Genetica **102/103:** 49–60, 1998.

Mutation pressure, natural selection, and the evolution of base composition in *Drosophila*

Hiroshi Akashi[1], Richard M. Kliman[2] & Adam Eyre-Walker[3]
[1] *Section of Evolution and Ecology, University of California at Davis, Davis, CA 95616, USA (Phone: (916) 752-4253; Fax: (916) 752-1449, E-mail: hakashi@ucdavis.edu);* [2] *Department of Biology, Radford University, Radford, VA 24142, USA;* [3] *Centre for the Study of Evolution, University of Sussex, Brighton, BN1 7ET, UK*

Key words: base composition, codon bias, molecular evolution, population genetics

Abstract

Genome sequencing in a number of taxa has revealed variation in nucleotide composition both among regions of the genome and among functional classes of sites in DNA. Mutational biases, biased gene conversion, and natural selection have been proposed as causes of this variation. Here, we review patterns of base composition in *Drosophila* DNA. Nucleotide composition in *Drosophila melanogaster* varys regionally, and base composition is correlated between introns and exons. *Drosophila* species also show striking patterns of non-random codon usage. Patterns of synonymous codon usage and the biochemistry of translation suggest that natural selection may act at 'silent' sites. A relationship between recombination rates and codon usage and comparisons of the evolutionary dynamics of silent mutations within and between species support natural selection discriminating among synonymous codons. The causes of regional base composition variation are less clear. Progress in functional studies of non-coding DNA, further investigations of genome patterns, and statistical tests based on evolutionary theory will lead to a greater understanding of the contributions of mutational processes and natural selection in patterning genome-wide nucleotide composition.

Introduction

The accumulation of DNA sequence data from the genomes of a number of organisms has revealed some intriguing patterns of nucleotide composition. Large-scale compositional variation along chromosomes has been described in mammals and birds (Bernardi, 1989), yeast (Sharp & Lloyd, 1993) , bacteria (Deschavanne & Filipski, 1995), and *Drosophila* (Carulli et al., 1993). The genomes of warm-blooded vertebrates appear to be organized into mosaic patterns of long stretches (>300 kb) of compositionally homogeneous 'isochores' (reviewed in Bernardi, 1995). The G+C content of isochores varies from 30 to 60%, and the base composition of coding regions, introns, and flanking regions for a given gene is strongly correlated with that of the isochore in which it resides. The evolutionary forces governing isochore evolution remain contentious; these patterns have been attributed to mutational biases (Bulmer, 1987; Filipski, 1987; Sueoka, 1988; Wolfe, Sharp & Li, 1989) , biased gene conversion (Holmquist, 1992; Eyre-Walker, 1993), and selection pressure for G+C content (Bernardi & Bernardi, 1986; DeBry & Marzluff, 1994; Eyre-Walker, unpublished results).

In contrast to regional base composition patterns in warm-blooded vertebrates, variation in nucleotide composition at silent sites in a number of microorganisms has been attributed to natural selection discriminating among synonymous codons (reviewed in Ikemura, 1985; Andersson & Kurland, 1990; Sharp et al., 1993). In *Escherichia coli* and *Saccharomyces cerevisiae*, codons recognized by the abundant tRNA(s) for a given amino acid tend to be used preferentially. The degree to which codon usage is biased is strongly related to gene expression levels. These patterns suggest that selection for translational efficiency may affect synonymous codon usage.

In this review, we will describe patterns of base composition in *Drosophila* and discuss attempts to determine their causes. The *Drosophila melanogaster* genome shows both regional heterogeneity in base composition and preferential usage of a subset of synonymous codons. We will first discuss regional patterns of base composition and base composition relationships among functional classes of sites in DNA. We then describe several population genetic approaches that attempt to distinguish between the effects of mutational biases and natural selection on synonymous codon usage. Finally, we discuss how statistical inference from within- and between-species DNA sequence data can reveal the processes underlying base composition evolution in different regions of the nuclear genome. *Drosophila* mitochondrial genomes show different patterns of base composition than the nuclear genome (see Clary & Wolstenholme, 1985; Ballard & Kreitman, 1994; Rand, Dorfsman & Kann, 1994) and will not be discussed here. We will focus primarily on base composition evolution in *D. melanogaster*, where the most data are available, and will make some comparisons to other *Drosophila* species.

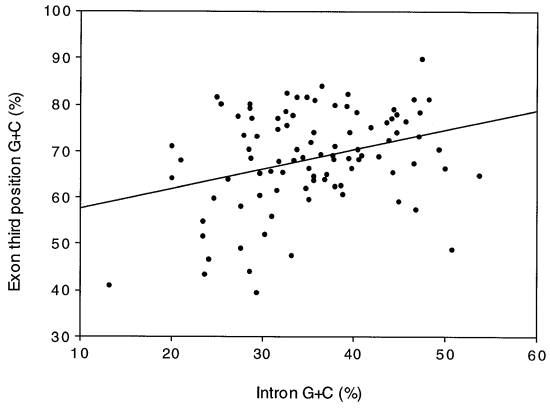

Figure 1. Codon third base G+C content *vs*. intron G+C content among *D. melanogaster* genes. Linear regression line is plotted as a visual aid. Data are plotted for 94 genes which satisfy three criteria: (1) genomic DNA has been completely sequenced between the start and stop codons; (2) there is no evidence for alternative transcripts with respect to the protein coding sequence; and (3) the intron/exon boundaries have been determined experimentally. Data from (Kliman & Hey, 1994).

Patterns of base composition in *D. melanogaster*

Patterns of base composition can be roughly divided into two categories: trends across the genome (at various scales) and distinctions among functional classes of sites within a region. In *D. melanogaster*, the overall base composition is approximately 43% G+C (Shapiro, 1976; Ashburner, 1989) and initial analyses provided little evidence for regional variation in base composition (Shields et al., 1988). In a study of YAC-cloned segments of the *D. melanogaster* genome, Carrulli et al. (1993) found compositional heterogeneity among segments. However, there was no correlation between third codon position G+C content and overall G+C content of the YAC-cloned region in which a gene was located. The authors noted that the fraction of a given YAC-cloned segment comprising amino acid-coding sequence was unknown and that this, in part, might explain heterogeneity among segments in overall G+C content; coding regions tend to have a higher G+C content than non-coding regions of the *D. melanogaster* genome.

Moriyama and Hartl (1993) took a different approach; they analyzed the DNA sequences of 40 *D. melanogaster* genes, directly comparing the base composition of introns to that of four-fold degenerate silent sites within exons. If mutational biases vary on a scale larger than the size of most genes, as proposed to explain mammalian isochores, then intron and silent-site base composition should be positively correlated. They did not find a significant correlation.

In a larger sample of 142 *D. melanogaster* genes, however, Kliman and Hey (1993) found a significant correlation among genes between codon bias and intron G+C content (Figure 1). Effective Number of Codons (Wright, 1990), a measure of codon bias that decreases with departures from equal synonymous codon usage, was negatively correlated with intron G+C content (Spearman's $r = -0.325$, 138 d.f., $P<0.0001$). This study differed from that of Moriyama and Hartl in two relevant ways. First, more genes had become available for analysis, making it possible to ascribe statistical significance to similar values of r^2. Second, by using codon bias as a covariate, rather than G+C content at four-fold degenerate sites, base composition at all but tryptophan and methionine codons was incorporated into the analysis. (In *D. melanogaster*, all codons that increase in frequency as codon bias increases end in either G or C (Akashi, 1995); thus, codon bias in this species is strongly correlated with third codon position G+C content).

In a follow-up analysis, Kliman and Hey (1994) looked more carefully at the among-gene correlation in intron vs. exon G+C content at 155 *D. melanogaster* loci. In this sample, significant positive correlations

between intron G+C content and that of either third (r^2 = 0.138) or first codon position (r^2 = 0.060) were found. The correlation between intron G+C content and second codon position G+C content (r^2 = 0.028) was significant (but only before correction for multiple comparisons). Using the 79 genes with more than one intron, a significant effect of gene on individual intron third position G+C was also detected ($F_{78,234}$ = 1.644, P = 0.0025). Thus, heterogeneity among genes, and, therefore, across the *D. melanogaster* genome, in base composition was supported by both types of analyses.

Establishing the causes of nucleotide variation at silent sites

Patterns of synonymous codon usage in *D. melanogaster* appear to parallel those found in a number of microorganisms. In *E. coli* and *S. cerevisiae*, synonymous codon usage is biased toward codons, which may enhance the efficiency of protein synthesis (reviewed in Andersson & Kurland, 1990; Sharp et al., 1993). Synonymous codon usage in these organisms is biased toward 'major' codons, which generally complement the most abundant tRNA(s) for each amino acid (Ikemura, 1981, 1982; Bennetzen & Hall, 1982; Grosjean & Fiers, 1982) . Among codons recognized by the same tRNA, the correct Watson-Crick pairing codon is generally most common. The degree to which codon usage is biased varies among *E. coli* and yeast genes and correlates strongly with protein abundance (Bennetzen & Hall, 1982; Gouy & Gautier, 1982; Ikemura, 1985; Sharp & Cowe, 1991).

These patterns are consistent with a model of 'major codon preference' under which natural selection discriminates among synonymous codons to enhance the efficiency and/or the accuracy of protein synthesis. During polypeptide chain elongation in *E. coli*, the arrival time of a cognate tRNA at the translational apparatus is inversely proportional to the tRNA's abundance (Varenne et al., 1984; Curran & Yarus, 1989). Faster arrival times at major codons could confer fitness benefits by enhancing translational elongation rates, by lowering the energetic cost of proofreading (rejecting non-cognate tRNAs), or by reducing the rate of amino acid misincorporation (Bulmer, 1988). Under this scenario, the fitness advantage to encoding a translationally superior codon should be a function of the rate at which it is translated; selection intensity for codon bias will be stronger in highly expressed genes. Silent DNA divergence between *E. coli* and *Salmonella typhemurium* is inversely related to codon usage bias, consistent with a relationship between the strength of selection and levels of codon bias (Sharp & Li, 1987; Berg & Martelius, 1995; but see Eyre-Walker & Bulmer, 1995 for contrary evidence).

Patterns of codon usage and silent DNA evolution in *D. melanogaster* parallel those found in *E. coli* and yeast (Shields et al., 1988; Sharp & Li, 1989). Codon usage is biased toward a subset of synonymous codons for each amino acid. In multi-cellular animals, tRNA abundances and gene expression levels can be tissue- and developmental-stage specific and are thus difficult to quantify. However, major codons correspond to abundant tRNAs for the three amino acids for which data are available (Shields et al., 1988). In addition, anecdotal evidence suggests a relationship between expression levels and codon bias; highly expressed genes encoding ribosomal proteins and glycolytic enzymes show greater codon bias than genes with limited or low expression such as those encoding developmental regulatory proteins. Similarly to prokaryotes, silent divergence between *Drosophila* species is inversely related to codon usage bias (Sharp & Li, 1989; Moriyama & Gojobori, 1992; Carulli et al., 1993).

The major codon preference model suggests a balance among the forces of mutation pressure, natural selection, and genetic drift at silent sites (Sharp & Li, 1986; Li, 1987; Bulmer, 1988, 1991). Major codons confer fitness benefits, but the magnitude of selection is small enough that non-major codons persist through mutation pressure and genetic drift. However, directional mutation pressure can also bias base composition in the absence of fitness differences (Freese, 1962; Sueoka, 1962, 1988). Under a mutational model, the equilibrium codon bias depends only on the ratio of the forward and backward mutation rates.

Major codon preference can be distinguished from mutational biases by contrasting predictions for their effects on DNA variation within the genome. Population genetics theory predicts a reduction in the efficacy of natural selection when genetic linkage exists among multiple sites affected by either negative or positive selection (Muller, 1964; Hill & Robertson, 1966; Felsenstein, 1974; Li, 1987; Barton, 1995; Charlesworth, Charlesworth & Morgan, 1995) . The effect of natural selection at a given site essentially accelerates genetic drift at linked sites. If mutational biases, selection coefficients at silent sites, and the distribution of fitness effects of newly-arising mutations do not differ systematically among genetic regions,

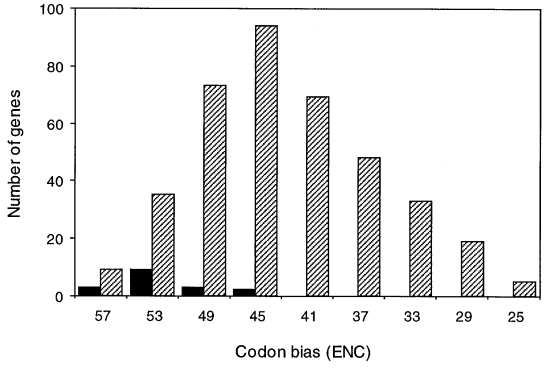

Figure 2. Frequency distributions of codon bias in regions of restricted and 'normal' levels of recombination in *D. melanogaster*. The codon bias index, ENC (Wright, 1990), is plotted for 17 genes in regions of restricted recombination (blackened) and 385 genes in regions of presumably 'normal' crossing-over (striped bars). ENC is inversely related to levels of codon bias and the x-axis shows the lower limit of ENC values for each category. Regions of reduced recombination were defined as polytene-chromosome sections 1, 40-41, 80-81, and 102 and regions of normal crossing-over were defined as sections 2-19, 22-39, 42-59, 62-79, 82-99. Genes from telomeric regions were not included because recombination rates in such regions are ambiguous. Data from (Kliman & Hey, 1993, 1994).

then major codon preference predicts that levels of codon bias should be a function of regional recombination rates. Kliman and Hey (1993) found such a relationship in the *D. melanogaster* genome; genes in regions of reduced crossing-over show lower codon bias relative to genes located in regions of higher recombination (Figure 2). The quantitative relationship between codon bias and genetic recombination, however, may depend on the particular causes of the reduction of the efficacy of natural selection; i.e., 'background selection' predicts a relatively smooth monotonic change in the efficacy of natural selection (Charlesworth, 1996) , whereas Muller's ratchet may predict a threshold effect at certain levels of recombination (Kliman & Hey, 1993). In addition, rates of recombination vary among closely-related species of *Drosophila* (True, Mercer & Laurie, 1996); measures of recombination may not accurately reflect the long-term rates of recombination that affect codon usage. A quantitative prediction for the relationship between recombination rates and codon bias may be difficult to test.

Other approaches can distinguish the effects of mutation pressure and major codon preference. Selection for codon bias predicts differences in the evolutionary dynamics of putative fitness classes of silent mutations interspersed within a region of DNA (Akashi, 1995). Under major codon preference, silent mutations from non-major to major codons, referred to as 'preferred' mutations, should confer a small fitness benefit to the organism. Mutations in the reverse direction, referred to as 'unpreferred' mutations, should incur a fitness cost of the same magnitude. Although both mutational biases and natural selection can bias codon usage, only major codon preference predicts that deterministic forces influence the history of mutations once they arise.

Comparing the evolutionary behavior of preferred and unpreferred mutations requires both the identification of candidates for major codons and inference of the direction of mutations (ancestral and derived states) in DNA. Although tRNA abundances have not been quantified in *Drosophila*, candidates for major codons can be identified as those whose frequencies (within a synonymous family) show a positive correlation with the degree of bias at other codons in the same gene (see Akashi, 1995 for a table of major codons). In *D. melanogaster*, at least one codon in each synonymous family shows such a pattern.

Major codon preference requires weak evolutionary forces that remain relatively constant over evolution time. Such a model conforms with Kimura and Ohta's view of molecular evolution; gene frequency changes within species and the accumulation of fixed differences between species reflect a single process of mutation, genetic drift, and (for some mutations) unidirectional selection (Kimura & Ohta, 1971; Kimura, 1983) . The process is governed by mutation rates and transition probabilities for changes in gene frequency. The transition probabilities depend on the product of effective population size and selection coefficient, $N_e s$ (Kimura, 1983). Relative to the neutral expectation, positive selection will increase the expected number of segregating sites in a sample of alleles, the frequencies with which the sites are segregating within the sample, and the number of fixed differences between alleles sampled from distantly related populations. Negative selection coefficients will have the opposite effect on each of these data. Under weak evolutionary forces, the time-scale of this process is too great to observe differences directly in the evolutionary trajectories of mutations in laboratory or in natural populations. Within- and between- species comparisons of DNA sequence data, however, provides a means of identifying weak selection.

Kimura's studies of the behavior of mutations in the neighborhood of neutrality ($|N_e s| \approx 1$) showed that

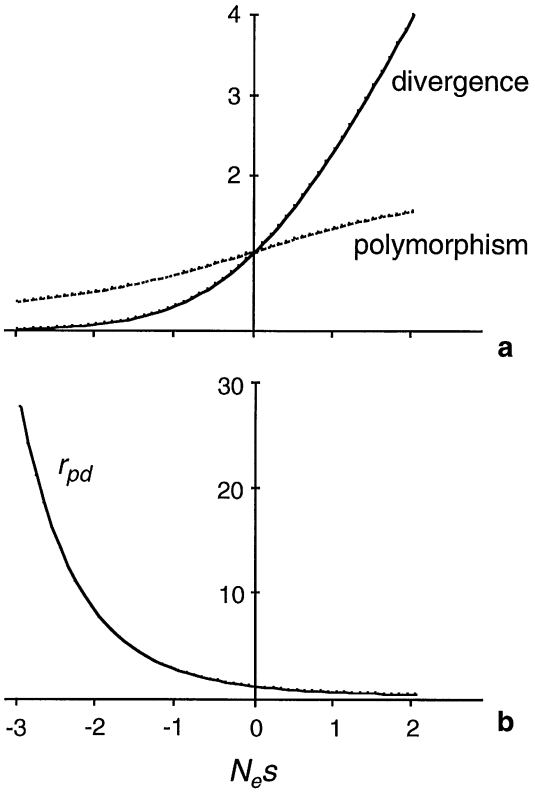

Figure 3. Polymorphism and divergence of nearly neutral mutations. (a) Expected numbers of polymorphic (dashed line) and fixed (solid line) mutations in a sample of DNA sequences as a function of $N_e s$, the product of effective population size and selection coefficient (see Kimura, 1983). Sawyer & Hartl's sampling formulae (1992) were used to calculate expected numbers of polymorphic and fixed changes relative to the expected numbers for neutral changes for parameter values, t_{div} = 1.9, m = 6, typical for the *D. melanogaster* and *D. simulans* data. (b) The ratio of expected numbers of polymorphic and fixed differences, r_{pd}, as a function of $N_e s$ for the same parameter values. From (Akashi, 1995).

the probability that a non-neutral change will rise to an intermediate frequency within a population is less affected by selection than the probability that it will go to fixation (Kimura, 1983; Figure 3a). Thus, the *ratio* of polymorphism within a species to divergence between species, referred to as r_{pd}, decreases monotonically as a function of $N_e s$ and is sensitive to even very weak selection (Akashi, 1995; Figure 3b). McDonald and Kreitman (1991) suggested comparison of ratios of polymorphism to divergence to test for fitness differences among classes of mutations interspersed in DNA. Under major codon preference, deleterious unpreferred ($N_e s<0$) mutations will show higher ratios of polymorphism to divergence than advantageous preferred

Table 1. Polymorphism and divergence of synonymous DNA changes in *D. melanogaster* and *D. simulans*

	mel (n=6)		*sim (n=5)*	
	unpref	pref	unpref	pref
polymorphic	47	1	87	24
fixed	54	4	14	13
r_{pd}	0.87	0.25	6.2	1.8

2×2 contingency tables for unpreferred and preferred synonymous mutations found polymorphic and fixed in *D. melanogaster* and *D. simulans* at eight loci. *n* refers to the number of alleles examined in each species. GenBank accession numbers or references for these sequences are given in (Akashi & Schaeffer, 1997).

changes ($N_e s> 0$). The null hypothesis, in this comparison, is no difference in the average fitness effects between the classes of mutations; neutral evolution for both classes is a subset of this null hypothesis.

Akashi (1995, in press) examined codon bias evolution in the sibling species *D. melanogaster* and *D. simulans* since their split from a common ancestor. A total of eight genes, for which six alleles have been sequenced within *D. melanogaster* and five alleles have been sequenced in *D. simulans*, were examined. Parsimony assumptions and outgroup sequences were used to infer the lineage in which silent mutations occurred (Akashi, 1995). For each codon having different nucleotides within or between the *D. melanogaster* and *simulans* alleles, mutations were assigned to minimize the number of changes in the phylogenetic tree.

Table 1 shows the ratios of polymorphism to divergence for preferred and unpreferred mutations for the genes examined. As predicted by mutation-selection-drift, the ratio of the number of synonymous changes segregating within *D. simulans* to the number of mutations that have fixed in this lineage is significantly higher for unpreferred than for preferred mutations (Fisher's exact test, P = 0.004, one-tailed). r_{pd}'s for the two classes of synonymous mutations differ in the predicted direction in *D. melanogaster,* but the null hypothesis of no fitness differences between classes of synonymous changes is not rejected (P = 0.19).

Comparisons of ratios of polymorphism to divergence incorporate the number of mutations segregating in a sample of sequences but do not take into account the number of alleles in the sample that carry newly-arisen mutations. The frequency distribution of segregating mutations can also reveal directional selection (note that 'frequency' here refers to the frequency of mutations within the sample of sequences rather than

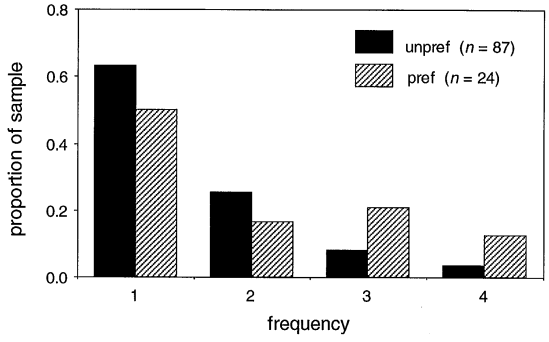

Figure 4. Frequency distributions of preferred and unpreferred synonymous DNA mutations in *D. simulans*. The proportion of unpreferred (black) and preferred (striped) mutations segregating at the given frequencies are shown. Frequency refers to the number of alleles (in a sample of five sequences) in which a mutation is found. Pooled data from eight *D. simulans* genes from (Akashi & Schaeffer, 1997).

the frequency of major codons at a given locus). Positive selection elevates the likelihood that a change will increase in frequency within a population, whereas negative selection decreases the likelihood for changes to spread (Fisher, 1930; Wright, 1938). The proportion of mutations segregating at higher frequencies increases under positive selection and the proportion of rares increase under $s < 0$; even small fitness differences between preferred and unpreferred mutations can differentiate their frequency distributions (Akashi & Schaeffer, 1997).

Sawyer, Dykhuizen and Hartl (1987) suggested comparisons of frequency distributions between classes of interspersed mutations to test for differences in their effect on fitness. Under major codon preference, advantageous preferred mutations will segregate at higher frequencies than deleterious unpreferred changes. In *D. melanogaster*, only one preferred silent mutation was found segregating among six alleles of each of eight genes; no comparisons of frequency spectra could be made in this species. Figure 4 shows the frequencies of preferred and unpreferred synonymous DNA changes in *D. simulans*. The 24 preferred changes are, on average, more common in the samples than the 87 unpreferred mutations (Mann-Whitney test, $z = 1.71$, $P = 0.044$, one-tailed). Although the statistical test is only marginally significant at the 5% level, in combination with the observed ratios of polymorphism to divergence, these analyses provide compelling evidence for differences in the evolutionary trajectories of preferred and unpreferred synonymous DNA mutations.

Sawyer, Dykhuizen and Hartl (1987) argued that comparisons between categories of mutations are relatively free of assumptions concerning the evolutionary histories of regions within the sampled data. Statistical comparisons between classes of closely-linked mutations do not require that populations have reached equilibrium or that all sites are evolving independently. If the two types of mutations are interspersed within genes, different evolutionary histories of local regions should have an equivalent effect on both the number of polymorphic sites (and consequently, the r_{pd}'s) and the frequency distributions of both classes of mutations.

Comparisons of the frequency spectra and r_{pd}'s of synonymous DNA changes also distinguish between major codon preference and mutational models of codon bias (Akashi, 1995). In *Drosophila*, higher mutation rates from A/T $\rightarrow$ C/G than in the reverse direction could explain codon usage bias (Freese, 1962; Sueoka, 1962, 1988). Such mutational biases, if they have remained constant over the time period examined, will have an equivalent impact on the numbers of polymorphic and fixed mutations and will not affect their ratios. Similarly, unequal mutation rates will affect the numbers of segregating sites in a sample but will not impact the proportion of sites segregating at different frequencies. Recent *changes* in mutation rates, however, could affect these comparisons (Eyre-Walker, 1997). If the ratio of mutation rates, *u/v*, has increased since the most recent common ancestor to the polymorphism segregating within a population, then the number of polymorphic unpreferred mutations will be higher and the frequency spectra of such mutations will be skewed toward rares. However, differences in the frequency distributions of preferred and unpreferred mutations have been observed in *D. pseudoobscura* as well as in *D. simulans*; it is unlikely that the same recent change in mutation pressure has occurred independently in these species (Akashi & Schaeffer, 1997). Similar patterns in a larger number of independent lineages would strongly support selection at silent sites.

Neither departures from stationarity and linkage equilibrium nor mutational biases appear to explain the observed difference between r_{pd}'s and frequency spectra of preferred and unpreferred silent changes in *D. simulans*. Major codon preference appears to be the predominant explanation for the maintenance of codon bias in this lineage. These findings do not, however, exclude the possibility that selection pressures other than major codon preference also contribute to patterns of codon bias. In *E. coli*, reduced codon bias at the start of genes (Bulmer, 1988; Eyre-Walker & Bulmer,

1993) and the evolutionary persistence of non-major codons (Maynard Smith & Smith, 1996) suggest that selection may favor non-major codons at some sites. In addition, major codon usage decreases at the 3′ end of *E. coli* genes; this appears to be due, in part, to the fact that many genes overlap the Shine-Dalgarno, or coding sequence of the next gene on the chromosome (Eyre-Walker, 1996). The extent to which conflicting selection pressures act at silent sites in *Drosophila* remains to be established.

Biased gene conversion or selection at silent sites?

Although mutational biases do not appear to be sufficient to explain patterns of codon usage evolution in *Drosophila*, a combination of gene conversion and biased DNA repair makes many predictions that overlap those of mutation-selection-drift. Gene conversion events have been inferred in *D. melanogaster* (reviewed in Hilliker, Clark & Chovnick, 1988), but we know little about DNA repair in *Drosophila*. In primate cell lines, DNA repair shows a bias toward incorporating G and C nucleotides (Brown & Jiricny, 1988). A combination of gene conversion and biased repair in Drosophila could cause preferential usage of G+C. Because biased gene conversion has effects similar to weak directional selection on the evolutionary dynamics of mutations (Nagylaki, 1983), it is consistent with the observed differences in ratios of polymorphism to divergence and frequency distributions of preferred and unpreferred mutations in *Drosophila*. Hickey, Wang and Magoulas (1994) have also noted that duplicated genes have synonymous codon bias towards G and C ending codons compared to non-duplicated genes, an observation they interpreted as being due to biased gene conversion.

A number of observations suggest that biased gene conversion, in the absence of natural selection, cannot explain synonymous codon bias in Drosophila. First, reduced rates of gene conversion in regions of low recombination should have an equal impact on intron and silent site base composition, whereas less effective selection in regions of reduced recombination predicts no difference in base composition at neutral sites. In *D. melanogaster*, base composition within introns shows a small difference in G+C content: 37.1% G+C among 130 introns in regions of presumed normal crossing-over and 34.5% G+C in 12 introns in regions of reduced recombination. The possibility of differences in intron base composition, perhaps due to differences in levels of gene conversion, is difficult to assess without more data. However, the difference in base composition at silent sites appears to be greater than that within introns (Kliman & Hey, 1993); the magnitude of reduction of codon bias in regions of low recombination suggests the action of natural selection at silent sites in *Drosophila*. In addition, Kliman and Hey (1994) found that silent sites in exons show consistently higher G+C content than sites within introns of the same genes (Figure 1). This finding is consistent with translational selection but is difficult to explain under biased gene conversion. Finally, in *D. melanogaster*, codon bias is correlated with functional constraint at the protein level (Akashi, 1994). Such a relationship was predicted under codon bias selection for translational accuracy because the fitness cost of misincorporations should be stronger at codons where amino acid changes have a deleterious impact on protein function. It is unclear how biased gene conversion could cause this pattern. Although biased gene conversion may contribute to base composition evolution in *Drosophila*, weak selection in favor of translationally superior codons appears to be required to explain many of the patterns we observe in the *D. melanogaster* genome.

Interpreting the relationship between silent and intron base composition

Although evidence for a role of natural selection in governing base composition evolution at silent sites in *Drosophila* appears strong, the cause(s) of the correlation between intron and exon (especially third codon position) base composition remains unresolved. Perhaps the most plausible explanation is differences in mutational biases across the genome (Kliman & Hey, 1994). For the universal genetic code, most of the possible silent substitutions are found at the third codon position, with a few possible silent substitutions at the first position; no second position substitutions are silent. If mutational pressures compete with (or act synergistically with) selection on base composition, codon usage or amino acid usage (i.e., protein function constraint), the level of base composition correlation between intron and exon sites is expected to increase as various selection pressures decrease. That is, the proportion of variance in base composition attributable to influences that act on introns and exons alike should decrease as other sources of variance increase.

The different values of r^2 for the three codon positions vs. introns (described above) might reflect constraint on amino acid sequence. However, the especially strong correlation involving third codon position was somewhat surprising in light of Moriyama and Hartl's (1993) findings. The difference might reflect variation among synonymous families; Moriyama and Hartl limited their analysis to four-fold degenerate sites. Although Kliman and Hey (1994) found a significant positive correlation between intron G+C content and that of four-fold degenerate sites (r^2 =0.073), the correlation between intron G+C content and that of two-fold degenerate sites was somewhat greater (r^2 = 0.165).

If the differences in the correlations were due to difference in the strength of selection, it follows that stronger correlations should be found in genes with lower codon bias. Kliman and Hey ranked genes by Codon Adaptation Index (Sharp & Li, 1986) , then divided them into two sets of 77 loci. For four-fold degenerate sites, the correlation was significantly weaker in the high biased group (Spearman's $r = -0.139$) than in the low biased group (Spearman's $r = 0.396$). On the other hand, for the correlations involving two-fold degenerate sites, little difference was found between the high biased group (Spearman's $r = 0.272$) and the low biased group (Spearman's $r = 0.399$). While the correlations involving two-fold and four-fold degenerate sites are obviously not different in the high biased group, they are significantly different in the low biased group ($r^2 = 6.50$, 1 d.f., $P = 0.011$ [Sokal & Rohlf, 1981, pp. 588–589]). These patterns suggest that selection may be acting more strongly on four-fold degenerate sites than on two-fold degenerate sites.

Natural selection could also account for the correlations between intron and exon base composition. Intron sequences are known to play roles in splice site recognition (Mount et al., 1992; Leicht et al., 1995), mRNA structures (Schaeffer & Miller, 1993; Stephan & Kirby, 1993, Kirby, Muse & Stephan, 1995), and gene regulation (Bruhat et al., 1993; Huang et al., 1993; McCullogh & Schuler, 1993; Tourmente et al., 1993; Laurie & Stam, 1994; Kapoun & Kaufman, 1995; Haerry & Gehring, 1996; Kohler, Schafer-Preuss & Buttgereit, 1996; McKenzie & Brennan, 1996) . However, relatively high rates of interspecific nucleotide (Hudson et al., 1994) and insertion/deletion evolution (Akashi, 1996) suggest that many, perhaps most, nucleotide sites within introns do not play a functional role. The question is whether selected sites in introns account for the relationship between intron and coding region G+C content (i.e., if both are related to gene expression levels).

Evaluating the roles of mutational biases and natural selection in governing the base composition of introns will require biological models from which to make evolutionary predictions. For example, selection for RNA structure has been established by examining patterns of genetic linkage among mutations segregating within populations. At the *Drosophila Adh* locus, 'compensatory' mutations that preserve predicted mRNA structures show stronger linkage disequilibrium than other mutations in the region (Schaeffer & Miller, 1993; Kirby, Muse & Stephan, 1995) . In contrast, at the alternatively spliced *Mlc1* gene, the predicted folding free energy of the pre-mRNA is significantly higher (less stable) than randomly permuted sequences, suggesting selection against secondary structures at this locus (Leicht et al., 1995). Gene conversion and biased DNA repair can also be tested by examining the relationship between gene conversion rates and G+C content. For example, gene conversion may occur across well-defined regions of DNA (Hickey et al., 1991; Shibata & Yamazaki, 1995) . Biased gene conversion suggests that base composition across such regions should show similarly abrupt changes. Greater knowledge of the functional significance of intron sequences will allow us to determine the contribution of selected sites to the correlation between intron and silent base composition.

Between-species comparisons of codon bias and intron base composition

Although the *D. melanogaster* genome has been by far the most extensively studied *Drosophila* genome, comparisons to other *Drosophila* species have revealed some interesting differences in base composition. In particular, the major codon preference model suggests methods to determine the causes of codon bias differences between species.

Akashi (1995, 1996) noted a consistent decline in major codon usage in the *D. melanogaster* lineage since its split from the common ancestor to *D. simulans*. Although these species are closely-related (silent divergence in low codon bias genes is about 10%) and the numbers of silent fixations within a given gene are low, the trend is consistent across loci. Under major codon preference, this genome-wide decline in codon bias could be due to either a stronger mutational bias

toward A+T or a reduction in the effectiveness of natural selection at silent sites.

Changes in mutation pressure and natural selection make distinguishable predictions for base composition evolution under different selection pressures (Akashi, 1996) . Regions under the weakest selection pressure for base composition show the greatest sensitivity to changes in mutational biases. Because all major codons end in either G or C, the mutational hypothesis predicts a greater decrease in G+C content at neutrally evolving sites than at silent sites in coding regions in *D. melanogaster*. Changes in $N_e s$ predict the opposite pattern, a larger excess of G/C $\rightarrow$ A/T changes in coding regions. Because introns were difficult to align with outgroup sequences, comparisons were made at sites fixed in each species but differing between species. Under the null hypothesis of no difference in base composition, the number of sites at which *D. melanogaster* encodes G or C and *D. simulans* encodes A or T ($mel_{GC}sim_{AT}$) should equal the number of sites in the opposite configuration ($mel_{AT}sim_{GC}$). The 16 introns examined (from the 8 genes of Table 1) show little evidence for a change in base composition: 30 intron sites are $mel_{AT}sim_{GC}$ and 26 are $mel_{GC}sim_{AT}$. Silent sites within exons, however, show significant differences in base compositions between these species: 83 sites are $mel_{AT}sim_{GC}$, whereas 25 are $mel_{GC}sim_{AT}$ (Fisher's exact test, P = 0.004). Changes in selection intensity, rather than in mutational biases, appear to explain the reduction of codon bias in the *D. melanogaster* lineage. Simulations of codon bias evolution under reduced selection show that the ratio of preferred to unpreferred fixations in *D. melanogaster* can be explained by a five-fold or greater reduction in $N_e s$; silent DNA evolution in this lineage is indistinguishable from that of neutrally evolving mutations.

Reduced codon bias has also been observed at the *Adh* locus in Hawaiian *Drosophila* (Thomas & Hunt, 1991; Moriyama & Gojobori, 1992) and in a number of genes in *D. willistoni* (Anderson, Carew & Powell, 1993). The causes of these differences and how often codon bias changes over long periods of evolutionary time will require larger numbers of gene sequences from a number of species. In particular, major codons differ for approximately half of the amino acids among *E. coli*, yeast, and *Drosophila* (Sharp, 1989) , but changes in major codons have not been documented in closely-related species. Under major codon preference, shifts in major codons are unlikely in a lineage showing biased codon usage because genome-wide codon usage is adapted to the abundant tRNA. Reversals in the relative abundances of tRNA's will entail fitness consequences equivalent to thousands of deleterious unpreferred mutations. This argument is analogous to Crick's 'frozen accident', which preserves the universal genetic code (Crick, 1968). A mutation that alters the amino acid encoded by a given codon will be strongly deleterious, because a different amino acid will be incorporated at this codon in hundreds of proteins. Shifts in major codons should be most likely following a state of silent base composition near mutational equilibrium; it will be of great interest to document such a change among closely-related species.

Future research in base composition evolution

We have reviewed some of the base composition patterns in *Drosophila* and the attempts to address the evolutionary forces governing these patterns, especially at silent sites in coding regions. For the most part, patterns of base composition have been observed on a relatively coarse scale (between genes and genetic regions). It is possible that many revealing patterns could be found on a finer (within-gene) scale. For example, Kliman and Eyre-Walker (unpublished data) have found reduced codon bias at the 5′ end of *D. melanogaster* genes. The pattern appears similar to that found in *E. coli*, but the causes of this reduction have not been established. They also found within-gene correlations in base composition of introns and third codon positions of flanking exons, apparently associated with changes in base composition along the length of genes. Interestingly, gene conversion shows polarity gradients in fungi: gene conversion is highest near transcription initiation sites and decreases in both the 3′ and 5′ directions (reviewed in Nicolas & Petes, 1994). Polarity gradients of G+C content along *Drosophila* genes could reflect polarity gradients of biased gene conversion. Another intriguing finding, noted by Powell and coworkers (1993), is higher codon bias in early expressed than in later expressed *Drosophila* genes. White (1973) observed differences in some tRNA pools between larvae and adult; whether tissue and developmental-stage specific patterns of tRNA abundance explain variation in codon bias in multi-cellular organisms remains to be established.

Our understanding of intron and intergenic region evolution is more ambiguous. How much of the genomes of organisms reflect the action of natural selection remains an open question. Although within-

and between-species DNA comparisons can reveal the action of natural selection, even when the functional effects of mutations may be too small to test in the lab, such tests often require *a priori* predictions for the fitness effects of particular mutations. A combination of DNA sequence comparisons, evolutionary theory, and functional studies of mechanisms of gene regulation, protein translation, and mutational biases will be required to understand the forces governing base composition evolution in the genomes of *Drosophila* and other taxa.

Acknowledgments

We are grateful to an anonymous reviewer for many helpful suggestions. H.A. is supported by a National Science Foundation/Alfred P. Sloan Foundation Postdoctoral Fellowship in Molecular Evolution. R.K. is supported by the Radford University College of Arts and Sciences and by a grant from the Jeffress Memorial Trust. A.E.W. is a Royal Society University Research Fellow.

References

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., in press. Codon bias evolution in *Drosophila*: Population genetics of mutation-selection-drift. Gene.

Akashi, H. & S.W. Schaeffer, 1997. Natural selection and the frequency distributions of 'silent' DNA polymorphism in *Drosophila*. Genetics 146: 295–307.

Anderson, C.L., E.A. Carew & J.R. Powell, 1993. Evolution of the Adh locus in the *Drosophila willistoni* group: the loss of an intron, and shift in codon usage. Mol. Biol. Evol. 10: 605–618.

Andersson, G.E. & C.G. Kurland, 1990. Codon preferences in free-living microorganisms. Microbiol. Rev. 54: 198–210.

Ashburner, M., 1989. *Drosophila*. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Ballard, J.W. & M. Kreitman, 1994. Unraveling selection in the mitochondrial genome of *Drosophila*. Genetics 138: 757–772.

Barton, N.H., 1995. Linkage and the limits to natural selection. Genetics 140: 821–841.

Bennetzen, J.L. & B.D. Hall, 1982. Codon selection in yeast. J. Biol. Chem. 257: 3026–3031.

Berg, O.G. & M. Martelius, 1995. Synonymous substitution-rate constants in *Escherichia coli* and *Salmonella typhimurium* and their relationship to gene expression and selection pressure. J. Mol. Evol. 41: 449–456.

Bernardi, G., 1989. The isochore organization of the human genome. Annu. Rev. Genet. 23: 637–661.

Bernardi, G., 1995. The human genome: organization and evolutionary history. Annu. Rev. Genet. 29: 445–476.

Bernardi, G. & G. Bernardi, 1986. Compositional constraints and genome evolution. J. Mol. Evol. 24: 1–11.

Brown, T.C. & J. Jiricny, 1988. Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. Cell 54: 705–711.

Bruhat, A., D. Dreau, M.E. Drake, S. Tourmente, S. Chapel, J.L. Couderc & B. Dastugue, 1993. Intronic and 5′ flanking sequences of the *Drosophila* beta 3 tubulin gene are essential to confer ecdysone responsiveness. Mol. Cell. Endocrinol. 94: 61–71.

Bulmer, M., 1987. A statistical analysis of nucleotide sequences of introns and exons in human genes. Mol. Biol. Evol. 4: 395–405.

Bulmer, M., 1988. Are codon usage patterns in unicellular organisms determined by selection-mutation balance. J. Evol. Biol. 1: 15–26.

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

Carulli, J.P., D.E. Krane, D.L. Hartl & H. Ochman, 1993. Compositional heterogeneity and patterns of molecular evolution in the *Drosophila* genome. Genetics 134: 837–845.

Charlesworth, B., 1996. Background selection and patterns of genetic diversity in *Drosophila* melanogaster. Genet. Res. 68: 131–149.

Charlesworth, D., B. Charlesworth & M.T. Morgan, 1995. The pattern of neutral molecular variation under the background selection model. Genetics 141: 1619–1632.

Clary, D.O. & D.R. Wolstenholme, 1985. The mitochondrial DNA molecular of *Drosophila* yakuba: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22: 252–271.

Crick, F.H.C., 1968. The origin of the genetic code. J. Mol. Biol. 38: 367–379.

Curran, J.F. & M. Yarus, 1989. Rates of aminoacyl–tRNA selection at 29 sense codons *in vivo*. J. Mol. Biol. 209: 65–77.

DeBry, R.W. & W.F. Marzluff, 1994. Selection on silent sites in the rodent H3 histone gene family. Genetics 138: 191–202.

Deschavanne, P. & J. Filipski, 1995. Correlation of GC content with replication timing and repair mechanisms in weakly expressed *E. coli* genes. Nucleic Acids Res. 23: 1350–1353.

Eyre-Walker, A., 1993. Recombination and mammalian genome evolution. Proc. R. Soc. Lond. B. Biol. Sci. 252: 237–243.

Eyre-Walker, A., 1996. The close proximity of *Escherichia coli* genes: consequences for stop codon and synonymous codon use. J. Mol. Evol. 42: 73–78.

Eyre-Walker, A., 1997. Differentiating between selection and mutation bias. [letter] Genetics 147: 1983–1987.

Eyre-Walker, A. & M. Bulmer, 1995. Synonymous substitution rates in enterobacteria. Genetics 140: 1407–1412.

Felsenstein, J., 1974. The evolutionary advantage of recombination. Genetics 78: 737–756.

Filipski, J., 1987. Correlation between molecular clock ticking, codon usage fidelity of DNA repair, chromosome banding and chromatin compactness in germline cells. FEBS Lett. 217: 184–186.

Fisher, R. A., 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford, England.

Freese, E., 1962. On the evolution of base composition of DNA. J. Theor. Biol. 3: 82–101.

Gouy, M. & C. Gautier, 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10: 7055–7074.

Grosjean, H. & W. Fiers, 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene 18: 199–209.

Haerry, T.E. & W.J. Gehring, 1996. Intron of the mouse Hoxa–7 gene contains conserved homeodomain binding sites that can function as an enhancer element in *Drosophila*. Proc. Natl. Acad. Sci. USA 93: 13884–13889.

Hickey, D.A., L. Bally–Cuif, S. Abukashawa, V. Payant & B.F. Benkel, 1991. Concerted evolution of duplicated protein–coding genes in *Drosophila*. Proc. Natl. Acad. Sci. USA 88: 1611–1615.

Hickey, D.A., S. Wang & C. Magoulas, 1994. Gene duplication, gene conversion, and codon bias, pp. 199–207. In Non-Neutral Evolution: Theories and Molecular Data, edited by B. Golding. Chapman & Hall, New York.

Hill, W.G. & A. Robertson, 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8: 269–294.

Hilliker, A. F., S. H. Clark & A. Chovnick, 1988. Genetic analysis of intragenic recombination in *Drosophila*, pp. 73–90. In The Recombination of Genetic Material, edited by K.B. Low. Academic Press, San Diego.

Holmquist, G.P., 1992 Chromosome bands, their chromatin flavors, and their functional features. Am. J. Hum. Genet. 51: 17–37.

Huang, J.D., D.H. Schwyter, J.M. Shirokawa & A.J. Courey, 1993. The interplay between multiple enhancer and silencer elements defines the pattern of decapentaplegic expression. Genes Dev. 7: 694–704.

Hudson, R.R., K. Bailey, D. Skarecky, J. Kwiatowski & F.J. Ayala, 1994. Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. Genetics 136: 1329–1340.

Ikemura, T., 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. 151: 389–409.

Ikemura, T., 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. 158: 573–597.

Ikemura, T., 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2: 13–34.

Kapoun, A.M. & T.C. Kaufman, 1995. A functional analysis of 5′, intronic and promoter regions of the homeotic gene proboscipedia in *Drosophila melanogaster*. Development 121: 2127–2141.

Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

Kimura, M. & T. Ohta, 1971. Protein polymorphism as a phase of molecular evolution. Nature, Lond. 229: 467–469.

Kirby, D.A., S.V. Muse & W. Stephan, 1995. Maintenance of pre–mRNA secondary structure by epistatic selection. Proc. Natl. Acad. Sci. USA 92: 9047–9051.

Kliman, R.M. & J. Hey, 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. 10: 1239–1258.

Kliman, R.M. & J. Hey, 1994. The effects of mutation and natural selection on codon bias in the genes of *Drosophila*. Genetics 137: 1049–1056.

Kohler, J., S. Schafer-Preuss & D. Buttgereit, 1996. Related enhancers in the intron of the beta1 tubulin gene of *Drosophila melanogaster* are essential for maternal and CNS-specific expression during embryogenesis. Nucleic Acids Res. 24: 2543–2550.

Laurie, C.C. & L.F. Stam, 1994. The effect of an intronic polymorphism on alcohol dehydrogenase expression in *Drosophila melanogaster*. Genetics 138: 379–385.

Leicht, B.G., S.V. Muse, M. Hanczyc & A.G. Clark, 1995. Constraints on intron evolution in the gene encoding the myosin alkali light chain in *Drosophila*. Genetics 139: 299–308.

Li, W.H., 1987. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. 24: 337–345.

Maynard Smith, J. & N.H. Smith, 1996. Site–specific codon bias in bacteria. Genetics 142: 1037–1043.

McCullough, A.J. & M.A. Schuler, 1993. AU-rich intronic elements affect pre-mRNA 5′ splice site selection in *Drosophila melanogaster*. Mol. Cell. Biol. 13: 7689–7697.

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

McKenzie, R.W. & M.D. Brennan, 1996. The two small introns of the *Drosophila affinidisjuncta* Adh gene are required for normal transcription. Nucleic Acids Res. 24: 3635–3642.

Moriyama, E.N. & T. Gojobori, 1992. Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. Genetics 130: 855–864.

Moriyama, E.N. & D.L. Hartl, 1993. Codon usage bias and base composition of nuclear genes in *Drosophila*. Genetics 134: 847–858.

Mount, S.M., C. Burks, G. Hertz, G.D. Stormo, O. White & C. Fields, 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. Nucleic Acids Res. 20: 4255–4262.

Muller, H.J., 1964. The relation of recombination to mutational advance. Mutation Research 1: 2–9.

Nagylaki, T., 1983. Evolution of a finite population under gene conversion. Proc. Natl. Acad. Sci. USA 80: 6278–6281.

Nicolas, A. & T.D. Petes, 1994. Polarity of meiotic gene conversion in fungi: contrasting views. Experimentia 50: 242–252.

Powell, J.R., A. Caccone, J.M. Gleason & L. Nigro, 1993. Rates of DNA evolution in *Drosophila* depend on function and developmental stage of expression. Genetics 133: 291–298.

Rand, D.M., M. Dorfsman & L.M. Kann, 1994. Neutral and nonneutral evolution of *Drosophila* mitochondrial DNA. Genetics 138: 741–756.

Sawyer, S.A., D.E. Dykhuizen & D.L. Hartl, 1987. Confidence interval for the number of selectively neutral amino acid polymorphisms. Proc. Natl. Acad. Sci. USA 84: 6225–6228.

Sawyer, S.A. & D.L. Hartl, 1992. Population genetics of polymorphism and divergence. Genetics 132: 1161–1176.

Schaeffer, S.W. & E.L. Miller, 1993. Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. Genetics 135: 541–552.

Shapiro, H.S., 1976. Distribution of purine and pyrimidines in deoxyribonucleic acids, pp. 241–281. In CRC Handbook of Biochemistry and Molecular Biology, edited by G.D. Fasman. CRC Press, Cleveland.

Sharp, P.M., 1989. Evolution at 'silent' sites in DNA, pp. 23–32. In Evolution and Animal Breeding: Reviews in Molecular and Quentitative Approaches in Honour of Alan Robertson, edited by W.G. Hill, and T.F.C. Mackay. CAB International, Wallingford, U.K.

Sharp, P.M. & E. Cowe, 1991. Synonymous codon usage in *Saccharomyces cerevisiae*. Yeast 7: 657–678.

Sharp, P.M. & W.H. Li, 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. J. Mol. Evol. 24: 28–38.

Sharp, P.M. & W.H. Li, 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. 4: 222–230.

Sharp, P.M. & W.H. Li, 1989. On the rate of DNA sequence evolution in *Drosophila*. J. Mol. Evol. 28: 398–402.

Sharp, P.M. & A.T. Lloyd, 1993. Regional base composition variation along yeast chromosome III: evolution of chromosome primary structure. Nucleic Acids Res. 21: 179–183.

Sharp, P.M., M. Stenico, J.F. Peden & A.T. Lloyd, 1993. Codon usage: mutational bias, translational selection, or both? Biochem. Soc. Trans. 21: 835–841.

Shibata, H. & T. Yamazaki, 1995. Molecular evolution of the duplicated Amy locus in the *Drosophila melanogaster* species subgroup: concerted evolution only in the coding region and an excess of nonsynonymous substitutions in speciation. Genetics 141: 223–236.

Shields, D.C., P.M. Sharp, D.G. Higgins & F. Wright, 1988. 'Silent' sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. 5: 704–716.

Sokal, R.R. & F.J. Rohlf, 1981. Biometry. W.H. Freeman & Co., New York.

Stephan, W. & D.A. Kirby, 1993. RNA folding in *Drosophila* shows a distance effect for compensatory fitness interactions. Genetics 135: 97–103.

Sueoka, N., 1962. On the genetic basis of variation and heterogeneity of DNA base composition. Proc. Natl. Acad. Sci. USA 48: 582–592.

Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA 85: 2653–2657.

Thomas, R.H. & J.A. Hunt, 1991. The molecular evolution of the alcohol dehydrogenase locus and the phylogeny of Hawaiian *Drosophila*. Mol. Biol. Evol. 8: 687–702.

Tourmente, S., S. Chapel, D. Dreau, M.E. Drake, A. Bruhat, J.L. Couderc & B. Dastugue, 1993. Enhancer and silencer elements within the first intron mediate the transcriptional regulation of the beta 3 tubulin gene by 20-hydroxyecdysone in Drosophila Kc cells. Insect Biochem. Mol. Biol. 23: 137–143.

True, J.R., J.M. Mercer & C.C. Laurie, 1996. Differences in crossover frequency and distribution among three sibling species of *Drosophila*. Genetics 142: 507–523.

Varenne, S., J. Buc, R. Lloubes & C. Lazdunski, 1984. Translation is a non-uniform process: effect of tRNA availability on the rate of elongation of nascent polypeptide chains. J. Biol. Chem. 180: 549–576.

White, B.N., G.M. Tener, J. Holden & D.T. Suzuki, 1973. Analysis of tRNAs during the development of *Drosophila*. Dev. Biol. 33: 185–195.

Wolfe, K.H., P.M. Sharp & W.H. Li, 1989. Mutation rates differ among regions of the mammalian genome. Nature, London 337: 283–285.

Wright, F., 1990. The 'effective number of codons' used in a gene. Gene 87: 23–29.

Wright, S., 1938. The distribution of gene frequencies under irreversible mutation. Proc. Natl. Acad. Sci. USA 24: 253–259.