

# Adaptive Evolution of Genes Duplicated from the *Drosophila pseudoobscura* neo-X Chromosome

Richard P. Meisel,<sup>\*</sup>1–3 Benedict B. Hilldorfer,<sup>†</sup>2 Jessica L. Koch,<sup>2</sup> Steven Lockton,<sup>4</sup> and Stephen W. Schaeffer<sup>1,2,5</sup>

<sup>1</sup>Intercollege Graduate Program in Genetics, The Pennsylvania State University

<sup>2</sup>Department of Biology, The Pennsylvania State University

<sup>3</sup>Department of Molecular Biology and Genetics, Cornell University

<sup>4</sup>Department of Integrative Biology, University of California, Berkeley

<sup>5</sup>Institute of Molecular Evolutionary Genetics, The Pennsylvania State University

<sup>†</sup>Present address: Interdisciplinary Biomedical Program, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania.

<sup>\*</sup>Corresponding author: E-mail: meisel@cornell.edu.

Associate editor: John True

## Abstract

*Drosophila* X chromosomes are disproportionate sources of duplicated genes, and these duplications are usually the result of retrotransposition of X-linked genes to the autosomes. The excess duplication is thought to be driven by natural selection for two reasons: X chromosomes are inactivated during spermatogenesis, and the derived copies of retroposed duplications tend to be testis expressed. Therefore, autosomal derived copies of retroposed genes provide a mechanism for their X-linked paralogs to “escape” X inactivation. Once these duplications have fixed, they may then be selected for male-specific functions. Throughout the evolution of the *Drosophila* genus, autosomes have fused with X chromosomes along multiple lineages giving rise to neo-X chromosomes. There has also been excess duplication from the two independent neo-X chromosomes that have been examined—one that occurred prior to the common ancestor of the *willistoni* species group and another that occurred along the lineage leading to *Drosophila pseudoobscura*. To determine what role natural selection plays in the evolution of genes duplicated from the *D. pseudoobscura* neo-X chromosome, we analyzed DNA sequence divergence between paralogs, polymorphism within each copy, and the expression profiles of these duplicated genes. We found that the derived copies of all duplicated genes have elevated nonsynonymous polymorphism, suggesting that they are under relaxed selective constraints. The derived copies also tend to have testis- or male-biased expression profiles regardless of their chromosome of origin. Genes duplicated from the neo-X chromosome appear to be under less constraints than those duplicated from other chromosome arms. We also find more evidence for historical adaptive evolution in genes duplicated from the neo-X chromosome, suggesting that they are under a unique selection regime in which elevated nonsynonymous polymorphism provides a large reservoir of functional variants, some of which are fixed by natural selection.

**Key words:** X chromosome, gene duplication, *Drosophila*, natural selection.

## Introduction

X chromosomes and autosomes are under different selection pressures (Vicoso and Charlesworth 2006). For example, the female-biased transmission of the X chromosome and its hemizyosity in males are predicted to result in unequal distributions of sexually antagonistic genes on the X and the autosomes (Rice 1984; Patten and Haig 2009). Indeed, there is a deficiency of genes with male-biased expression—a proxy for genes that have been under sexually antagonistic selection—on *Drosophila* X chromosomes (Parisi et al. 2003; Ranz et al. 2003; Connallon and Knowles 2005; Sturgill et al. 2007). This deficiency may also be a by-product of the dosage compensation mechanism in *Drosophila*—the X chromosome is hypertranscribed in males, which may prevent further increases in the transcription of X-linked genes in males (Vicoso and Charlesworth 2009).

In addition to different transmission dynamics of X chromosomes and autosomes, X chromosomes are also inactivated during the meiotic stage of spermatogenesis (Kelly et al. 2002; Hense et al. 2007; Turner 2007; Vibranovski et al. 2009). This poses problems for X-linked genes that would confer a fitness benefit if expressed during meiosis because they cannot merely acquire an appropriate regulatory sequence to gain expression on the inactivated chromosome. However, X-linked genes can be duplicated to an autosome, allowing them to “escape” X inactivation (Betrán et al. 2002). These X-to-autosome duplications usually occur via retrotransposition (i.e., reverse transcription of mRNA and insertion into the genome); retroposed genes in *Drosophila melanogaster* are almost always testis expressed (regardless of their chromosome of origin), essentially priming the autosomal retrocopies to allow their X-linked paralogs to gain testis expression by proxy (Meisel

et al. 2009). Indeed, the autosomal derived copies of genes retroposed from the X chromosome appear to compensate for the meiotic inactivation of their X-linked paralogs (Potrzebowski et al. 2008; Vibranovski et al. 2009).

There has been an excess of gene traffic from the X to the autosomes throughout the *Drosophila* genus (Betrán et al. 2002; Dai et al. 2006; Meisel et al. 2009; Vibranovski et al. 2009). In the case of duplication events in which the ancestral copy is retained, the excess X-to-autosome duplication is primarily driven by retroposition (Meisel et al. 2009). Along the lineage leading to *D. pseudoobscura*, however, both retroposition and a DNA-mediated mechanism are responsible for the excess duplication from the X to the autosomes (Meisel et al. 2009). Two selective mechanisms have been proposed to explain the excess X-to-autosome duplication: spermatogenic X inactivation (Betrán et al. 2002) and sexually antagonistic selection (Wu and Xu 2003). Case studies indicate that genes duplicated (often via retroposition) from the X to the autosomes tend to be testis or germ line expressed and often evolve under positive selection (Betrán and Long 2003; Betrán et al. 2006; Kalamegham et al. 2007; Tracy et al. 2010). Among duplicated genes in the *D. melanogaster* genome, an excess of paralogous pairs have a testis-biased copy and a copy that is downregulated in testis (Mikhaylova et al. 2008), suggesting that coregulation of the testis expression of paralogs is common. Additionally, the derived copies of genes duplicated from the *Caenorhabditis elegans* X chromosome to the autosomes have germ line-specific functions (Maciejowski et al. 2005). Furthermore, there is functional evidence that genes retroposed from mammalian X chromosomes to the autosomes have essential roles in spermatogenesis (Wang and Page 2002; Bradley et al. 2004; Rohozinski and Bishop 2004; Dass et al. 2007), and many of the genes retroposed from the mouse X chromosome are under positive selection (Shiao et al. 2007).

A unique feature of *Drosophila* chromosomal evolution is the fusion of the ancestral X chromosome with autosomal chromosomes along multiple independent lineages (Powell 1997). We refer to these ancestrally autosomal chromosome arms that are currently X linked as “neo-X chromosomes.” As with the ancestral X chromosome, an excess of genes has been duplicated from the *D. pseudoobscura* and *D. willistoni* neo-X chromosomes to the autosomes (Meisel et al. 2009). Interestingly, many of the same genes were independently duplicated from these two neo-X chromosomes, suggesting that X linkage is disadvantageous for certain functions of some genes. A burst of duplication from the *D. pseudoobscura* neo-X followed shortly after that chromosome arm became X linked (Meisel et al. 2009). The duplications appear to have been fixed by positive selection, but we do not know what role natural selection has played in their evolution subsequent to fixation. In other examples of X-to-autosome duplications, the derived copies have a narrower expression profile than the ancestral copies (Potrzebowski et al. 2008; Meisel et al. 2009). Genes with narrow expression profiles are expected to be under less selective constraints (Duret and

Mouchiroud 2000; Zhang and Li 2004; Liao et al. 2006; Larracuenta et al. 2008), suggesting that they are also under less pleiotropic constraints, which could increase the likelihood of adaptive fixations (Fisher 1930).

To understand the evolutionary dynamics of genes duplicated from the *D. pseudoobscura* neo-X chromosome, we collected DNA sequence polymorphism and gene expression data from the ancestral and derived copies of genes duplicated from one chromosome arm to another in the *D. pseudoobscura* genome. Most of these duplications were the result of retrotransposition, and the majority of the derived copies are testis expressed (regardless of their chromosome arm of origin). The derived copies are under relaxed selective constraints and show more evidence for adaptive evolution than the ancestral copies. Interestingly, derived copies duplicated from the neo-X to an autosome have experienced more positive selection and are under less selective constraints than genes duplicated from the autosomes or the ancestral X. We propose that the relaxed selective constraints on genes duplicated from the neo-X allow for elevated levels of nonsynonymous polymorphism that provide more standing functional variation that may be adaptively fixed when selection pressures change.

## Materials and Methods

### Identification of Duplicated Genes

Duplicated genes in the *D. pseudoobscura* genome were identified as described in the [supplementary methods](#) (Supplementary Material online) based on a previously published approach (Meisel 2009b). Briefly, one-to-one best-hit orthologs to *D. melanogaster* genes had been annotated in the *D. pseudoobscura* genome (Richards et al. 2005). Intergenic regions in the *D. pseudoobscura* genome were searched against those genic sequences using nucleotide Blast (Altschul et al. 1990). Additionally, the intergenic regions were also searched against a database of all known *D. melanogaster* proteins using BlastX (Altschul et al. 1997). Intergenic regions that matched an orthologous *D. pseudoobscura* gene and *D. melanogaster* protein are said to contain a gene that had been duplicated at some point after the divergence between the two species' lineages. The ancestral and derived copies were annotated as described in the [supplementary methods](#) (Supplementary Material online). The amino acid sequence of the ancestral copy, derived copy, and *D. melanogaster* ortholog were aligned using the ClustalW implementation in MEGA4 (Tamura et al. 2007), and the nucleotide sequences were overlaid on the amino acid alignment. Tests for unequal rates of evolution between paralogs were performed on the translated amino acid sequences using the *D. melanogaster* sequence as an outgroup (Tajima 1993).

### Sequencing Strategy

*Drosophila* genomes are organized into five major chromosome arms and a small dot chromosome, and each arm is referred to as a Muller element (Muller 1940; Powell 1997). The ancestral karyotype consists of an acrocentric

X chromosome (Muller element A), four acrocentric major autosomes (Muller elements B–E), and the dot chromosome (Muller element F). Throughout the evolution of the genus, various fusions of Muller elements have occurred, giving rise to metacentric chromosomes (Powell 1997; Schaeffer et al. 2008). The duplicated genes sequenced for this study were selected from the pool of duplicates in the *D. pseudoobscura* genome based on two criteria: 1) the paralogs were located on different Muller elements and 2) the ancestral and derived copies could be distinguished (see [supplementary methods](#), Supplementary Material online). We limited ourselves to genes with intact open-reading frames to prevent including pseudogenes in our data set.

Genes were chosen based on whether the paralogs had a nucleotide divergence threshold of approximately 20% so that unique primers could be designed for each copy; however, we did not set a firm limit, choosing instead to allow primer design to be the limiting factor. Additionally, this level of divergence should decrease the likelihood of gene conversion between paralogs (Teshima and Innan 2004). Polymerase chain reaction (PCR) primer pairs were chosen to amplify fragments of approximately 400–800 nucleotides and to be specific for each paralog of interest. To ensure unique amplification of one copy, we required that each primer have sites unique to one copy at its 3' end, span an insertion/deletion or intron-exon boundary unique to one copy, or be located outside the duplicated region. Overlapping PCR products were used to sequence across the entire length of each gene without gaps. In the case of the ancestral copy of GA24652, a gap was permitted within a long intron that is not present in the derived copy of the gene. Additionally, in some cases, we were unable to design primers or amplify PCR products from the 5' or 3' end of the gene. We sequenced the ancestral and derived copies of 14 duplicated genes (28 genes in total), 6 with ancestral copies on the neo-X chromosome. All 14 of the ancestral copies were included in previously published analyses of gene family evolution in *Drosophila* (Hahn et al. 2007; Meisel et al. 2009), but only 5 of the derived copies were included in those analyses. The sequencing primers for the 28 genes are listed in the [supplementary table S1](#) (Supplementary Material online).

To polarize nonsynonymous and synonymous substitutions along the lineages leading to the ancestral and derived copies requires an outgroup species. An ideal outgroup would be distantly related enough to have split from the species of interest prior to the duplication events (i.e., the outgroup would not share the duplicated genes) but closely related enough to allow for the polarization of synonymous changes along the lineages leading to the ancestral and derived copies. None of the species with sequenced genomes are appropriate outgroups for analyzing the 14 duplicated genes described here; *D. persimilis* is too closely related to *D. pseudoobscura*, and synonymous sites are saturated between *D. pseudoobscura* and *D. melanogaster* (Richards et al. 2005) or any of the other nine sequenced species (*Drosophila* 12 Genomes Consortium

2007). The *obscura* species subgroup is an ideal outgroup for three reasons: synonymous sites are not saturated between genes sampled from the *obscura* and *pseudoobscura* subgroups (Russo et al. 1995; Wells 1996), the *obscura* subgroup is the closest relative to the *pseudoobscura* subgroup that does not share the *D. pseudoobscura* neo-X chromosome (Patterson and Stone 1952; Steinemann et al. 1984), and the burst of duplication that followed the creation of the neo-X chromosome along the *D. pseudoobscura* lineage likely occurred after the split between the *pseudoobscura* and *obscura* subgroups (Meisel et al. 2009). There are no whole-genome sequences available for any species in the *obscura* subgroup, so we sequenced the orthologs of 10 of the 14 duplicated genes in a strain of *D. guanche* collected from the Canary Islands, Spain, in 1971 (*Drosophila* Species Stock Center number 14011-0095.00). We designed PCR primers in coding sequence that is conserved between *D. pseudoobscura* and *D. melanogaster*, maximizing the length of the portion of the coding region sequenced ([supplementary table S2](#), Supplementary Material online). Two genes were not included (the homologs of GA17928 and GA24652) because we obtained sequences for two different copies of the *D. guanche* genes, suggesting that these genes were duplicated prior to the split between the *D. guanche* and *D. pseudoobscura* lineages. Additionally, we were unable to successfully amplify and sequence two additional genes in *D. guanche* (the homologs of GA17756 and GA29002).

### Strains Selected for Resequencing

We selected isofemale strains of *D. pseudoobscura* from three populations: Mesa Verde National Park (MV), Colorado (37°18'0"N; 108°24'58"W, collected by S.W. Schaeffer in 2005); Kaibab National Park (KB), Arizona (36°12'30"N; 112°3'30"W, collected by S.W. Schaeffer in 2005); and Santa Cruz Island (SC), California (34°0'42"N, 119°48'42"W, collected by Luciano Matzkin in 2004). Each strain was inbred for 10–11 generations using single-pair sibmatings to purge heterozygosity (balancer chromosomes only exist for Muller element C in *D. pseudoobscura*). We initially selected six strains from KB and six strains from MV, and we isolated DNA from these strains using a CsCl protocol (Bingham et al. 1981). We attempted to sequence each gene from these 12 strains. If we were unable to amplify a PCR product from a strain or a strain was heterozygous for a particular gene, that strain was removed from the sample for that duplicated gene (both the ancestral and derived copies were removed). At least 1 of the 12 strains (and no more than 6) was removed for every sampled gene, and 2.7 strains were removed on average for each gene. In some cases, the strain was replaced with another inbred strain from the same population or with an inbred strain from another population. DNA was isolated from additional strains from KB, MV, and SC using a protocol requiring a small sample of flies (Gloor et al. 1993). All sampled genes required 1–3 replacement strains, and 1.7 strains were added on average. A complete list of strains sequenced for each duplicated gene is provided



in [supplementary table S3](#) (Supplementary Material online). Because we removed the strains from both the ancestral and derived copies, this should not affect our comparisons of polymorphism between ancestral and derived copies. However, this may bias our comparisons of genes duplicated from the neo-X chromosome with those duplicated from other chromosomes. To test for any bias, the number of excluded strains was compared between genes duplicated from the neo-X and the other duplications.

### Sequencing, Assembly of Sequence Traces, and Alignment of Sequences

We directly sequenced the PCR products after they were cleaned with ExoSAP-IT (USB Corporation, Cleveland, OH). Sequencing of *D. pseudoobscura* genes was carried out on an ABI 3720XL machine at the Penn State Nucleic Acids Facility. Sequencing of *D. guanche* genes was carried out on the same machine and on an ABI 3730 DNA Analyzer at the Cornell University Life Sciences Core Laboratories Center. Each PCR product was sequenced on both strands. The SeqMan program in the DNASTAR Lasergene package (Madison, WI) was used to remove lower-quality sequences at the 5' and 3' ends of the traces. SeqMan was also used to assemble sequence traces into contigs, a process that was done separately for each copy of each paralog from each strain. A gene sampled from a strain was retained if unambiguous calls were found across the entire length of the gene (i.e., no heterozygosity) with coverage of at least two traces per nucleotide. If a strain did not meet these criteria, it was removed from the sample for both the ancestral and derived copy of that duplicated gene, and it was replaced as described above. The assembled sequences were aligned by eye to the reference sequences identified in the complete genome sequence. The copies of the genes from the genome sequence (Richards et al. 2005) were included in our analysis of polymorphism (the genome strain MV2-25 was collected by Wyatt W. Anderson from MV). The final alignment for each duplicated gene contains all sampled alleles from both the ancestral and derived copies of the paralog (alignments available as Supplementary Material online).

### Gene Expression Assays

Strain MV2-25 was sequenced by the *D. pseudoobscura* genome project (Richards et al. 2005). We reared this strain on a yeast, agar, and dextrose medium at 18 °C. Virgin males and females were collected and aged 5–7 days, and live flies were dissected in Ringer's solution. Heads, thoraxes, and abdomens were isolated and flash frozen in liquid nitrogen. Additionally, testes and ovaries were isolated from males and females, respectively, in separate dissections and flash frozen. Total RNA was extracted from male and female heads, thoraxes, and abdomens using the TRIzol Reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. Total RNA was isolated from testes and ovaries using an RNeasy Mini Kit (Qiagen, Valencia, CA). First-strand cDNA synthesis was performed for each RNA sample using oligo(dT)<sub>16</sub>-primed reverse

transcription-PCR, and single-stranded cDNA was purified with the QIAquick PCR Purification Kit (Qiagen).

Primer pairs were designed to uniquely amplify each copy of the 14 duplicated genes from which we collected DNA sequence polymorphism data ([supplementary table S4](#), Supplementary Material online). Single-stranded cDNA was used as a template in a 35 cycle PCR with an annealing temperature of 60 °C for each primer pair. For each PCR product, 12 µl were run on a 2% agarose gel and visualized using the GelStar Nucleic Acid Stain (Lonza) under an ultraviolet light. In some cases, different combinations of PCR primers were needed to amplify cDNA of the genes of interest from different body parts. We were unable to successfully amplify the ancestral or derived copies of GA17928 using cDNA from any body part.

Illumina RNA sequencing was performed on testis tissue dissected from approximately fifty 12-day-old virgin male *D. pseudoobscura*. Tissue was stored in RNAlater solution (Ambion, Inc., Austin TX) until extraction with the PicoPure RNA Extraction Kit (Molecular Devices, Sunnyvale, CA). RNA-sequencing libraries were prepared following the standard Illumina protocol. Briefly, mRNA was isolated using DYNAL oligo(dT) beads (Invitrogen) and fragmented using fragmentation buffer (Ambion, Inc.). Synthesis of cDNA was performed using random hexamer primers. The cDNA fragments were then subjected to an end repair reaction followed by the addition of an adenine base to facilitate subsequent adaptor ligation using Illumina's single end adaptor kit (Illumina, Hayward, CA). These cDNA templates were purified on a 2% agarose gel, and a band corresponding to a 200-bp fragment ( $\pm 25$  bp) band was excised from the gel. After gel extraction with the QIAquick Gel Extraction Kit (Qiagen), the cDNA templates were PCR enriched using Phusion DNA polymerase (New England BioLabs, Ipswich, MA) and adaptor-specific primers (Illumina). This single testis library was run on two lanes of an Illumina Genome Analyzer (GA pipeline version 1.4), generating a total of 27,102,864 high-quality 76 nucleotide single-end reads. We trimmed each read by three bases at the 3' end. We used the TopHat software package (Trapnell et al. 2009) to both align the reads to the reference genome and to calculate reads per kilobase of exon model per million mapped reads (RPKM). For TopHat, we used a custom GFF annotation file comprised of *D. pseudoobscura* release 2.4 exons plus the exons of the duplicated genes described in this study, with the options "fill-gaps," "coverage-search," and "microexon-search" enabled.

### Analysis of DNA Sequence Polymorphism

The strains we sampled came primarily from two populations, MV and KB. There is no evidence for structure in North American *D. pseudoobscura* populations (Riley et al. 1989; Schaeffer and Miller 1992; Kovacevic and Schaeffer 2000; Schaeffer et al. 2003), so we should be able to treat samples from these two populations as coming from a single population. However, to ensure that there is no evidence for population structure in our data, a diverse array of metrics were estimated using DnaSP 4.10 (Rozas

et al. 2003):  $H_{ST}$ ,  $K_{ST}$ ,  $K_{ST}^*$ ,  $Z$ ,  $Z^*$ ,  $S_{nn}$  (Hudson et al. 1992; Hudson 2000). Additionally, we tested the duplicated genes for evidence of gene conversion tracts between paralogs using the method of Betrán et al. (1997). Evidence of gene conversion between paralogs would mean that we could not treat each copy as an independent gene in our analysis.

An excess of nonsynonymous fixed differences along a lineage is suggestive of positive selection or relaxed selective constraints along that lineage (Li 1997). We used the *D. guanche* sequences as outgroups for estimating nonsynonymous divergence ( $d_N$ ), synonymous divergence ( $d_S$ ), and the ratio of the two ( $\omega$ ) along the lineages leading to the ancestral and derived copies of the ten duplicated genes in which we were able to sequence *D. guanche* orthologs. These estimates were computed using model 1 in the CODEML program within the PAML 4.2 package (Yang 2007). Sequences from the sampled alleles of each copy were provided to the program, along with the sequence of the *D. guanche* ortholog, to ensure that polymorphic sites were not included in the lineage-specific estimates of  $d_N$ ,  $d_S$ , and  $\omega$ .

Although inferences of the historical effects of natural selection are possible using divergence data alone, analyses of patterns of DNA sequence polymorphism provide much greater power (Zhai et al. 2009). We polarized the fixed differences between paralogs along the lineages leading to the ancestral and derived copies of the duplicated genes using *D. guanche* as the outgroup (fig. 1). We were only able to perform this analysis on the ten duplicated genes for which we sequenced a single ortholog in *D. guanche*. We classified each variable site as either a nonsynonymous polymorphism in the ancestral copy ( $P_{NA}$ ), synonymous polymorphism in the ancestral copy ( $P_{SA}$ ), nonsynonymous polymorphism in the derived copy ( $P_{ND}$ ), or synonymous polymorphism in the derived copy ( $P_{SD}$ ). We classified nucleotide sites with a fixed difference between the ancestral and derived copies as nonsynonymous substitutions along the lineage leading to the ancestral copy ( $D_{NA}$ ), synonymous substitutions in the lineage leading to the ancestral copy ( $D_{SA}$ ), nonsynonymous substitutions in the lineage leading to the derived copy ( $D_{ND}$ ), or synonymous substitutions in the lineage leading to the derived copy ( $D_{SD}$ ). For assigning lineage-specific polymorphic sites and fixed differences, we required that the *D. guanche* sequence match the nucleotide in either the ancestral or derived copy; otherwise, the nucleotide site was ignored. Alignment gaps present in the ancestral copy, derived copy, or *D. guanche* ortholog were excluded from the analysis of all coding sequences.

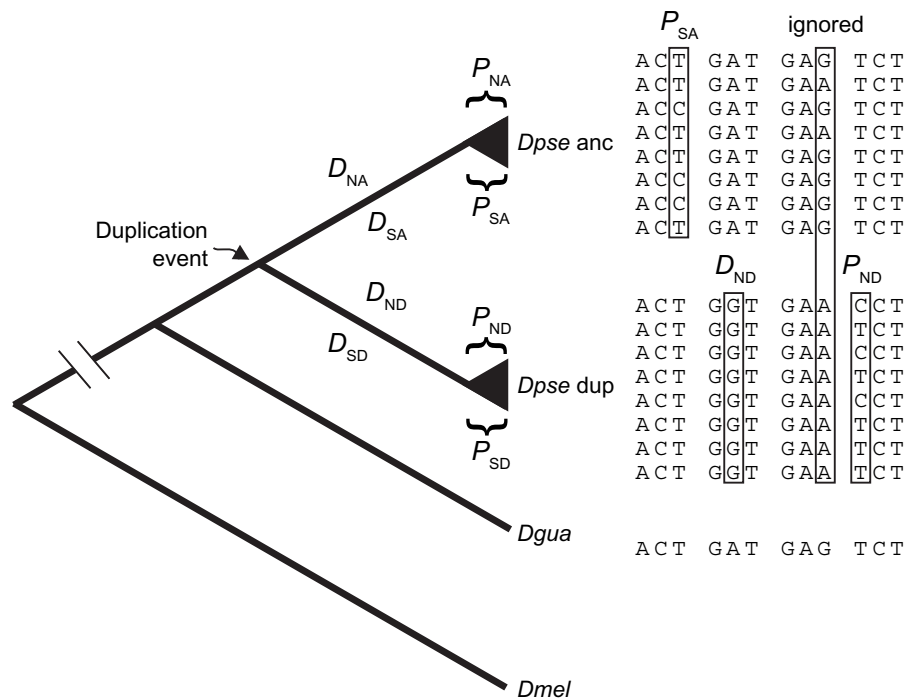
We performed a lineage-specific McDonald–Kreitman test (McDonald and Kreitman 1991; Akashi 1995) using Fisher's exact test to evaluate whether the ratios of nonsynonymous to synonymous polymorphism and divergence differ from those predicted under neutrality for each gene. An excess of  $D_{NA}$  or  $D_{ND}$  is indicative of historical positive selection along the ancestral or derived lineages, respectively. Additionally, we tested for deviations from neutrality using counts of nonsynonymous and synonymous polymorphic and divergent sites pooled

across all ancestral copies and all derived copies separately. The proportion of amino acid substitutions driven by adaptive evolution ( $\alpha$ ) (Smith and Eyre-Walker 2002) was estimated for ancestral and derived copies as follows:

$$\alpha_{anc} = 1 - \frac{\overline{D_{SA}}}{\overline{D_{NA}}} \left( \frac{P_{NA}}{P_{SA} + 1} \right); \alpha_{dup} = 1 - \frac{\overline{D_{SD}}}{\overline{D_{ND}}} \left( \frac{P_{ND}}{P_{SD} + 1} \right).$$

We also estimated  $\alpha_{anc}$  and  $\alpha_{dup}$  separately for genes duplicated from the neo-X and those duplicated from other chromosome arms. To test for deviations from neutrality, we permuted the McDonald–Kreitman table for each gene 1,000 times (Patefield 1981) and then recalculated  $\alpha_{anc}$  and  $\alpha_{dup}$  for each permutation. These permutations were used to generate a 95% confidence interval (CI) for  $\alpha_{anc}$  and  $\alpha_{dup}$  under neutral expectations.

Recent positive selection is expected to leave a characteristic signature in the pattern of polymorphism at sites linked to the region under selection (Przeworski 2002). An excess of rare polymorphisms is suggestive of the recent selective sweep of an advantageous mutation, but it can also be caused by background selection against deleterious mutations or population expansion. An excess of intermediate-frequency polymorphisms, on the other hand, suggests balancing selection, a shrinking population, or population substructure. Noncoding polymorphism was measured in the region in and around each copy of a duplicated gene using sites 5' and 3' of the coding sequence, introns, and synonymous mutations in the coding sequence. Ancestral and derived copies were analyzed separately, and alignment gaps present in sequences from one copy were excluded only in the analysis of that copy. Noncoding polymorphism was measured using both the average pairwise differences between sequences ( $k$ ) and the number of segregating sites ( $S$ ) in DnaSP 4.10 (Rozas et al. 2003). These values were used to estimate Tajima's (1989)  $D$ , as well as the ratio of  $D$  to its theoretical minimum ( $D/D_{min}$ ) (Schaeffer 2002). A significantly positive  $D$  indicates a deficiency of  $S$  relative to  $k$ , which results from an excess of intermediate-frequency variants. An excess of rare alleles leads to a significantly negative  $D$  statistic.  $D < 0$  could result from a selective sweep, background selection, or population expansion. The major difference between these explanations is that selective sweeps will affect only one or a few loci in the genome. The  $D/D_{min}$  ratio will be similar among all loci in the genome even if sampled loci have different numbers of segregating sites or sample sizes (Schaeffer 2002). In addition, the magnitude of the  $D/D_{min}$  ratio is directly related to the strength of population expansion. *Drosophila pseudoobscura* populations have experienced a modest population expansion based on an expansion parameter of  $Nr = 7$  (Schaeffer 2002), where  $N$  is the effective population size and  $r$  is the expansion rate (Slatkin and Hudson 1991). Coalescent simulations with population expansion parameter  $Nr = 7$  were used to determine whether the  $D/D_{min}$  ratio for the sampled loci departs from neutral expectations under a demographic model of an exponentially growing population.



**Fig. 1.** Estimating polymorphism within and divergence between paralogs. This schematic represents some of the measures of polymorphism and divergence estimated from the data. The phylogeny shows the relationships of the ancestral (anc) and derived (dup) copies of a *Drosophila pseudoobscura* (*Dpse*)–duplicated gene, along with the *D. guanache* (*Dgua*) and *D. melanogaster* (*Dmel*) orthologs. Nonsynonymous polymorphism was measured in the ancestral ( $P_{NA}$ ) and derived copies ( $P_{ND}$ ), and synonymous polymorphism was measured in the ancestral ( $P_{SA}$ ) and derived copies ( $P_{SD}$ ). The number of nonsynonymous fixed differences along the lineage leading to the ancestral copy ( $D_{NA}$ ) and derived copy ( $D_{ND}$ ) were estimated as described in the text. We also estimated the number of synonymous fixed differences along the ancestral ( $D_{SA}$ ) and derived ( $D_{SD}$ ) lineages. To the right is an alignment of four hypothetical codons that illustrates the configuration of variation at nucleotide sites for the different types of sites.

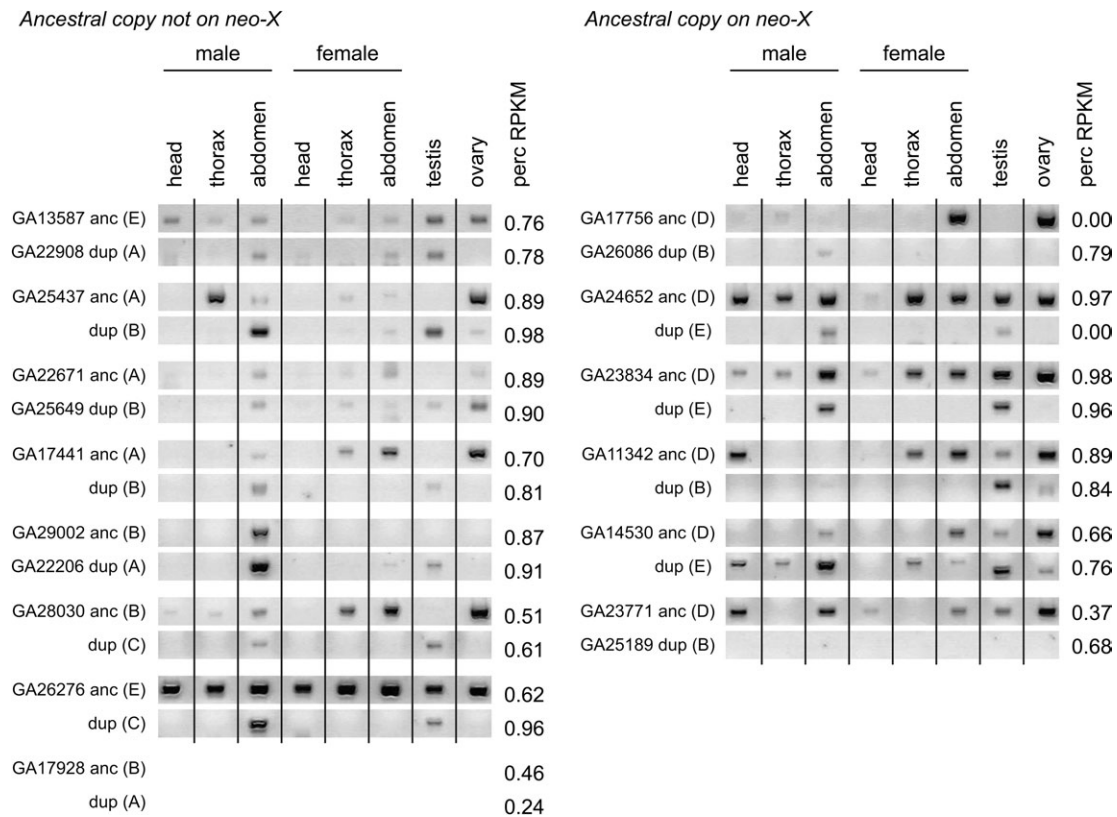
An excess of nonsynonymous polymorphism ( $P_N$ ) in one class of sequences may indicate relaxed selective constraints on that class of genes (Li 1997; Nei and Kumar 2000; Wang et al. 2004). If there is also an excess of synonymous polymorphism ( $P_S$ ), however, high  $P_N$  could be explained by a higher mutation rate in that class of genes.  $P_S$  and  $P_N$  were estimated for the protein-coding regions of the ancestral and derived copies of the duplicated genes separately using the number of mutations required to give rise to the haplotypes ( $\eta$ ) in DnaSP 4.10 (Rozas et al. 2003). Alignment gaps present in either the ancestral or derived copy were excluded from the analysis of both copies. We estimated  $\Theta$  (Watterson 1975) for each gene using synonymous ( $\Theta_S$ ) and nonsynonymous ( $\Theta_N$ ) sites separately, standardizing for the length of the coding region. These estimates were used in comparisons of polymorphism between genes. We also estimated the average pairwise differences between sequences at nonsynonymous sites ( $k_N$ ) for the ancestral and derived copies of each duplicated gene separately. As with our estimates of Tajima’s  $D$  at noncoding sites, we compared  $\Theta_N$  and  $k_N$  (per nonsynonymous site) to determine if there is an excess or deficiency of intermediate-frequency nonsynonymous variants. Finally, X chromosomes and autosomes have different effective population sizes ( $N_e$ ). Assuming equal numbers of mating males and females, the effective number of X chromosomes in a population will be 75% that of the autosomes. To

correct for the expected differences in  $N_e$  of the X and the autosomes, we replicated our analysis using  $\Theta$  estimates for X-linked genes multiplied by 4/3. This correction also assumes no selection on linked sites, and there is some evidence that the correction may not be appropriate for *Drosophila* populations (Andolfatto 2001; Hutter et al. 2007; Singh et al. 2007), so we also report our results without the correction.

## Results

### General Properties of Duplicated Genes

We analyzed 14 genes that were duplicated from one chromosome arm to another in the *D. pseudoobscura* genome, 6 of which arose from the neo-X chromosome. Interchromosome-arm duplications in the *D. pseudoobscura* genome are usually the result of retrotransposition, rather than DNA duplication (Meisel 2009a). These two mechanisms can be distinguished because a retroposed duplication will be missing any introns present in the ancestral copy, whereas a DNA duplication will have all introns found in the ancestral copy. However, if the ancestral copy is an intron-less gene or if the derived copy is only missing some introns, the mechanism of duplication will be ambiguous—although it appears that most ambiguously duplicated genes in the *D. pseudoobscura* genome were retroposed (Meisel 2009a). The majority of duplications



**Fig. 2.** Expression of ancestral and derived copies of duplicated genes. Agarose gels of the products of PCR of cDNA are shown for the ancestral and derived copies of 13 duplicated genes; cDNA was extracted from eight body parts in males and females. The frequency of genes with lower RPKM (perc RPKM) from the RNA-seq of testis-derived mRNA is given for each gene. For each duplicated gene, the ancestral copy (anc) is presented first and the derived copy (dup) is presented second. The gene identifiers of the ancestral copies (and the derived copies, when previously annotated) are given, with the Muller element location of each gene in parentheses. We were unable to amplify cDNA from either the ancestral or derived copies of GA17928, and only perc RPKM is given for those genes.

examined here were retroposed (10/14), including nearly all those that arose from the neo-X chromosome (5/6).

### Expression Profiles of Duplicated Genes

We measured the sex-specific expression of the ancestral and derived copies of the 14 duplicated genes. Previously published data on sex-biased expression from whole flies (Zhang et al. 2007; Jiang and Machado 2009) were not used because the microarrays included probes for only 5/14 derived copies in our data set. We performed PCR on cDNA extracted from heads, thoraxes, and abdomens from males and females separately and on cDNA extracted from testes and ovaries. Reactions for which a band is visible on an agarose gel indicate the expression of a gene in a particular body part (fig. 2). We were unable to amplify the ancestral or derived copy of GA17928 and the derived copy of GA23771 using cDNA from any of the body parts. This method does not allow for quantification of expression differences, and we only treat it as a qualitative assay. However, we also quantitatively measured genome-wide expression in testis with RNA-seq, and we compared the RPKM values for the ancestral and derived copies with the distribution of RPKM of all genes.

Both the ancestral and derived copies of duplicated genes tend to be testis expressed. The majority of both ancestral and derived copies have RPKM values greater

than at least 60% of all genes, and most have RPKM values greater than 70% of all genes (fig. 2 and supplementary table S5, Supplementary Material online). The derived copies also tend to have narrower expression profiles than their ancestral paralogs (fig. 2). Of the 12 derived copies for which we were able to confirm expression in our qualitative assay, 6 have testis- or male abdomen-specific expression (the paralogs of GA17441, GA28030, GA26276, GA17756, GA24652, and GA23834) (fig. 2). There is some disagreement between the RPKM values and the qualitative measures from PCR and gel electrophoresis, which may be a result of differences in the chemistries of the two methodologies. Regardless of these differences, it is clear that many of the derived copies of these 14 duplicated genes have a higher level of testis expression than the genomic average, and they are often limited to male-specific tissues (either testis or somatic accessory tissues).

### Tests for Population Differentiation and Gene Conversion between Paralogs

We resequenced the ancestral and derived copies of the 14 duplicated genes in chromosomes sampled from inbred isofemale lines. We included 9–13 lines per gene (supplementary table S3, Supplementary Material online); the number of lines included depended on our success in



PCR amplification and sequencing. Nearly all inbred lines were created from individuals collected from the MV and KB populations from Colorado and Arizona, respectively. If these populations were genetically differentiated, we would have to treat them separately in our analysis. Of the ancestral copies we sequenced, only GA29002 has evidence for significant differentiation between the MV and KB samples (supplementary table S6, Supplementary Material online), but the differentiation is no longer significant after a Bonferroni correction for multiple tests. Three derived copies also show evidence for population subdivision in at least one of the tests performed (the paralogs of GA13587, GA17756, and GA29002), but these differences are not significant after correcting for multiple tests. Therefore, we treat sequences sampled from the different populations as if they came from a single population.

There is also minimal evidence for gene conversion between the paralogs. Small potential gene conversion tracts ( $\leq 11$  nucleotides) were identified between ancestral and derived copies of three duplicated genes (GA29002, GA11342, and GA23771). We used a maximum likelihood method to estimate the true tract length (Betrán et al. 1997) and found that the longest expected tract length is  $< 16$  nucleotides (between the paralogs of GA23771). The short length of these conversion tracts suggests that they are false positives, and excluding the sequences with observed gene conversion tracts in the following analyses does not affect our results. Importantly, the lack of gene conversion between paralogs (as well as their lack of genetic linkage) allows us to treat the sequences of the two copies of a duplicated gene as independent loci in our analyses of polymorphism and divergence.

### Positive Selection on the Derived Copies of Duplicated Genes

Previous work has shown that the amino acid sequences of the derived copies of duplicated genes located on different chromosomes from the ancestral copies tend to evolve at faster rates than the ancestral copies (Cusack and Wolfe 2007). This is generally true for the 14 interchromosome-arm-duplicated genes we examined (fig. 3). If a gene is evolving under neutral expectations,  $\omega \leq 1$ . Nonsynonymous and synonymous substitutions can be polarized along individual lineages using outgroup sequences, allowing one to test the neutral hypothesis for the lineages leading to the ancestral and derived copies of duplicated genes separately. Most lineages have  $\omega < 1$ , but the lineages leading to the derived copies have higher  $\omega$  than those leading to the ancestral copies ( $P < 0.01$ , Mann–Whitney test) (table 1), consistent with the accelerated rate of amino acid evolution along the derived lineages (fig. 3). The differences in evolutionary rates may be attributable to either increased positive selection or relaxed selective constraints on the derived copies (relative to the ancestral copies). To test these hypotheses, we incorporated DNA sequence polymorphism into the analysis (fig. 1).

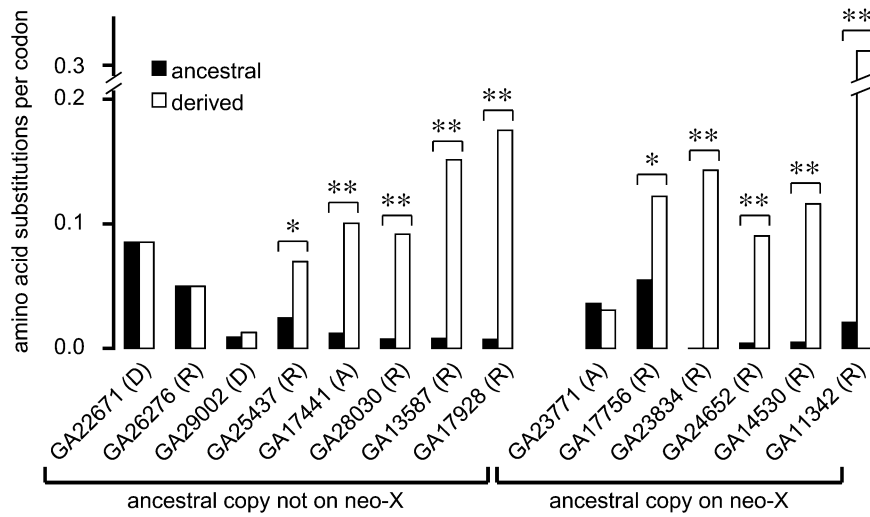
Under neutral expectations, the ratio of nonsynonymous polymorphism to nonsynonymous divergence

should equal the ratio of synonymous polymorphism to synonymous divergence (McDonald and Kreitman 1991). Our estimates of nonsynonymous and synonymous polymorphism and divergence in the ancestral copies of the duplicated genes do not deviate from neutrality (table 1). However, four of the ten derived copies have a significant excess of nonsynonymous fixed differences, whereas the other six derived copies are not statistically distinguishable from neutral expectations (table 1). If we apply a correction for multiple tests (Benjamini and Hochberg 1995), the excess nonsynonymous substitutions on the lineage leading to the derived copy of GA14530 are no longer significant ( $P = 0.092$ ). Elevated nonsynonymous divergence is indicative of historical positive selection fixing multiple beneficial amino acid changing mutations, suggesting that the derived copies of duplicated genes experience more adaptive evolution than the ancestral copies. Additionally, if we analyze all ancestral and derived copies separately, there is evidence that a larger fraction of amino acid fixations were driven by positive selection in the derived copies than the ancestral copies ( $\alpha_{\text{anc}} = 0.635$ ,  $\alpha_{\text{dup}} = 0.835$ ); our estimate of  $\alpha_{\text{dup}}$  falls outside the 95% CI under neutral expectations, whereas our estimate of  $\alpha_{\text{anc}}$  is within the 95% CI.

The comparison of polymorphism and divergence in the ancestral and derived copies of the duplicated genes will detect adaptive evolution that occurred prior to the expected neutral coalescence time of the sampled alleles (Przeworski 2002). More recent selective sweeps will affect the patterns of polymorphism around the sites under selection. If the faster rates of amino acid evolution in the derived copies of duplicated genes are the result of recurrent selective sweeps that are still occurring, there should be evidence for these sweeps in the site-frequency spectra of these genes. Tajima's (1989)  $D$  was calculated for the noncoding sequence and synonymous sites within and flanking the ancestral and derived copies of the duplicated genes, and  $D$  was standardized using its theoretical minimum,  $D_{\text{min}}$  (Schaeffer 2002). There is not a significant difference in  $D/D_{\text{min}}$  between the ancestral and derived copies (median ancestral =  $-0.2209$ , derived =  $-0.3058$ ;  $P = 0.804$ , Mann–Whitney test) (table 2), indicating that there is no evidence for more recent selective sweeps in the derived copies. Coalescent simulations using a neutral model with population expansion (Slatkin and Hudson 1991) show that 7/28 loci reject the null model (table 2). Six loci reject the null model because the frequency spectra have an excess of intermediate-frequency variants given the recent population expansion, whereas one locus shows a significant excess of rare variants. Three of the departures from neutrality are in derived copies of duplicated genes, but all three genes have more intermediate-frequency variants than expected (table 2). Thus, there is no evidence for recent selective sweeps in the sequences of the derived copies of duplicated genes.

Interestingly, three of the four derived copies with a significant excess of amino acid substitutions were duplicated from the neo-X chromosome, and of the four genes duplicated from the neo-X we examined, only one conformed to





**Fig. 3.** Relative rates of amino acid evolution for ancestral and derived copies of duplicated genes. The number of amino acid substitutions per codon is shown for substitutions along the lineages leading to the ancestral copies (closed bars) and derived copies (open bars). Significant differences in the number of substitutions between ancestral and derived copies using a chi-square test (Tajima 1993) are indicated by asterisks (\* $P < 0.05$ , \*\* $P < 0.005$ ). The mechanism responsible for each duplicated gene is given with the gene identifier of the ancestral copy: DNA duplication (D), retroposed duplications (R), and ambiguous duplications (A).

neutral expectations (table 1). The fraction of amino acid substitutions driven by positive selection is greater for the derived copies duplicated from the neo-X than those duplicated from other chromosome arms ( $\alpha_{\text{dup(not-neo-x)}} = 0.808$ ,  $\alpha_{\text{dup(neo-x)}} = 0.888$ ), but both are significantly different than the neutral expectation ( $P < 0.001$ ). There is also not a significant difference in  $D/D_{\text{min}}$  between the derived

copies of genes duplicated from the neo-X chromosome and those duplicated from other chromosome arms (median neo-X =  $-0.2793$ , not-neo-X =  $-0.3122$ ;  $P = 0.950$ , Mann-Whitney test) (table 2). Therefore, there is evidence from individual genes that those duplicated from the neo-X chromosome experience more positive selection on amino acid substitutions than genes duplicated from other

**Table 1.** Lineage-Specific McDonald–Kreitman Tests and Estimates of  $\omega$  for Ancestral and Derived Copies of Duplicated Genes.

Gene <sup>a</sup>	<i>n</i> <sup>b</sup>	Codons	A/D <sup>c</sup>	ME <sup>d</sup>	<i>d</i> <sub>N</sub>	<i>d</i> <sub>S</sub>	$\omega$	<i>D</i> <sub>NA</sub> / <i>D</i> <sub>ND</sub>	<i>D</i> <sub>SA</sub> / <i>D</i> <sub>SD</sub>	<i>P</i> <sub>NA</sub> / <i>P</i> <sub>ND</sub>	<i>P</i> <sub>SA</sub> / <i>P</i> <sub>SD</sub>	<i>P</i> <sup>e</sup>
<b>Ancestral copy not on neo-X</b>												
GA13587	13	439	anc	E	0.007	0.164	0.043	4	24	1	16	0.6353
GA22908			dup	A	0.112	0.201	0.559	91	36	1	9	0.0002
GA25437	12	201	anc	A	0.018	0.116	0.157	3	11	4	5	0.3630
			dup	B	0.047	0.320	0.147	15	22	2	7	0.4503
GA22671	13	36	anc	A	0.046	0.156	0.294	3	1	0	0	NA
GA25649			dup	B	0.072	0.315	0.228	5	4	0	1	1.0000
GA17441	13	109	anc	A	0.004	0.209	0.019	1	10	0	0	NA
			dup	B	0.090	0.192	0.468	17	10	1	5	0.0701
GA28030	12	177	anc	B	0.008	0.046	0.166	3	3	0	2	0.4643
			dup	C	0.037	0.110	0.333	13	5	10	9	0.3133
GA26276	13	253	anc	E	0.006	0.142	0.042	2	15	3	8	0.3531
			dup	C	0.071	0.567	0.126	35	41	2	9	0.1076
<b>Ancestral copy on neo-X</b>												
GA23834	9	97	anc	D	0.000	0.015	0.000	0	2	0	1	NA
			dup	E	0.058	0.155	0.374	11	8	0	2	0.2143
GA11342	11	288	anc	D	0.006	0.099	0.063	4	9	1	5	0.6311
			dup	B	0.230	0.283	0.813	113	27	13	12	0.0039
GA14530	11	78	anc	D	0.005	0.324	0.016	1	8	0	0	NA
			dup	E	0.051	0.000	$\infty$	8	2	2	8	0.0230
GA23771	13	221	anc	D	0.046	0.071	0.648	17	6	1	2	0.2154
GA25189			dup	B	0.041	0.017	2.389	15	2	6	12	0.0016

NOTE.—NA, not applicable.

<sup>a</sup> Identifier of the *Drosophila pseudoobscura* ancestral copy; the identifier of the derived copy is also presented if the gene has been previously annotated.

<sup>b</sup> Number of chromosomes sampled.

<sup>c</sup> Whether estimates are from the ancestral (anc) or derived (dup) copy.

<sup>d</sup> Muller element location of gene.

<sup>e</sup> *P* value of Fisher's exact test.

**Table 2.** Noncoding Polymorphism in Ancestral and Derived Copies of Duplicated Genes and Tajima's *D* Statistic.

Gene <sup>a</sup>	<i>n</i> <sup>b</sup>	Ancestral					Derived				
		ME <sup>c</sup>	Sites <sup>d</sup>	<i>S</i> <sub>nc</sub> <sup>e</sup>	<i>k</i> <sub>nc</sub> <sup>f</sup>	<i>D</i> / <i>D</i> <sub>min</sub> <sup>g</sup>	ME <sup>c</sup>	Sites <sup>d</sup>	<i>S</i> <sub>nc</sub> <sup>e</sup>	<i>k</i> <sub>nc</sub> <sup>f</sup>	<i>D</i> / <i>D</i> <sub>min</sub> <sup>g</sup>
<b>Ancestral copy not on neo-X</b>											
GA13587	13	E	626.9	46	11.8	−0.360	A	453.1	19	5.3	−0.241
GA25437	12	A	774.8	16	3.5	−0.632	B	353.2	25	7.8	−0.098 <sup>ei</sup>
GA22671	13	A	560.8	51	8.7	−0.837 <sup>er</sup>	B	363.7	11	3.0	−0.265
GA17441	13	A	411.0	11	3.2	−0.190	B	865.6	83	18.6	−0.545
GA17928	12	B	775.2	52	15.1	−0.232	A	550.7	15	3.3	−0.624
GA29002	13	B	902.0	48	15.0	−0.059 <sup>ei</sup>	A	490.3	33	8.5	−0.366
GA28030	12	B	1261.5	43	12.6	−0.210	C	615.7	89	23.8	−0.359
GA26276	13	E	646.1	39	8.9	−0.523	C	386.8	16	5.3	0.060 <sup>ei</sup>
<b>Ancestral copy on neo-X</b>											
GA17756	12	D	745.4	50	15.5	−0.120 <sup>ei</sup>	B	153.3	4	1.0	−0.470
GA24652	12	D	1267.0	11	3.5	−0.068	E	457.4	36	10.6	−0.212
GA23834	9	D	758.2	13	5.0	0.083 <sup>ei</sup>	E	294.4	17	5.7	−0.198
GA11342	11	D	600.3	11	2.4	−0.696	B	576.2	68	19.1	−0.346
GA14530	11	D	443.2	3	1.0	−0.076	E	229.5	35	11.8	−0.031 <sup>ei</sup>
GA23771	13	D	266.1	6	1.4	−0.454	B	641.7	48	9.8	−0.657

NOTE.—“ei” indicates that the null model was rejected because of an excess of intermediate-frequency variant and “er” indicates that the null model was rejected because of an excess of rare variants.

<sup>a</sup> Identifier of the *Drosophila pseudoobscura* ancestral copy.

<sup>b</sup> Number of chromosomes sampled.

<sup>c</sup> Muller element.

<sup>d</sup> Number of noncoding sites analyzed.

<sup>e</sup> Number of noncoding segregating sites.

<sup>f</sup> Average pairwise differences at noncoding sites.

<sup>g</sup> Departures from neutral expectations with modest population expansion  $Nr = 7$  (Slatkin and Hudson 1991).

chromosome arms, but the evidence for increased adaptive evolution is not conclusive and the adaptive evolution does not appear to be recent.

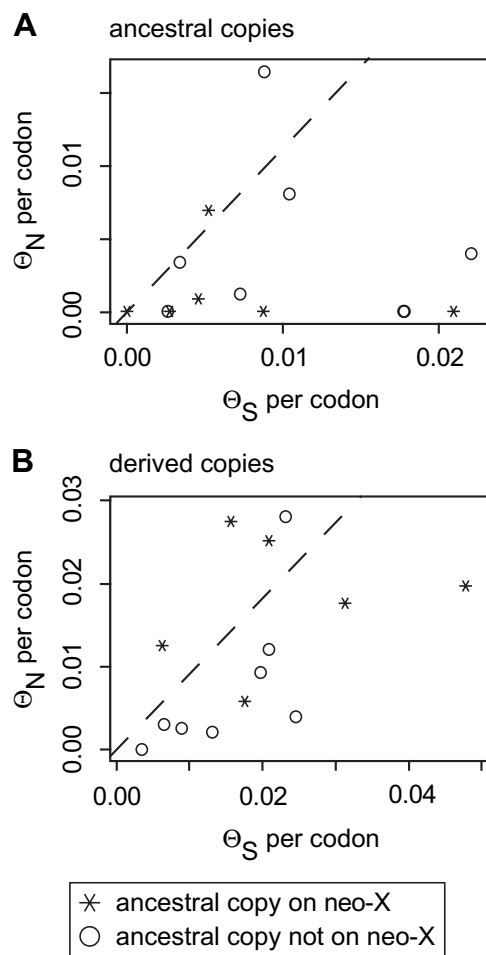
### Relaxed Selective Constraints on the Derived Copies of Duplicated Genes

Four derived copies showed evidence for historical positive selection in the McDonald–Kreitman test (table 1), but there is no evidence for recent adaptive evolution from the site-frequency spectra of these genes (table 2). This lead us to investigate what the current selection pressures are on the derived copies of duplicated genes. Genes that are under relaxed selective constraints should have elevated levels of nonsynonymous polymorphism when compared with the synonymous polymorphism in those genes (Li 1997; Nei and Kumar 2000; Wang et al. 2004). We tested for differences in the selective constraints between the ancestral and derived copies of the 14 duplicated genes by estimating  $\theta$  (Watterson 1975) at nonsynonymous ( $\theta_N$ ) and synonymous ( $\theta_S$ ) sites separately as a measure of polymorphism within each gene. There is elevated nonsynonymous polymorphism in the derived copies (median  $\theta_{N(\text{ancestral})} = 0.0005$ ,  $\theta_{N(\text{derived})} = 0.0107$ ;  $P < 0.05$ , paired Mann–Whitney test), but there is not significantly elevated synonymous polymorphism in the derived copies (median  $\theta_{S(\text{ancestral})} = 0.0080$ ,  $\theta_{S(\text{derived})} = 0.0186$ ;  $P = 0.069$ , paired Mann–Whitney test) (fig. 4). This same pattern holds if we correct for the potentially lower  $N_e$  of the X chromosome. Balancing selection can also explain elevated levels of polymorphism (Strobeck 1983; Hudson and Kaplan 1988; Charlesworth 2006), but this is not likely for our data because the derived copies do not have an excess of

intermediate-frequency variants at synonymous and noncoding sites (table 2).

We tested for differences in selective constraints between genes duplicated from the neo-X chromosome and those duplicated from other chromosome arms by comparing the amount of nonsynonymous polymorphism in these genes. Interestingly, the derived copies of genes duplicated from the neo-X chromosome have significantly higher  $\theta_N$  than those duplicated from other chromosome arms (median neo-X = 0.0188, not-neo-X = 0.0034;  $P < 0.05$ , Mann–Whitney test) (fig. 4). This same pattern holds if we correct for the potentially lower  $N_e$  of the X chromosome. Additionally, we compared  $k_N$  for each gene with our estimate of  $\theta_N$  from the number of polymorphic sites, allowing us to determine if the excess  $\theta_N$  is due to low- or intermediate-frequency alleles. Both for genes duplicated from the neo-X and for those duplicated from other chromosome arms, the average difference between  $k_N$  and  $\theta_N$  is  $>0$  (median neo-X =  $1.040 \times 10^{-3}$ , not-neo-X =  $3.133 \times 10^{-4}$ ), and this difference is greater for genes duplicated from the neo-X chromosome ( $P < 0.05$ , Mann–Whitney test). Therefore, the excess nonsynonymous polymorphism in genes duplicated from the neo-X chromosome is the result of an excess of intermediate-frequency nonsynonymous variants.

The difference in  $\theta_N$  between genes duplicated from the neo-X and those duplicated from other chromosome arms may be the result of relaxed constraints on genes duplicated from the neo-X chromosome, balancing selection on genes duplicated from the neo-X or higher rates of mutation in those genes. Balancing selection is unlikely to explain this pattern because only one gene duplicated from the neo-X has an excess of intermediate-frequency variants



**FIG. 4.** Nonsynonymous and synonymous polymorphism in paralogs. Each asterisk represents a duplicated gene with an ancestral copy on the neo-X chromosome, whereas each circle represents a duplicated gene with an ancestral copy on another chromosome arm. Data for ancestral (A) and derived (B) copies are graphed separately. Polymorphism ( $\theta$ ) is measured as the number of mutations per-codon standardized for the number of sequences in the sample. The dashed line indicates values at which nonsynonymous and synonymous polymorphism are equal.

(when we examine noncoding sites), whereas two genes duplicated from other chromosome arms have an excess of intermediate-frequency variants (table 2). We reject the mutation rate hypothesis because  $\theta_S$  is not significantly different between genes duplicated from the neo-X and those duplicated from other chromosome arms (median neo-X = 0.0192, not-neo-X = 0.0164;  $P = 0.207$ , Mann–Whitney test) (fig. 4). Differences in nonsynonymous polymorphism may also reflect nascent properties of the types of genes duplicated from the neo-X chromosome. However,  $\theta_N$  in the ancestral copies on the neo-X chromosome is not higher than  $\theta_N$  in ancestral copies on other chromosome arms (median neo-X = 0.0000, not-neo-X = 0.0023;  $P = 0.241$ , Mann–Whitney test) (fig. 4). Therefore, it appears that the derived copies of genes duplicated from the neo-X chromosome in *D. pseudoobscura* are currently under less selective constraints than those duplicated from other chromosome arms.

Finally, it is unlikely that our sequencing strategy affected these results. For each duplicated gene with an ancestral copy on the neo-X chromosome, a median of three strains were replaced because of PCR failure or heterozygosity in the sequences, compared with two strains on average for genes duplicated from the other chromosome arms ( $P < 0.005$ , Mann–Whitney test). Both PCR failure and heterozygosity are expected to be more common in genes with greater polymorphism. Therefore, the exclusion of more strains for genes duplicated from the neo-X chromosome is consistent with the elevated  $\theta_N$  in the derived copies of genes duplicated from the neo-X chromosome. Also, excluding sequences in the manner we did should decrease the measured amount of polymorphism in the sampled genes, and the effects should be greatest in the class of genes with the most strains excluded (because genes with more heterozygosity should have more strains removed). Thus, the finding of excess  $\theta_N$  in the derived copies of genes duplicated from the neo-X chromosome is conservative because these genes had more strains excluded; it is unclear how many nucleotide or indel changes were responsible for the exclusions. There is not a significant difference in the number of strains added to replace the excluded strains between genes duplicated from the neo-X with those duplicated from other chromosome arms ( $P = 0.292$ , Mann–Whitney test).

## Discussion

We examined expression profiles of, polymorphism within, and divergence between 14 genes that were duplicated in the *D. pseudoobscura* genome after the split with the *D. melanogaster* lineage. Most of the ancestral copies are broadly expressed and/or testis/ovary expressed, which is consistent with retrotransposition as the primary mechanism of duplication of these genes (Langille and Clark 2007). The elevated nonsynonymous polymorphism in the derived copies (fig. 4) suggests that they are under less selective constraints than their ancestral paralogs. Additionally, the derived copies have narrower expression profiles than the ancestral copies (fig. 2), suggesting that the derived copies perform fewer functions. It has been observed that the protein-coding sequences of more broadly expressed genes are under more selective constraints than those with more tissue-specific expression profiles (Duret and Mouchiroud 2000; Zhang and Li 2004; Liao et al. 2006; Larracuenté et al. 2008). Therefore, the derived copies could be under relaxed selective constraints because they are not as pleiotropic as the ancestral copies.

There is also some evidence for historical positive selection on the derived copies. The amino acid sequences of the derived copies evolve at faster rates than the ancestral copies (fig. 3 and table 1), and this rate acceleration is partially attributable to positive selection (table 1). Interestingly, there is no evidence for recent selective sweeps on the derived copies (table 2), suggesting that any positive selection has occurred prior to the coalescence time of the sampled alleles (Przeworski 2002). Alternatively,



selection could have acted on standing genetic variation (Barrett and Schluter 2008), which is not expected to leave the characteristic signature of a deficiency of polymorphism that we tested for in the sight-frequency spectrum (Przeworski et al. 2005). Consistent with this scenario, we observed elevated nonsynonymous polymorphism in the derived copies (fig. 4), which may act as a pool of variation upon which natural selection could act.

A neutral explanation for excess nonsynonymous substitutions in the McDonald–Kreitman test has been presented in the context of duplicated genes: the derived copies may have been under relaxed selective constraints shortly after the duplication events, followed by greater selective constraints more recently (Hahn 2009). However, the genes in our data set with a significant excess of nonsynonymous fixations do not have low levels of nonsynonymous polymorphism (table 1). In general, the derived copies are currently under relaxed constraints, which would require them to be under even more relaxed constraints in the past if the excess amino acid fixations were not driven by positive selection. Such extremely relaxed constraints are unlikely, and we conclude that recent increases in selective constraints probably do not explain the excess nonsynonymous fixations along the lineages leading to the derived copies.

Why are the derived copies of these duplicated genes experiencing so much adaptive evolution? First, pleiotropic constraints are predicted to limit adaptive evolution (Fisher 1930), and the relaxed selective constraints on the duplicated genes may expose them to more positive selection. It also is worth noting that the derived copies tend to be testis expressed (fig. 2) because male-biased genes in *Drosophila* are known to evolve under positive selection (Swanson et al. 2001; Meiklejohn et al. 2003; Zhang et al. 2004). Additionally, male-biased genes have higher rates of amino acid evolution than unbiased and female-biased genes in *D. pseudoobscura* (Jiang and Machado 2009). The testis expression of the derived copies suggests that their adaptive evolution may be attributable to a yet-to-be-characterized role they play in reproduction (Swanson and Vacquier 2002). Most of these genes have not been functionally characterized and many do not have predicted functions, which limits our ability to speculate as to what phenotypes were under selection. For example, two of the genes with strong signatures of positive selection, the paralogs of GA13587 and GA23771 (table 1), are uncharacterized (<http://flybase.org/reports/FBgn0039601.html>) or predicted to be involved in mRNA export from the nucleus (<http://flybase.org/reports/FBgn0052135.html>), respectively.

Although most of these genes have not been functionally characterized, the ones that have reveal interesting possibilities regarding the evolutionary dynamics of these duplicated genes. For example, *Pros28.1* (*Pro $\alpha$ 4*) is X linked in *D. melanogaster*, and it encodes the broadly expressed  $\alpha$ 4 proteasome subunit (Haass et al. 1990). *Pros28.1* has two paralogs in the *D. melanogaster* genome, *Pros28.1A* (*Pro $\alpha$ 4-t1*) and *Pros28.1B* (*Pro $\alpha$ 4-t2*), which are both

autosomal and encode spermatogenesis-specific isoforms of the  $\alpha$ 4 subunit (Yuan et al. 1996). *Pros28.1A* was created by the retroposition of *Pros28.1*, and it is under relaxed selective constraints (Torgerson and Singh 2004). The *D. pseudoobscura* homolog of *Pros28.1* (GA17441) is on the ancestral X chromosome and also gave rise to an autosomal retrogene that appears to be under relaxed constraints—it is rapidly evolving relative to its ancestral paralog (fig. 3), but the excess amino acid substitutions are not significantly greater than expected based on the amount of polymorphism in the gene (table 1). Although there is no significant signature of positive selection in the protein-coding sequence of these duplicated genes in *D. melanogaster* or *D. pseudoobscura*, the independent duplication of this gene along multiple evolutionary lineages (Belote et al. 1998; Belote and Zhong 2009), and subsequent testis expression of one copy suggests that it is advantageous to have testis-specific copies of protease subunits. The autosomal derived copies of *Pros28.1* may allow for expression of this subunit during spermatogenic X inactivation (Betrán et al. 2002).

Two of the genes duplicated from the neo-X chromosome (GA11342 and GA14530) encode proteins involved in chromosome segregation. The gene product of the GA11342 ortholog *Cdc37* is involved in cell cycle control (Cutforth and Rubin 1994), and mutations in *Cdc37* lead to defects in chromosome segregation in both mitosis and male meiosis (Lange et al. 2002). The autosomal derived copy of GA11342 in *D. pseudoobscura* is highly expressed in testis (fig. 2) and has experienced a significant excess of amino acid substitutions (fig. 3 and table 1). We hypothesize that the derived copy of this duplicated gene has evolved a spermatogenesis-specific function to compensate for the inactivation of the ancestral copy when it became X linked.

Another gene duplicated from the *D. pseudoobscura* neo-X chromosome, GA14530 (the ortholog of *mad2*), encodes a mitotic spindle assembly checkpoint protein (Li and Murray 1991; Shah and Cleveland 2000). The derived copy is rapidly evolving (fig. 3 and table 1), and it is highly expressed in testis, whereas the ancestral copy seems to be enriched in ovary (fig. 2). Interestingly, unlike the other duplicated genes described here, the derived copy is more broadly expressed than the ancestral copy (fig. 2). There is evidence that the *mad2* gene product is involved in the meiotic checkpoint (Nicklas et al. 2001), and it appears that the two copies of *mad2* in *D. pseudoobscura* have been specialized for sex-specific meiotic functions (the ancestral copy for oogenic meiosis and the derived copy for spermatogenic meiosis). We hypothesize that the ancestral copy is female biased because of the unique properties of the spindle apparatus in female meiosis (Orr-Weaver 1995). The derived copy on the autosomes would encode the *mad2* protein used in male meiosis (because the X chromosome is inactivated in late spermatogenesis) and in mitosis (because the mitotic spindle resembles the male meiotic spindle), whereas the neo-X-linked ancestral copy would encode the *mad2* protein that performs the oogenesis-specific

functions. The *D. melanogaster* copy of *mad2* is primarily expressed in ovary but also moderately expressed in testis (Chintapalli, Wang, and Dow 2007). If optimizing the function of this gene in ovary is in conflict with optimization of the gene in testis and soma, the duplication of the gene from the neo-X to an autosome in *D. pseudoobscura* may have allowed this gene to overcome an adaptive conflict (Piatigorsky and Wistow 1991; Hughes 1994; Hittinger and Carroll 2007; Des Marais and Rausher 2008). This scenario is expected to lead to adaptive evolution in both copies (Des Marais and Rausher 2008), but we only see evidence for positive selection in the derived copy, whereas the ancestral copy is highly constrained (fig. 3 and table 1). It is possible that the ancestral copy retained a function that was optimized for oogenic meiosis, whereas the derived copy evolved male meiotic and mitotic specialization. Functional experimentation on these genes in *D. pseudoobscura* is necessary to further evaluate this hypothesis.

A burst of duplication followed the creation of the *D. pseudoobscura* neo-X chromosome (Meisel et al. 2009) which may have been driven by the need to escape X inactivation (Betrán et al. 2002) or to resolve sexually antagonistic conflicts (Wu and Xu 2003). Examples of individual genes described above seem to support both hypotheses. Additionally, multiple genes duplicated from the neo-X have signatures of adaptive evolution (table 1), and genes duplicated from the neo-X appear to be under relaxed selective constraints relative to genes duplicated from other chromosome arms (fig. 4). We hypothesize that the relaxed constraints on the genes duplicated from the neo-X chromosome are in part responsible for the elevated rate of adaptive fixation of amino acid substitutions in these genes. This works within Fisher's (1930) model of natural selection, but the elevated nonsynonymous polymorphism may also prime the genes for adaptive evolution. If, as we propose, selection is acting on standing genetic variants, it is reasonable to assume that the more variation a gene harbors, the more potential targets for positive selection there are. The genes duplicated from the neo-X have more intermediate-frequency nonsynonymous variants, which could provide targets for natural selection if there are changes in selection pressures (including changes in sexual selection pressures as expected in scenarios of sexually antagonistic selection [Rice and Holland 1997]). Unfortunately, this hypothesis does not explain why genes duplicated from the neo-X are under relaxed selective constraints in the first place. If selective constraints do correlate with functional scope, then we predict that the derived copies of genes duplicated from the neo-X would have more specific functions than genes duplicated from other chromosome arms. There is no evidence that genes duplicated from the neo-X have more sex-biased expression or a narrower range of expression than those duplicated from the other chromosome arms (fig. 2), but gene expression is a very coarse measure of gene function. Further experimentation is necessary to determine the functional scope of genes duplicated from the neo-X and those duplicated from other chromosome arms in order to test this hypothesis.

## Supplementary Material

Supplementary tables S1–S6 and other materials are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

DNA sequencing was carried out at the Penn State University Nucleic Acids Facility and Cornell University Life Sciences Core Laboratories Center, and we thank A. Wallace for assistance with sequencing. Members of the Clark laboratory and the Cornell M.B.G. department provided useful discussions that improved the quality of this article, and John True and multiple anonymous reviewers aided in producing a manuscript suitable for publication. This work was supported by the National Science Foundation (DEB 0608186 to R.P.M. and S.W.S.) and the National Institutes of Health (F32 GM087611 to R.P.M. and A.G. Clark and 5R01 GM076007-5 to D. Bachtrög). Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or National Institutes of Health. All sequences described in this paper have been deposited in the GenBank database (accession nos. HM017513–HM017824).

## References

- Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* 139:1067–1076.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Andolfatto P. 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol Biol Evol.* 18:279–290.
- Barrett RDH, Schluter D. 2008. Adaptation from standing genetic variation. *Trends Ecol Evol.* 23:38–44.
- Belote JM, Miller M, Smyth KA. 1998. Evolutionary conservation of a testes-specific proteasome subunit gene in *Drosophila*. *Gene* 215:93–100.
- Belote JM, Zhong L. 2009. Duplicated proteasome subunit genes in *Drosophila* and their roles in spermatogenesis. *Heredity* 103:23–31.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B.* 57:289–300.
- Betrán E, Bai Y, Motiwale M. 2006. Fast protein evolution and germ line expression of a *Drosophila* parental gene and its young retroposed paralog. *Mol Biol Evol.* 23:2191–2202.
- Betrán E, Long M. 2003. Dntf-2r, a young *Drosophila* retroposed gene with specific male expression under positive Darwinian selection. *Genetics* 164:977–988.
- Betrán E, Rozas J, Navarro A, Barbadilla A. 1997. The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. *Genetics* 146:89–99.
- Betrán E, Thornton K, Long M. 2002. Retroposed new genes out of the X in *Drosophila*. *Genome Res* 12:1854–1859.

- Bingham PM, Levis R, Rubin GM. 1981. Cloning of DNA sequences from the white locus of *D. melanogaster* by a novel and general method. *Cell* 25:693–704.
- Bradley J, Baltus A, Skaletsky H, Royce-Tolland M, Dewar K, Page DC. 2004. An X-to-autosome retrogene is required for spermatogenesis in mice. *Nat Genet.* 36:872–876.
- Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* 2:e64.
- Chintapalli VR, Wang J, Dow JAT. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet.* 39:715–720.
- Connallon T, Knowles LL. 2005. Intergenomic conflict revealed by patterns of sex-biased gene expression. *Trends Genet.* 21:495–499.
- Cusack BP, Wolfe KH. 2007. Not born equal: increased rate asymmetry in relocated and retrotransposed rodent gene duplicates. *Mol Biol Evol.* 24:679–686.
- Cutforth T, Rubin GM. 1994. Mutations in *Hsp83* and *cdc37* impair signaling by the sevenless receptor tyrosine kinase in *Drosophila*. *Cell* 77:1027–1036.
- Dai H, Yoshimatsu TF, Long M. 2006. Retrogene movement within- and between-chromosomes in the evolution of *Drosophila* genomes. *Gene* 385:96–102.
- Dass B, Tardif S, Park JY, Tian B, Weitlauf HM, Hess RA, Carnes K, Griswold MD, Small CL, MacDonald CC. 2007. Loss of polyadenylation protein  $\tau$ CstF-64 causes spermatogenic defects and male infertility. *Proc Natl Acad Sci U S A.* 104:20374–20379.
- Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454:762–765.
- Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.
- Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol.* 17:68–70.
- Fisher RA. 1930. The genetical theory of natural selection. Oxford: Clarendon Press.
- Gloor GB, Preston CR, Johnson-Schlitz DM, Nassif NA, Phillis RW, Benz WK, Robertson HM, Engels WR. 1993. Type I repressors of P element mobility. *Genetics* 135:81–95.
- Haass C, Pesold-Hurt B, Multhaup G, Beyreuther K, Kloetzel PM. 1990. The *Drosophila* PROS-28.1 gene is a member of the proteasome gene family. *Gene* 90:235–241.
- Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered.* 100:605–617.
- Hahn MW, Han MV, Han S-G. 2007. Gene family evolution across 12 *Drosophila* genomes. *PLoS Genet.* 3:e197.
- Hense W, Baines JF, Parsch J. 2007. X chromosome inactivation during *Drosophila* spermatogenesis. *PLoS Biol.* 5:e273.
- Hittinger CT, Carroll SB. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449:677–681.
- Hudson R, Boos D, Kaplan N. 1992. A statistical test for detecting geographic subdivision. *Mol Biol Evol.* 9:138–151.
- Hudson RR. 2000. A new statistic for detecting genetic differentiation. *Genetics* 155:2011–2014.
- Hudson RR, Kaplan NL. 1988. The coalescent process in models with selection and recombination. *Genetics* 120:831–840.
- Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B Biol Sci.* 256:119–124.
- Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W. 2007. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosome-wide single nucleotide polymorphism data. *Genetics* 177:469–480.
- Jiang Z-F, Machado CA. 2009. Evolution of sex-dependent gene expression in three recently diverged species of *Drosophila*. *Genetics* 183:1175–1185.
- Kalamegham R, Sturgill D, Siegfried E, Oliver B. 2007. *Drosophila* *mojless*, a retroposed GSK-3, has functionally diverged to acquire an essential role in male fertility. *Mol Biol Evol.* 24:732–742.
- Kelly WG, Schaner CE, Dernburg AF, Lee M-H, Kim SK, Villeneuve AM, Reinke V. 2002. X-chromosome silencing in the germline of *C. elegans*. *Development* 129:479–492.
- Kovacevic M, Schaeffer SW. 2000. Molecular population genetics of X-linked genes in *Drosophila pseudoobscura*. *Genetics* 156:155–172.
- Lange BM, Rebollo E, Herold A, Gonzalez C. 2002. *Cdc37* is essential for chromosome segregation and cytokinesis in higher eukaryotes. *EMBO J.* 21:5364–5374.
- Langille MGI, Clark DV. 2007. Parent genes of retrotransposition-generated gene duplicates in *Drosophila melanogaster* have distinct expression profiles. *Genomics* 90:334–343.
- Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG. 2008. Evolution of protein-coding genes in *Drosophila*. *Trends Genet.* 24:114–123.
- Li R, Murray AW. 1991. Feedback control of mitosis in budding yeast. *Cell* 66:519–531.
- Li W-H. 1997. Molecular evolution. Sunderland (MA): Sinauer.
- Liao B-Y, Scott NM, Zhang J. 2006. Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. *Mol Biol Evol.* 23:2072–2080.
- Maciejowski J, Ahn JH, Cipriani PG, et al. (12 co-authors). 2005. Autosomal genes of autosomal/X-linked duplicated gene pairs and germ-line proliferation in *Caenorhabditis elegans*. *Genetics* 169:1997–2011.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654.
- Meiklejohn CD, Parsch J, Ranz JM, Hartl DL. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. *Proc Natl Acad Sci U S A.* 100:9894–9899.
- Meisel RP. 2009a. Repeat mediated gene duplication in the *Drosophila pseudoobscura* genome. *Gene* 438:1–7.
- Meisel RP. 2009b. Evolutionary dynamics of recently duplicated genes: selective constraints on diverging paralogs in the *Drosophila pseudoobscura* genome. *J Mol Evol.* 69:81–93.
- 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.
- Mikhaylova LM, Nguyen K, Nurminsky DI. 2008. Analysis of the *Drosophila melanogaster* testes transcriptome reveals coordinate regulation of paralogous genes. *Genetics* 179:305–315.
- Muller HJ. 1940. Bearings of the 'Drosophila' work on systematics. In: Huxley J, editor. The new systematics. Oxford: Clarendon Press. p. 185–268.
- Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. New York: Oxford University Press.
- Nicklas RB, Waters JC, Salmon ED, Ward SC. 2001. Checkpoint signals in grasshopper meiosis are sensitive to microtubule attachment, but tension is still essential. *J Cell Sci.* 114:4173–4183.
- Orr-Weaver TL. 1995. Meiosis in *Drosophila*: seeing is believing. *Proc Natl Acad Sci U S A.* 92:10443–10449.
- 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.
- Patefield WM. 1981. An efficient method of generating r x c tables with given row and column totals. *Appl Stat.* 30:91–97.



- 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.
- Patterson JT, Stone WS. 1952. Evolution in the Genus *Drosophila*. New York: The Macmillan Company.
- Piatigorsky J, Wistow G. 1991. The recruitment of crystallins: new functions precede gene duplication. *Science* 252:1078–1079.
- Potrzebowski L, Vinckenbosch N, Marques AC, Chalmel F, Jégou B, Kaessmann H. 2008. Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biol.* 6:e80.
- Powell JR. 1997. Progress and prospects in evolutionary biology: the *Drosophila* model. New York: Oxford University Press.
- Przeworski M. 2002. The signature of positive selection at randomly chosen loci. *Genetics* 160:1179–1189.
- Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic variation. *Evolution* 59:2312–2323.
- 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.
- Rice WR, Holland B. 1997. The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behav. Ecol. Sociobiol.* 41:1–10.
- Richards S, Liu Y, Bettencourt BR, et al. (52 co-authors). 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* 15:1–18.
- Riley MA, Hallas ME, Lewontin RC. 1989. Distinguishing the forces controlling genetic variation at the *Xdh* locus in *Drosophila pseudoobscura*. *Genetics* 123:359–369.
- Rohozinski J, Bishop CE. 2004. The mouse juvenile spermatogonial depletion (*jsd*) phenotype is due to a mutation in the X-derived retrogene, *mUtp14b*. *Proc Natl Acad Sci U S A.* 101:11695–11700.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
- Russo CA, Takezaki N, Nei M. 1995. Molecular phylogeny and divergence times of drosophilid species. *Mol Biol Evol.* 12:391–404.
- Schaeffer SW. 2002. Molecular population genetics of sequence length diversity in the *Adh* region of *Drosophila pseudoobscura*. *Genet Res Camb.* 80:163–175.
- Schaeffer SW, Bhutkar A, McAllister BF, et al. (38 co-authors). 2008. Polytene chromosomal maps of 11 *Drosophila* species: the order of genomic scaffolds inferred from genetic and physical maps. *Genetics* 179:1601–1655.
- Schaeffer SW, Goetting-Minesky MP, Kovacevic M, Peoples JR, Graybill JL, Miller JM, Kim K, Nelson JC, Anderson WW. 2003. Evolutionary genomics of inversions in *Drosophila pseudoobscura*: evidence for epistasis. *Proc Natl Acad Sci U S A.* 100:8319–8324.
- Schaeffer SW, Miller EL. 1992. Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. *Genetics* 132:471–480.
- Shah JV, Cleveland DW. 2000. Waiting for anaphase: mad2 and the spindle assembly checkpoint. *Cell* 103:997–1000.
- Shiao M-S, Khil P, Camerini-Otero RD, Shiroishi T, Moriwaki K, Yu H-T, Long M. 2007. Origins of new male germ-line functions from X-derived autosomal retrogenes in the mouse. *Mol Biol Evol.* 24:2242–2253.
- Singh N, Macpherson JM, Jensen J, Petrov D. 2007. Similar levels of X-linked and autosomal nucleotide variation in African and non-African populations of *Drosophila melanogaster*. *BMC Evol Biol.* 7:202.
- Slatkin M, Hudson RR. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555–562.
- Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* 415:1022–1024.
- Steinemann M, Pinsker W, Sperlich D. 1984. Chromosome homologies within the *Drosophila obscura* group probed by in situ hybridization. *Chromosoma* 91:46–53.
- Strobeck C. 1983. Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics* 103:545–555.
- Sturgill D, Zhang Y, Parisi M, Oliver B. 2007. Demasculinization of X chromosomes in the *Drosophila* genus. *Nature* 450:238–241.
- Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc Natl Acad Sci U S A.* 98:7375–7379.
- Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nat Rev Genet.* 3:137–144.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tajima F. 1993. Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* 135:599–607.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596–1599.
- Teshima KM, Innan H. 2004. The effect of gene conversion on the divergence between duplicated genes. *Genetics* 166:1553–1560.
- Torgerson DG, Singh RS. 2004. Rapid evolution through gene duplication and subfunctionalization of the testes-specific  $\alpha 4$  proteasome subunits in *Drosophila*. *Genetics* 168:1421–1432.
- Tracy C, Rio J, Motiwale M, Christensen SM, Betran E. 2010. Convergently recruited nuclear transport retrogenes are male biased in expression and evolving under positive selection in *Drosophila*. *Genetics Advance Access published January 11, 2010*, doi:10.1534/genetics.109.113522.
- Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25:1105–1111.
- Turner JMA. 2007. Meiotic sex chromosome inactivation. *Development.* 134:1823–1831.
- Vibrantovski MD, Lopes HF, Karr TL, Long M. 2009. Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. *PLoS Genet.* 5:e1000731.
- Vibrantovski MD, Zhang Y, Long M. 2009. General gene movement off the X chromosome in the *Drosophila* genus. *Genome Res.* 19:897–903.
- Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet.* 7:645–653.
- 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.
- Wang PJ, Page DC. 2002. Functional substitution for TAFII250 by a retroposed homolog that is expressed in human spermatogenesis. *Hum Mol Genet.* 11:2341–2346.
- Wang X, Thomas SD, Zhang J. 2004. Relaxation of selective constraint and loss of function in the evolution of human bitter taste receptor genes. *Hum Mol Genet.* 13:2671–2678.
- Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 7:256–276.
- Wells RS. 1996. Nucleotide variation at the *Gpdh* locus in the genus *Drosophila*. *Genetics* 143:375–384.
- Wu C-I, Xu EY. 2003. Sexual antagonism and X inactivation—the SAXI hypothesis. *Trends Genet.* 19:243–247.

- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
- Yuan X, Miller M, Belote JM. 1996. Duplicated proteasome subunit genes in *Drosophila melanogaster* encoding testes-specific isoforms. *Genetics* 144:147–157.
- Zhai W, Nielsen R, Slatkin M. 2009. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. *Mol Biol Evol.* 26:273–283.
- Zhang L, Li W-H. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Mol Biol Evol.* 21:236–239.
- Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. 2007. Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature* 450:233–237.
- Zhang Z, Hambuch TM, Parsch J. 2004. Molecular evolution of sex-biased genes in *Drosophila*. *Mol Biol Evol.* 21:2130–2139.