Molecular Evolution in Nonrecombining Regions of the Drosophila melanogaster Genome

GBE

Genome Biology and Evolution

SMBE

Molecular Evolution in Nonrecombining Regions of the *Drosophila melanogaster* Genome

José L. Campos*, Brian Charlesworth, and Penelope R. Haddrill

Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, United Kingdom

*Corresponding author: E-mail: j.campos@ed.ac.uk.

Accepted: 17 January 2012

Abstract

We study the evolutionary effects of reduced recombination on the *Drosophila melanogaster* genome, analyzing more than 200 new genes that lack crossing-over and employing a novel orthology search among species of the *melanogaster* subgroup. These genes are located in the heterochromatin of chromosomes other than the dot (fourth) chromosome. Noncrossover regions of the genome all exhibited an elevated level of evolutionary divergence from *D. yakuba* at nonsynonymous sites, lower codon usage bias, lower GC content in coding and noncoding regions, and longer introns. Levels of gene expression are similar for genes in regions with and without crossing-over, which rules out the possibility that the reduced level of adaptation that we detect is caused by relaxed selection due to lower levels of gene expression in the heterochromatin. The patterns observed are consistent with a reduction in the efficacy of selection in all regions of the genome of *D. melanogaster* that lack crossing-over, as a result of the effects of enhanced Hill–Robertson interference. However, we also detected differences among nonrecombining locations: The X chromosome seems to exhibit the weakest effects, whereas the fourth chromosome and the heterochromatic genes on the autosomes located most proximal to the centromere showed the largest effects. However, signatures of selection on both nonsynonymous mutations and on codon usage persist in all heterochromatic regions.

Key words: recombination, crossing-over, background selection, Hill–Robertson interference, codon usage bias, heterochromatin.

Introduction

Levels of variation and rates of evolution in different regions of the genome may be greatly affected by differences in the frequency of recombination, as a result of the process of Hill–Robertson interference (HRI) (Hill and Robertson 1966; Felsenstein 1974; Gordo and Charlesworth 2001; Comeron et al. 2008; Charlesworth et al. 2010), whereby evolutionary processes at a given site in the genome are influenced by selection acting on closely linked sites. To a good approximation, this can be viewed as a reduction in effective population size (N_e) at the site in question, caused by the variance in fitness at linked sites subject to selection (Charlesworth et al. 2010). This effect is likely to be maximal in regions with little or no genetic recombination because recombination reduces the intensity of HRI, increasing the N_e of a region, and hence the efficacy of selection.

These theoretical predictions are consistent with the observed lower rates of adaptive evolution, higher levels of fixation of deleterious mutations, and reduced levels of neutral or nearly neutral variability in genomic regions or populations with little or no genetic recombination (Comeron et al. 1999; Charlesworth 2003; Bachtrog 2005; Presgraves 2005; Bartolomé and Charlesworth 2006; Haddrill et al. 2007; Bachtrog et al. 2008; Arguello et al. 2010; Qiu et al. 2011).

In *Drosophila*, studies have focused on comparisons of regions of the genome that apparently lack crossing-over with regions where crossovers are known to occur (Comeron et al. 1999; Bachtrog 2005; Presgraves 2005; Bartolomé and Charlesworth 2006; Haddrill et al. 2007; Bachtrog et al. 2008; Arguello et al. 2010). It is unclear at present whether gene conversion is also lacking in these regions (Betancourt et al. 2009; Arguello et al. 2010); for simplicity, we refer to them here as "nonrecombining regions." The following features of such nonrecombining regions in *Drosophila* have been found: elevated between-species sequence divergence at nonsynonymous sites and in long introns, reduced

codon usage bias, increased gene length, an increased level of nonsynonymous polymorphism relative to synonymous polymorphism, and a reduced incidence of positive selection.

However, in *Drosophila melanogaster* and its relatives, these studies have mostly focused on the small dot (fourth) chromosome, where recombination is minimal or completely absent (Haddrill et al. 2007; Arguello et al. 2010). This is because sequence data for most of the other nonrecombining genes were not available because they are in heterochromatic regions that pose problems for sequencing and assembly. The discovery of at least 230 protein-coding genes in the centromeric heterochromatin as a result of the *Drosophila* Heterochromatin Genome Project (DHGP, http://www.dhgp.org/) provides new nonrecombining regions with which to test the predictions of HRI in genomic regions outside the fourth chromosome, because many of these heterochromatic genes are known to have orthologs in other *Drosophila* and Dipteran species (Smith et al. 2007). This provides unique and abundant material for an in-depth analysis of the effects of the nonrecombining environment on patterns of molecular evolution. This should enable us to exclude the possibility that the features of nonrecombining genes described in Haddrill et al. (2007) and Larracuente et al. (2008) reflect peculiarities of the set of genes residing on the fourth chromosome rather than the effect of reduced recombination. In addition, recent RNAseq data on gene expression in *D. melanogaster* provide new information with which to assess the effects of gene expression on the patterns of molecular evolution in nonrecombining regions (Haddrill et al. 2008; Larracuente et al. 2008).

We have used a data set of more than 10,000 genes from *D. melanogaster* and a comparison with the related species *D. yakuba*, to examine the effects of recombinational environment on rates and patterns of evolution in coding genes and on measures of adaptation at the molecular level. We have thereby extended previous analyses to include genes in nonrecombining regions on all the autosomal arms and on the X chromosome.

Materials and Methods

Gene Partitioning

We divided the *D. melanogaster* genome into four recombination categories: high (H), intermediate (I), low (L), and no recombination (N). These divisions are based on their cytological location (Charlesworth 1996) and the recently annotated heterochromatic regions (Smith et al. 2007) (supplementary table 1, Supplementary Material online), which include a much larger set of nonrecombining genes than those in Haddrill et al. (2007). We used release 5.34 of the *D. melanogaster* genome, available in FlyBase (Tweedie et al. 2009), to download all the currently annotated genes.

We also partitioned the whole data set by chromosome type: Autosomal (A) and X chromosome (X). Within the nonrecombining genes, we subdivided genes into three categories: nonfourth autosomal (No), fourth chromosome (N4), and the X chromosome genes (NX). The latter was subdivided into nonrecombining genes near the centromere and nonrecombining genes near the telomere. We also contrasted the patterns observed in genes located in the beta-heterochromatin (contiguous to the euchromatin, see Miklos and Cotsell 1990), to those located in the alpha-heterochromatin, which constitutes the majority of the centromeric heterochromatin (Miklos and Cotsell 1990). The latter was termed "scaffold heterochromatin" in the DHGP and is comprised of internal scaffolds that have been cytologically localized to an arm and are located proximal to the centromere relative to the beta-heterochromatin. Finally, we also compared distal and proximal autosomal beta-heterochromatin genes using the mid-coordinate within each region as a boundary (see supplementary table 1, Supplementary Material online).

For each gene with multiple transcripts, we chose the one that showed the highest transcript score (available for each transcript reported in FlyBase). In the case of equal scores, we randomly selected an isoform.

Search for Homologous Heterochromatic Sequences

We found 401 coding genes in the heterochromatic/nonrecombining regions in release 5.34 of the *D. melanogaster* genome (regions shown in supplementary table 1, Supplementary Material online). To detect orthologs of these genes, we first carried out gene annotations of sequences homologous to these genes in another five species of the *melanogaster* group (*D. ananassae*, *D. erecta*, *D. sechelia*, *D. simulans*, and *D. yakuba*). We used genBlastG to perform the search and gene annotation, which uses a homology-based gene predictor approach (She et al. 2011). We used the protein sequence of the 401 protein-coding genes located in the heterochromatin as the input query; as the target, we used each of the "chromosome" DNA Fasta files available in FlyBase for each *melanogaster* group species. We used an *e* value cutoff of 10^{-20}, the filtering option (-f T), coverage of 20% (-c 0.2), and the default setting for the remaining options. Any newly annotated gene obtained with genBlastG that overlapped with a coding gene present in FlyBase was excluded.

Ortholog Assignments

We used orthomcl (Li et al. 2003) to cluster the proteomes of the six *Drosophila* species, in order to assemble groups of orthologous and paralogous sequences. We used an *e* value cutoff of 10^{-20} and an inflation value of 1.5. The proteomes were obtained from FlyBase, and we added the newly annotated genes obtained with genblastG

that showed a high homology (*e* value of 10^{-20}) to any heterochromatic gene.

Expression Data

We used RNAseq gene expression available for *D. melanogaster* in FlyBase (Gelbart and Emmert 2010). For each *D. melanogaster* gene, we calculated the mean RNAseq expression value (expressed as RPKM, reads per kilobase of exon per million mapped reads) across the 27 temporal stages of the data set. For each of the adult stages, in the autosomal genes, we calculated the average of the two sexes, whereas for genes located on the X chromosome, we used a weighted average of 2/3 for females and 1/3 for males, which reflects the mean time an X chromosome spends in each sex.

We also used an alternative measure of gene expression for the same data set: maximum gene expression at any of the 27 temporal stages and sexes. This is to take into account the possibility that a stage/sex-specific estimate of gene expression could have an important effect on gene evolution that may not have been detected by using an overall measure of gene expression. We report these two measures of gene expression (overall expression and maximum expression) data for each gene as log2 (RPKM + 1).

Parameters Estimated and Final Data Set

We selected *D. yakuba* as an outgroup to estimate K_A/K_S because its divergence from *D. melanogaster* is sufficiently large enough that we avoid any major influence of ancestral polymorphisms, and its genome is well annotated with a high coverage (9.1×) (Clark et al. 2007). We chose only 1:1 orthologous genes and performed amino acid sequence alignments using MAFFT (Katoh et al. 2002). Using the protein alignment and the coding sequence (CDS) obtained from FlyBase, we made an in-frame CDS alignment using custom scripts in PERL. All sequence alignments used in this study are available from the corresponding author upon request.

We calculated K_A/K_S using the method of Comeron (1995) implemented in Gestimator (Thornton 2003). Estimates of the level of codon usage bias from the frequency of optimal codons, *Fop*, were calculated using CodonW (Peden 1999). GC content was estimated for the third positions of codons (GC_3) and for the introns of the selected isoform, following removal of 8/30 bp at the beginning/end of the introns and masking of possible exonic sequences to exclude any sites that may be subject to selective constraints within the selected introns. We divided introns into short (≤80 bp) and long (>80 bp) following Halligan and Keightley (2006). As measures of gene length, we used the lengths of the CDS, short introns (in bp) and long introns (in Kb) for the selected transcript.

The final data set included only genes with a K_S above 0 and below 0.5, amino acid length above 29, percentage of amino acid sequence identity above 50%, less than 50% gaps, presence of a single orthologous gene in *D. yakuba*, and with gene expression data (RPKM > 0). We excluded the few genes present on the Y chromosome and the U genes (unmapped to any chromosome arm), so we report only the nonrecombining genes present on a known chromosome. We also excluded nine genes from the scaffold heterochromatin with the gene status of "incomplete."

Statistical Analyses

We used nonparametric Mann–Whitney *U* (two-tailed) and Kruskal–Wallis tests to compare data sets. We controlled for the false discovery rate (FDR) by using the method of Benjamini and Hochberg (1995), implemented in the package multtest (Pollard et al. 2005), using a FDR threshold of 0.05, and report only the adjusted *P* values. For each data set and parameter, we report the mean and confidence interval (CI; calculated by bootstrapping across genes), except for the length variables (CDS length and intron length) where we report medians, because they are less sensitive to outliers. Patterns of correlations among divergence measures, codon usage, expression levels, and gene length were analyzed using partial correlations. We calculated partial correlations between two given variables (*Fop* and K_S, expression and *Fop*, *Fop* and CDS length, expression and K_A and *Fop* and K_A) while controlling for their covariates (the variables other than the pair involved in the correlation), using R function *pcor.test* (variance–covariance matrix method) available at http://www.yilab.gatech.edu/pcor (Kim and Yi 2006); we report Spearman's nonparametric correlation coefficients and their 95% CIs obtained from bootstrapping across genes.

In order to compare pairs of partial correlations between two data sets of interest, we calculated the CIs of the absolute difference between partial correlations for the nonrecombining and recombining data sets by resampling the two sets 1,000 times without replacement from the pooled data set and considered an observed difference between partial correlations to be significant if it fell outside its 95% CI. For partial correlations that showed the same trend among the four independent nonrecombining regions analyzed (second, third, fourth, and X chromosome), we tested if such a common trend was significant by combining their probabilities using Fisher's combined probability test. *P* values were combined by adding $-2\ln(P)$ for the four nonrecombining data sets. This follows a chi-squared distribution with eight degrees of freedom (df), which was used to determine the combined *P* value.

Results

The final data set contained 10,642 genes. We divided them into autosomal (A) genes and X chromosome (X) genes. Within the nonrecombining genes, after filtering the initial

Table 1
Comparisons of Recombining Genes and Nonrecombining Genes for the Autosomes and X Chromosome

	Autosomal Genes			X Chromosome Genes		
	RA	NA	*P*	RX	NX	*P*
Number of genes	8,729	226		1,645	42	
K_A	0.039 (0.038–0.040)	0.056 (0.050–0.061)	***	0.041 (0.039–0.043)	0.047 (0.035–0.056)	ns
K_S	0.263 (0.262–0.265)	0.276 (0.268–0.284)	**	0.258 (0.254–0.262)	0.259 (0.243–0.276)	ns
K_A/K_S	0.138 (0.136–0.141)	0.198 (0.180–0.215)	***	0.145 (0.139–0.152)	0.173 (0.134–0.211)	ns
Fop	0.517 (0.515–0.519)	0.344 (0.326–0.360)	***	0.551 (0.547–0.555)	0.452 (0.428–0.475)	***
GC_3	0.641 (0.639–0.643)	0.441 (0.420–0.460)	***	0.688 (0.684–0.692)	0.572 (0.547–0.594)	***
GC_S	0.354 (0.351–0.356)	0.290 (0.269–0.309)	***	0.394 (0.388–0.401)	0.351 (0.320–0.382)	ns
GC_L	0.373 (0.372–0.375)	0.347 (0.339–0.354)	***	0.377 (0.373–0.381)	0.339 (0.322–0.360)	***
Overall exp.	9.79 (9.74–9.84)	10.24 (9.97–10.55)	**	9.90 (9.80–10.00)	9.22 (8.64–9.86)	ns
Max. exp.	11.76 (11.71–11.80)	11.82 (11.57–12.07)	ns	11.83 (11.73–11.93)	11.44 (10.80–12.05)	ns
CDS length (aa)	387 (380–393)	454.5 (360–502)	***	407 (389–423)	424.5 (248–517.5)	ns
Length short (bp)	61.33 (61.16–61.66)	59 (58.33–59.71)	***	65.00 (64.67–65.30)	64 (62.67–67)	ns
Length long (Kb)	0.49 (0.46–0.51)	1.21 (0.80–1.52)	***	0.51 (0.44–0.56)	1.22 (0.19–2.13)	ns

Note.—For each variable in a row, we report the mean and the 95% CI in parentheses, except for the length variables where we use medians. The genome is divided into four categories: RA, recombining genes in the autosomal regions; NA, nonrecombining genes in autosomal regions; RX, recombining genes on the X chromosome; and NX, nonrecombining genes on the X chromosome. *P*: adjusted *P* value of Mann–Whitney *U* test (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; and ns, not significant); GC_3, GC content of third codon positions; GC_S, GC content in short introns (≤80 bp); GC_L, GC content in long introns (>80 bp); Overall exp., total RNAseq expression across all tissues as log2 (mean RPKM + 1); Max. exp., maximum RNAseq level expression level across all tissues; CDS length, CDS length in number of amino acids; Length short (bp), length of short introns in base pairs; and Length long (Kb), length of long introns in kilobases.

401 gene data set, we had 268 genes for the analysis. We subdivided these genes into three categories: nonfourth chromosome autosomal (No, $N = 159$), fourth chromosome (N4, $N = 67$), and X chromosome genes (NX, $N = 42$). Of the nonfourth chromosome genes, 131 genes were located in the beta-heterochromatin (contiguous to the euchromatin) and 28 genes were in the alpha-heterochromatin (scaffold heterochromatin). Among the nonrecombining X chromosome genes, 19 were in the centromeric region and 23 were at the telomere.

As in Haddrill et al. (2007), the most striking differences were generally seen in the comparison of recombining regions versus nonrecombining regions. We observed small differences among regions experiencing high, intermediate, and low rates of recombination (supplementary table 2, Supplementary Material online). However, given that the magnitude of these differences was small compared with those between recombining and nonrecombining regions, within each of the autosomal and X-linked data sets, these three recombining categories were combined into a single group of recombining genes.

Divergence

The autosomal nonrecombining (NA) regions showed much higher levels of nonsynonymous divergence (K_A), slightly higher synonymous divergence (K_S), and higher K_A/K_S than the recombining autosomal (RA) regions (table 1 and fig. 1). In the X chromosome data set, there were no significant differences between recombining (RX) and nonrecombining (NX) regions for K_A, K_S, and K_A/K_S, although K_A/K_S showed a similar trend to the autosomal genes, being higher for the nonrecombining genes (table 1 and fig. 1). When we contrasted the three groups of genes within nonrecombining regions (No, N4, and NX), we observed higher synonymous divergence for No than for NX or N4, which had similar K_S values (table 2 and fig. 1). N4 had an apparently higher K_A/K_S than No and NX, and this difference was significant against the K_A/K_S of NX by virtue of the N4 genes having a higher K_A and slightly lower K_S than the NX genes.

Codon Usage Bias and GC Content

Codon usage bias, as measured by *Fop*, was significantly lower for the nonrecombining genes for both the autosomal and X chromosome data sets (table 1). We also found significant differences among the three nonrecombining categories (table 2); N4 showed the lowest mean *Fop* and NX the highest, whereas No was intermediate between N4 and NX. The GC content at third position coding sites (GC_3) showed the same patterns as for *Fop* (tables 1 and 2).

We also examined levels of GC content in introns, separating them into short (GC_S; ≤80 bp) and long introns (GC_L; >80 bp). In the autosomal but not the X chromosome data set, GC_S was lower in the nonrecombining regions than in the recombining category (table 1). There were also significant differences among the nonrecombining regions: NX and No showed similar levels of GC_S content, but both were higher than for N4 (table 2). For GC_L, there was a significantly lower GC content in the nonrecombining category in both the autosomal and the X-linked data sets (table 1). Among the three nonrecombining categories, there were highly significant differences between N4 and both of the other two nonrecombining groups (table 2).

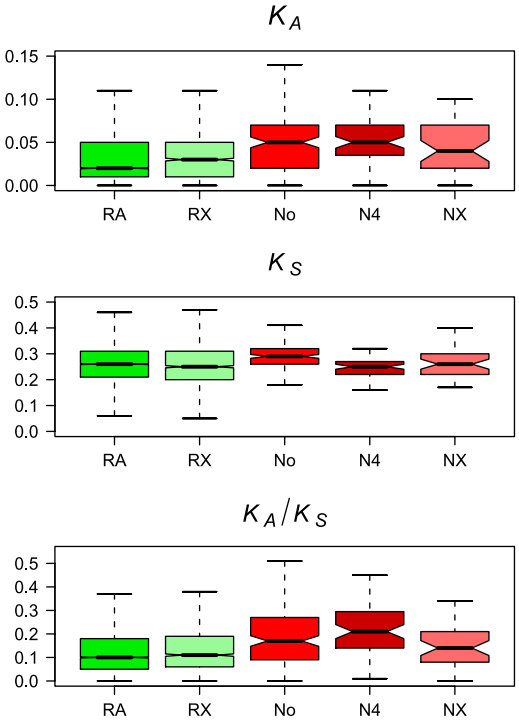

Fig. 1.—Notched box-plots of K_A, K_S, and K_A/K_S for: RA, recombining autosomal genes; RX, recombining X chromosome genes; No, nonfourth nonrecombining genes; N4, fourth chromosome genes; and NX, nonrecombining X chromosome genes. The box extends from the lower to the upper quartile, with a line in the middle at the median. The dotted bars represent the 5th and 95th percentiles. The notches represent 95% CIs for the medians.

Gene Expression and Gene Length

There were marginally significantly higher levels of overall gene expression for nonrecombining compared with RA genes, but not for maximum expression (table 1). We did not observe gene expression differences between recombining and nonrecombining X-linked genes (table 1). Within the three different nonrecombining regions, we observed marginally significantly lower overall levels of expression for NX than N4 or No (table 2). Similarly, for the autosomes, but not the X chromosome, there were highly significant differences with respect to the gene length variables. The CDS length and long introns were longer in the nonrecombining genes, whereas short introns were slightly shorter on average (table 1). Within the nonrecombining regions, N4 genes had the longest CDSs, NX the longest short introns, and No the longest long introns (table 2).

Individual Nonrecombining Regions

When comparing all nonrecombining regions separately, most parameters (except K_A, K_A/K_S, and expression) showed significant differences among the six nonrecombining regions (chromosomal arms 2L, 2R, 3L, 3R, 4, and X; supplementary table 3, Supplementary Material online). When we compared genes located in the alpha-heterochromatin against genes located in the beta-heterochromatin, for the autosomes, only *Fop*, GC_3, and GC_S showed a significant difference (table 3 and fig. 2), being lower in the alpha-heterochromatin. No such differences were found for the X chromosome, but we only had seven genes in the X alpha-heterochromatin (data not shown). We found no differences between distal and proximal autosomal beta-heterochromatin or between telomeric and centromeric nonrecombining genes on the X chromosome (data not shown).

Partial Correlation Analyses

We contrasted well established genome-wide correlations among divergence, codon usage bias, and expression

Table 2
Comparisons of the Three Nonrecombining Regions

	Region				MW *P*		
	No	N4	NX	KW *P*	No versus N4	No versus NX	N4 versus NX
Number of genes	159	67	42				
K_A	0.056 (0.049–0.063)	0.055 (0.047–0.062)	0.047 (0.035–0.056)	ns	ns	ns	ns
K_S	0.287 (0.278–0.297)	0.249 (0.238–0.259)	0.259 (0.243–0.276)	***	***	***	ns
K_A/K_S	0.190 (0.169–0.211)	0.217 (0.190–0.246)	0.173 (0.134–0.211)	ns	ns	ns	*
Fop	0.379 (0.358–0.399)	0.260 (0.247–0.272)	0.452 (0.428–0.475)	***	***	**	***
GC_3	0.480 (0.456–0.505)	0.348 (0.330–0.363)	0.572 (0.547–0.594)	***	***	**	***
GC_S	0.321(0.294–0.345)	0.228 (0.206–0.251)	0.351 (0.320–0.382)	***	***	ns	***
GC_L	0.364 (0.354–0.375)	0.320 (0.311–0.329)	0.339 (0.322–0.360)	***	***	ns	**
Overall exp.	10.28 (9.93–10.63)	10.15 (9.68–10.66)	9.224 (8.64–9.86)	*	ns	**	*
Max. exp.	12.00 (11.63–12.37)	11.61 (11.18–12.04)	11.44 (10.81–12.06)	ns	ns	ns	ns
CDS length (aa)	384 (322–420)	729 (615–875)	424.5 (248–517.5)	***	***	ns	**
Length short (bp)	59 (58.33–60)	59.33 (57.99–60.41)	64 (62.67–67)	***	ns	***	**
Length long (Kb)	2.03 (0.74–2.88)	0.89 (0.57–1)	1.22 (0.19–2.13)	*	*	ns	ns

Note.—The nonrecombining genes are divided into three regions: No, autosomal genes excluding the fourth chromosome; N4, fourth chromosome; and NX, nonrecombining genes on the X chromosome. KW *P*: Kruskal–Wallis adjusted *P* values for No, N4, and NX comparison. MW *P*, Mann–Whitney *U* test adjusted *P* values (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; and ns, not significant).

Table 3
Comparisons of Autosomal Alpha- and Beta-Heterochromatin Genes

	Region		
	Beta-Heterochromatin	Alpha-Heterochromatin	*P*
Number of Genes	131	28	
K_A	0.055 (0.047–0.063)	0.061 (0.046–0.076)	ns
K_S	0.285 (0.274–0.295)	0.299 (0.273–0.325)	ns
K_A/K_S	0.188 (0.161–0.212)	0.201 (0.155–0.245)	ns
Fop	0.404 (0.379–0.426)	0.264 (0.239–0.285)	***
GC_3	0.509 (0.484–0.536)	0.341 (0.314–0.369)	***
GC_S	0.343 (0.318–0.370)	0.228 (0.189–0.260)	***
GC_L	0.368 (0.357–0.378)	0.352 (0.332–0.381)	ns
Overall exp.	10.29 (9.88–10.67)	10.257 (9.346–11.162)	ns
Max. exp.	12.06 (11.67–12.45)	11.75 (10.89–12.60)	ns
CDS length (aa)	407 (326–455)	261 (104–324.50)	ns
Length short (bp)	59 (58.17–60)	59 (57–62)	ns
Length long (Kb)	1.65 (0.32–2.28)	5.73 (0–8.48)	ns

NOTE.—*P*, Mann–Whitney *U* test adjusted *P* value (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$; and ns, not significant).

(Bierne and Eyre-Walker 2006; Drummond and Wilke 2008; Marion de Procé et al. 2009; Haddrill et al. 2011) between recombining and nonrecombining regions because we might expect an absence of recombination to cause these relationships to break down (Haddrill et al. 2007, 2008). In general, all partial correlations among these variables (holding all variables constant other than the pair under consideration) were significant in the recombining groups, whereas most of these associations were not significant in the nonrecombining data sets (table 4; for raw correlations, see supplementary fig. 1, Supplementary Material online). However, we still found a strong association between gene expression and codon usage bias (table 4) for the nonrecombining autosomal data set (NA) and for the fourth chromosome (N4). In addition, for the NA, No, and N3 (nonrecombining third chromosome) genes, there was a strong negative correlation between gene expression and K_A and between *Fop* and K_A (table 4), whereas the NX genes only showed a significantly strong negative correlation between *Fop* and K_A. All the partial correlations between expression and K_A and between *Fop* and K_A for the four independent nonrecombining data sets (second, third, fourth, and X chromosome) showed a negative trend (table 4). These common negative trends were significant when using Fisher's method for combining probabilities from independent tests (Exp $\sim K_A$, $\chi^2 = 23.64$, df = 8, $P = 0.0026$; *Fop* $\sim K_A$, $\chi^2 = 48.16$, df = 8, $P = 0$).

Using the alternative measure maximum gene expression, instead of the overall level of gene expression, we found the same trends and significant results as above (supplementary table 4, Supplementary Material online), with the exception of the partial correlation between expression and K_A, which showed no significant results, although the trend was in the same direction. In addition, Fisher's method showed the same results as above in the nonrecombining regions for the association between expression and K_A (Exp $\sim K_A$, $\chi^2 = 19$, df = 8, $P = 0.0149$) and between codon usage bias and K_A (*Fop* $\sim K_A$, $\chi^2 = 49.62$, $P = 0$).

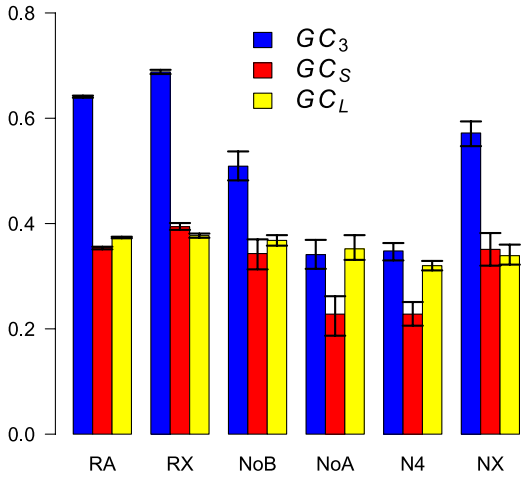

FIG. 2.—GC content of the third position of codons (GC_3), short introns (GC_S), and long introns (GC_L) for: RA, recombining autosomal genes; RX, recombining X chromosome genes; NoB, nonfourth beta-heterochromatin genes; NoA, nonfourth alpha-heterochromatin genes; N4, fourth chromosome genes; and NX, nonrecombining X chromosome genes. Values reported are means; error bars indicate 95% CIs obtained by bootstrapping.

Discussion

Nonsynonymous Divergence

As expected from previous analyses of *Drosophila* (Bachtrog 2005; Bartolomé and Charlesworth 2006; Haddrill et al. 2007; Bachtrog et al. 2008; Larracuente et al. 2008; Arguello et al. 2010), K_A and K_A/K_S are significantly higher in the autosomal nonrecombining regions than in the recombining regions (table 1). These results are consistent with less effective selection against weakly deleterious mutations in nonrecombining regions due to increased HRI when crossing-over is absent (Charlesworth et al. 2010). Our results largely confirm and extend the conclusions of Haddrill et al. (2007), although they only found significant effects of a lack of recombination for the fourth chromosome. In contrast, Arguello et al. (2010) found that autosomal heterochromatic genes behaved similarly to fourth chromosome genes, in agreement with our results (it is, however, unclear how many heterochromatic genes were included in that study). Interestingly, there is no significant difference in K_A or K_A/K_S between recombining and nonrecombining regions of the X chromosome (table 1), which may suggest smaller effects of HR interference on this chromosome. Nonetheless, K_A and K_A/K_S are slightly higher for the NX genes compared with recombining X-linked genes,

Table 4
Partial Correlations among Divergence Levels, Codon Usage Bias, and Expression

	Variables				
Region	***Fop*** **~** K_S	**Exp ~** ***Fop***	***Fop*** **~ CDS length**	**Exp ~** K_A	***Fop*** **~** K_A
RA	−0.103***	0.298***	−0.188***	−0.116***	−0.307***
	(−0.130/−0.077)	(0.272/0.323)	(−0.214/−0.162)	(−0.140/−0.090)	(−0.333/−0.285)
RX	−0.085*	0.240***	−0.276***	−0.116***	−0.246***
	(−0.150/−0.015)	(0.180/0.303)	(−0.330/−0.216)	(−0.181/−0.052)	(−0.309/−0.188)
NA	0.139 (0.242)	0.213** (0.085)	−0.046 (0.142)	−0.232** (0.116)	−0.205** (0.102)
No	0.010 (0.113)	0.117 (0.181)	0.058 (0.246)	−0.238* (0.122)	−0.335** (0.028)
N2	0.050 (0.153)	0.257 (0.041)	0.034 (0.222)	−0.114 (0.002)	−0.139 (0.168)
N3	0.051 (0.154)	−0.118 (0.416)	−0.042 (0.146)	−0.428*** (0.312)	−0.539*** (0.232)
N4	−0.008 (0.095)	0.375** (0.077)	0.204 (0.392)	−0.223 (0.107)	−0.201 (0.106)
NX	−0.087 (0.002)	−0.242 (0.482)	0.019 (0.295)	−0.226 (0.110)	−0.708*** (0.462)
Covariates	Exp, K_A, GC_S, CDS length	K_A, K_S, GC_S, CDS length	K_A, K_S, GC_S, Exp	K_S, *Fop*, GC_S, CDS length	Exp, K_S, GC_S, CDS length

Note.—We examined eight regions: RA, recombining autosomal genes; RX, recombining X chromosomal genes; NA, nonrecombining autosomal genes; No, nonrecombining autosomal genes excluding the fourth chromosome; N2, nonrecombining genes on the second chromosome; N3, nonrecombining genes on the third chromosome; N4, fourth chromosome genes; and NX, nonrecombining X chromosome genes. We show Spearman's rank partial correlation coefficient and its significance level (***$P < 0.001$; **$P < 0.01$; and *$P < 0.05$). For the recombining genes, we show the 95% CIs of the Spearman partial correlations in parentheses. For the nonrecombining genes, we show the absolute difference between partial correlations for the recombining versus nonrecombining categories in parentheses; the differences that are outside the 95% CI limit are shown as underlined values.

especially in centromeric regions, so that the lack of significance may simply reflect the small number of genes involved.

The similarities in K_A and K_A/K_S among the high, intermediate, and low recombination bins, also found by Haddrill et al. (2007) and Larracuente et al. (2008), suggest that relatively low levels of recombination are enough to counteract any HRI effects. This result suggests that rates of adaptive protein sequence evolution are not higher in regions of high recombination, as has been proposed (Betancourt and Presgraves 2002; Presgraves 2005). Overall, there seems to be a similar pattern of faster protein sequence evolution in every nonrecombining region of the genome, with the exception of the X chromosome, which also shows different patterns from the autosomes with respect to codon usage bias and K_S in recombining regions (Singh et al. 2005b, 2008; Vicoso et al. 2008).

Synonymous Divergence and Codon Usage Bias

Synonymous divergence, as measured by K_S is also slightly but significantly higher for nonrecombining than RA genes, whereas codon usage bias is significantly lower in the nonrecombining regions of both X and autosomes (table 1). Given the lower levels of codon usage bias and the lack of a negative correlation between K_S and *Fop* in nonrecombining regions, in contrast to the negative partial correlation between them in recombining regions (table 4), this suggests that selection on codon usage bias is reduced when crossing-over is absent, resulting in a higher level of synonymous divergence.

Weakly selected sites are especially susceptible to HRI effects (Charlesworth et al. 2010); as HRI increases in intensity, they should therefore approach neutrality. The lowest *Fop* values are on the fourth chromosome, suggesting that selection is least efficient on this chromosome, possibly due to a longer history of no crossing-over in this region than other regions that lack crossing-over (Haddrill et al. 2007), or a larger concentration of genes in a single genomic region that lacks crossing-over (K_A/K_S is also highest for the fourth chromosome). Surprisingly, despite its lower *Fop*, K_S is lower for the fourth chromosome than for the other nonrecombining autosomal genes. This may be due to higher synonymous substitution rates in regions that have slightly larger N_e values than on the fourth chromosome. An increase in the rate of sequence evolution as $N_e s$ increases away from zero is expected when there is a strong mutational bias toward slightly deleterious variants (Eyre-Walker 1992; McVean and Charlesworth 1999; Lawrie et al. 2011). This means that it is possible that slightly higher values of N_e for the other nonrecombining autosomal genes could result in a higher rate of synonymous evolution compared with the fourth chromosome.

The magnitude of the expected reduction in *Fop* on the fourth chromosome can be roughly estimated as follows. Using polymorphism data for an autosomal set of normally recombining genes, Zeng and Charlesworth (2010) estimated that the mean scaled intensity of selection on preferred codons ($\gamma = 4N_e s$, where s is the selection coefficient for heterozygotes for a preferred variant) was 1.29 and the mutational bias (κ) toward nonpreferred codons was 2.84 (table 2 in Zeng and Charlesworth [2010]). Using the Li–Bulmer formula for equilibrium codon usage (Bulmer 1991), the expected value of *Fop* is $1/(1 + \kappa \exp[-\gamma])$. The predicted *Fop* for recombining genes is then 0.56, slightly higher than the value in table 1. Studies of silent diversity on the fourth chromosome suggest that its N_e is approximately 10% of that for normally recombining regions (Arguello et al. 2010; Charlesworth et al. 2010); this reduces γ to 0.13, so that the predicted value of *Fop* for the

fourth chromosome is 0.29, again slightly higher than observed. With no selection, the predicted *Fop* is 0.26, very close to the fourth chromosome value (table 2).

Fop is also significantly lower for nonrecombining X-linked genes than in recombining X chromosome regions, but the reduction in *Fop* for NX genes is half of that observed for the autosomal counterparts. *Fop* is elevated on the X chromosome relative to that on the autosomes (Singh et al. 2005b), and this effect is even observed in nonrecombining X-linked versus autosomal regions (table 2). This result differs from that of Singh et al. (2005a), who report a negative effect of recombination on X chromosome codon usage, and probably reflects the fact that our X-linked data set covers a broader range of recombination rates, including genes in nonrecombining regions, whereas the lowest recombination rate category in their study was 0.27 cM/Mb. There is evidence from population genetic analyses that the scaled intensity of selection on preferred versus unpreferred codons, γ, is somewhat higher for the freely recombining part of the X chromosome than that for the autosomes in *D. melanogaster* (Zeng and Charlesworth 2010). However, if γ is reduced in the nonrecombining regions of the X chromosome by the same factor as for the fourth chromosome, the predicted equilibrium value of *Fop* from the estimates of κ and γ in table 2 of Zeng and Charlesworth (2010) is around 30%, far lower than the observed value. This strongly suggests a smaller reduction in N_e for the nonrecombining part of the X than for the autosomes.

GC Content

GC content at third coding positions in nonrecombining regions is higher than for introns (fig. 2), which suggests that some selection is still acting in favor of preferred codons, which mostly end in GC in *Drosophila* (Akashi 1994). Zeng and Charlesworth (2010) also estimated selection and mutational parameters for GC versus AT in intronic sites (mostly from short introns), obtaining $\gamma = 0.60$ and $\kappa = 3.42$, giving a predicted equilibrium value of 0.35, fairly close to the observed short intronic GC content for RA genes in table 1. The corresponding predicted value for the fourth chromosome is 0.24, similar to the observed value of 0.23 (table 2). However, the GC content of most other nonrecombining regions is higher, suggesting that for much of their evolutionary history, their effective population sizes have been higher than for the fourth chromosome.

It is possible that there is a gradient in the efficiency of selection on synonymous sites within the nonrecombining regions. The genes present in the alpha-heterochromatin, which are located closer to the centromere, have a significantly lower GC_3 content than in the beta-heterochromatin (adjacent to the euchromatin). It therefore seems likely that the base composition at third position coding sites in the alpha-heterochromatin is closer to neutral equilibrium than in other nonrecombining regions. However, there is no significant difference among the heterochromatic genes adjacent to the euchromatin (beta-heterochromatin) when we divided these into two groups according to their proximity to the centromere. The larger effect observed in alpha- compared with beta-heterochromatin genes might therefore be due to the location of the former in regions where recombination is totally absent. It is possible that some recombination due to gene conversion or low residual levels of crossing-over may occur in the beta-heterochromatin adjacent to the euchromatin, as has been found for the dot chromosomes (Betancourt et al. 2009; Arguello et al. 2010).

Furthermore, the difference in GC_3 between recombining and nonrecombining regions cannot be caused solely by lower GC-biased gene conversion in these regions or a higher AT mutational bias because the drop in GC content in nonrecombining genes is much higher for GC_3 (30%) than for short introns (18% for GC_S). Interestingly, the significantly lower short intron GC content in nonrecombining compared with RA genes suggests that short introns in recombining regions are under some selective constraints, contrary to what is often assumed (Halligan and Keightley 2006; Parsch et al. 2010).

Gene Expression

There is no evidence for lower expression of genes in any of the nonrecombining regions in this analysis (except the U genes, which were excluded from our analyses). Indeed, if anything, there are slightly higher levels of gene expression in the nonrecombining compared with the recombining genes. These results are in the same direction as those observed by Haddrill et al. (2008) who reported significantly higher gene expression in the genes that lack crossing-over, using ESTs. However, in contrast to their results, gene expression on the fourth chromosome is similar to that on the other autosomes, as was also found for the dot chromosome of *D. virilis* (Betancourt et al. 2009). It follows, therefore, that the higher nonsynonymous divergence and lower *Fop* of nonrecombining genes cannot be explained by the well-documented negative correlation between expression level and nonsynonymous divergence (Marais et al. 2004; Drummond and Wilke 2008; Larracuente et al. 2008).

Gene Length

The greater length of long introns in the nonrecombining versus the recombining genes is consistent with the findings of Smith et al. (2007), who observed that heterochromatic introns are enriched in fragmented transposable element (TE) sequences and show less length conservation in interspecies comparisons. Nonrecombining regions tend to have longer introns, which are probably due to an

accumulation of TE-derived sequences as a result of weakened counterselection, either due to HRI or a lack of ectopic exchange for TEs (Bartolomé et al. 2002); overall, TE density is very high in the heterochromatin (Smith et al. 2007).

Long introns are especially long in the nonfourth chromosome autosomal nonrecombining genes (No), particularly in the alpha-heterochromatin. The fact that the long introns on the fourth chromosome are shorter than in other nonrecombining regions is consistent with the lower fraction of TE-derived DNA, especially retroviral-like elements, on the fourth chromosome compared with the nonrecombining regions of chromosome 2R and the X chromosome (Bartolomé et al. 2002; Kaminker et al. 2002), the reasons for which remain obscure. In contrast, short introns in nonrecombining autosomal regions are slightly smaller than in autosomal recombining regions, which could be due to relaxed selection on minimal intron length for correct splicing (Comeron and Kreitman 2000).

Interestingly, only the fourth chromosome shows evidence for longer CDS length (table 2), with genes that are on average twice as long as in other parts of the genome. Since longer proteins entail a greater metabolic cost (Akashi 1996), protein length can increase when selective constraints are relaxed, so that small insertions into coding regions that have very small effects on fitness can then be fixed more frequently. Weaker translational selection has been proposed as the cause of the longer genes observed in *D. melanogaster* than in *D. simulans*, as a result of the apparently smaller effective population size along the *D. melanogaster* lineage compared with that in *D. simulans* (Akashi 1996). It is possible that this process has affected the fourth chromosome but not the other nonrecombining regions, consistent with the other evidence that it has been subject to more intense HRI.

The Effect of a Lack of Recombination on Genome-Wide Relationships among Variables

In a completely nonrecombining block of genes, all genes must be subject to the same intensity of HRI effects. Correlations induced by HRI effects that act specifically within the gene, such as the postulated effect on codon usage of amino acid fixations driven by positive selection (Betancourt and Presgraves 2002; Presgraves 2005), should therefore be largely absent, especially if there is little or no fixation of positively selected variants in these regions, as suggested by recent results for the dot chromosome (Betancourt et al. 2009; Arguello et al. 2010). Similarly, if selection on synonymous sites is greatly reduced by HRI effects throughout a nonrecombining region, relationships between genomic features that reflect such weak selection (e.g., the negative correlation between *Fop* and K_S) should be greatly reduced in strength (Haddrill et al. 2007, 2008).

Some of these genome-wide relationships are nonsignificant in nonrecombining regions and significantly smaller than in the corresponding recombining genes (table 4); for example, the negative partial correlations between *Fop* and K_S and *Fop* and CDS length. Nevertheless, within the nonrecombining regions, some footprints of selection are observed because there are significant partial correlations between expression and *Fop*, expression and K_A, and *Fop* and K_A. There is also a higher GC content in third position coding sites than introns in nonrecombining regions.

Paradoxically, the negative partial correlation between expression level and K_A is larger in magnitude for nonrecombining than recombining genes, for both the X chromosome and most of the autosomes. This may reflect selection on translational accuracy, provided that highly expressed genes are more selectively constrained than lowly expressed genes (Akashi 1994; Bierne and Eyre-Walker 2006; Drummond and Wilke 2008), because the accelerated fixation of slightly deleterious nonsynonymous mutations is probably most marked for genes under weak selective constraints, corresponding to those with lower expression levels. This is because of the strongly nonlinear dependence of the fixation probability of a deleterious mutation on the product of N_e and the selection coefficient s against a deleterious mutation (Kimura 1983, p. 44). When $N_es >> 1$, an increase in N_es has little effect, but with N_es around 1, there can be a substantial decrease in the fixation probability of a deleterious mutation as N_es increases.

There has been some controversy regarding the causes of the negative correlation between K_A and codon usage bias, and whether it is caused by more intense HRI resulting from more frequent selective sweeps of favorable amino acid mutations in genes with higher K_A values (Betancourt and Presgraves 2002; Andolfatto 2007) or by stronger selection on translational accuracy in genes with more highly constrained protein sequences (Akashi 1994; Bierne and Eyre-Walker 2006; Drummond and Wilke 2008).

The fact that *Fop* and K_A remain negatively associated in the absence of crossing-over suggests that this relationship is at least partly caused by correlated differences in levels of selective constraint on protein sequence and codon usage across genes and that selection is still partially effective in the absence of crossing-over. Since we have used partial correlations, these constraints must be independent of gene expression levels, at least as measured by the methods used to generate the data we have employed, in contradiction to the predictions of the model of Drummond and Wilke (2008). This may simply reflect differences in the overall importance of proteins for cellular functions, with more important proteins showing both higher levels of constraint on their amino acid sequence and stronger selection for codon usage.

A significant genome-wide negative partial correlation between *Fop* and K_A was previously found by Marion de Procé et al. (2009) and Haddrill et al. (2011), who interpreted it as potentially indicating an effect of the

fixation of adaptive mutations on the efficacy of selection on codon usage bias, but its existence in nonrecombining regions cannot be explained in this way, unless there is sufficient residual recombination in the nonrecombining regions to allow different genes in the same region to evolve independently.

It is, however, unlikely that this is the case. Kim (2004) has modeled the effect of adaptive substitutions on levels of codon usage bias in normally recombining *Drosophila* genes. He concluded that fairly strong selection on favorable mutations ($N_e s$ values of around 100) is required to explain the observed pattern for these genes, on this basis. However, the highest estimate by Arguello et al. (2010) of the product of N_e and recombination rate for the entire fourth chromosome was about 20 (for *D. simulans*). Even allowing for the reduced N_e of the fourth chromosome (bringing an $N_e s$ of 100 for a recombining region down to 10), a selective sweep due to selection of the intensity proposed by Kim (2004) would have an effect extending well beyond a single gene (for which the $N_e r$ value for the fourth chromosome is around 0.5 or less) because a recombination rate of the order of the selection coefficient is needed to remove the effect of a sweep (see fig. 2 of Kim 2004).

The partial negative correlations between *Fop* and K_A that are observed for the four independent nonrecombining regions suggest that neither HRI caused by selective sweeps nor selective constraints dependent on gene expression are responsible for the association between *Fop* and K_A in these genes. However, we cannot rule out the possibility that levels of gene expression in a particular tissue or set of tissues, not captured in our measure of overall level of gene expression, could be strong enough to affect selection on codon usage bias. In any case, this association suggests that some selection is still acting on both codon usage and on nonsynonymous mutations in these regions. Future analyses using polymorphism data to estimate selection intensities may help to test this possibility.

Conclusion

We have found only small differences among genes with different frequencies of crossing-over in the regions of the genome that have crossing-over. All regions that lacked crossing-over showed at least some effects of the type expected from increased HRI: an accelerated rate of protein sequence evolution, lower codon usage bias, and lower GC content in both coding regions and introns. However, there were some significant differences among nonrecombining regions. In general, the nonrecombining genes on the X chromosome show less severe effects, whereas the fourth chromosome and the autosomal genes located most proximal to the centromeres exhibited the most intense effects of HRI. Nonetheless, there was evidence for some residual effects of selection acting on nonrecombining genes.

Supplementary Material

Supplementary tables 1–4 and figure 1 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

We gratefully acknowledge Pablo Librado and Filipe Vieira for help with the bioinformatic and statistical analyses and for kindly providing computer code and Dan Halligan for help with the statistical analyses. We thank members of the Charlesworth lab group for helpful discussions and comments. We thank C. Smith and G. Karpen for advice on how to annotate the heterochromatic genes. We are also grateful to two anonymous reviewers for their comments on the manuscript. This work was supported by the UK Biotechnology and Biological Sciences Research Council (grant number BB/H006028/1 to B.C.). P.R.H. is supported by a Fellowship from the Natural Environment Research Council (grant number NE/G013195/1).

Literature Cited

Akashi H. 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. Genetics 136:927–935.

Akashi H. 1996. Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. Genetics 144:1297–1307.

Andolfatto P. 2007. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. Genome Res. 17:1755–1762.

Arguello JR, et al. 2010. Recombination yet inefficient selection along the *Drosophila melanogaster* subgroup's fourth chromosome. Mol Biol Evol. 27:848–861.

Bachtrog D. 2005. Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in *Drosophila*. Genome Res. 15:1393–1401.

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.

Bartolomé C, Maside X, Charlesworth B. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. Mol Biol Evol. 19:926–937.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Biol Sci. 57:289–300.

Betancourt A, Presgraves DC. 2002. Linkage limits the power of natural selection in *Drosophila*. Proc Natl Acad Sci U S A. 99:13616–13620.

Betancourt A, Welch JJ, Charlesworth B. 2009. Reduced effectiveness of selection caused by a lack of recombination. Curr Biol. 19:655–660.

Bierne N, Eyre-Walker A. 2006. Variation in synonymous codon use and DNA polymorphism within the *Drosophila* genome. J Evol Biol. 19:1–11.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics 129:897–907.

Charlesworth B. 1996. Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet Res. 68:131–149.
Charlesworth B, Betancourt A, Kaiser VB, Gordo I. 2010. Genetic recombination and molecular evolution. Cold Spring Harb Symp Quant Biol. 74:177–186.
Charlesworth D. 2003. Effects of inbreeding on the genetic diversity of populations. Philos Trans R Soc B Biol Sci. 358:1051–1070.
Clark AG, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450:203–218.
Comeron JM. 1995. A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J Mol Evol. 41:1152–1159.
Comeron JM, Kreitman M. 2000. The correlation between intron length and recombination in *Drosophila*: dynamic equilibrium between mutational and selective forces. Genetics 156:1175–1190.
Comeron JM, Kreitman M, Aguadé M. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. Genetics 151:239–249.
Comeron JM, Williford A, Kliman RM. 2008. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity 100:19–31.
Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134:341–352.
Eyre-Walker A. 1992. The effect of constraint on the rate of evolution in neutral models with biased mutation. Genetics 131:233–234.
Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 78:737–756.
Gelbart WM, Emmert DB. 2010. FlyBase Reference Report: Gelbart and Emmert, 2010.10.13, FlyBase High Throughput Expression Pattern Data Beta Version; [cited 2011 Mar]. Available from: http://flybase.org/reports/FBrf0212041.html
Gordo I, Charlesworth B. 2001. Genetic linkage and molecular evolution. Curr Biol. 11:R684–R686.
Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the *Drosophila genome* that lack crossing over. Genome Biol. 8:R18.
Haddrill PR, Waldron FM, Charlesworth B. 2008. Elevated levels of expression associated with regions of the *Drosophila* genome that lack crossing over. Biol Lett. 4:758–761.
Haddrill PR, Zeng K, Charlesworth B. 2011. Determinants of synonymous and nonsynonymous variability in three species of *Drosophila*. Mol Biol Evol. 28:1731–1743.
Halligan DL, Keightley PD. 2006. Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. Genome Res. 16:875–884.
Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res. 8:269–294.
Kaminker JS, et al. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. Genome Biol. 3:research0084. 1–84.2.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
Kim SH, Yi SV. 2006. Understanding relationship between sequence and functional evolution in yeast proteins. Genetica 131:151–156.
Kim Y. 2004. Effect of strong directional selection on weakly selected mutations at linked sites: implication for synonymous codon usage. Mol Biol Evol. 21:286–294.
Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.
Larracuente AM, et al. 2008. Evolution of protein-coding genes in *Drosophila*. Trends Genet. 24:114–123.
Lawrie DS, Petrov DA, Messer PW. 2011. Faster than neutral evolution of constrained sequences: the complex interplay of mutational biases and weak selection. Genome Biol Evol. 3:383–395.
Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189.
Marais G, Domazet-Loso T, Tautz D, Charlesworth B. 2004. Correlated evolution of synonymous and nonsynonymous sites in *Drosophila*. J Mol Evol. 59:771–779.
Marion de Procé S, Halligan DL, Keightley PD, Charlesworth B. 2009. Patterns of DNA-sequence divergence between *Drosophila miranda* and *D. pseudoobscura*. J Mol Evol. 69:601–611.
McVean GAT, Charlesworth B. 1999. A population genetic model for the evolution of synonymous codon usage: patterns and predictions. Genet Res. 74:145–158.
Miklos GL, Cotsell JN. 1990. Chromosome structure at interfaces between major chromatin types: *alpha*- and *beta*-heterochromatin. Bioessays 12:1–6.
Parsch J, Novozhilov S, Saminadin-Peter SS, Wong KM, Andolfatto P. 2010. On the utility of short intron sequences as a reference for the detection of positive and negative selection in *Drosophila*. Mol Biol Evol. 27:1226–1234.
Peden JF. 1999. Analysis of codon usage [PhD thesis]. [Nottingham (UK)]: University of Nottingham. CodonW: Correspondence analysis of codon usage; [cited 2011 Feb]. Available from: http://codonw.sourceforge.net/.
Pollard K, Dudoit S, Van der Laan MJ. 2005. Multiple testing procedures: R multtest package and applications to genomics. In: Bioinformatics and computational biology solutions using R and Bioconductor. New York: Springer-Verlag. p. 251–272.
Presgraves DC. 2005. Recombination enhances protein adaptation in *Drosophila melanogaster*. Curr Biol. 15:1651–1656.
Qiu S, Zeng K, Slotte T, Wright S, Charlesworth D. 2011. Reduced efficacy of natural selection on codon usage bias in selfing *Arabidopsis* and *Capsella* species. Genome Biol Evol. 3:868–880.
She R, et al. 2011. GenBlastG: extending BLAST to be a high performance gene finder. Bioinformatics 27:2141–2143.
Singh ND, Davis JC, Petrov DA. 2005a. Codon bias and noncoding GC content correlate negatively with recombination rate on the *Drosophila* X chromosome. J Mol Evol. 61:315–324.
Singh ND, Davis JC, Petrov DA. 2005b. X-linked genes evolve higher codon bias in *Drosophila* and *Caenorhabditis*. Genetics 171:145–155.
Singh ND, Larracuente AM, Clark AG. 2008. Contrasting the efficacy of selection on the X and autosomes in *Drosophila*. Mol Biol Evol. 25:454–467.
Smith C, Shu S, Mungall CJ, Karpen GH. 2007. The Release 5.1 annotation of *Drosophila melanogaster* heterochromatin. Science 316:1586–1591.
Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics 19:2325–2327.
Tweedie S, et al. 2009. FlyBase: enhancing *Drosophila* gene ontology annotations. Nucleic Acids Res. 37:D555–D559.
Vicoso B, Haddrill PR, Charlesworth B. 2008. A multispecies approach for comparing sequence evolution of X-linked and autosomal sites in Drosophila. Genet Res. 90:421–431.
Zeng K, Charlesworth B. 2010. Studying patterns of recent evolution at synonymous sites and intronic sites in *Drosophila melanogaster*. J Mol Evol. 70:116–128.

Associate editor: Marta Wayne