rows) is plotted for single-copy non-relocated genes, the ancestral copies of dispersed duplicates, derived copies, and relocated genes. Error bars are 95% CIs from 1,000 bootstrap replicates of the data. Asterisks indicate correlations that are significantly different from zero.



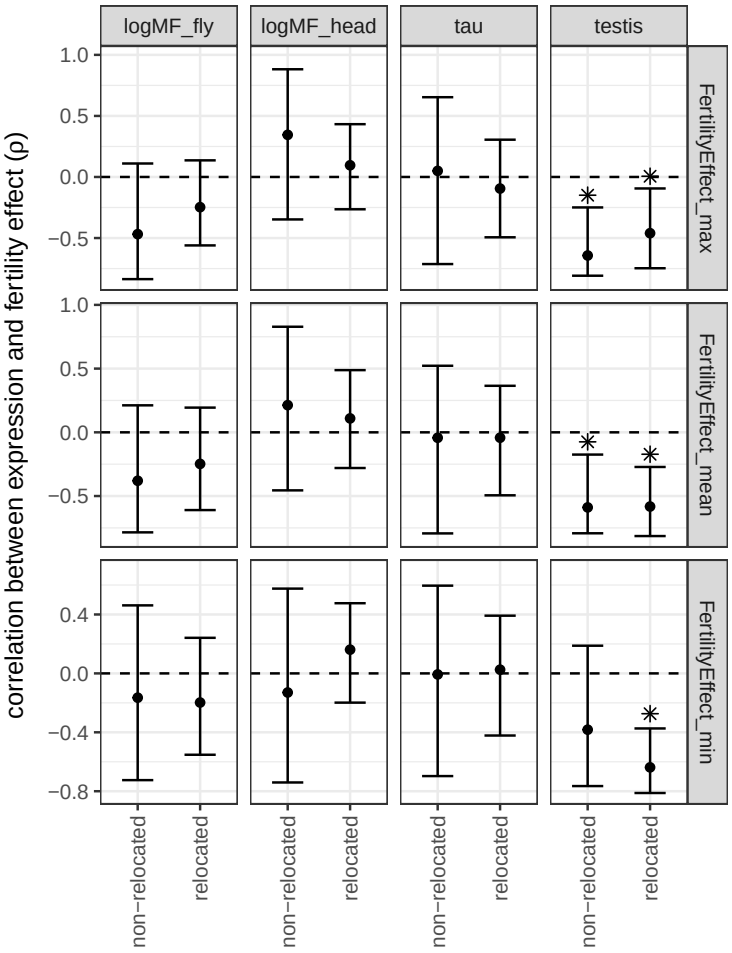
Supplemental Fig S2: Spearman's non-parametric rank order correlation (ρ) between expression breadth (τ) and effect of RNAi knockdown on viability for single-copy non-relocated genes and relocated genes. Viability effect for each gene is either the maximum of all RNAi lines for the gene (left), mean of all RNAi lines (middle), or minimum of all lines (right). Those viability effects were calculated for all lines (top) or excluding KK lines (bottom). Error bars are 95% CIs from 1,000 bootstrap replicates of the data. No correlations are significantly different from zero.



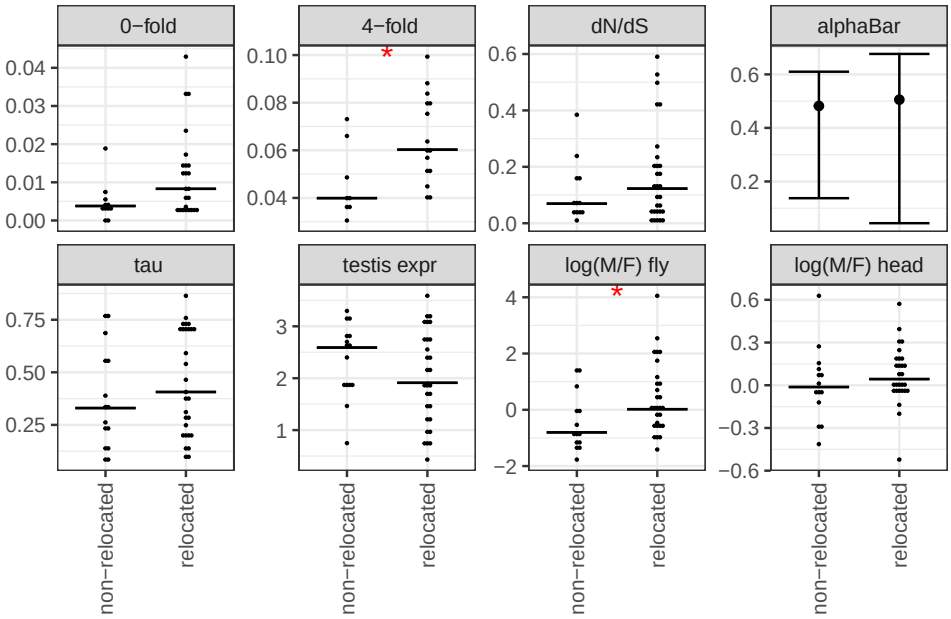
Supplemental Fig S3: Divergence measures, tests for selection, and gene expression data are plotted for genes used in the RNAi assays of viability effects. Estimates of $\overline{\alpha}$ are plotted as a point, along with the 95% CI. Significant differences in values when comparing single-copy non-relocated genes with genes in the other three classes are shown by red asterisks (*$P < 0.05$, **$P < 0.005$, ***$P < 0.0005$, and ****$P < 0.00005$ in a Mann-Whitney U test).



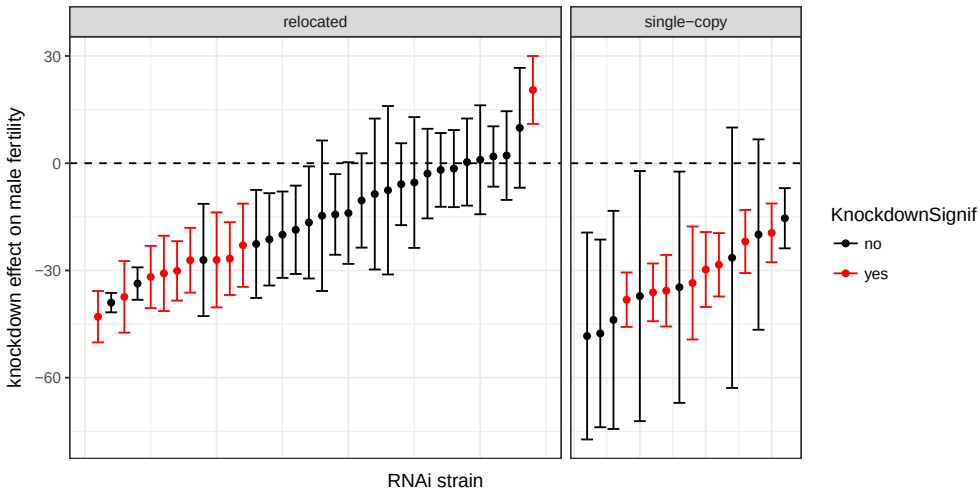
Supplemental Fig S4: Effect of ubiquitous RNAi knockdown on viability using only GD lines for relocated genes (left) and non-relocated single-copy genes (right) are plotted. Dots indicate the mean effect of knockdown across replicates, and the lines show the standard error. Each point is a RNAi line. Lines colored red are significantly different from zero.



Supplemental Fig S5: Spearman's non-parametric rank order correlation (ρ) between gene expression measures (columns) and effect of RNAi knockdown on fertility for single-copy non-relocated genes and relocated genes. Fertility effect for each gene is either the maximum of all RNAi lines for the gene (top), mean of all RNAi lines (middle), or minimum of all lines (bottom). Gene expression is either $\log_2 \frac{M}{F}$ for whole flies, $\log_2 \frac{M}{F}$ for heads, τ, or testis expression level. Error bars are 95% CIs from 1,000 bootstrap replicates of the data. Asterisks indicate correlations that are significantly different from zero.



Supplemental Fig S6: Divergence measures, tests for selection, and gene expression data are plotted for genes used in the RNAi assays of fertility effects. For all metrics except $\overline{\alpha}$, each dot is an individual gene, and the median across genes is indicated by a horizontal line. Estimates of $\overline{\alpha}$ are plotted as a point, along with the 95% CI. Significant differences in values when comparing non-relocated genes with relocated genes are shown by red asterisks (*$P < 0.05$, **$P < 0.005$, ***$P < 0.0005$, and ****$P < 0.00005$ in a Mann-Whitney U test).



Supplemental Fig S7: Effect of germline RNAi knockdown on male fertility using only GD lines for relocated genes (left) and non-relocated single-copy genes (right) are plotted. Dots indicate the mean effect of knockdown across replicates, and the lines show the standard error. Each point is a RNAi line. Lines colored red are significantly different from zero.

	lethal	non-lethal	perc lethal
Relocated	4	20	16.7%
Derived dups	3	17	15.0%
Single-copy	4	11	26.7%

Supplemental Table S1: Counts of relocated genes, derived copies of dispersed duplicates, and single-copy non-relocated genes that are lethal or non-lethal to knockdown. The criterion for lethality is no knockdown progeny recovered in 90% of replicate experiments.

Reference Cited

Abascal, F., A. Corpet, Z. A. Gurard-Levin, D. Juan, F. Ochsenbein, *et al.*, 2013 Subfunctionalization via adaptive evolution influenced by genomic context: the case of histone chaperones ASF1a and ASF1b. Mol. Biol. Evol. 30: 1853–1866.

Baker, R. H., A. Narechania, P. M. Johns, and G. S. Wilkinson, 2012 Gene duplication, tissue-specific gene expression and sexual conflict in stalk-eyed flies (Diopsidae). Philos Trans R Soc Lond B Biol Sci 367: 2357–2375.

Bateman, J. R., A. M. Lee, and C.-t. Wu, 2006 Site-specific transformation of Drosophila via ϕC31 integrase-mediated cassette exchange. Genetics 173: 769–777.

Betrán, E., Y. Bai, and M. Motiwale, 2006 Fast protein evolution and germ line expression of a Drosophila parental gene and its young retroposed paralog. Molecular biology and evolution 23: 2191–202.

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

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

Bhutkar, A., S. M. Russo, T. F. Smith, and W. M. Gelbart, 2007 Genome-scale analysis of positionally relocated genes. Genome Res. 17: 1880–1887.

Bikard, D., D. Patel, C. Le Mette, V. Giorgi, C. Camilleri, *et al.*, 2009 Divergent evolution of duplicate genes leads to genetic incompatibilities within A. thaliana. Science 323: 623–626.

Chen, D., and D. M. McKearin, 2003 A discrete transcriptional silencer in the bam gene determines asymmetric division of the *Drosophila* germline stem cell. Development 130: 1159–1170.

Chen, S., Y. E. Zhang, and M. Long, 2010 New genes in *Drosophila* quickly become essential. Science 330: 1682–1685.

Chintapalli, V. R., J. Wang, and J. A. T. Dow, 2007 Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. Nat. Genet. 39: 715–720.

Ciomborowska, J., W. Rosikiewicz, D. Szklarczyk, W. Makalowski, and I. Makalowska, 2013 "Orphan" retrogenes in the human genome. Mol. Biol. Evol. 30: 384–396.

Conant, G. C., and A. Wagner, 2003 Asymmetric sequence divergence of duplicate genes. Genome Res. 13: 2052–2058.

Conant, G. C., and K. H. Wolfe, 2008 Turning a hobby into a job: how duplicated genes find new functions. Nat. Rev. Genet. 9: 938–950.

Connallon, T., and A. G. Clark, 2011 The resolution of sexual antagonism by gene duplication. Genetics 187: 919–937.

Des Marais, D. L., and M. D. Rausher, 2008 Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature 454: 762–765.

Díaz-Castillo, C., and J. M. Ranz, 2012 Nuclear chromosome dynamics in the *Drosophila* male germline contribute to the nonrandom genomic distribution of retrogenes. Mol. Biol. Evol. 29: 2105–2108.

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

Dittmar, K. A., and D. A. Liberles, 2010 *Evolution after Gene Duplication*, Wiley-Blackwell.

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

Fay, J. C., G. J. Wyckoff, and C.-I. Wu, 2001 Positive and negative selection on the human genome. Genetics 158: 1227–1234.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y.-l. Yan, *et al.*, 1999 Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151: 1531–1545.

Gallach, M., and E. Betrán, 2011 Intralocus sexual conflict resolved through gene duplication. Trends Ecol. Evol. 26: 222–228.

Green, E. W., G. Fedele, F. Giorgini, and C. P. Kyriacou, 2014 A *Drosophila* RNAi collection is subject to dominant phenotypic effects. Nat. Methods 11: 222–223.

Groth, A. C., M. Fish, R. Nusse, and M. P. Calos, 2004 Construction of transgenic *Drosophila* by using the site-specific integrase from phage ϕC31. Genetics 166: 1775–1782.

Hahn, M. W., M. V. Han, and S.-G. Han, 2007 Gene family evolution across 12 Drosophila genomes. PLoS Genet. 3: 0.

Han, M. V., J. P. Demuth, C. L. McGrath, C. Casola, and M. W. Hahn, 2009 Adaptive evolution of young gene duplicates in mammals. Genome Res. 19: 859–867.

Han, M. V., and M. W. Hahn, 2012 Inferring the history of interchromosomal gene transposition in *Drosophila* using n-dimensional parsimony. Genetics 190: 813–825.

Hittinger, C. T., and S. B. Carroll, 2007 Gene duplication and the adaptive evolution of a classic genetic switch. Nature 449: 677–681.

Hu, T. T., M. B. Eisen, K. R. Thornton, and P. Andolfatto, 2013 A second-generation assembly of the *Drosophila simulans* genome provides new insights into patterns of lineage-specific divergence. Genome Res. 23: 89–98.

Hughes, A. L., 1994 The evolution of functionally novel proteins after gene duplication. Proc. Biol. Sci. 256: 119–124.

Innan, H., and F. Kondrashov, 2010 The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11: 97–108.

Kaessmann, H., N. Vinckenbosch, and M. Long, 2009 RNA-based gene duplication: mechanistic and evolutionary insights. Nat. Rev. Genet. 10: 19–31.

Kondrashov, F., I. Rogozin, Y. Wolf, and E. Koonin, 2002 Selection in the evolution of gene duplications. Genome Biol. 3: research0008.1.

Larracuente, A. M., T. B. Sackton, A. J. Greenberg, A. Wong, N. D. Singh, *et al.*, 2008 Evolution of protein-coding genes in *Drosophila*. Trends Genet. 24: 114–123.

Lee, T., and L. Luo, 1999 Mosaic analysis with a repressible neurotechnique cell marker for studies of gene function in neuronal morphogenesis. Neuron 22: 451–461.

Lynch, M., and A. Force, 2000 The probability of duplicate gene preservation by subfunctionalization. Genetics 154: 459–473.

Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, *et al.*, 2012 The *Drosophila melanogaster* genetic reference panel. Nature 482: 173–178.

Marques, A. C., I. Dupanloup, N. Vinckenbosch, A. Reymond, and H. Kaessmann, 2005 Emergence of young human genes after a burst of retroposition in primates. PLoS Biology 3: e357.

Marques-Bonet, T., S. Girirajan, and E. E. Eichler, 2009 The origins and impact of primate segmental duplications. Trends Genet. 25: 443–454.

Masly, J. P., C. D. Jones, M. A. F. Noor, J. Locke, and H. A. Orr, 2006 Gene transposition as a cause of hybrid sterility in *Drosophila*. Science 313: 1448–1450.

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

Meiklejohn, C. D., E. L. Landeen, J. M. Cook, S. B. Kingan, and D. C. Presgraves, 2011 Sex chromosome-specific regulation in the *Drosophila* male germline but little evidence for chromosomal dosage compensation or meiotic inactivation. PLoS Biol. 9: e1001126.

Meisel, R. P., 2009 Evolutionary dynamics of recently duplicated genes: selective constraints on diverging paralogs in the Drosophila pseudoobscura genome. J. Mol. Evol. 69: 81–93.

Meisel, R. P., 2011 Towards a more nuanced understanding of the relationship between sex-biased gene expression and rates of protein coding sequence evolution . Mol. Biol. Evol. 28: 1893–1900.

Meisel, R. P., M. V. Han, and M. W. Hahn, 2009 A complex suite of forces drives gene traffic from *Drosophila* X chromosomes. Genome Biol. Evol. 1: 176–188.

Meisel, R. P., B. B. Hilldorfer, J. L. Koch, S. Lockton, and S. W. Schaeffer, 2010 Adaptive evolution of genes duplicated from the *Drosophila pseudoobscura* neo-X chromosome. Mol. Biol. Evol. 27: 1963–1978.

Meisel, R. P., J. H. Malone, and A. G. Clark, 2012 Disentangling the relationship between sex-biased gene expression and X-linkage. Genome Res. 22: 1255–1265.

Metta, M., and C. Schlotterer, 2010 Non-random genomic integration - an intrinsic property of retrogenes in *Drosophila*? BMC Evol Biol 10: 114.

Moyle, L. C., C. D. Muir, M. V. Han, and M. W. Hahn, 2010 The contribution of gene movement to the "two rules of speciation". Evolution 64: 1541–1557.

Muller, H. J., 1940 Bearings of the 'Drosophila' work on systematics, in *The New Systematics*, edited by J. Huxley, pp. 185–268, Clarendon Press, Oxford, Oxford.

Ohno, S., 1970 *Evolution by Gene Duplication*, Springer-Verlag, New York, NY.

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

Potrzebowski, L., N. Vinckenbosch, A. C. Marques, F. Chalmel, B. Jégou, *et al.*, 2008 Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. PLoS Biology 6: e80.

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

R Core Team, 2015 *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria.

Ràmia, M., P. Librado, S. Casillas, J. Rozas, and A. Barbadilla, 2012 PopDrowser: the population *Drosophila* Browser. Bioinformatics 28: 595–596.

Rice, W. R., 1984 Sex chromosomes and the evolution of sexual dimorphism. Evolution 38: 735–742.

Rosso, L., A. C. Marques, M. Weier, N. Lambert, M.-A. Lambot, *et al.*, 2008 Birth and rapid subcellular adaptation of a hominoid-specific CDC14 protein. PLoS Biol. 6: e140.

Sartain, C. V., J. Cui, R. P. Meisel, and M. F. Wolfner, 2011 The poly(A) polymerase GLD2 is required for spermatogenesis in *Drosophila melanogaster*. Development 138: 1619–1629.

Schaeffer, S. W., A. Bhutkar, B. F. McAllister, M. Matsuda, L. M. Matzkin, *et al.*, 2008 Polytene chromosomal maps of 11 Drosophila species: the order of genomic scaffolds inferred from genetic and physical maps. Genetics 179: 1601–1655.

Schrider, D. R., K. Stevens, C. M. Cardeno, C. H. Langley, and M. W. Hahn, 2011 Genome-wide analysis of retrogene polymorphisms in *Drosophila melanogaster*. Genome Res. 21: 2087–2095.

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

Spofford, J. B., 1969 Heterosis and the evolution of duplications. Am. Nat. 103: 407 – 432.

Stanley, C. E., and R. J. Kulathinal, 2016 flyDIVaS: a comparative genomics resource for *Drosophila* divergence and selection. G3 6: 2355–2363.

Torgerson, D. G., and R. S. Singh, 2004 Rapid evolution through gene duplication and subfunctionalization of the testes-specific a4 proteasome subunits in Drosophila. Genetics 168: 1421–1432.

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

Vibranovski, M. D., H. F. Lopes, T. L. Karr, and M. Long, 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, M. D., Y. Zhang, and M. Long, 2009b General gene movement off the X chromosome in the Drosophila genus. Genome Res. 19: 897–903.

Vinckenbosch, N., I. Dupanloup, and H. Kaessmann, 2006 Evolutionary fate of retroposed gene copies in the human genome. Proc. Natl. Acad. Sci. U.S.A. 103: 3220–3225.

Vissers, J. H. A., S. A. Manning, A. Kulkarni, and K. F. Harvey, 2016 A *Drosophila* RNAi library modulates Hippo pathway-dependent tissue growth. Nat. Commun. 7: 10368.

Wicker, T., J. P. Buchmann, and B. Keller, 2010 Patching gaps in plant genomes results in gene movement and erosion of colinearity. Genome Research 20: 1229–1237.

Wu, C.-I., and E. Y. Xu, 2003 Sexual antagonism and X inactivation - the SAXI hypothesis. Trends Genet. 19: 243–247.

Yanai, I., H. Benjamin, M. Shmoish, V. Chalifa-Caspi, M. Shklar, *et al.*, 2005 Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21: 650–659.

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