Locus-Specific Decoupling of Base Composition Evolution at Synonymous Sites and Introns along the *Drosophila melanogaster* and *Drosophila sechellia* Lineages

Vanessa L. Bauer DuMont, Nadia D. Singh, Mark H. Wright, and Charles F. Aquadro

Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York

Selection is thought to be partially responsible for patterns of molecular evolution at synonymous sites within numerous *Drosophila* species. Recently, "per-site" and likelihood methods have been developed to detect loci for which positive selection is a major component of synonymous site evolution. An underlying assumption of these methods, however, is a homogeneous mutation process. To address this potential shortcoming, we perform a complementary analysis making gene-by-gene comparisons of paired synonymous site and intron substitution rates toward and away from the nucleotides G and C because preferred codons are G or C ending in Drosophila. This comparison may reduce both the false-positive rate (due to broadscale heterogeneity in mutation) and false-negative rate (due to lack of power comparing small numbers of sites) of the per-site and likelihood methods. We detect loci with patterns of evolution suggestive of synonymous site selection pressures predominately favoring unpreferred and preferred codons along the *Drosophila melanogaster* and *Drosophila sechellia* lineages, respectively. Intron selection pressures do not appear sufficient to explain all these results as the magnitude of the difference in synonymous and intron evolution is dependent on recombination environment and chromosomal location in a direction supporting the hypothesis of selectively driven synonymous fixations. This comparison identifies 101 loci with an apparent switch in codon preference between *D. melanogaster* and *D. sechellia*, a pattern previously only observed at the *Notch* locus.

Introduction

In Drosophila, the nonrandom use of synonymous codons (codon bias) is well documented. This phenomenon is primarily thought to be due to selective pressures related to translational accuracy and/or efficiency (e.g., Akashi 1994, 1995; Akashi et al. 1998; Kliman 1999; Drummond and Wilke 2008). Other phenotypic attributes associated with synonymous mutations include protein and mRNA folding and stability (e.g., Duan et al. 2003; Oresic et al. 2003; Stenøien and Stephan 2005; Biro 2006; Kimchi-Sarfaty et al. 2007; Drummond and Wilke 2008), mutagenesis (Hoede et al. 2006), protein expression (e.g., Carlini and Stephan 2003; Duan et al. 2003; Baines et al. 2004; Carlini 2004), and maintenance of exon/intron splice conjunctions (Warnecke and Hurst 2007).

By convention, "preferred" codons in Drosophila refer to those identified as being used significantly more often in genes with highly biased codon usage compared with those with low levels of codon bias (Akashi 1995; Bachtrog 2007). All other codons are labeled "unpreferred." A longstanding observation in Drosophila is that preferred codons end in the nucleotides G and C (e.g., Shields et al. 1988; Akashi 1995; Bachtrog 2007; Vicario et al. 2007). Although this pattern is observed across numerous Drosophila species, there are notable differences in the degree of codon bias toward the GC-rich preferred codons (Ko et al. 2006; Akashi et al. 2007; Vicario et al. 2007). Two examples are the increased use of A- and T-ending codons along the *Drosophila melanogaster* (e.g., Akashi 1996; McVean and Vieira 2001; Bauer DuMont et al. 2004) and *Drosophila willistoni* lineages (e.g., Rodríguez-Trelles et al. 1999; Powell et al. 2003; Singh et al. 2006; Vicario et al. 2007). The increased rate of fixation of unpreferred codons along the *D. melanogaster* lineage has long been attributed to a lineage-specific relaxation of selective pressures favoring GC-ending preferred codons (Akashi 1995, 1996; McVean and Vieira 2001). Although this hypothesis explains many features of the data, recent findings suggest that for some loci, unpreferred codons are being driven to fixation by natural selection in the *D. melanogaster* and the *Drosophila simulans* clade (Bauer DuMont et al. 2004; Neafsey and Galagan 2007; Nielsen et al. 2007; Singh et al. 2007; Holloway et al. 2008).

In this study we compare, on a gene-by-gene basis, substitution rates of GC-increasing (AT to GC) and GC-decreasing (GC to AT) mutations between introns and synonymous sites (a combination of 2-fold and 4-fold degenerate sites). We consider these rate comparisons to be less affected by the observed regional fluctuations in mutation processes in Drosophila (Singh, Arndt, and Petrov 2005; Singh, Davies, and Petrov 2005; Singh et al. 2007) than the previously mentioned "per-site" likelihood comparisons given the proximity of exons and introns and their joint transcription (although there is no evidence of transcription dependent mutation in *D. melanogaster*; i.e., Sekelsky et al. 2000). The GC bias of preferred codons also allows us to make a priori substitution pattern predictions regarding synonymous site selection pressures. For example, selection favoring preferred codons is expected to increase G and C substitutions at synonymous sites relative to neighboring introns. On the other hand, selection favoring unpreferred codons is expected to decrease G and C substitutions specifically at synonymous sites. Our comparisons not only detect loci in *D. melanogaster* and *Drosophila sechellia* consistent with natural selection fixing preferred codons but also expand the list of genes for which selection appears to favor unpreferred codons.

Methods

Sequence Acquisition

We consider the same set of loci as Singh et al. (2007) and focus our analysis on the following species: *D. melanogaster*, *D. sechellia*, and *Drosophila yakuba*. These loci correspond to a set of protein-coding genes with 1:1 orthologs across six *D. melanogaster* group species. We

obtained aligned intron sequence for these open reading frames (where available) from the whole-genome alignments available at the UCSC Genome Browser (http://genome-test. cse.ucsc.edu/cgi-bin/hgGateway; Karolchik et al. 2003). These alignments have had ambiguous regions masked, such that we only analyzed intron regions that could be reliably aligned across these three species. In an effort to reduce bias associated with small numbers, a locus was only included in the analysis if we observed both at least five synonymous and five intron substitutions.

A program was written to map the downloaded intron sequence fragments to their corresponding open reading frame using the whole-genome assembly of *D. melanogaster* (release 4.2). For cases where there is a small open reading frame within an intron of another longer locus, we assigned the intron to the longer transcript. Using multiple Blast routines, intron regions that at times are also exons (due to alternative splicing or embedded open reading frames) were masked and not included in subsequent analyses. To remove potentially constrained splice sites, 6 bp from both ends of each intron were eliminated.

For the coding regions, we only considered synonymous changes at 2-fold and 4-fold degenerate sites that were designated as such in the most recent common ancestor of *D. melanogaster* and *D sechellia*, which was deduced using parsimony. The rate comparisons were performed considering 2-fold and 4-fold sites separately and combined. When analyzed separately, there is moderate overlap (10–38%) in the loci identified as rejecting the rate comparisons between 2-fold and 4-fold sites. When they are analyzed together (using only transitions at 4-fold and intron sites), most of the loci that reject overlap with those detected in the separate analysis (64–87%). Thus, we only present the results from the combined analysis.

GC-Increasing or GC-Decreasing Substitution Rate Comparisons

For both synonymous and intron sites, *D. yakuba* was used as an outgroup to reconstruct the ancestral sequence of *D. melanogaster* and *D. sechellia* under parsimony. Nucleotide positions for which parsimony reconstruction was ambiguous (i.e., 3 nt present in three species) were removed from further analysis. Using this ancestral sequence, we calculate what we term "GC rate" and "AT rate" at synonymous sites and intron sites, separately. The GC rate is the number of substitutions along a lineage from an A or T to a G or C divided by the total number of A's and T's in the ancestral sequence. Correspondingly, the AT rate is the number of G or C to A or T substitutions divided by the total number of G's and C's in the ancestor. Both the GC rate and AT rate were compared between synonymous sites and intron sites at each locus individually. The hypothesis of equal rates was evaluated using the 2×2 contingency table Chi-square statistic with 1 degree of freedom. False-discovery rate (FDR) was corrected for using the method of Storey (2002). Gene Ontology (GO) analysis was performed using the web-based program GOstat (Beissbarth and Speed 2004).

Comparisons with Chromosomal Location and Recombination Rate

We investigate how lineage-specific synonymous and intron divergence along with GC and AT rates are affected by a locus' chromosomal location and recombination environment. For these comparisons, we use what we call the GC- and AT-"adjusted" rates. The rates are adjusted in an attempt to account for reported differences between the X chromosome and autosomes in the mutation process (e.g., Singh et al. 2007). It is well documented (Petrov and Hartl 1999) that mutation in Drosophila is AT biased. Although true for both sex and nonsex chromosomes, in *D. melanogaster*, the mutational bias toward A and T appears to be stronger on the X chromosome (Nielsen et al. 2007; Singh et al. 2007). Using the mutation matrices given in Singh et al. (2007), we adjusted the ancestral GC content (and thus the denominator of the rate calculation) for both *D. melanogaster* and *D. sechellia* by multiplying the number of ancestral G's, C's, A's, and T's by the values given in supplementary table 1 (Supplementary material online). In essence, these values depict the relative increase or decrease in mutation potential for each nucleotide from the expectation of equal rate of change across the 12 possible mutations following Bauer DuMont et al. (2004). Total synonymous and intron divergence were similarly adjusted (see Bauer DuMont et al. 2004). All comparisons are made using relative rates; in other words, the GC or AT rate at a locus is given relative to the total substitution rate at that locus.

The recombination comparisons are only performed with substitutions along the *D. melanogaster* lineage as this is the only species with a detailed genetic map. Relationships were evaluated through nonparametric partial correlations using the R library pcor.R. Estimates of recombination were calculated based on the relative location of genetic markers on both the genetic and physical maps. We used two approaches, the regression polynomial (RP) method and the adjusted coefficient of exchange (ACE; Kindahl 1994). With RP, a third order polynomial curve was fit across loci with the genetic position as a function of physical position (based on Release 4.3 of the *D. melanogaster* genome). The recombination rate (in units of cM/Mb) is estimated as the derivative of this polynomial at a given nucleotide coordinate. For our analyses, we estimated recombination rate at the midpoint of each gene. ACE was calculated using a set of 494 markers, which had precise genetic map positions and were consecutively placed (with reference to each other) along the published genome sequence (based on Release 4.3 of the *D. melanogaster* genome). The rate of crossing-over (in units of cM/Mb) was calculated for any pair of markers that were within 1 Mb of each other. For any position in the genome, the ACE estimate is calculated by averaging the cM/Mb values for all pairs of markers that span the region of interest.

These two measures of local recombination rate are highly correlated, though they differ in the details of local heterogeneity because of the smoothing effect of the RP method. Our results are largely independent of the recombination rate estimator used. Given uncertainty in the *D. melanogaster* sex ratio, the recombination rates (RP and ACE) on the X chromosome were not multiplied by
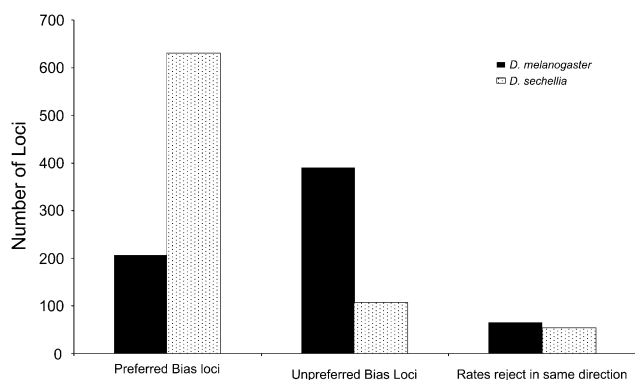
Fig. 1.—Number of loci rejecting the rate comparisons between synonymous and intron sites along the *D. melanogaster* and *D. sechellia* lineages after 10% FDR correction.

4/3 (which would be the appropriate correction if the sex ratio was known to be 1:1).

Results

We make gene-by-gene comparisons between paired synonymous sites (2-fold and 4-fold degenerate) and intron in what we term GC rate and AT rate. In brief, the GC rate is the number of substitutions from A or T to G or C divided by the number of A's and T's in the ancestor. The AT rate is the number of G or C to A or T substitutions divided by the number of G's and C's in the ancestor. The purpose of these comparisons is to deduce underlying selective pressures acting on synonymous sites presumably in a manner not compromised by the assumption of a homogeneous mutation process. We note, it is unlikely that this comparison will detect all loci with synonymous site selection pressure. We are necessarily only examining loci containing introns, and within these genes, only those that have accumulated at least five synonymous and five intron substitutions, which is a subset of genes in the Drosophila genome. Moreover, because our analysis is conducted in a gene-by-gene fashion, we may have limited power to detect selection. However, our goal is to determine a high-confidence set of loci for which a large fraction of their codons experiences synonymous site selection pressure, so we believe that our approach, although conservative, is appropriate. We limited our analysis to the Singh et al. (2007) loci that have introns and met the substitution cut off, resulting in a set of 4,682 and 4,313 loci in *D. melanogaster* and *D. sechellia*, respectively.

Per-locus Synonymous and Intron GC and AT Rate Comparisons

A strictly neutral expectation is that synonymous and intron sites will have similar rates of evolution toward and away from the nucleotides G and C (assuming locus-scale mutation homogeneity). Given that preferred codons are GC biased in Drosophila, we can make a priori predictions of how synonymous site selection pressures can decouple synonymous and intron rates of evolution. For example,

loci selectively maintaining preferred codon usage could have a significantly faster GC rate (due to positive selection for preferred codons) and/or a significantly slower AT rate (due to negative selection on unpreferred codons) at synonymous sites compared with neighboring introns. In contrast, selection favoring unpreferred codon usage is predicted to result in a significantly faster AT rate (due to positive selection for unpreferred codons) and/or a significantly slower GC rate (due to negative selection on preferred codons) at synonymous sites compared with introns. These comparisons should be robust to any recent shift in the mutation process between these species (Nielsen et al. 2007; Singh et al. 2007) because we are comparing actual lineage-specific rates, in other words, number of differences calibrated by ancestral GC content.

Along the *D. sechellia* lineage, 631 loci reject the rate comparison consistent with a selective bias for preferred codons, which we label "preferred biased loci" after correcting for multiple testing (Storey 2002) at the 10% level (517 with a faster GC rate only, 111 a slower AT rate only, and 3 with both a faster GC rate and slower AT rate; fig. 1). In contrast, only 107 loci are consistent with a selective bias toward unpreferred codons in this species, which we label "unpreferred biased loci" (fig. 1; 76 faster AT rate only and 31 slower GC rate only). Along the *D. melanogaster* lineage, we identify 207 preferred biased loci (fig. 1; 102 faster GC rate only and 105 slower AT rate only) and 390 unpreferred biased loci (fig. 1; 372 faster AT rate only and 18 slower GC rate only). A list of these loci can be found in supplementary table 2 (Supplementary Material online).

Comparisons between the two species (fig. 1) reveals an interesting contrast. Significantly, more loci show an acceleration of substitution toward preferred codons in *D. sechellia* (207 of 4,682 vs. 631 of 4,313 in *D. melanogaster* and *D. sechellia*, respectively; 2×2 contingency table *P* value < 0.0001). On the other hand, significantly more loci appear to have an acceleration of unpreferred codon substitutions in *D. melanogaster* (390 of 4,682 vs. 107 of 4,313 in *D. melanogaster* and *D. sechellia*, respectively; 2×2 contingency table *P* value < 0.0001).

We also observe loci with significant differences between synonymous and intron GC rates and AT rates in the same direction (i.e., both rates faster or slower at synonymous sites; fig. 1). Along the *D. sechellia* lineage, there were 54 such loci: 49 with both rates faster and 5 with both rates slower at synonymous sites. In *D. melanogaster*, there were 65 such loci: 63 with both rates faster and 2 with both rates slower at synonymous sites.

Effects of Chromosomal Location and Recombination Environment on GC and AT Rates

We detect loci with significant differences between synonymous site and intron rates of substitution toward or away from the nucleotides G and C. Our a priori prediction, given that preferred codons are GC biased in Drosophila, is that the observed rate decoupling are due to selection acting on synonymous sites. We note, however, that selection may modulate intron evolution as well (e.g., Halligan et al. 2004; Andolfatto 2005; Kern and Begun 2005; Wang et al. 2007;

Table 1
Median (mean) Relative Adjusted Rates on the X Chromosome and Autosome and *P* Value for the Wilcoxon Rank-Sum Test Comparing Rates across Chromosomes

	X Chromosome	Autosome	*P* Value
Drosophila melanogaster lineage			
Total synonymous	0.046 (0.049)	= 0.047 (0.048)	0.901
Total intron	0.043 (0.052)	= 0.043 (0.050)	0.081
AT rate, synonymous	0.412 (0.445)	> 0.399 (0.422)	0.007
AT rate, intron	0.352 (0.356)	< 0.363 (0.364)	0.007
GC rate, synonymous	0.385 (0.367)	< 0.408 (0.393)	0.001
GC rate, intron	0.358 (0.359)	> 0.349 (0.350)	0.018
Drosophila sechellia lineage			
Total synonymous	0.040 (0.042)	< 0.042 (0.043)	<0.0001
Total intron	0.041 (0.049)	= 0.041 (0.048)	0.267
AT rate, synonymous	0.294 (0.316)	< 0.323 (0.335)	<0.0001
AT rate, intron	0.347 (0.348)	< 0.357 (0.363)	0.010
GC rate, synonymous	0.527 (0.524)	> 0.501 (0.498)	0.0002
GC rate, intron	0.381 (0.387)	> 0.361 (0.362)	<0.0001

Table 2
Results of Partial Correlations between Two Estimates of Recombination and Relative GC- and AT-Adjusted Rates after Considering Secondary Relationships with Codon Bias (Effective Number of Codons), Coding Length, Intron Length, Total Synonymous Divergence, and Total Intron Divergence

	ACE	RP
Autosome (*n* = 4011)		
AT rate, synonymous	0.048 (<0.0001)	0.066 (<0.0001)
AT rate, introns	0.028 (0.008)	0.031 (0.004)
GC rate, synonymous	−0.040 (0.0002)	−0.054 (<0.0001)
GC rate, introns	−0.059 (<0.0001)	−0.050 (<0.0001)
X chromosome (*n* = 671)		
AT rate, synonymous	0.061 (0.0019)	0.069 (0.008)
AT rate, introns	−0.002 (0.946)	−0.054 (0.038)
GC rate, synonymous	−0.070 (0.007)	−0.060 (0.020)
GC rate, introns	−0.014 (0.595)	0.028 (0.281)

NOTE.—Under each recombination estimate is the corresponding Kendall's Tau and *P* value (in parenthesis) for each comparison.

Haddrill et al. 2008). Therefore, the next challenge is to elucidate the relative role of synonymous and intron selection pressures in producing the rate decoupling. In an attempt to do this, we investigated the effect of recombination rate and chromosomal location on substitution rate. We made these comparisons after adjusting total divergence and GC and AT rates (termed GC- and AT-adjusted rates) to take into account differences in the mutation process between the X chromosome and autosomes (Singh et al. 2007; see Methods). We also make these comparisons using relative rates (relative to total substitution rate) to factor out differences in mutation rate and constraint across loci.

Comparisons of rates of substitution between autosomes and the X chromosome can be used as a proxy of selection under a number of assumptions. If selected mutations tend to be recessive and the effect of selection is similar in males and females, selection is expected to be more efficient on the X chromosome compared with the autosomes (e.g., Charlesworth et al. 1987; Vicoso and Charlesworth 2006; Singh et al. 2008). Therefore, we predict a faster or slower rate of substitution on the X chromosome versus the autosomes for advantageous and deleterious mutations, respectively.

In Table 1, we show how chromosomal location affects rates of evolution for these species. In *D. melanogaster*, there is no significant difference across chromosomes in total synonymous and intron divergence, in agreement with previous studies (Bauer and Aquadro 1997; Betancourt et al. 2002; Singh et al. 2008). Yet, synonymous AT-adjusted rates are significantly faster on the X chromosome, whereas intron AT-adjusted rates are significantly slower. In contrast, the GC-adjusted rates are significantly slower and faster on the X chromosome for synonymous and intron sites, respectively.

In *D. sechellia*, total synonymous divergence is significantly slower on the X chromosome in agreement with previous work (Singh et al. 2008), but there is no difference between chromosomes in total intron divergence. X-linked loci also have a significantly slower AT-adjusted rate but significantly faster GC-adjusted rate at synonymous sites.

Overall, intron rates mimic the synonymous rates but the difference between chromosomes is greater for synonymous sites. For both species, there is no difference between the chromosomes in intron length, which has been shown to affect substitution rates (e.g., Halligan and Keightley 2006), suggesting that this variable is not confounding our results.

Comparisons between substitution and recombination rate can also be used as a proxy for selection. Selection (positive and negative) is thought to be more efficient in regions of high versus low recombination due to the Hill–Robertson effect (e.g., Begun and Aquadro 1992; Charlesworth et al. 1993; Presgraves 2005). Therefore, a simple prediction is observing a positive relationship between recombination and substitution rate if positive selection, or a negative relationship if negative selection, plays a dominant role in the molecular evolution of a particular class of mutations. The X chromosome and autosomes were analyzed separately given recent observations that the substitution and mutation processes are different between these classes of chromosomes (Singh et al. 2006, 2007). Table 2 documents how relative substitution rates are affected by recombination along the *D. melanogaster* lineage (the only species with a well-defined genetic map). We used partial correlations to account for influences of coding and intron length and levels of codon bias on rates of evolution in this species (i.e., Sharp and Li 1989; Comeron and Guthrie 2005). For synonymous sites, the results are similar between the chromosomes with the relative AT-adjusted rate being significantly positively and the relative GC-adjusted rate significantly negatively correlated with recombination. The effect of recombination on intron rate of substitution differs between the chromosomes. On the autosomes, we observe significant positive and negative correlations for the AT- and GC-adjusted rates, respectively. However, on the X chromosome, the GC-adjusted rate is not correlated with recombination. The AT-adjusted rate is not correlated with ACE but is negatively correlated with RP on the X.

These associations with recombination are consistent with a model in which selection is subtly favoring AT and GC rates at synonymous sites in *D. melanogaster* and

D. sechellia, respectively, which supports the hypothesis that synonymous site selection pressure can play a role in the rate decoupling. Under such a model, we may expect to observe loci classified as preferred biased in *D. sechellia* and unpreferred biased in *D. melanogaster*. To date, this has only been observed at the *Notch* locus (Bauer DuMont et al. 2004; Nielsen et al. 2007; Singh et al. 2007). We now identify 101 loci (including *Notch*) with a selective bias for preferred codons in *D. sechellia* and unpreferred bias in *D. melanogaster*. GO analysis of these loci suggests that they are overrepresented in the plasma membrane and in biological processes related to cell communication, system and cell development, and anatomical and cellular structure morphogenesis and neurogenesis (results of 27 most significant terms; P values ≤ 0.00007). There are no underrepresented terms.

Discussion

In this study, we compare GC increasing and decreasing substitution rates between synonymous and intron sites along the *D. melanogaster* and *D. sechellia* lineages with the goal of inferring synonymous site selection pressure. For most loci, there is no detectable difference between such sites in what we term GC rate or AT rate (82% *D. sechellia*, 86% *D. melanogaster*). However, for some loci, synonymous and intron base composition evolution appears to be decoupled. When interpreting these results, we must consider the following. First, has improper ancestral reconstruction biased our results? Second, is nonequilibrium evolution affecting our results? Third, is the decoupling predominately due to intron or synonymous site selection pressures?

With regard to ancestral reconstruction and nonequilibrium evolution, Akashi et al. (2007) demonstrate that ancestral inference is biased by many parameters. For example, base composition disequilibrium, which is likely in *Drosophila* species (e.g., Akashi et al. 2006; Ko et al. 2006; Singh et al. 2007), appears to bias both parsimony and likelihood ancestral inference by causing an overestimation of the ancestral frequency and number of mutations away from the common nucleotides (Eyre-Walker 1998; Akashi et al. 2007). By comparing the number of substitutions calibrated by ancestral base composition, which are similarly affected by ancestral inference and nonequilibrium, we have hopefully minimized any effect of these factors on our results. Also, they do not appear to be systematically affecting our conclusions given that we observe significantly different numbers of loci classified as unpreferred and preferred biased between *D. melanogaster* and *D. sechellia* even though they share our inferred ancestor and are both likely to be in base composition nonequilibrium. Also, our rate comparisons are not likely to be affected by fine-scale differences in mutation between synonymous sites and introns (i.e., due to neighbor-dependent mutation), given that coding and intron regions have been shown to have similar dinucleotide bias characteristics in *D. melanogaster* (Liu and Li 2008) and dinucleotide biases have been shown to be stable across this species' genome (Gentles and Karlin 2001).

Given both synonymous and intron sites in Drosophila are thought to experience some degree of selective pressures (e.g., Halligan et al. 2004; Andolfatto 2005; Kern and Begun 2005; Neafsey and Galagan 2007; Singh et al. 2007; Wang et al. 2007; Haddrill et al. 2008; Holloway et al. 2008), selection at either or both types of sites could contribute to our observed rate decoupling. We made comparisons between substitution rates and chromosomal location and local recombination rate to deduce the relative roles of synonymous site and intron selection pressures in producing the observed rate decoupling.

The validity of comparing substitution rates between chromosomes to deduce the nature of selection depends on demographic history, an equal sex ratio, and no male- or female-biased mutation. There is no evidence for sex-biased mutation in *D. melanogaster* or *D. simulans* (a close relative of *D. sechellia*; Bauer and Aquadro 1997; Betancourt et al. 2002; but see Bachtrog 2008). Demographic history and unequal sex ratios differentially affect levels of diversity on the X chromosome and autosomes (Charlesworth 2001; Pool and Nielsen 2007, 2008), which can impact chromosomal comparisons of pairwise divergence because it includes both fixed and polymorphic variants. We infer that violations of these conditions, and our use of pairwise divergence, are only marginally influencing our *D. melanogaster* results as we observe no difference between chromosomes in total synonymous or intron divergence in this species. Within *D. sechellia*, we observe a difference between the chromosomes for total synonymous divergence. These results could indicate that synonymous evolution is constrained in *D. sechellia*, but further investigation into the life history of this species is warranted.

When we consider synonymous and intron GC- and AT-adjusted rates, we observe significant differences between the chromosomes in both species. In *D. sechellia*, the synonymous site and intron rates respond similarly to chromosomal location suggesting regional pressures on overall GC content across the *D. sechellia* genome, as has also been noted for its close relative *D. simulans* (Haddrill and Charlesworth 2008). In general, there appears to be a bias favoring GC fixations in this species. This GC bias could be due to selection or biased gene conversion. Gene conversion is suspected to be biased toward G and C in Drosophila (Galtier et al. 2006; Galtier and Duret 2007) and may be more frequent on the X chromosome (assuming gene conversion and recombination rates are colinear). This phenomenon is expected to affect synonymous and intron sites equally. As such, biased gene conversion does not appear sufficient to explain both the greater effect of chromosomal location on synonymous site rates and the locus-specific decoupling of substitution rates between synonymous and intron sites in this species.

In *D. melanogaster*, the GC and AT rates at synonymous sites and introns are affected differently by chromosomal location. This could indicate that there is weak or no regional pressure on base composition. Given that the pattern of substitution in *D. sechellia* more closely resembles that of the most recent common ancestor between these species (Singh N, Arndt P, Clark A, and Aquadro C, personal communication); another hypothesis is that synonymous sites have responded more quickly to a change in selective or nonselective pressures

on base composition along the *D. melanogaster* lineage. In particular, the between-chromosome comparisons suggest the existence of a fixation bias toward A and T as has previously been suggested for this species (e.g., Bauer DuMont et al. 2004; Holloway et al. 2008).

The nature of the relationship between a locus' substitution rate and recombination rate in *D. melanogaster* largely agree with the rate comparisons between the X chromosome and autosome. Within both the X chromosome and autosomes, we observe a significant positive relationship between synonymous site AT-adjusted rate and local recombination rate. However, regions of the genome with high recombination rate are expected to have a deeper ancestral gene genealogy compared with regions with low recombination due to positive and/or background selection (see Wang et al. 2007). As a result, pairwise divergence of neutral variants, along with those experiencing positive selection, may exhibit a positive relationship with recombination rate. However, we also observe a significantly negative relationship between recombination and synonymous site GC-adjusted rate for these same loci, which would not be expected if pairwise divergence totally explained our results. These observations suggest a selective advantage for AT-rich codons and an actual deleterious effect of GC-rich codons, at least for the loci studied here.

Thus, our consideration of the effect of recombination and chromosomal location on the rates of substitution supports the hypothesis that selection is affecting base composition in both *D. melanogaster* and *D. sechellia*. However, the substitution bias is in opposite directions in the two species. In general, we observe evidence for selection pressures that would increase GC-ending codons in *D. sechellia* and AT-ending codons in *D. melanogaster*. Thus, these comparisons suggest that selection on synonymous sites could play a role in the decoupling of synonymous and intron rates of evolution toward and away from the nucleotides G and C in both species. In addition, previous studies of within-species nucleotide variability independently supports the hypothesis that selection is shaping synonymous mutations within two loci identified in the rate comparison (i.e., *Notch* and *diminutive*; Bauer DuMont et al. 2004; Jensen et al. 2007).

Given the recombination and chromosomal location results, one may have expected that a greater number of loci would have rejected with the rate comparison in these species. This dichotomy may illustrate that the rate comparison has little power to detect synonymous site selection pressure. However, it may also suggest that only a subset of loci contain a large enough number of codons experiencing such selection pressure to be detectable in our comparison. We do not suggest that we have identified all loci experiencing synonymous site selection pressure but rather believe that we have identified a high-confidence set for which future analyses can be performed to 1) confirm the rate comparison signal and 2) elicit the functional consequences of the selective fixations of unpreferred codons.

We also note considerable overlap between the *D. melanogaster* candidate loci identified here with the rate comparison and the regions of accelerated synonymous site evolution identified specifically along this species lineage by Holloway et al. (2008). Of their 24 DMARss

(*D. melanogaster* Accelerated Regions—synonymous site) that were included in our comparison, 10 rejected as unpreferred biased and 2 preferred biased loci.

The accumulation of unpreferred synonymous substitutions along the *D. melanogaster* lineage was previously hypothesized to be due solely to relaxation of selective constraint maintaining codon bias (Akashi 1995, 1996; McVean and Vieira 2001). We do observe that significantly fewer loci have differences between synonymous and intron GC rate in *D. melanogaster* compared with *D. sechellia*. However, we also present evidence that positive selection favoring unpreferred substitutions is acting at a number of loci, with a greater proportion being seen in *D. melanogaster* relative to *D. sechellia*. Thus, a simple relaxation of constraint does not appear to entirely explain the accumulation of unpreferred substitutions in this species.

Consistent with our earlier study, synonymous site sequence evolution is most extreme at *Notch*, with a strong fixation bias toward unpreferred codons on the *D. melanogaster* lineage and a contrastingly strong fixation bias toward preferred codons on the *D. sechellia*/*D. simulans* lineage (Bauer DuMont et al. 2004; Singh et al. 2007). Previously, *Notch* had appeared to be the only locus with opposite selection pressures between these lineages. However, our rate comparisons (after FDR correction) presented here allows us to identify 100 other loci with an apparent selective advantage of unpreferred codons along the *D. melanogaster* lineage and preferred codons along the *D. sechellia* lineage. Interestingly, GO analysis suggests these loci are overrepresented in biological processes related to morphogenesis and development. Thus, this apparent switch in codon preference has occurred within loci that are generally considered to be functionally crucial, consistent with *Notch*.

An increasing number of studies in organisms as diverse as bacteria, yeast, humans, and Drosophila have demonstrated functional roles for synonymous variation, including the stability of pre-mRNA, mRNA, and DNA secondary structures in the following: mutability, mRNA stability, protein folding, and levels and patterns of protein expression (e.g., Carlini 2004; Stenøien and Stephan 2005; Biro 2006; Hoede et al. 2006; Kimchi-Sarfaty et al. 2007; Drummond and Wilke 2008). Whether a combination of a change in mutation pressure together with selection to conserve secondary structures along the *D. melanogaster* lineage could lead to the patterns of molecular evolution reported here warrants further investigation as does the potential down regulation of expression in *D. melanogaster* as might be expected given the presence of more unpreferred codons. For a highly conserved protein-like *Notch*, with extensive critical roles in many developmental and cellular processes, the striking reversal of selection in the *D. melanogaster* lineages is particularly surprising. Having identified a set of additional loci showing this same pattern of synonymous site evolution may help the formation of testable hypotheses.

Supplementary Material

Supplementary tables 1 and 2 are available at *Genome Biology and Evolution* online (http://www.oxfordjournals.org/our_journals/gbe/).

Funding

Acknowledgments

Literature Cited

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

Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in Drosophila DNA. Genetics. 139:1067–1076.

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.

Akashi H, Goel P, John A. 2007. Ancestral inference and the study of codon bias evolution: implications for molecular evolutionary analyses of the *Drosophila melanogaster* subgroup. PLoS ONE. 2(10):e1065.

Akashi H, Kliman RM, Eyre-Walker A. 1998. Mutation pressure, natural selection, and the evolution of base composition in Drosophila. Genetica. 102–103:49–60.

Akashi H, et al. 2006. Molecular evolution in the *Drosophila melanogaster* species subgroup: frequent parameter fluctuations on the timescale of molecular divergence. Genetics. 172:1711–1726.

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature. 437:1149–1152.

Bachtrog D. 2007. Reduced selection for codon usage bias in *Drosophila miranda*. J Mol Evol. 64:586–590.

Bachtrog D. 2008. Evidence for male-driven evolution in Drosophila. Mol Biol Evol. 25:617–619.

Baines JF, Parsch J, Stephan W. 2004. Pleiotropic effect of disrupting a conserved sequence involved in a long-range compensatory interaction in the Drosophila *Adh* gene. Genetics. 166:237–242.

Bauer VL, Aquadro CF. 1997. Rates of DNA sequence evolution are not sex-biased in *Drosophila melanogaster* and *D. simulans*. Mol Biol Evol. 14:1252–1257.

Bauer DuMont V, Fay JC, Calabrese PP, Aquadro CF. 2004. DNA variability and divergence at the notch locus in *Drosophila melanogaster* and *D. simulans*: a case of accelerated synonymous site divergence. Genetics. 167:171–185.

Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature. 356:519–520.

Beissbarth T, Speed TP. 2004. GOstat: find statistically over-represented Gene Ontologies within a group of genes. Bioinformatics. 20:1464–1465.

Betancourt AJ, Presgraves DC, Swanson WJ. 2002. A test for faster X evolution in Drosophila. Mol Biol Evol. 19:1816–1819.

Biro JC. 2006. Indication that "codon boundaries" are physico-chemically defined and that protein-folding information is contained in the redundant exon bases. Theor Biol Med Model. 3:28.

Carlini DB. 2004. Experimental reduction of codon bias in the *Drosophila alcohol dehydrogenase* gene results in decreased ethanol tolerance of adult flies. J Evol Biol. 17:779–785.

Carlini DB, Stephan W. 2003. *In vivo* introduction of unpreferred synonymous codons into the Drosophila *Adh* gene results in reduced levels of ADH protein. Genetics. 163:239–243.

Charlesworth B. 2001. The effect of life-history and mode of inheritance on neutral genetic variability. Genet Res. 77:153–166.

Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 130:113–146.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics. 134:1289–1303.

Comeron JM, Guthrie TB. 2005. Intragenic Hill-Robertson interference influences selection intensity on synonymous mutations in Drosophila. Mol Biol Evol. 22:2519–2530.

Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell. 134:341–352.

Duan J, et al. 2003. Synonymous mutations in the *human dopamine receptor D2 (DRD2)* affect mRNA stability and synthesis of the receptor. Hum Mol Genet. 12:205–216.

Eyre-Walker A. 1998. Problems with parsimony in sequences of biased base composition. J Mol Evol. 47:686–690.

Galtier N, Bazin E, Bierne N. 2006. GC-biased segregation of noncoding polymorphisms in Drosophila. Genetics. 172:221–228.

Galtier N, Duret L. 2007. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Trends Genet. 23:273–277.

Gentles AJ, Karlin S. 2001. Genome-scale compositional comparisons in eukaryotes. Genome Res. 11:540–546.

Haddrill PR, Bachtrog D, Andolfatto P. 2008. Positive and negative selection on noncoding DNA in *Drosophila simulans*. Mol Biol Evol. 25:1825–1834.

Haddrill PR, Charlesworth B. 2008. Non-neutral processes drive the nucleotide composition of non-coding sequences in Drosophila. Biol Lett. 4:438–441.

Halligan DL, Eyre-Walker A, Andolfatto P, Keightley PD. 2004. Patterns of evolutionary constraints in intronic and intergenic DNA of *Drosophila*. Genome Res. 14:273–279.

Halligan DL, Keightley PD. 2006. Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. Genome Res. 16:875–884.

Hoede C, Denamur E, Tenaillon O. 2006. Selection acts on DNA secondary structures to decrease transcriptional mutagenesis. PLoS Genet. 2:1–5.

Holloway AK, Begun DJ, Siepel A, Pollard KS. 2008. Accelerated sequence divergence of conserved genomic elements in *Drosophila melanogaster*. Genome Res. 18:1592–1601.

Jensen JD, Bauer DuMont VL, Ashmore AB, Gutierrez A, Aquadro CF. 2007. Patterns of sequence variability and divergence at the diminutive gene region of *Drosophila melanogaster*: complex patterns suggest an ancestral selective sweep. Genetics. 177:1071–1085.

Karolchik D, et al. 2003. The UCSC Genome Browser Database. Nucl Acids Res. 31(1):51–54.

Kern AD, Begun DJ. 2005. Patterns of polymorphism and divergence from noncoding sequences of *Drosophila melanogaster* and *D. simulans*: evidence for nonequilibrium processes. Mol Biol Evol. 22:51–62.

Kimchi-Sarfaty C, et al. 2007. A "silent" polymorphism in the MDR1 gene changes substrate specificity. Science. 315:525–528.

Kindahl EC. 1994. Recombination and DNA polymorphism on the third chromosome of *Drosophila melanogaster*. [Ithaca (NY)]: Cornell University.[PhD dissertation]

Kliman RM. 1999. Recent selection on synonymous codon usage in *Drosophila*. J Mol Evol. 49:343–351.

Ko W-Y, Piao S, Akashi H. 2006. Strong regional heterogeneity in base composition evolution on the Drosophila X chromosome. Genetics. 174:349–362.

Liu G, Li H. 2008. The correlation between recombination rate and dinucleotide bias in *Drosophila melanogaster*. J Mol Evol. 67:358–367.

McVean GAT, Vieira J. 2001. Inferring parameters of mutation, selection, and demography from patterns of synonymous site evolution in Drosophila. Genetics. 157:245–257.

Neafsey DE, Galagan JE. 2007. Positive selection for unpreferred codon usage in eukaryotic genomes. BMC Evol Biol. 7:119.

Nielsen R, Bauer DuMont VL, Hubisz MJ, Aquadro CF. 2007. Maximum likelihood estimation of ancestral codon usage bias parameters in *Drosophila*. Mol Biol Evol. 24:228–235.

Oresic M, Dehn MHH, Korenblum D, Shalloway D. 2003. Tracing specific synonymous codon-secondary structure correlations through evolution. J Mol Evol. 56:473–484.

Petrov DA, Hartl DL. 1999. Patterns of nucleotide substitution in Drosophila and mammalian genomes. Proc Natl Acad Sci USA. 96:1475–1479.

Presgraves DC. 2005. Recombination enhances protein adaptation in *Drosophila melanogaster*. Curr Biol. 15:1651–1656.

Pool JE, Nielsen R. 2007. Population size changes reshape genomic patterns of diversity. Evolution. 61:3001–3006.

Pool JE, Nielsen R. 2008. The impact of founder events on chromosomal variability in multiply mating species. Mol Biol Evol. 25:1728–1736.

Powell JR, Sezzi E, Moriyama EN, Gleason JM, Caccone A. 2003. Analysis of a shift in codon usage in Drosophila. J Mol Evol. 57:214–225.

Rodríguez-Trelles F, Tarrío R, Ayala FJ. 1999. Switch in codon bias and increased rates of amino acid substitution in the *Drosophila saltans* species group. Genetics. 153:339–350.

Sekelsky JJ, Brodsky MH, Burtis KC. 2000. DNA repair in *Drosophila*: insights from the *Drosophila* genome sequence. J Cell Biol. 150:F31–F36.

Sharp PM, Li W-L. 1989. On the rate of DNA sequence evolution in Drosophila. J Mol Biol. 28:398–402.

Shields DC, Sharp PM, Higgins DG, Wright F. 1988. Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol Biol Evol. 5:704–716.

Singh ND, Arndt PF, Petrov DA. 2005. Genomic heterogeneity of background substitutional patterns in *Drosophila melanogaster*. Genetics. 169:709–722.

Singh ND, Arndt PF, Petrov DA. 2006. Minor shift in background substitutional pattern in the *Drosophila saltans* and *willistoni* lineages is insufficient to explain GC content of coding sequences. BMC Biol. 4:37.

Singh ND, Bauer DuMont VL, Hubisz MJ, Nielsen R, Aquadro CF. 2007. Patterns of mutation and selection at synonymous sites in Drosophila. Mol Biol Evol. 24:2687–2697.

Singh ND, Davis JC, Petrov DA. 2005. Codon bias and noncoding GC content correlate negatively with recombination rate on the Drosophila X chromosome. J Mol Evol. 61:315–324.

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.

Stenøien HK, Stephan W. 2005. Global mRNA structure is not associated with levels of gene expression in *Drosophila melanogaster* but shows a negative correlation with codon bias. J Mol Evol. 61:306–314.

Storey JD. 2002. A direct approach to false discovery rates. J R Stat Soc Ser B. 64:479–498.

Vicario S, Moriyama EN, Powell JR. 2007. Codon usage in twelve species of Drosophila. BMC Evol Biol. 7:226.

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

Wang J, Keightley PD, Halligan DL. 2007. Effect of divergence time and recombination rate on molecular evolution of *Drosophila* INE-1 transposable elements and other candidates for neutrally evolving sites. J Mol Evol. 65:627–639.

Warnecke T, Hurst LD. 2007. Evidence for a trade-off between translational efficiency and splicing regulation in determining synonymous codon usage in *Drosophila melanogaster*. Mol Biol Evol. 24:2755–2762.