# Nonrandom Gene Loss from the *Drosophila miranda* Neo-Y Chromosome

Vera B. Kaiser*, Qi Zhou, and Doris Bachtrog

Department of Integrative Biology, University of California–Berkeley

*Corresponding author: E-mail: vera.kaiser@berkeley.edu.

**Accepted:** 2 October 2011

## Abstract

A lack of recombination leads to the degeneration of an evolving Y chromosome. However, it is not known whether gene loss is largely a random process and primarily driven by the order in which mutations occur or whether certain categories of genes are lost less quickly than others; the latter would imply that selection counteracts the degeneration of Y chromosomes to some extent. In this study, we investigate the relationship between putative ancestral expression levels of neo-Y–linked genes in *Drosophila miranda* and their rates of degeneration. We use RNA-Seq data from its close relative *Drosophila pseudoobscura* to show that genes that have become nonfunctional on the *D. miranda* neo-Y had, on average, lower ancestral transcript levels and were expressed in fewer tissues compared with genes with intact reading frames. We also show that genes with male-biased expression are retained for longer on the neo-Y compared with female-biased genes. Our results imply that gene loss on the neo-Y is not a purely random, mutation-driven process. Instead, selection is—at least to some extent—preserving the function of genes that are more costly to lose, despite the strongly reduced efficacy of selection on the neo-Y chromosome.

**Key words:** sex chromosomes, evolution, *Drosophila pseudoobscura*, sex-biased gene expression, tissue specificity, degeneration.

## Introduction

After recombination becomes suppressed between a pair of proto-sex chromosomes, the Y undergoes a process of degeneration and loses most of its functional genes, due to a reduced efficacy of natural selection in a nonrecombining genomic region. This results in small and gene-poor Y chromosomes, such as the human or Drosophila Y (Charlesworth B and Charlesworth D 2000; Skaletsky et al. 2003; Carvalho et al. 2009). In the initial stages of sex chromosome evolution, there exists a gene dosage problem for genes that are lost on the Y because males only produce functional gene products from their X-linked copies. In the long run, mechanisms of dosage compensation have evolved in a range of taxa to equalize expression levels between the X and autosomes (Bull 1983; Charlesworth 1996). However, in the initial phases of Y degeneration, losing Y-linked genes is likely to be costly, and reduced gene dose might be more deleterious for certain classes of sex-linked genes compared with others. However, with a strongly reduced effective population size of an evolving Y chromosome (Felsenstein 1974; Charlesworth et al. 2009; Kaiser and Charlesworth 2010), it is unclear whether selection is efficiently discriminating between genes that are more or less costly to lose. Instead, the order of gene loss on the Y could mostly be a random process, depending primarily on which genes happen to mutate first (Bachtrog 2006).

There are reasons to assume that ancestral gene expression levels may be a crucial factor influencing the rate of gene loss on a degenerating Y chromosome. In particular, genes expressed at low levels or genes with female-specific function may be lost more quickly. First, the loss of genes expressed at high levels could be associated with an increased metabolic cost. Indeed, highly expressed genes are retained more often after whole-genome duplications in organisms such as Paramecium (Gout et al. 2010) and yeast (Seoighe and Wolfe 1999), whereas genes with low expression are lost more easily (Gout et al. 2010). Second, the few genes remaining on old Y chromosomes tend to be male specific in function, expressed predominantly in testis, and are often recruited secondarily onto the Y (Skaletsky et al. 2003; Carvalho et al. 2009). Conversely, because females never inherit a Y chromosome, genes with female

© The Author(s) 2011. Published by Oxford University Press on behalf of the *Society for Molecular Biology and Evolution*.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

function should, in the long run, be lost from the Y without any fitness consequences. However, we do not know whether this process is simply part of the overall gene loss (i.e., whether female-biased genes are degenerating at the same rate as random genes) or whether loss of female genes is actively selected for. If selection is a substantial force, we expect genes that are beneficial for females—and either detrimental or neutral to males—to be among the first to be eliminated from an evolving Y chromosome.

The *Drosophila miranda* neo-sex chromosomes are an ideal system to study these questions (Steinemann M and Steinemann S 1998; Yi and Charlesworth 2000; Bachtrog and Charlesworth 2002). The neo-Y chromosome arose when an autosomal chromosome arm (Muller C) became fused to the existing Y, leading to a chromosome-wide loss of recombination. Over a timescale of only about 1.5 myr (Bachtrog and Charlesworth 2002; Bartolomé and Charlesworth 2006), the neo-Y has lost function at nearly half of the genes that it originally contained (Bachtrog et al. 2008). This chromosomal fusion is absent from *D. pseudoobscura*, the sister species of *D. miranda* (Dobzhansky 1935). Thus, homologues of genes that are neo-sex linked in *D. miranda* are autosomal in *D. pseudoobscura*, and, as demonstrated below, expression levels of genes on the Muller C element in *D. pseudoobscura* can be used as a close proxy for estimates of ancestral expression levels of neo-Y–linked genes.

Here, we make use of *D. pseudoobscura* RNA-Seq data to investigate whether gene loss from the neo-Y chromosome of *D. miranda* is a nonrandom process. We show that certain gene classes are being retained preferentially on this degenerating Y chromosome, including homologues of genes that show high ancestral transcript abundances, broadly expressed genes, as well as genes with male-biased expression.

## Materials and Methods

### RNA-Seq Data

We dissected testes, accessory glands, and ovaries from about 100 virgin males and females of *D. pseudoobscura* and *D. miranda* in phosphate-buffered saline solution under the microscope. At least 10 μg total RNA were prepared from these tissues as well as from whole flies or gonadectomized male and female flies using the RNAeasy (Qiagen Inc.) Kit. Messenger RNA (mRNA) paired-end libraries were prepared for all tissues except for *D. miranda* ovaries, for which a single-end library was prepared; sequencing was done following the Illumina protocol. Additionally, RNA-Seq data from *D. pseudoobscura* male and female head were downloaded from National Center for Biotechnology Information (NCBI)'s sequence read archive (accession numbers SRX016182 and SRX016183; Malone JH and Oliver B, unpublished data).

Transcript abundance levels were calculated using the Tophat and Cufflinks programs (Trapnell et al. 2009; Trapnell et al. 2010): Tophat was used to map the *D. pseudoobscura* reads against the *D. pseudoobscura* NCBI gene annotation file "dpse-all-r2.12.gff" (available at www.flybase.org), and based on this mapping, transcript abundances were calculated using cufflinks. Similarly, the *D. miranda* reads were mapped against the genome assembly of *D. miranda* (Zhou and Bachtrog, forthcoming). The cufflinks "genes.expr" output files containing the *D. pseudoobscura* or *D. miranda* gene IDs, genomic locations, and transcript abundances (measured as the number of fragments per kilobase of exon per million fragments mapped [FPKM]; Trapnell et al. 2010) were used in the analyses.

To confirm that *D. pseudoobscura* transcript abundance levels can be used as a proxy for ancestral neo-Y expression levels, we calculated the Pearson's product–moment correlation, $r$, between FPKM values of homologous autosomal genes in *D. pseudoobscura* and *D. miranda*.

### Comparing Transcript Abundance Levels among Neo-Y Functional and Nonfunctional Genes

The neo-Y contains 1,805 putative functional and 1,374 nonfunctional genes, the latter of which include genes that were either deleted from the neo-Y or contain internal stop codons and/or frameshift mutations (Zhou et al., in preparation). RNA transcripts of *D. pseudoobscura* were classified into three groups based on the location and state of degeneration of their *D. miranda* homologues: 1) all transcripts expressed in a given tissue, 2) transcripts with functional, or 3) transcripts with nonfunctional neo-Y–linked homologues (Zhou et al., in preparation). Median FPKM values were calculated for all *D. pseudoobscura* genes expressed in a given tissue, and these "ancestral" transcript abundance levels were compared between genes that are still functional on the neo-Y and those that have been lost, with the Wilcoxon test using R (www.r-project.org).

### Tissue Specificity

To test whether tissue specificity had an impact on a gene's probability of being lost or retained on the neo-Y, we calculated τ for each gene (Yanai et al. 2005):

$$\tau = \frac{\sum_{i=1}^{N}\left(1 - \frac{x_i}{x_{\max}}\right)}{N - 1},$$

where $N$ is the number of tissues investigated (i.e., seven in this study, treating "gonadectomized carcass" as a single tissue), $x_i$ is a gene's FPKM value in the $i$th tissue, and $x_{\max}$ is the gene's maximum level of expression in any of the tissues investigated. Genes with τ values equal to or larger than 0.9 were classified as "tissue specific," and the chance of degeneration on the neo-Y was compared between genes preferentially expressed in testis, accessory gland, and ovary; the

impact of expression levels on the rate of gene loss was studied for each tissue separately.

### Sex-Biased Genes

Based on the data in the supplementary table S1 (Supplementary Material online) of Jiang and Machado (2009), we categorized genes as being male or female biased in *D. pseudoobscura* if their expression levels differed at least 2-fold in the two sexes, and *q*, the measure of statistical significance of this difference, was less than $10^{-3}$. We tested whether male- and female-biased genes were lost from the neo-Y chromosome at a similar rate and whether transcript abundance levels differed between functional and nonfunctional neo-Y homologues that were categorized as being sex biased.

### Functional Analysis

Patterns of gene loss that are seemingly caused by differences in transcript abundance levels might be confounded by transcript abundances differing systematically between functional gene categories. To exclude this possibility, we extracted all genes that fell into gene ontology (GO) categories overrepresented among either functional or nonfunctional genes, respectively. Transcript abundance levels were calculated for all *D. pseudoobscura* samples separately within the two groups of GO categories, and FPKM values between homologues of functional versus nonfunctional genes were compared.

### Gene Connectivity

We made use of the data at the Drosophila Interactions Database (Yu et al. 2008) to calculate the number of annotated protein interaction partners for each *Drosophila melanogaster* gene and extracted all genes with a *D. miranda* neo-Y–linked homologue, using the file "gene_orthologs_fb_2010_07-1" at www.flybase.org. The median number of interaction partners was compared between *D. melanogaster* homologues with functional versus nonfunctional homologues using a Wilcoxon sign ranked test in R.

### Probit Regression

To disentangle the effects of gene expression levels and other factors that may influence gene loss on the neo-Y chromosome, we used probit regression to model the binary outcome variable "gene functionality," which could take the values "0" (for nonfunctional genes) or "1" (functional genes). As predictor variables, we used *D. pseudoobscura* transcript abundance levels in the seven tissues and whole-body samples examined, the breadth of gene expression, gene connectivity, and the factor "sex bias" (which could take three levels: unbiased, male-biased, or female-biased genes). The Wald test was used in R to test for the significance of the respective predictor variables when gene expression levels were also part of the model.

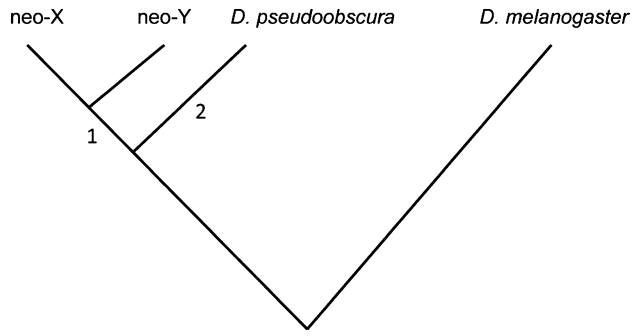


**Fig. 1.**—A phylogenetic tree showing the relationship between the neo-sex chromosomes in *Drosophila miranda* and autosomal genes in *D. pseudoobscura* and *D. melanogaster*. The two branches used to calculate the rate of sequence evolution are labeled as "1" and "2." Branch lengths are not drawn to scale.

### Rates of Evolution

To test whether selective constraints on the coding sequences differed among neo-Y–linked genes at the time of sex chromosome formation in a systematic manner, we compared the "ancestral" rate of sequence evolution between genes that are still functional versus those that have already become pseudogenized on the neo-Y. For this analysis, we removed frameshift mutations and premature stop codons from the coding sequence of degenerate neo-Y genes; next, the coding sequences of neo-sex–linked genes were aligned with homologous sequences of *D. pseudoobscura* and *D. melanogaster* using TranslatorX (Abascal et al. 2010). This program aligns divergent nucleotide sequences based on the corresponding protein alignment; hence, it maintains a codon structure, even if gaps are present in one or several of the sequences.

Because the neo-sex chromosome formation occurred only shortly after the split between *D. miranda* and *D. pseudoobscura* and we needed to exclude any sequences from the analysis with zero synonymous divergences, we expect little power to compare the rate of evolution between different classes of neo-Y–linked genes on this short branch (from the common ancestor of *D. miranda* and *D. pseudoobscura* to the split of the neo-X and neo-Y chromosome; branch 1 in fig. 1). Hence, we estimated the rate of evolution along the branch leading to the neo-sex chromosome formation after the split with *D. pseudoobscura*, as well as the *D. pseudoobscura* branch itself (branches labeled "1" and "2" in fig. 1). d*N*/d*S*, the ratio of the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site, was compared between homologues/ancestral branches of nonfunctional versus functional neo-Y–linked genes.

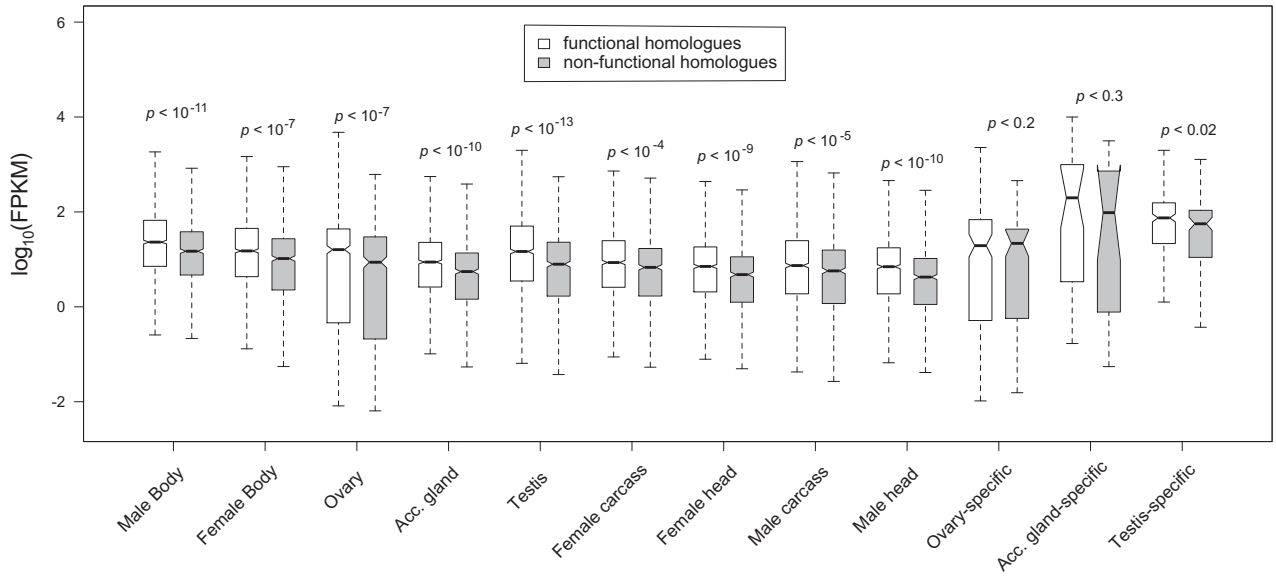


**Fig. 2.**—Transcript abundance (FPKM) of *Drosophila pseudoobscura* genes with functional and nonfunctional neo-Y–linked homologues. The *P* value indicates the significance of the Wilcoxon test, comparing median FPKM levels between the two gene categories; the criterion for tissue specificity is a $\tau$ value of at least 0.9.

## Results

### Expression Levels Matter

To test if natural selection is influencing the dynamics of gene loss on the neo-Y chromosome, we obtained RNA-Seq data from *D. pseudoobscura* genes expressed in male and female whole body, as well as genes expressed in seven different *D. pseudoobscura* tissues (supplementary table S1, Supplementary Material online, GenBank submission number SRP009072). To quantify transcript abundance, we used the FPKM statistics (Trapnell et al. 2010). We found that in each sample investigated, the median FPKM value of transcripts with a functional homologue on the neo-Y was significantly higher compared with genes with a degenerate neo-Y homologue (fig. 2). These results strongly suggest that highly expressed genes are retained on a degenerating Y chromosome for longer, on average, compared with lowly expressed genes.

This nonrandomness of gene loss on the neo-Y chromosome indicates that purifying selection is operating on the neo-Y chromosome and discriminating between different categories of genes. Levels of gene expression, however, correlate with several other known features influencing genome evolution. To determine the importance of gene expression levels in preserving neo-Y genes, we explored if other factors were contributing to this pattern, including gene function, breadth of expression, or protein connectivity. To this end, we either compared transcript abundance of *D. pseudoobscura* genes with functional versus nonfunctional neo-Y homologues or compared the "proportion" of nonfunctional genes among different gene categories.

### GO Categories

One confounding variable that could influence rates of gene loss on the neo-Y is gene function: If genes that belong to different functional categories differ systematically in their levels of expression and if certain GO categories are lost or retained more often on the neo-Y chromosome, gene functionality, rather than gene expression levels, might have caused the pattern of nonrandom gene loss that we observe. Interestingly, our results remained unchanged when we considered only genes belonging to GO categories that were overrepresented either among degenerate or putative functional neo-Y genes (24 and 35 GO categories, respectively; supplementary table S2, Supplementary Material online). For both sets of GO categories, the transcript abundances of *D. pseudoobscura* homologues differed significantly in each of the samples, with homologues of functional genes consistently showing higher transcript abundances compared with homologues of nonfunctional genes (table 1). Because gene loss correlates with transcript abundance also within GO categories, we can exclude the possibility that differential gene loss is only driven by gene function.

### Selection on Sex Tissue–Specific Genes

The neo-Y chromosome is never passed through females, and we expect genes that are expressed predominantly in male or female sex tissues to be under distinct selective regimes. For example, because ovary-specific genes are never expressed in males, gene expression levels should have no impact on their degeneration on the neo-Y. Conversely, if

**Table 1**
Median Transcript Abundance Levels (FPKM) of *Drosophila pseudoobscura* Genes in the 24 GO Categories Overrepresented among Nonfunctional Neo-Y–Linked Genes (Top Part) and of Genes in the 35 GO Categories Overrepresented among Functional Genes (Bottom Part)

| | Male Whole Body | Female Whole Body | Ovary | Testis | Accessory Gland | Male Head | Female Head | Male Carcass | Female Carcass |
|---|---|---|---|---|---|---|---|---|---|
| GO overrepresentation nonfunctional | | | | | | | | | |
| Transcripts with functional neo-Y homologues ($N$ = 315) | 25.5 | 20.0 | 12.8 | 9.2 | 6.4 | 6.5 | 6.9 | 9.5 | 10.2 |
| Transcripts with nonfunctional neo-Y homologues ($N$ = 307) | 12.2 | 11.0 | 1.2 | 4.7 | 3.2 | 2.7 | 3.3 | 5.8 | 5.8 |
| $P$ | $<10^{-5}$ | $<10^{-4}$ | $<10^{-3}$ | $<10^{-2}$ | $<10^{-3}$ | $<10^{-6}$ | $<10^{-4}$ | $<10^{-3}$ | $<10^{-2}$ |
| GO overrepresentation functional | | | | | | | | | |
| Transcripts with functional neo-Y homologues ($N$ = 663) | 22.4 | 17.3 | 21.9 | 11.2 | 7.2 | 5.9 | 6.1 | 7.5 | 7.8 |
| Transcripts with nonfunctional neo-Y homologues ($N$ = 482) | 14.3 | 11.8 | 8.8 | 7.3 | 4.6 | 3.1 | 3.6 | 5.3 | 5.3 |
| $P$ | $<10^{-6}$ | $<10^{-4}$ | $<10^{-7}$ | $<10^{-2}$ | $<10^{-3}$ | $<10^{-4}$ | $<10^{-3}$ | $<0.02$ | $<0.06$ |

Note.—The $P$ value indicates the statistical significance of the Wilcoxon test, comparing FPKM values between functional and nonfunctional homologues.

the loss of highly expressed male-specific genes is associated with a high fitness cost, we expect those genes to be preferentially retained on the neo-Y.

Our analyses revealed that genes expressed predominantly in testis and accessory glands showed very high absolute transcript abundance levels in those tissues compared with nonspecific genes (supplementary table S1, Supplementary Material online). However, even among these male sex tissue–specific genes, transcripts with functional homologues were found at higher levels compared with those with nonfunctional homologues (fig. 2). Although median FPKM values differ about 2-fold for accessory gland–specific functional and nonfunctional genes, this difference is not significant—possibly due to the small number of genes in this category—and hence reduced power to detect significant differences (see supplementary table S1, Supplementary Material online). Ovary-specific genes, on the other hand, did not show this trend, that is, FPKM values did not differ between genes with functional versus nonfunctional neo-Y homologues (fig. 2). This is expected because selection should not act on ovary-specific (female limited) neo-Y–linked genes. However, we note that the proportion of degenerate genes did not differ between ovary-specific and male sex tissue–specific genes, that is, genes in these categories are lost at similar rates (chi-square test; $\chi^2 = 1.12$, $P < 0.6$, degrees of freedom [df] = 5).

## Sex-Biased Genes

Similar to genes expressed predominantly in sex tissues, genes that are sex biased in expression may also be under different selective pressures on the neo-Y chromosome. In particular, sexually antagonistic genes (i.e., genes that are beneficial to one sex and detrimental to the other; Rice 1984; Innocenti and Morrow 2010) are of interest because we may expect an active removal of female-beneficial/male-detrimental genes from the neo-Y if selection is an important force. We classified genes as either male or female biased based on gene expression data from four different *D. pseudoobscura* isofemale lines (Jiang and Machado 2009). Overall, female-biased genes on Muller C showed lower ancestral transcript abundance levels in females compared with male-biased genes in males (table 2; Wilcoxon test: $W = 132{,}918$, $P < 10^{-15}$), and female-biased genes have become nonfunctional at a higher rate compared with male-biased genes (42% vs. 34%; chi-square test with Yate's correction; $\chi^2 = 5.24$, $P < 0.03$, df = 1; table 2). Importantly, the fit of a probit regression model is significantly increased when the factor sex bias is included to predict the outcome variable gene functionality (table 3), suggesting that the differential loss of female- and male-biased genes is not merely caused by a systematic difference in transcript levels between these gene categories. If we control for transcript abundance levels by holding the FPKM values at their median values, the probit model predicts the proportion of degenerate genes to be 50%, 39%, and 37% for female-biased, male-biased, and unbiased genes, respectively. This indicates that the neo-Y has become "defeminized," mainly as a result of reduced purifying selection to prevent the degeneration of female-biased genes compared with selection on male-biased and unbiased genes (or because female genes have been actively removed). On the other hand, purifying selection on male-biased genes only appears to be stronger than selection on unbiased genes when transcript abundance levels are not taken into account, suggesting that preferential conservation of male-biased genes on the neo-Y is mainly a consequence of their higher expression levels in males (table 2).

**Table 2**
Median Transcript Abundance Levels (FPKM) of Sex-Biased Genes in *Drosophila pseudoobscura*, Classified by the State of Degeneration of Their *D. miranda* Neo-Y–Linked Homologues

| | Female-Biased Genes in Female Whole Body (*N*) | Male-Biased Genes in Male Whole Body (*N*) |
|---|---|---|
| All Muller C transcripts | 13.0 (393) | 68.3 (410) |
| Transcripts with functional neo-Y homologues | 14.2 (226) | 75.4 (269) |
| Transcripts with nonfunctional neo-Y homologues | 10.6 (167) | 55.7 (141) |
| *P* | $<10^{-3}$ | $<0.05$ |
| Proportion nonfunctional | 0.42 | 0.34 |

Note.—The *P* value indicates the significance of the Wilcoxon test, comparing FPKM values between transcripts with functional versus nonfunctional homologues in *D. miranda*.

Female-biased genes with functional homologues showed higher ancestral transcript abundances compared with those with nonfunctional homologues (table 2). However, highly expressed female-biased genes are also expressed at a higher level in males compared with lowly expressed female-biased genes (Pearson's correlation comparing transcript abundance of female-biased genes in female whole body with transcript abundance of female-biased genes in male whole body = 0.82, $P < 10^{-15}$). This implies that at least some (highly expressed) female-biased genes also have important function in males and are thus actively maintained on the neo-Y and that not all sex-biased genes are necessarily sexually antagonistic (Innocenti and Morrow 2010).

## Gene Connectivity

Genes with a higher number of interaction partners cause a greater fitness reduction in yeast knockout experiments (Papp et al. 2003), possibly because the number of epistatic interactions and thus possible deleterious effects of mutations increases with a gene's connectivity (Papp et al. 2003; Qian and Zhang 2008). In line with this, we found that functional genes on the neo-Y chromosome of *D. miranda* have, on average, a higher number of interaction partners compared with genes that have lost their function (mean of 32 vs. 26 partners, based on gene interactions measured in *D. melanogaster*; Yu et al. 2008) (Wilcoxon test, $W = 2{,}144{,}931$, *P* value $< 2.2 \times 10^{-16}$). However, highly connected genes also tend to be highly expressed (Pearson's correlation coefficients of the number of interaction partners with transcript abundance in male and female whole body were 0.25 and 0.27, respectively; $P < 10^{-15}$ in both comparisons). Including the number of interaction partners into the probit regression model does not significantly improve its fit to the data (table 3), suggesting that transcript abundance and connectivity cannot be treated as independent factors influencing the rate of gene loss on the neo-Y chromosome. Instead, the influence of gene connectivity on the fate of neo-Y–linked genes can mainly be explained by its effect on gene expression levels.

## Breadth of Gene Expression

Classifying all neo-Y genes based on the number of tissues in which their *D. pseudoobscura* homologues were expressed, the proportion of degenerate genes was always about 35–40% (supplementary table S3, Supplementary Material online; chi-square test, NS, df = 6); this result alone would suggest that the breadth of gene expression does not have a strong impact on gene degeneration on the neo-Y. However, including breadth of gene expression into the regression model significantly increases its fit (table 3)—possibly because expression breadth is only moderately correlated with transcript levels (supplementary table S4, Supplementary Material online). If we control for transcript abundance levels by holding FPKM levels at their median values, the predicted proportion of functional neo-Y genes increases monotonically with gene expression breadth (fig. 3). Figure 3 is based on a probit model using only whole-body male and female expression and expression breadth as predictor

**Table 3**
Fitting a Probit Regression Model to Predict the Outcome Variable "Gene Functionality"

| Predictor Variables | $\chi^2$ | df | *P* | Wald test |
|---|---|---|---|---|
| Transcript levels + sex bias | 53.3 | 11 | $<10^{-7}$ | $\chi^2 = 8.4$, df = 2, $P < 0.02$ |
| Transcript levels + number of interaction partners | 72.3 | 10 | $<10^{-11}$ | $\chi^2 = 3.0$, df = 1, $P < 0.1$ |
| Transcript levels + number of tissues expressed | 177.5 | 10 | $<10^{-32}$ | $\chi^2 = 47.4$, df = 1, $P < 10^{-11}$ |

Note.—The first four columns show the combination of parameters tested and the significance of the respective models. The Wald test shows the significance of the predictor variables "sex bias," "number of interaction partners," and "number of tissues expressed," respectively, given that transcript levels in all seven tissues and the whole-body samples are part of the model.

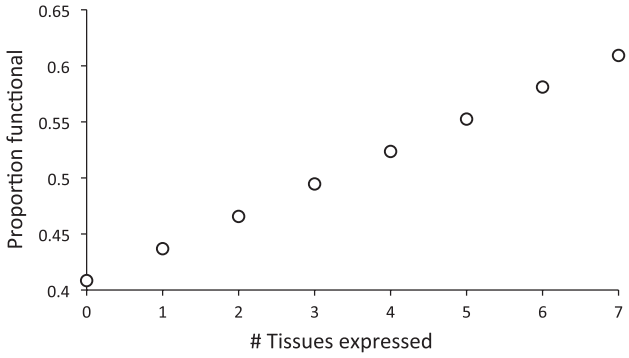


**Fig. 3.**—The predicted proportion of functional neo-Y genes under a probit regression model, accounting for transcript abundance in male and female whole body.

variables; the same trend is observed when we correct for transcript levels in all tissues examined (results not shown).

### Rate of Sequence Evolution

To test for differences in the efficacy of purifying selection on the coding sequences of neo-Y genes, we calculated the ancestral rate of constraint, given by d*N*/d*S*, for neo-sex–linked genes that have known homologues in both *D. pseudoobscura* and *D. melanogaster*. We found that d*N*/d*S* values did not differ significantly between neo-Y genes that are still functional compared with those that have become pseudogenized. For the branch that is ancestral to the neo-sex chromosome formation (branch 1 in fig. 1), median d*N*/d*S* values were 0.06 for functional neo-Y genes ($N = 256$) and 0.07 for nonfunctional genes ($N = 166$) (two-sample Kolmogorov–Smirnov test; $D = 0.0976$; not significant [NS]). Similarly, considering both this ancestral *D. miranda* branch as well as the branch leading to *D. pseudoobscura* (branch "2" in fig. 1), the number of functional and nonfunctional neo-Y genes with at least one synonymous substitution increased to 426 and 264, respectively, but median d*N*/d*S* values did not differ significantly between the two groups (0.10 for functional genes and 0.12 for nonfunctional genes; two-sample Kolmogorov–Smirnov test; $D = 0.0929$, NS). Furthermore, adding the factor "d*N*/d*S*" to the probit model does not increase its fit to the data, given that transcript levels are already part of the model (Wald test; $\chi^2 = 2.9$, df = 1, NS). This suggests that constraint on the coding sequence, as captured by d*N*/d*S*, has not played as large a role in preventing neo-Y gene degeneration relative to other factors that we discussed. Interestingly, a recent study of de novo Y chromosome evolution in *Silene latifolia* (Chibalina and Filatov 2011) found a negative correlation between the constraint on the coding sequence and X–Y divergence, considering putative functional (i.e., transcribed) Y-linked genes only. Furthermore, in contrast to earlier studies that found a negative correlation between the rate of protein evolution and mRNA abundance (Lemos et al. 2005) or gene connectivity (Fraser et al. 2002) in Drosophila and yeast, respectively, there was no significant correlation when d*N*/d*S* along branches 1 and 2 was compared with either male or female whole-body transcript abundance (Pearson's correlation coefficient = −0.01 [NS] in both cases) or the number of interaction partners in *D. melanogaster* (Pearson's correlation coefficient = −0.04, NS).

**Table 4**
Pearson's Product–Moment Correlation, *r*, between the Transcript Abundance of Homologous Autosomal Genes in *Drosophila miranda* and *D. pseudoobscura*

| Tissue | *r* | *P* |
|---|---|---|
| Male body | 0.74 | $<10^{-15}$ |
| Female body | 0.52 | $<10^{-15}$ |
| Ovary | 0.82 | $<10^{-15}$ |
| Testis | 0.83 | $<10^{-15}$ |
| Accessory gland | 0.76 | $<10^{-15}$ |
| Female carcass | 0.71 | $<10^{-15}$ |
| Male carcass | 0.77 | $<10^{-15}$ |

### Ancestral Gene Expression Patterns

Our study assumes that transcript abundance in *D. pseudoobscura* can be used as a proxy for the ancestral mRNA levels of neo-sex chromosome–linked genes in *D. miranda*. Indeed, we find a strong correlation between FPKM values of autosomal genes in the two species (table 4) in all tissues examined. Assuming that both species have diverged in expression since their common ancestor, the correlation between ancestral neo-Y and current transcript levels in *D. pseudoobscura* is expected to be even higher. Importantly, our results are conservative with respect to any expression divergence between the two species, which would mask any effects of transcript levels on neo-Y degeneration.

## Discussion

We have shown that degeneration on the neo-Y is not a simple, mutation-driven process. Instead, purifying selection is acting to prevent the loss of genes that are more costly to lose, including genes with male function, broadly expressed genes, and genes that are highly expressed. In contrast, neo-Y genes appear to degenerate randomly with respect to protein constraint, as measured by rates of protein evolution (d*N*/d*S*).

There might be a direct cost associated with losing highly expressed genes because the reduction in gene product is greater for such genes. For example, experiments in yeast have shown that the fitness reduction caused by a heterozygous knockout mutation is higher if the deleted gene is normally highly expressed (Gout et al. 2010). On the neo-sex chromosome, the evolution of dosage compensation (i.e., balancing the X to autosome expression ratio in males) adds an additional layer of complexity to this issue. Dosage compensation in *D. melanogaster* is mediated by the male-

specific lethal (MSL) complex. The MSL complex binds to high-affinity chromatin entry sites on the single X chromosome in males, spreads in *cis*, and leads to the upregulation of X-linked genes through enhanced transcriptional elongation (Larschan et al. 2011; Straub and Becker 2011). The neo-X of *D. miranda* has been shown to have evolved partial dosage compensation (Bone and Kuroda 1996), but we know little about the extent of this upregulation, or which individual genes on the neo-X chromosome are dosage compensated. About 80% of the 58 gene pairs investigated by Bachtrog (2006) were expressed at lower levels from the neo-Y compared with the neo-X, which may be due to either dosage compensation of neo-X genes or downregulation of neo-Y–linked genes. Eight genes in that study were expressed more highly from the neo-Y compared with the neo-X, suggesting that dosage compensation does not act chromosome wide on the neo-X and possibly depends on the state of degeneration of the neo-Y–linked genes. Also, because the dosage compensation machinery in Drosophila spreads in *cis* along the X chromosome from the chromatin entry sites, some genes may be upregulated on the neo-X even though their homologues are still functional on the neo-Y. Accordingly, one might expect to see a selective advantage in shutting down neo-Y–linked genes that are dosage compensated on the neo-X and thus expressed at too high levels in males. It will be interesting to link patterns of Y degeneration to dosage compensation of neo-X–linked genes and to test, for example, if the upregulation of neo-X–linked genes in males is harder to achieve if the corresponding homologue is highly expressed in *D. pseudoobscura* or if dosage-compensated genes are becoming downregulated from the neo-Y.

We show that female-biased genes, but not ovary-specific genes, are lost at an increased rate from the neo-Y (supplementary table S1, Supplementary Material online). This observation is puzzling, given that ovary-specific genes are, by definition, not expressed in males, and hence we do not expect any purifying selection acting on these genes. One explanation for this might be that our transcript data are all based on tissues of adult flies, and we might have classified some genes as ovary specific even though they have important function in other tissues at earlier life stages. Additionally, the number of genes classified as ovary specific in our study may be too low to detect any increased rate of gene loss. In *D. melanogaster*, genes expressed in ovary do not show increased levels of sexual antagonism compared with genes expressed in other sex tissues (Innocenti and Morrow 2010). Thus, selection may not actively remove ovary-specific genes from the neo-Y. In contrast, a larger fraction of genes classified as female biased might have sexually antagonistic effects in males, contributing to their faster rate of loss relative to ovary-specific genes.

Our data are consistent with the idea that, in the early stages of Y chromosome evolution, differential purifying selection on sex-biased genes may lead to a gradual "defeminization" or "masculinization" of the neo-Y chromosome; this process may occur at the same time or precede the active recruitment of male-beneficial genes onto the Y, which would lead to further specialization in male function (Carvalho et al. 2009). The ultimate fate of the neo-Y chromosome is likely to further lose most of its current genes and to maintain only a few genes that are male specific in function and/or highly expressed, as observed for the human or *D. melanogaster* Y chromosome (Skaletsky et al. 2003).

## Supplementary Material

Supplementary tables S1–S4 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

## Acknowledgments

We thank Beatriz Vicoso and Eric Durand for useful discussions, Steve Lockton for the *Drosophila miranda* ovary RNA-Seq data, and John Malone and Brian Oliver for providing the *D. pseudoobscura* head data. This research was funded by NIH grants (R01GM076007 and R01GM093182) and a Packard Fellowship to D.B.

## Literature Cited

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38:W7–W13.

Bachtrog D. 2006. Expression profile of a degenerating neo-Y chromosome in Drosophila. Curr Biol. 16:1694–1699.

Bachtrog D, Charlesworth B. 2002. Reduced adaptation of a non-recombining neo-Y chromosome. Nature 416:323–326.

Bachtrog D, Hom E, Wong KM, Maside X, de Jong P. 2008. Genomic degradation of a young Y chromosome in *Drosophila miranda*. Genome Biol. 9:R30.

Bartolomé C, Charlesworth B. 2006. Evolution of amino-acid sequences and codon usage on the *Drosophila miranda* neo-sex chromosomes. Genetics 174:2033–2044.

Bone JR, Kuroda MI. 1996. Dosage compensation regulatory proteins and the evolution of sex chromosomes in Drosophila. Genetics 144:705–713.

Bull JJ. 1983. Evolution of sex determining mechanisms. Menlo Park (CA): Benjamin Cummings.

Carvalho AB, Koerich LB, Clark AG. 2009. Origin and evolution of Y chromosomes: Drosophila tales. Trends Genet. 25:270–277.

Charlesworth B. 1996. The evolution of chromosomal sex determination and dosage compensation. Curr Biol. 6:149–162.

Charlesworth B, Betancourt AJ, Kaiser VB, Gordo I. 2009. Genetic recombination and molecular evolution. Cold Spring Harb Symp Quant Biol. 74:177–186.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci. 355:1563–1572.

Chibalina MV, Filatov DA. 2011. Plant Y chromosome degeneration is retarded by haploid purifying selection. Curr Biol. 21:1475–1479.

Dobzhansky T. 1935. Drosophila miranda, a New Species. Genetics 20:377–391.

Felsenstein J. 1974. Evolutionary advantage of recombination. Genetics 78:737–756.

Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. 2002. Evolutionary rate in the protein interaction network. Science 296:750–752.

Gout JF, Kahn D, Duret L. 2010. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. PloS Genet. 6:e1000944.

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

Jiang ZF, Machado CA. 2009. Evolution of sex-dependent gene expression in three recently diverged species of Drosophila. Genetics 183:1175–1185.

Kaiser VB, Charlesworth B. 2010. Muller's Ratchet and the degeneration of the *Drosophila miranda* neo-Y chromosome. Genetics 185:339–348.

Larschan E, et al. 2011. X chromosome dosage compensation via enhanced transcriptional elongation in Drosophila. Nature 471:115–118.

Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. 2005. Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. Mol Biol Evol. 22:1345–1354.

Papp B, Pal C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. Nature 424:194–197.

Qian W, Zhang J. 2008. Gene dosage and gene duplicability. Genetics 179:2319–2324.

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

Seoighe C, Wolfe KH. 1999. Yeast genome evolution in the post-genome era. Curr Opin Microbiol. 2:548–554.

Skaletsky H, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837.

Steinemann M, Steinemann S. 1998. Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda* a model for sex chromosome evolution. Genetica 102–103:409–420.

Straub T, Becker PB. 2011. Transcription modulation chromosome-wide: universal features and principles of dosage compensation in worms and flies. Curr Opin Genet Dev. 21:147–153.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105–1111.

Trapnell C, et al. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 28:511–515.

Yanai I, et al. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics 21:650–659.

Yi S, Charlesworth B. 2000. Contrasting patterns of molecular evolution of the genes on the new and old sex chromosomes of *Drosophila miranda*. Mol Biol Evol. 17:703–717.

Yu J, Pacifico S, Liu G, Finley RL Jr. 2008. DroID: the Drosophila Interactions Database, a comprehensive resource for annotated gene and protein interactions. BMC Genomics 9:461.

**Associate editor:** Kateryna Makova