Reduced Natural Selection Associated with Low Recombination in *Drosophila melanogaster*[1]

Richard M. Kliman[2] *and Jody Hey*
Nelson Labs, Rutgers University

Synonymous codons are not used equally in many organisms, and the extent of codon bias varies among loci. Earlier studies have suggested that more highly expressed loci in *Drosophila melanogaster* are more biased, consistent with findings from several prokaryotes and unicellular eukaryotes that codon bias is partly due to natural selection for translational efficiency. We link this model of varying selection intensity to the population-genetics prediction that the effectiveness of natural selection is decreased under reduced recombination. In analyses of 385 *D. melanogaster* loci, we find that codon bias is reduced in regions of low recombination (i.e., near centromeres and telomeres and on the fourth chromosome). The effect does not appear to be a linear function of recombination rate; rather, it seems limited to regions with the very lowest levels of recombination. The large majority of the genome apparently experiences recombination at a sufficiently high rate for effective natural selection against suboptimal codons. These findings support models of the Hill-Robertson effect and genetic hitchhiking and are largely consistent with multiple reports of low levels of DNA sequence variation in regions of low recombination.

Introduction

Codon bias, the unequal usage of synonymous codons, occurs in many unicellular and multicellular organisms (Grantham et al. 1980, 1981; Ikemura 1985; Shields et al. 1988). The extent of codon bias varies widely among loci within species and is positively associated with the level of gene expression in *Escherichia coli* (Gouy and Gautier 1982), *Salmonella typhimurium* (Sharp and Li 1987*a*), *Saccharomyces cerevisiae* (Bennetzen and Hall 1982), and *Dictyostelium discoideum* (Sharp and Devine 1989). Previous analyses of synonymous codon usage indicate that a species-specific set of optimal codons (or, conversely, the use of a subset of more common or otherwise preferred tRNAs) is selectively favored in highly expressed genes, where translational efficiency is critical (Ikemura 1985). Less highly expressed genes have more even patterns of codon usage, reflecting a relatively reduced effect of natural selection. Shields et al. (1988) suggested, in summarizing several studies on gene expression, that this may also explain codon-bias variation among *Drosophila melanogaster* loci. These authors also found that variation among loci for the number of synonymous substitutions between *Drosophila melanogaster* and *Drosophila pseudoobscura* is negatively associated with codon bias levels.

The correspondence between the intensity of natural selection and the level of codon bias provides a test of a population-genetics prediction that natural selection

1. Key words: recombination, codon bias, population genetics, Hill-Robertson effect, hitchhiking, Muller's ratchet. Abbreviations: CAI = codon-adaptation index; ENC = effective number of codons.

2. Present address: Department of Organismic and Evolutionary Biology, The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138.

Address for correspondence and reprints: Jody Hey, Nelson Labs, Rutgers University, P.O. Box 1059, Piscataway, New Jersey 08855-1059.

Mol. Biol. Evol. 10(6):1239–1258. 1993.

0737-4038/93/1006-0008$02.00

is less effective against slightly disadvantageous mutations in regions lacking recombination. That is, the capacity of natural selection to determine nucleotide frequencies at multiple sites is predicted to be reduced if those positions are tightly linked (Felsenstein 1974). In *Drosophila melanogaster,* recombination is reduced near the centromeres and telomeres and across the fourth chromosome (Hochman 1976; Lindsley and Sandler 1977; Ashburner 1989, p. 453). Here, we show that codon bias is reduced, as expected, in these regions.

Material and Methods

We assessed codon bias in 385 *Drosophila melanogaster* loci for which the complete amino acid coding sequences were available in GenBank Release 72 and which have been mapped to specific polytene-chromosome regions. We calculated two codon bias indices for each gene: (1) the effective number of codons (ENC) (Wright 1990) and (2) the codon-adaptation index (CAI) (Sharp and Li 1987*b*). ENC measures deviation from equal usage of synonymous codons and can range from 20 (only one codon used throughout the gene for each amino acid) to 61 (the value that an infinitely long gene would have if all synonymous codons were used equally). CAI reflects the usage of preferred codons for each amino acid and reaches a maximum value of 1.0 when only preferred codons are used in the gene. To calculate CAI, one needs estimates of the relative frequencies of codons used in highly biased genes; we used the frequencies reported by Shields et al. (1988). For each amino acid, the codon used most frequently was assigned a relative frequency of 1, with all synonymous codons scaled accordingly. However, since we do not know whether insufficient sampling of sequences underestimated or missed certain codons, we have set the lower limit of relative frequency to 0.01 (Bulmer 1988). Because of restrictions inherent to the calculation of the index, ENC could not be calculated for two short genes, *MtnB* and *Mst87F.* The inability to calculate ENC is most likely to occur for short genes and does not necessarily reflect low codon bias.

The database of Ashburner (1992) was used to develop the relationship between recombination-map position and DNA content. All *D. melanogaster* loci that have been placed on both the recombination map and the polytene-chromosome map were included. Polytene-chromosome positions were converted to cumulative DNA content for the X, second, and third chromosomes, by using the estimates of Sorsa (1988, pp. 83–105). A curve was then generated for each chromosome by least-squares polynomial curve fitting. Recombination rate, as a function of distance along the chromosome, was estimated by taking the derivative of the polynomial for each chromosome.

Results

Codon-bias values and polytene-chromosome map positions of all genes used in this study are presented in the Appendix. The relationship between recombination rate and codon bias was examined first by comparing codon-bias levels in regions of known reduced recombination with the remainder of the genome. Regions of low recombination were defined as (*a*) the polytene-chromosome map sections adjacent to the centromeres of the X, second, and third chromosomes (sections 20, 40–41, and 80–81, respectively); (*b*) the areas adjacent to the telomeres of the X, second, and third chromosomes (sections 1, 21, 60–61, and 100); and (*c*) the entire fourth chromosome (sections 101–102) (Bridges 1935). The difference is highly significant for both ENC and CAI (table 1 and figure 1), consistent with the prediction that codon bias is reduced in regions of low recombination. The statistical significance of these tests is not due to reduced codon bias in one particular region of reduced recombination.

Table 1
Kruskal-Wallis Tests for Differences in Codon Bias in Regions of Low Recombination versus Regions of Higher Recombination

	Sample Size		Mean			
Codon Bias Index	N_{high}	N_{low}	N_{high}	N_{low}	H^a	P
ENC	343	40	44.9	49.6	14.528	≤0.0001
CAI	345	40	0.439	0.354	18.170	≤0.0001

[a] Kruskal-Wallis test statistic.

Average codon bias is noticeably reduced in the six centromeric genes (ENC = 55.2, CAI = 0.226), the eight genes at the tip of the X chromosome (ENC = 53.2, CAI = 0.310), and the three genes on the fourth chromosome (ENC = 55.6, CAI = 0.184). While average codon bias in the 23 genes in autosomal telomeric regions is much greater (ENC = 46.0, CAI = 0.425), estimates of recombination rate in several of the polytene-chromosome regions associated with telomeres are also higher (see fig. 1). If we remove the autosomal telomeric genes from the analysis, given the possibility that they have been misclassified as undergoing little recombination, codon bias in regions of low recombination remains significantly lower than codon bias in the remaining loci (ENC_{low} = 54.3, ENC_{high} = 44.9, Kruskal-Wallis H = 29.014, $P \leq 0.0001$; CAI_{low} = 0.258, CAI_{high} = 0.439, Kruskal-Wallis H = 32.129, $P \leq 0.0001$). It is interesting that genes located at the extreme tips of the autosomal telomeres (i.e., *l(2)gl, Kr, Map205,* and *mod*), where recombination rates may drop off, also show very little codon bias (see Appendix).

We also considered recombination rate as a continuous variable, and we estimated nonparametric correlations with codon bias (see Appendix). Recombination rate for this analysis was calculated from the polynomials generated for each chromosome as described in Material and Methods (fig. 2). It is unfortunate that estimates of recombination rate in those regions which we defined a priori as undergoing little recombination for the Kruskal-Wallis tests are probably inaccurate. As one can see in figure 2, the available data for loci located near autosomal telomeres do not indicate substantial reduction in recombination rate in these regions. While it is possible that finer mapping of the telomeres would verify low recombination, the curves modeled from the existing data produce estimates of telomeric recombination rates that overlap considerably with those from the rest of the genome (fig. 3). Recombination rates for autosomal centromeric loci are also probably overestimates. Regions of nonrecombining heterochromatic DNA near the centromeres do not appear in polytene chromosomes because of underreplication in salivary gland nuclei (Ashburner 1989, p. 42). Therefore, the smooth curves apparent near the centromeres of chromosomes 2 and 3 in figure 2 are misleading; if chromocenter DNA were included, it is likely that the curves would flatten more completely at the centromeres.

If we consider only the 345 loci in regions not thought to experience reduced recombination, there is no evidence of a quantitative relationship between codon bias and recombination rate [ENC: Spearman's r_s = −0.048, 341 degrees of freedom (df), P = 0.377; CAI: Spearman's r_s = −0.005, 343 df, P = 0.919] (fig. 3). If we choose to include all 385 loci in the analyses, ENC is correlated, in the expected direction, with

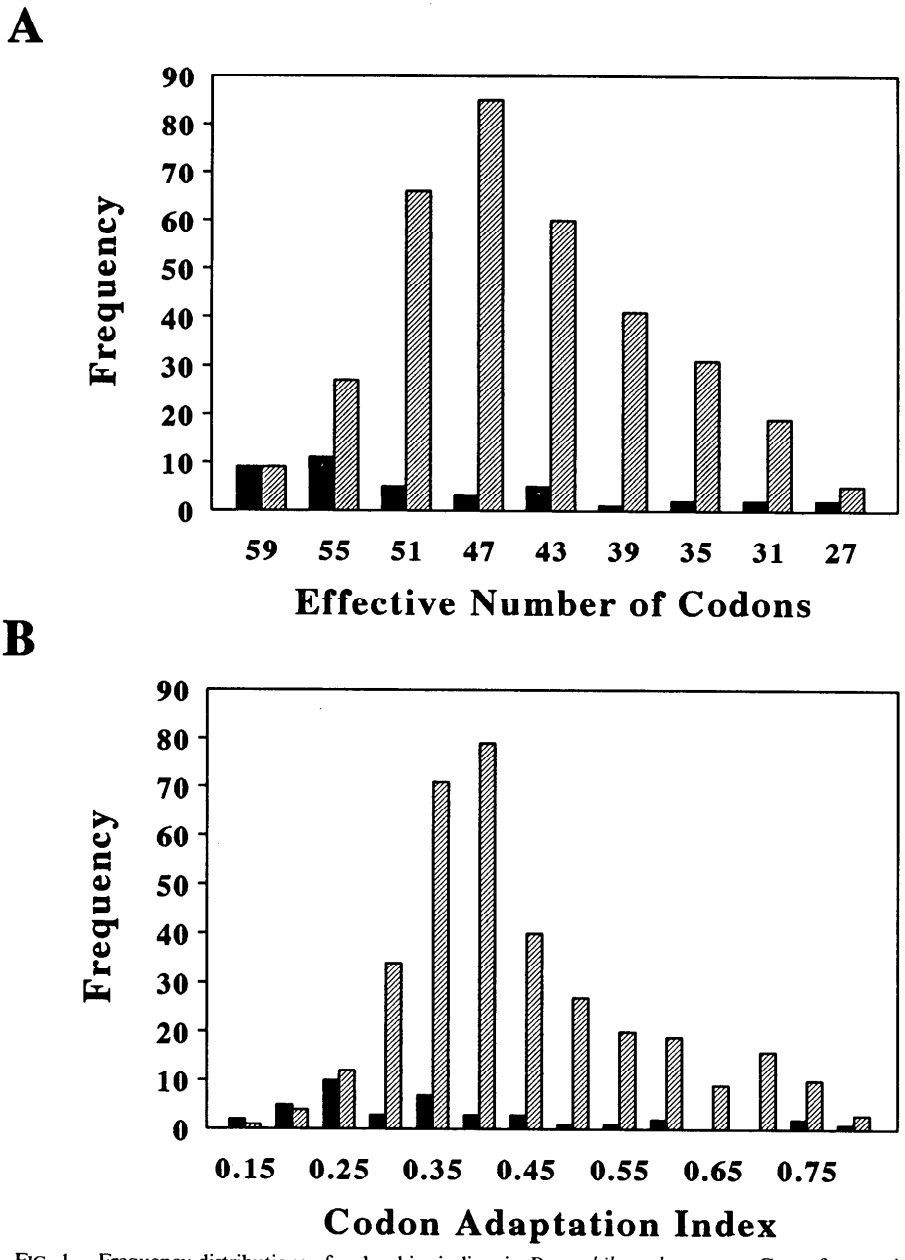
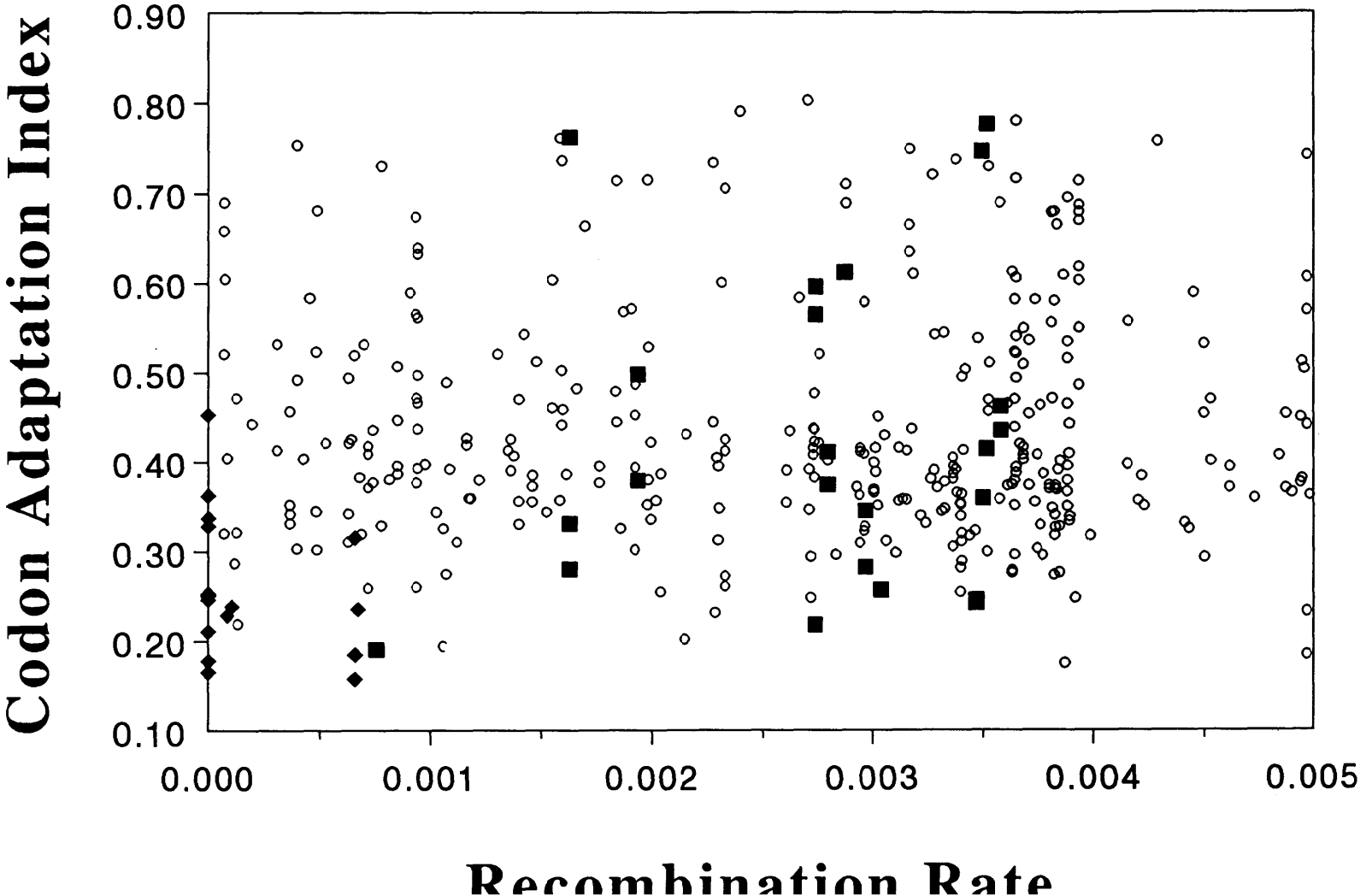


FIG. 1.—Frequency distributions of codon-bias indices in *Drosophila melanogaster*. Genes from regions of low recombination (polytene-chromosome sections 1, 21, 40–41, 60–61, 80–81, and 100–102) are represented by blackened bars; all others are represented by hatched bars. A, ENC, for which lower values indicate higher bias. B, CAI, for which higher values indicate higher bias.

recombination rate (Spearman's $r_s = -0.138$, 381 df, $P = 0.007$), and the correlation between CAI and recombination rate is very nearly significant (Spearman's $r_s = 0.099$, 383 df, $P = 0.053$). However, statistical significance of these correlations is probably an artifact of the inclusion of low-biased loci from the regions of low recombination (fig. 3). Treating recombination rate as a class variable, as in the Kruskal-Wallis tests,


Codon Adaptation Index
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.000
0.001
0.002
0.003
0.004
0.005
Recombination Rate

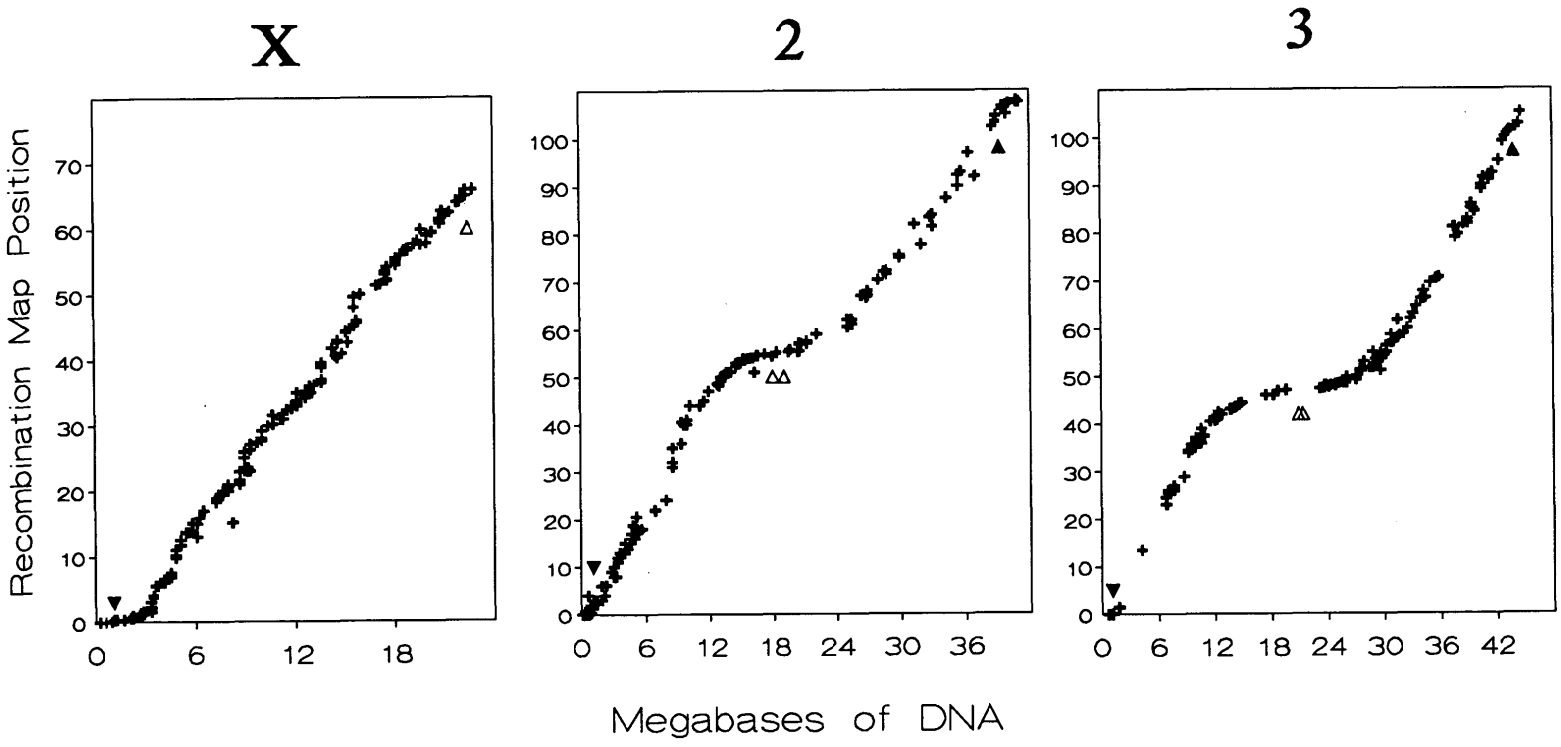


FIG. 2.—Recombination-map position vs. cumulative DNA content. Only genes that have been placed on both the recombination and polytene-chromosome maps are included (of the 385 loci used in this study, some are included in the fig., though most have not yet been placed on the recombination map; see Material and Methods). Blackened and unblackened triangles indicate the limits of telomere sections and centromere sections as they are defined in the text. The X chromosome is telocentric, and the telomere is placed at the origin of the figure. Chromosomes 2 and 3 are metacentric, and, in each case, loci near the telomeres of the left arm are at the origin of the figure.

is more appropriate in the absence of better estimates from the centromeres and telomeres.

Discussion

Evidence for Natural Selection

In several well-studied single-celled organisms, codon bias is positively associated with gene-expression levels (Bennetzen and Hall 1982; Gouy and Gautier 1982; Sharp and Li 1987*a*). In *Drosophila melanogaster* and other multicellular organisms with complex life histories, measurement of gene expression—or that component of gene expression that might be acted on by natural selection—is problematic; however, there are two observations suggesting that variation in natural selection on codon usage is at least partly responsible for variation in codon bias among genes. First, synonymous substitution rates between *Drosophila* species are negatively correlated with codon bias (Shields et al. 1988). In addition, Shields et al. (1988) found that the most common codons found within each amino acid class within highly biased *D. melanogaster* genes end in either G or C. Moriyama and Gojobori (1992) found that *Drosophila* genes with high G+C or C levels in the third codon position evolve more slowly. These observations are consistent with a model in which natural selection limits codon usage in highly expressed genes and, thus, limits the possible synonymous substitutions that can occur as species diverge. Second, support for an explanation based on gene expression levels in *D. melanogaster* emerges from a summary of accounts of actual measurements of gene-expression levels of several *D. melanogaster* loci (Shields et al. 1988).

Despite supportive evidence from *D. melanogaster,* some uncertainty persists. The negative correlation between synonymous substitution rate and codon bias is also expected under a more general model in which some genes are more variable for codon usage, regardless of the actual mechanism that creates limits on codon usage. Similarly, the significance of the measurements of gene-expression levels summarized by Shields et al. (1988) is difficult to assess, because there are a small number of primarily nonquantitative reports.

Our study addresses the role of natural selection by using a population-genetics approach. Several population-genetic models predict natural selection to be less effective under reduced recombination (Muller 1964; Hill and Robertson 1966; Felsenstein 1974; Li 1987; Charlesworth et al. 1993*a*). We confirm these predictions, and thus our finding of reduced codon bias in areas of low recombination supports the hypothesis that observed differences in codon bias among *D. melanogaster* genes result, at least in part, from variation in natural selection acting on codon usage. Our findings do not indicate how natural selection is acting, but, given knowledge from unicellular

FIG. 3.—Codon bias vs. recombination rate for 385 *Drosophila melanogaster* loci. For 382 of the loci included in the present study, the cumulative DNA position was determined from the polytene-chromosome position (e.g., 1A, 1B, and 1C). Recombination rate was estimated for each polytene-chromosome section, from the derivative of the polynomial from the appropriate chromosome (see Material and Methods). Some values for the tip of the X chromosome were slightly negative and were converted to zero. The remaining three loci are on chromosome 4 and were assigned a recombination rate of zero (Hochman 1976). Loci from polytene chromosome sections 1, 40–41, and 80–81 (i.e., the tip of the X chromosome, autosomal centromeres, and chromosome 4) are represented by blackened diamonds; loci from polytene-chromosome regions 21, 60–61, and 100 (i.e., autosomal telomeres) are represented by blackened squares; and the remaining loci are represented by circles.

organisms and supportive evidence from *D. melanogaster,* the most-parsimonious explanation is that natural selection limits codon usage in highly expressed genes via an effect of codon usage on the efficiency of translation.

Alternative explanations for the main result that do not invoke natural selection can also be considered. For example, highly expressed genes might be rare in regions of low recombination. It is unfortunate that this can only be poorly addressed until more is known about levels of gene expression in *D. melanogaster.* It does not appear, however, that the regions of low recombination contain only lowly expressed genes. The *60A* gene (located in map region 60) is reportedly expressed at high levels in third-instar larvae and pupae, as well as in adult males (Wharton et al. 1991), though it also has relatively high codon bias (ENC = 34.7, CAI = 0.564). Lai et al. (1991) report "prominent" expression of *zfh-1* (located in polytene-band region 100) and *zfh-2* (on chromosome 4) in the developing central nervous system; *zfh-1* is moderately biased, but *zfh-2* has very low codon bias (ENC = 59.1, CAI = 0.210). The *ase* gene, located near the tip of the X chromosome, is reported to be strongly expressed in the central nervous system (Gonzales et al. 1989) and also has low codon bias (ENC = 55.7, CAI = 0.246). Finally, Fitch and Strausbaugh (1993) have noted that the histone genes, also highly expressed, have low codon bias (see Appendix). These genes are located in polytene-band region 39, fairly close to the centromere of chromosome 2. While this evidence is anecdotal, it does suggest that the regions of low recombination contain genes that are highly expressed and low biased.

A second potential explanation for the relationship between codon bias and recombination rate is variation in G+C content. Given that the codons that occur disproportionately in highly biased loci of *D. melanogaster* end in G or C (Shields et al. 1988), our observation might be explained if areas of low recombination have generally lower G+C content than the remainder of the genome. To test this, we measured intron G+C content in all those genes for which intron sequences were available (see Appendix). Both ENC and CAI were found to be highly correlated to intron G+C content, with high intron G+C content associated with high codon bias (ENC: Spearman's $r_s = -0.325$, 138 df, $P < 0.0001$; CAI: Spearman's $r_s = 0.287$, 150 df, $P = 0.0005$). However, G+C content does not explain the codon-bias difference in regions of high versus low recombination. Although only 12 of the 142 genes with available intron sequences are in the regions of low recombination, they still show significantly lower codon bias than do genes found elsewhere. However, there is no significant regional difference in intron G+C content (table 2).

Table 2
Kruskal-Wallis Tests for Differences in Codon Bias and Intron G+C Content in Genes for Which Intron Data Are Available

	Sample Size		Mean			
Codon Bias Index	N_{high}	N_{low}	N_{high}	N_{low}	H[a]	P
ENC	128	12	44.1	49.5	4.251	0.039
CAI	130	12	0.457	0.361	5.169	0.023
$G+C_i$[b]	130	12	0.371	0.345	0.762	0.383

[a] Kruskal-Wallis test statistic.

[b] Proportion of G+C used in introns, not including the conserved GT from the 5′ end or AG from the 3′ end of each intron.

Models of Natural Selection

There remains uncertainty over exactly which of several population-genetic models best explain the observations. The most general description of the interaction between natural selection and recombination rate is the Hill-Robertson effect (Hill and Robertson 1966; Felsenstein 1974). In brief, natural selection, acting on a locus, will also affect the persistence of alleles at linked loci and will effectively inflate the variance in the reproductive success of linked loci. This accelerated rate of genetic drift can also be described as a reduction in effective population size. Standard population genetics holds that selection cannot proceed and overcome the effects of genetic drift if the product of the selective differential and the effective population size is $\ll 1$ (e.g., see Dobzhansky et al. 1977, p. 306). Thus, the Hill-Robertson effect can be described as a reduction in the effectiveness of natural selection in regions of low recombination.

A special, or extreme, case of the Hill-Robertson effect is genetic hitchhiking (Maynard Smith and Haigh 1974), whereby strong directional selection on a favorable mutation leads to rapid fixation within the population of that mutation and of all linked nucleotide sites. Under this model, the codon-usage pattern from a single gene copy becomes fixed in the population, even if that codon-usage pattern is not optimal.

The Hill-Robertson effect has been closely associated with another effect expected under low recombination: Muller's ratchet (Muller 1964; Felsenstein 1974). The "ratchet" refers to the successive loss from the population of genomes (or haplotypes, in regions of tight linkage) containing the fewest number of slightly deleterious mutations. By genetic drift, tight linkage inevitably leads to an increase in the number of slightly deleterious mutations per genome. While the ratchet leads to an increase in the number of mutations per genome, the Hill-Robertson effect leads to an increase in the fixation rate of mutations. Both effects are predicted for slightly deleterious mutations under tight linkage, but the magnitudes of each effect can vary somewhat independently, depending on the selection model and the population size (Charlesworth et al. 1993*b*). In terms of codon usage, the ratchet model alone predicts that individual sequences from genes in regions of low recombination should have low codon bias because of an elevated usage of "suboptimal" codons. The ratchet model also predicts that there should be considerable variation among gene copies for codon usage, since genealogical lineages should independently accumulate different sets of slightly deleterious mutations. In contrast, the Hill-Robertson effect would lead to both low codon bias and low variation among gene copies.

Another model, recently proposed, holds that regions of low recombination will have a reduced effective population size because of a high rate of deleterious mutations (Charlesworth et al. 1993*a*). In this "background selection" model, the overall mutation rate to harmful mutations for blocks of genes in tight linkage is a function of the number of loci in the region. If this mutation rate is high, then, by mutation-selection balance, a large fraction of all genomes may carry deleterious mutations and, thus, have no chance of fixation. The effective population size for the region is reduced to that subset of genomes that do not carry deleterious mutations.

Li (1987) has modeled the effectiveness of natural selection on codon usage and has found that even slight selective differences among codons can lead to strong codon bias. Li (1987) also modeled the reduced effectiveness of natural selection for codon usage under reduced recombination and found a large effect. Several different schemes were considered for combining the effects of multiple, very small selective differentials, with little effect on the results. These findings support the general Hill-Robertson effect

as well as Muller's ratchet (Li 1987), though hitchhiking, per se, and the background-selection model were not addressed.

To evaluate the different population genetic models, we must also consider the growing body of data on levels of DNA sequence variation in regions of low recombination in *D. melanogaster.* Loci near the telomere of the X chromosome (Begun and Aquadro 1991), near the centromere of the X chromosome (Langley et al. 1993), and on the fourth chromosome (Berry et al. 1991) have very little intraspecific sequence polymorphism. In general, loci in regions of reduced recombination are less polymorphic within *D. melanogaster,* though they have typical levels of interspecific divergence, as is expected if intraspecific polymorphism is limited by the effect of genetic hitchhiking (Begun and Aquadro 1992).

The near lack of variation in areas of low recombination is strong support for a model of reduced effective population size, and all of the studies reported have invoked genetic hitchhiking. The hitchhiking model is attractive because effective population size can be reduced to almost any level as a function of the strength and frequency of selective sweeps. Charlesworth et al. (1993*a*) discuss the background-selection model specifically within the context of the observation of low variation in *D. melanogaster.* The authors conclude that, while deleterious mutations may make a significant contribution to the reduced levels of variation, the model cannot account for the complete lack of variation found in some portions of the genome and in some populations.

Our findings on codon bias may be at odds, in one respect, with data on levels of sequence variation. We found codon bias to be significantly diminished in regions of lowest recombination, but we found no correlation between codon bias and recombination rate in the remainder of the genome. This suggests that, once recombination rate exceeds some low threshold, selection against slightly deleterious codons proceeds without hinderance. In contrast, Begun and Aquadro (1992) observed a relationship between polymorphism and recombination rate in a sample of 20 loci that was consistent with a linear model (that is, as recombination rate increases, fixation of selectively advantageous mutations brings about loss of polymorphism in increasingly smaller genomic regions). The data were very limited (Begun and Aquadro 1992), but this pattern is also apparent in more recent data from the X and third chromosomes (D. Begun, E. Kindahl, and C. Aquadro, personal communication). In a simplistic view, both reduced codon bias (via reduced natural selection) and reduced levels of variation are predicted for regions of reduced recombination, so it is not clear why they both exhibit an effect for recombination near zero but differ for intermediate levels of recombination. It is possible that Muller's ratchet contributes to this apparent incongruity. Charlesworth et al. (1993*b*) found that the ratchet slows sharply once recombination rate reaches some threshold value; that is, the speed of Muller's ratchet is not linearly related to recombination rate. It may be that low codon bias in regions of very low recombination came about largely through the action of the ratchet. In contrast, polymorphism levels may vary more as a function of local effective population size which, with either a general Hill-Robertson effect or hitchhiking, may vary linearly with recombination rate.

Acknowledgments

We thank M. Bulmer, A. C. Eyre-Walker, and D. Charlesworth for constructive comments. This research was supported by the Bureau of Biological Research, Rutgers University and by the National Science Foundation (support to J.H.).

APPENDIX

Table A1

***Drosophila melanogaster* Loci Used in the Present Study**

Locus Abbreviations (and Name)	AC[a]	Map[b]	ENC[c]	CAI[d]	G+C$_3^e$	Rec[f]
X chromosome:						
Appl (b-amyloid-protein-precursor)	J04516	1B	45.2	0.453		0.00000
ac (achaete)	M17120	1B	54.2	0.253		0.00000
ase (asense)	X52892	1B	55.7	0.246		0.00000
elav (embryonic-lethal)	M21152	1B	46.7	0.337		0.00000
1(1)sc (lethal-of-scute)	X12549	1B	52.3	0.362		0.00000
sc (scute)	M17119	1B	61.0	0.250		0.00000
su (s) (suppressor-of-sable)	M57889	1B	55.9	0.328	0.388	0.00000
y (yellow)	X04427	1B	54.5	0.251	0.332	0.00000
arm (armadillo)	X54468	2B	47.7	0.392	0.454	0.00109
Actn (a-actinin)	X51753	2C	36.6	0.603		0.00155
usp (ultraspiracle)	X53417	2C	41.5	0.460		0.00155
Pgd (Phosphogluconate-dehydrogenase)	M80598	2D	37.2	0.568	0.466	0.00187
fs(1)K10 (female-sterile-(1)-K10)	X12836	2E	54.8	0.357	0.398	0.00202
crn (crooked-neck)	X58374	2F	48.4	0.395		0.00230
phl (pole-hole)	X07181	2F	51.4	0.312	0.297	0.00230
gt (giant)	X61148	3A	46.8	0.422	0.493	0.00274
tko (technical-knock-out)	M19494	3A	42.6	0.476		0.00274
z (zeste)	X06743	3A	45.9	0.382	0.377	0.00274
fs(1)Ya (female-sterile-(1)-Ya)	M38442	3B	43.5	0.415		0.00294
per (period)	M30114	3B	41.2	0.411	0.523	0.00294
sgg (shaggy)	X53332	3B	51.8	0.309		0.00294
Fcp3C (Follicle-cell-protein-3C)	M18281	3C	52.7	0.281	0.449	0.00340
N (Notch)	K03508	3C	48.3	0.352	0.393	0.00340
Pig1 (pre-intermoult-gene-1)	X15760	3C	42.5	0.254		0.00340
w (white)	X51749	3C	47.7	0.353	0.397	0.00340
dnc (dunce)	M14982	3D	48.7	0.299	0.276	0.00352
Fas2 (Fasciclin-2)	M77165	4B	48.3	0.350		0.00388
norpA (no-receptor-potential-A)	J03138	4B	47.8	0.395		0.00388
omb (optomotor-blind)	M81796	4C	51.9	0.317		0.00399
Act5C (Actin-5C)	K00667	5C	32.4	0.713	0.353	0.00393
sqh (spaghetti-squash)	M67494	5D	36.7	0.608	0.423	0.00387
swa (swallow)	X56023	5E	46.1	0.373	0.337	0.00383
ogre (optic-ganglion-reduced)	X61180	6E	36.4	0.538		0.00348
nullo (nullo)	X65444	6F	44.9	0.413		0.00341
ct (cut)	X07985	7B	51.3	0.331		0.00324
RpS14B (Ribosomal-protein-S14B)	M21045	7C	35.8	0.634	0.386	0.00317
RpS14A (Ribosomal-protein-S14A)	M21045	7C	33.5	0.664	0.418	0.00317
mys (myospheroid)	J03251	7D	47.0	0.429		0.00305
Cp38 (Chorion-protein-38)	X05245	7F	37.2	0.578	0.356	0.00296
Cp36 (Chorion-protein-36)	X05245	7F	42.5	0.579	0.471	0.00296
Nrg (neuroglian)	M28231	7F	45.5	0.408		0.00296
otu (ovarian-tumors)	M30825	7F	53.6	0.322	0.331	0.00296
Yp1 (Yolk-protein-1)	V00248	8E	32.6	0.710	0.254	0.00288
Yp2 (Yolk-protein-2)	X01524	8E	33.1	0.688	0.375	0.00288
sev (sevenless)	X13665	10A	49.0	0.357	0.371	0.00316
v (vermilion)	M34147	10A	44.2	0.412	0.386	0.00316
dlg1 (discs-large-1)	M73529	10B	48.4	0.371		0.00329
RpII215 (RNA-polymerase-II-215KD-subunit)	M27431	10C	46.4	0.387	0.334	0.00337
nod (no-distributive-disjunction)	M36195	10C	51.5	0.305		0.00337
Ptp10D (protein-tyrosine-phosphatase-10D)	M80465	10D	46.2	0.364		0.00341

APPENDIX

Table A1 (*Continued*)

Locus Abbreviations (and Name)	AC[a]	Map[b]	ENC[c]	CAI[d]	G+C$_i$[e]	Rec[f]
X chromosome: (*Continued*)						
CkIIb (Casein-kinase-II-b-subunit)	M16535	10E	56.7	0.317		0.00344
Yp3 (Yolk-protein-3)	X04754	12A	33.2	0.679	0.357	0.00382
up (upheld)	X54504	12A	36.3	0.580		0.00382
rut (rutabaga)	M81887	12F	46.3	0.401		0.00385
eag (ether-a-go-go)	M61157	13A	50.6	0.326		0.00383
Ipou (Ipou)	X58436	13C	53.8	0.295		0.00377
Gapdh2 (Glyceral-3-phosphate-dehydrogenase-2)	M11255	13F	35.2	0.689		0.00358
Gb13F (G-protein-b-subunit-13F)	M22567	13F	50.7	0.358		0.00358
disco (disconnected)	X56232	14B	45.6	0.395		0.00337
nonA (no-action-potential-A)	X55902	14C	53.4	0.345	0.334	0.00331
if (inflated)	M19059	15A	44.7	0.450		0.00302
r (rudimentary)	X04813	15A	51.4	0.351	0.355	0.00302
B (Bar)	M73259	16A	47.4	0.390		0.00261
B-H2 (Bar-H2)	M82885	16A	45.0	0.354	0.552	0.00261
Sh (Shaker)	M17211	16F	58.0	0.202		0.00215
bnb (bangles-and-beads)	X63828	17E	40.2	0.571		0.00190
Zw (Zwischenferment)	M26673	18E	34.5	0.600	0.451	0.00231
AnnX (Annexin-X)	M34069	19A	36.8	0.583		0.00267
run (runt)	X56432	19E	36.9	0.471		0.00382
shakB (shaking-B)	X65103	19E	50.4	0.348		0.00382
sol (small-optic-lobes)	M64084	19F	44.8	0.397		0.00415
Chromosome 2:						
l(2)gl (lethal-(2)-giant-larvae)	X05426	21A	56.0	0.190	0.313	0.00076
M(2)21C (Minute-(2)-21C)	Y00504	21C	28.4	0.762		0.00163
Plc21C (Phospholipase-C-21C)	M60452	21C	54.1	0.331		0.00163
Rpl135 (RNA-polymerase-I-135KD-subunit)	X17298	21C	57.4	0.280	0.310	0.00163
Pkg21D (cGMP-dependent-protein-kinase-21D)	M27114	21D	47.6	0.379	0.365	0.00193
ninaA (neither-inactive-nor-action-potent-A)	M22851	21D	41.0	0.498	0.379	0.00193
Eno (Enolase)	X17034	22A	28.3	0.803		0.00270
Rrp1 (Recombination-repair-protein-1)	M62472	23B	55.8	0.303		0.00375
Syt (Synaptotagmin)	M55048	23B	47.2	0.407		0.00375
Tubg (Tubulin-g)	M61765	23C	48.5	0.391		0.00384
Shaw (Shaker-cognate-w)	M32661	24B	46.3	0.384		0.00422
ft (fat)	M80537	24D	52.4	0.331		0.00441
Bbbf2 (Box-B-binding-factor-2)	X64429	25C	44.2	0.371		0.00488
Cg25C (Collagen-type-IV)	M23704	25C	49.8	0.454		0.00488
Bsg25D (blastoderm-specific-gene-25D)	X04896	25D	49.2	0.366	0.427	0.00490
Glu-RII (Glutamate-receptor-II)	M73271	25F	48.7	0.381		0.00495
Lam (Lamin)	X16275	25F	38.1	0.512	0.388	0.00495
Gpdh (Glycerol-3-phosphate-dehydrogenase)	J04567	26A	37.5	0.569	0.403	0.00497
Mst26Ab (male-specific-RNA-26Ab)	Y00219	26A	52.2	0.232	0.351	0.00497
Mst26Aa (male-specific-RNA-26Aa)	Y00219	26A	61.0	0.184	0.308	0.00497
Vm26Aa (Vitelline-membrane-26Aa)	M20936	26A	34.9	0.606		0.00497
Vm26Ab (Vitelline-membrane-26Ab)	M18280	26A	27.0	0.742		0.00497
chi (chickadee)	M84528	26A	34.5	0.441		0.00497
H2.0 (homeodomain-protein-2.0)	Y00843	26B	47.4	0.362		0.00498
Hrb27C (heterogeneous-nuclear-ribonucleoprotein-27C)	X62639	27C	43.0	0.504		0.00496
Pcp (Pupal-cuticle-protein)	X06286	27D	42.8	0.376	0.507	0.00494
snRNP27D (sn-ribonucleoprotein-27D)	M31162	27D	36.7	0.450	0.467	0.00494

APPENDIX

Table A1 (*Continued*)

Locus Abbreviations (and Name)	AC[a]	Map[b]	ENC[c]	CAI[d]	G+C$_3^e$	Rec[f]
Chromosome 2: (*Continued*)						
wg (wingless)	J03650	28A	41.7	0.407		0.00485
LanB1 (Laminin-B1)	M19525	28D	51.4	0.359		0.00473
Src29A (src-oncogene-2)	M16599	29A	53.3	0.371		0.00462
Su(var)205 (Suppressor-of-variegation-205)	M57574	29A	48.5	0.394	0.365	0.00462
numb (numb)	M27815	30B	50.1	0.350		0.00423
Pka-C1 (cAMP-dependent-protein-kinase-A)	X16969	30C	35.4	0.556	0.394	0.00415
me31B (maternal-expression-at-31B)	M59926	31B	59.9	0.248		0.00392
cdc2 (cdc2)	X57485	31E	56.1	0.276		0.00363
da (daughterless)	Y00221	31E	44.8	0.374		0.00363
Vm32Ec (Vitelline-membrane-32Ec)	M27647	32E	54.9	0.381		0.00326
sala (spalt-accessory)	X57474	33A	47.3	0.298	0.226	0.00311
prd (paired)	M14548	33C	48.8	0.372	0.466	0.00293
dim (didymous)	M65016	33F	44.0	0.346		0.00271
twn (twain)	M65015	33F	42.8	0.391		0.00271
Sos (Suppressor-of-sevenless)	M77501	34D	47.5	0.348	0.366	0.00230
Adh (Alcohol-dehydrogenase)	M11290	35B	31.2	0.715	0.389	0.00198
Su(H) (Suppressor-of-Hairless)	X58393	35B	48.4	0.352		0.00198
esg (escargot)	M83207	35C	48.0	0.394		0.00192
vas (vasa)	X12946	35C	55.5	0.302	0.311	0.00192
sna (snail)	Y00288	35D	41.6	0.479		0.00183
Cyt-c1 (Cytochrome-c1)	X01761	36A	49.6	0.357		0.00158
Cyt-c2 (Cytochrome-c2)	M11381	36A	27.0	0.761		0.00158
Cyt-b5 (Cytochrome-b5-related)	X15008	36B	51.4	0.344	0.287	0.00153
BicD (Bicaudal-D)	X51652	36C	51.2	0.385		0.00146
Dlar (Dlar)	M27700	36C	49.0	0.373		0.00146
dl (dorsal)	M23702	36C	51.6	0.355		0.00146
Arr1 (Arrestin-A)	M30177	36D	41.6	0.543	0.316	0.00142
Fas3 (Fasciclin-3)	M27813	36E	44.8	0.407		0.00138
Dox-A2 (Diphenol-oxidase-A2)	M63010	37B	45.3	0.427	0.404	0.00117
amd (a-methyl-dopa-resistant)	X04695	37B	48.7	0.419	0.378	0.00117
l(2)37Cs (lethal-(2)-37Cs)	X05991	37C	54.5	0.310	0.466	0.00112
Sd (Segregation-distorter)	X60218	37D	52.0	0.274		0.00107
Top2 (Topoisomerase-2)	X61209	37D	45.5	0.489	0.374	0.00107
ref(2)P (refractory-to-sigma-P)	X16993	37E	51.6	0.344	0.335	0.00103
cad (caudal)	M21070	38E	44.5	0.380	0.506	0.00081
His1 (Histone-H1)	X04073	39D	51.1	0.259		0.00072
His2B (Histone-H2B)	X14215	39D	56.3	0.371		0.00072
His3 (Histone-H3)	X14215	39D	49.4	0.409		0.00072
His4 (Histone-H4)	X14215	39D	42.4	0.418		0.00072
Ef2b (elongation-factor-2B)	X15805	39E	41.9	0.531		0.00070
tsh (teashirt)	M57496	40A	57.9	0.235		0.00067
cta (concertina)	M63651	40F	56.7	0.184		0.00066
Gprk1 (G-protein-coupled-receptor-kinase-1)	M80493	41B	52.6	0.157		0.00066
ap (apterous)	X65158	41B	54.0	0.315		0.00066
Act42A (Actin-42A)	K00670	42A	45.2	0.494	0.437	0.00063
EcR (ecdysone-receptor)	M74078	42A	51.9	0.342		0.00063
mle (male-less)	M74121	42A	53.2	0.310		0.00063
Adf1 (transcription-factor-Adh1)	M37787	42C	46.1	0.425		0.00064
Gapdh1 (Glyceral-3-phosphate-dehydrogenase-1)	M11254	43E	31.1	0.730		0.00078
tor (torso)	X15150	43E	52.2	0.329	0.315	0.00078
Gg1 (G-protein-g-1)	M85042	44C	35.7	0.589		0.00091

APPENDIX

Table A1 (*Continued*)

Locus Abbreviations (and Name)	AC[a]	Map[b]	ENC[c]	CAI[d]	G+C$_3^e$	Rec[f]
Chromosome 2: (*Continued*)						
Lcp2 (Larval-cuticle-protein-2)	V00203	44D	48.9	0.561	0.431	0.00094
Lcp4 (Larval-cuticle-protein-4)	V00203	44D	35.2	0.639	0.396	0.00094
Lcp3 (Larval-cuticle-protein-3)	V00203	44D	37.3	0.632	0.442	0.00094
LvpL (Larval-visceral-protein-L)	V00204	44D	42.1	0.497	0.448	0.00094
LvpH (Larval-visceral-protein-H)	V00204	44D	45.4	0.466	0.309	0.00094
LvpD (Larval-visceral-protein-D)	V00204	44D	50.3	0.393		0.00094
tuf (tufted)	X17558	44D	42.1	0.437	0.337	0.00094
wnt2 (wnt-oncogene-analog-2)	X64735	45F	45.7	0.359		0.00118
trpl (trp-like)	M88185	46A	47.6	0.380		0.00122
eve (even-skipped)	M14767	46C	37.8	0.521	0.403	0.00130
D14-3-3 (D14-3-3-protein)	M77518	46E	47.7	0.390		0.00136
Jra (Jun-related-antigen)	M36181	46E	42.8	0.425		0.00136
Rab3 (Ras-related-protein)	M64621	47B	52.5	0.459		0.00160
Try (Trypsin)	X02989	47D	33.6	0.663		0.00169
inv (invected)	X05273	48A	48.3	0.325		0.00186
Ef1a48D (elongation-factor-1a48D)	X06870	48D	41.0	0.528	0.444	0.00198
Tk48D (Tyrosine-kinase-at-48D)	X63453	48D	48.7	0.380		0.00198
Cal (Calmodulin)	X05951	49A	46.5	0.431	0.355	0.00215
sca (scabrous)	M60065	49D	45.6	0.404		0.00229
Mdr49 (Multiple-drug-resistance-49)	M59076	49E	47.9	0.412		0.00233
Psc (Posterior-sex-combs)	X59275	49E	53.5	0.272		0.00233
Su(z)2 (Suppressor-of-zeste-2)	X56799	49E	52.9	0.261	0.349	0.00233
Mp20 (muscle-protein-20)	Y00795	49F	30.0	0.791	0.337	0.00240
mam (mastermind)	X54251	50C	48.0	0.434		0.00262
Pox-n (Paired-box-neural)	M86927	52D	52.0	0.356		0.00312
sli (slit)	X53959	52D	43.7	0.415		0.00312
Kin (Kinesin)	M24441	52F	45.2	0.437		0.00318
RpA1 (Ribosomal-protein-A1)	X05016	53C	36.6	0.720		0.00327
Pkc53E (Protein-C-kinase-53E)	X05283	53E	49.5	0.377	0.326	0.00333
inaC (inactivation-no-afterpotential-C)	J04845	53E	54.2	0.347		0.00333
Amy-d (Amylase-A)	X04569	54B	29.3	0.737		0.00338
Hsf (Heat-shock-factor)	M60070	55A	52.0	0.320		0.00341
PpY-55A (protein-phosphatase-Y-55A)	Y07510	55A	56.6	0.289		0.00341
pAbp (poly-A-binding-protein)	M38019	55A	45.6	0.495		0.00341
stau (staufen)	M69111	55A	50.2	0.311		0.00341
Ote (Otefin)	X17495	55C	54.0	0.339		0.00340
grh (grainy-head)	X15657	55E	48.1	0.365		0.00338
5HT-R2A (serotonin-receptor)	M55533	56A	45.1	0.381		0.00336
Dpt (Diptericin)	Z11728	56A	45.6	0.404		0.00336
Pcna (Proliferating-cell-nuclear-antigen)	M33950	56E	37.2	0.542	0.286	0.00328
Act57A (Actin-57A)	K00673	57A	29.7	0.749		0.00317
tud (tudor)	X62420	57C	55.1	0.311		0.00306
TfIId (TfIID)	M38082	57F	50.7	0.327		0.00296
CycB (Cyclin-B)	X55542	59A	45.3	0.401		0.00280
twi (twist)	X12506	59C	36.3	0.520		0.00276
bw (brown)	M20630	59E	40.3	0.436		0.00273
Chromosome 3:						
Ca-P60A (Calcium-ATPase)	M62892	60A	35.7	0.595		0.00274
G-sa60A (G-protein-sa-60a)	M33998	60A	61.0	0.218	0.414	0.00274
Tgfb-60A (Transforming-growth-factor-b-60A)	M77012	60A	34.7	0.564		0.00274
Acr60C (Acetylcholine-receptor-C)	M23412	60C	41.1	0.374		0.00280

APPENDIX

Table A1 (*Continued*)

Locus Abbreviations (and Name)	AC[a]	Map[b]	ENC[c]	CAI[d]	G+C$_i$[e]	Rec[f]
Chromosome 3: (*Continued*)						
Tubb60D (tubulin-b60D)	M22335	60D	32.7	0.611	0.409	0.00287
uzip (unzipped)	X07450	60E	60.5	0.282		0.00297
zip (zipper)	M35012	60E	51.9	0.345		0.00297
Kr (Kruppel)	X03414	60F	61.0	0.257	0.332	0.00304
emc (extra-macrochaetae)	M32637	61C	43.8	0.411	0.313	0.00280
Aprt (Adenine-phosphoribosyltransferase)	M18432	62B	48.4	0.402		0.00368
R (Roughened)	M80535	62B	39.6	0.509		0.00368
Spec-a (a-Spectrin)	M26400	62B	39.9	0.549		0.00368
Shab (Shaker-cognate-b)	M32659	63A	51.2	0.356		0.00420
Hsp83 (heat-shock-protein-83)	X03810	63B	32.1	0.757		0.00429
Ubi-p (Ubiquitin)	M22428	63F	35.2	0.589		0.00445
Acr64B (Acetylcholine-receptor-D)	M20316	64A	43.8	0.454	0.378	0.00450
Ras64B (ras-oncogene-2)	M10804	64A	36.4	0.531	0.327	0.00450
Src64B (src-oncogene-1)	M11917	64B	43.3	0.470	0.298	0.00453
Bj1 (Bj1-protein)	X58530	64F	48.7	0.401	0.391	0.00453
Mdr65 (Multiple-drug-resistance-65)	M59077	65A	55.2	0.293		0.00450
Cf1a (Cf1a-transcription-factor)	X58435	65D	47.0	0.325		0.00443
Arr2 (Arrestin-B)	M32141	66D	39.5	0.549	0.331	0.00393
Cp18 (Chorion-protein-18)	X02497	66D	35.4	0.669	0.285	0.00393
Cp15 (Chorion-protein-15)	X02497	66D	39.2	0.617	0.537	0.00393
Cp19 (Chorion-protein-19)	X02497	66D	32.8	0.686	0.447	0.00393
Cp16 (Chorion-protein-16)	X16715	66D	33.5	0.602	0.393	0.00393
Prm (Paramyosin)	X58722	66D	30.9	0.679		0.00393
h (hairy)	X15905	66D	41.0	0.486	0.441	0.00393
Rdl (cyclodiene-resistant)	M69057	66F	53.5	0.273		0.00382
Hsp22 (heat-shock-protein-22)	X03888	67B	47.8	0.438		0.00365
Hsp23 (heat-shock-protein-23)	V00210	67B	39.2	0.581		0.00365
Hsp26 (heat-shock-protein-26)	X03890	67B	39.5	0.523		0.00365
Hsp27 (heat-shock-protein-27)	X03891	67B	47.0	0.470		0.00365
Hsp67Ba (heat-shock-protein-67B-a)	M26267	67B	43.3	0.379		0.00365
Hsp67Bb (heat-shock-protein-67B-b)	X07311	67B	58.9	0.296	0.289	0.00365
Hsp67Bc (heat-shock-protein-67B-c)	X06542	67B	44.5	0.351		0.00365
LanB2 (Laminin-B2)	M58417	67C	43.8	0.456	0.389	0.00353
M(3)67C (Minute-(3)-67c)	M22142	67C	30.8	0.729	0.365	0.00353
Tuba67C (tubulin-a67C)	M14646	67C	42.9	0.469	0.293	0.00353
Sod (Superoxide-dismutase)	M24421	68A	37.6	0.610	0.356	0.00319
Sgs7 (salivary-gland-secretion-7)	X01918	68C	47.1	0.369	0.355	0.00301
Sgs8 (salivary-gland-secretion-8)	X01918	68C	44.6	0.399		0.00301
Sgs3 (salivary-gland-secretion-3)	X01918	68C	46.7	0.366	0.435	0.00301
CycA (Cyclin-A)	M24841	68D	51.4	0.362		0.00294
EstP (Esterase-P)	M33780	69A	55.7	0.248	0.308	0.00272
Est6 (Esterase-6)	M33780	69A	54.6	0.294	0.234	0.00272
Acp70A (Accessory-gland-protein-70A)	M21201	70A	46.1	0.231		0.00229
Eip71CD (Ecdysone-induced-protein-28/29KD)	X04521	71C	49.6	0.386		0.00161
Pka-C3 (cAMP-dependent-protein-kinase)	X16961	72A	51.0	0.330		0.00140
arl (arflike)	M61127	72A	39.8	0.470	0.393	0.00140
brm (brahma)	M85049	72A	52.1	0.356		0.00140
gil (giant-lens)	X65161	73A	49.3	0.325		0.00106
tra (transformer)	M17478	73A	55.5	0.194	0.226	0.00106
Nrt (Neurotactin)	X53837	73C	50.1	0.397		0.00098
Dbp73D (Dead-box-protein-73D)	M74824	73D	58.5	0.260	0.302	0.00094

APPENDIX

Table A1 (*Continued*)

Locus Abbreviations (and Name)	AC[a]	Map[b]	ENC[c]	CAI[d]	$G+C_i$[e]	Rec[f]
Chromosome 3: (*Continued*)						
Rh4 (Rhodopsin-4)	M17719	73D	46.4	0.377		0.00094
sina (seven-in-absentia)	M38384	73D	38.5	0.471		0.00094
Eip74EF (Ecdysone-induced-protein-74EF)	M37082	74E	43.6	0.383		0.00068
Pep (Protein-on-ecdysone-puffs)	X56689	74F	42.4	0.519		0.00066
Eip75B (Ecdysone-induced-protein-75B)	X51548	75B	42.4	0.421		0.00053
ftz-f1 (ftz-transcription-factor-1)	M63711	75C	49.4	0.345		0.00048
term (terminus)	M19140	75C	38.1	0.523		0.00048
Cat (Catalase)	X52286	75D	38.9	0.583		0.00046
Shal (Shaker-cognate-l)	M32660	76B	40.3	0.413		0.00031
polo (polo)	X63361	77A	43.7	0.442		0.00020
kni (knirps)	X13331	77E	40.7	0.471	0.419	0.00013
knrl (knirps-like)	X14153	77E	50.8	0.321		0.00013
Pc (Polycomb)	X55702	78C	45.7	0.404	0.452	0.00009
Edg78E (ecdysone-dependent-gene-78E)	M71247	78E	38.7	0.604	0.457	0.00007
Act79B (Actin-79B)	M18829	79B	32.0	0.690	0.326	0.00007
Egon (embryonic-gonad)	X16631	79B	47.0	0.320		0.00007
Ape (apurinic-endonuclease-3)	M25772	79C	34.1	0.658		0.00007
Dromsopa (Dromsopa)	X56491	79C	49.7	0.521		0.00007
CkIIa (Casein-kinase-II-a-subunit)	M16534	80A	56.1	0.228		0.00009
Dsk (Drosulfakinin)	J03957	81F	54.2	0.238		0.00011
tub (tube)	M59501	82A	53.0	0.287		0.00012
UbcD6 (Ubiquitin-conjugating-enzyme)	M63791	82C	56.9	0.219	0.414	0.00013
Rm62 (Rm62)	X52846	83E	41.7	0.532		0.00031
Dfd (Deformed)	X05136	84A	52.0	0.331		0.00037
ama (amalgam)	M23561	84A	41.8	0.457	0.272	0.00037
bcd (bicoid)	X07870	84A	49.6	0.342	0.336	0.00037
pb (proboscipedia)	X63728	84A	52.8	0.352		0.00037
Scr (Sex-combs-reduced)	X14475	84B	44.4	0.303		0.00040
Tuba84B (tubulin-a84D)	M14643	84B	32.1	0.754	0.392	0.00040
ftz (fushi-tarazu)	X00854	84B	39.4	0.492	0.288	0.00040
Gld (Glucose-dehydrogenase)	M29298	84C	49.0	0.404	0.401	0.00043
Tuba84D (tubulin-a84B)	M14645	84D	33.7	0.681		0.00049
lds (lodestar)	X62629	84D	56.5	0.302		0.00049
hb (hunchback)	Y00274	85A	45.2	0.421	0.419	0.00063
osk (oskar)	M63492	85B	50.5	0.319	0.262	0.00069
neu (neuralized)	X61617	85C	46.2	0.436		0.00074
pum (pumilio)	X62589	85C	48.4	0.377		0.00074
Dl (Dl-chromosomal-protein)	J04725	85D	53.5	0.395		0.00085
Fps85D (fps-oncogene-analog)	X52844	85D	46.9	0.387		0.00085
Ras85D (ras-oncogene-1)	M16429	85D	44.7	0.446		0.00085
Tubb85D (tubulin-b85D)	M20420	85D	40.8	0.507	0.364	0.00085
MtnA (Metallothionein-A)	M69015	85E	32.0	0.674	0.326	0.00093
Tuba85E (tubulin-a85E)	M14644	85E	39.4	0.565	0.316	0.00093
Takr86C (Tachykinin-like-receptor)	M77168	86C	46.0	0.359		0.00118
pros (prospero)	Z11743	86E	43.0	0.413		0.00135
Hsp70A (heat-shock-protein-70A)	J01103	87A	40.9	0.512		0.00148
Gst (Glutathione-S-transferase)	X14233	87B	28.9	0.736		0.00159
Pp1-87B (protein-phosphatase-1-87B)	X15583	87B	41.3	0.502		0.00159
svp (seven-up)	M28863	87B	43.7	0.441		0.00159
Hsp70B (heat-shock-protein-70B)	J01104	87C	43.8	0.482		0.00166
ry (rosy)	Y00308	87D	49.8	0.377	0.370	0.00176

APPENDIX

Table A1 (*Continued*)

Locus Abbreviations (and Name)	AC[a]	Map[b]	ENC[c]	CAI[d]	$G+C_i^e$	Rec[f]
Chromosome 3: (*Continued*)						
snk (snake)	X04513	87D	48.9	0.395		0.00176
Ace (Acetylcholine-esterase)	X17572	87E	43.3	0.444		0.00184
Act87E (Actin-87E)	X12452	87E	32.0	0.714	0.333	0.00184
Mst87F (male-specific-RNA-87F)	Y00831	87F		0.452	0.276	0.00192
SR55 (serine-arginine-protein-55)	X58720	87F	44.7	0.487		0.00192
ems (empty-spiracles)	X51653	88A	42.9	0.422		0.00199
su(Hw) (suppressor-of-Hairy-wing)	Y00228	88A	52.1	0.336	0.389	0.00199
RpII140 (RNA-polymerase-II-140KD-subunit)	X05709	88B	49.0	0.386	0.246	0.00204
trx (trithorax)	M31617	88B	57.3	0.255		0.00204
CycC (Cyclin-C)	X62948	88E	46.9	0.444		0.00228
Hsc70-4 (heat-shock-protein-cognate-4)	M36114	88E	32.1	0.734	0.364	0.00228
Act88F (Actin-88F)	M18826	88F	32.0	0.705	0.421	0.00233
ea (easter)	J03154	88F	44.7	0.424		0.00233
Abd-B (Abdominal-B)	X16134	89E	46.3	0.408		0.00273
Fas1 (Fasciclin-1)	M32311	89E	44.1	0.437	0.409	0.00273
abd-A (abdominal-A)	X54453	89E	49.9	0.415		0.00273
Pros35 (proteasome-35KD-subunit)	X15497	89F	47.8	0.421		0.00276
Sgs5 (salivary-gland-secretion-5)	X04269	90B	60.7	0.296	0.234	0.00283
Edg91 (ecdysone-dependent-gene-91)	M71250	91A	39.5	0.415	0.508	0.00301
gl (glass)	X15400	91A	45.7	0.388	0.317	0.00301
Rh2 (Rhodopsin-2)	M12896	91D	48.9	0.358	0.318	0.00314
nos (nanos)	M72421	91F	47.8	0.339	0.364	0.00322
Dl (Delta)	Y00222	92A	50.0	0.390		0.00328
ninaE (neither-inactive-nor-afterpotent-E)	K02320	92B	37.1	0.544	0.327	0.00332
Rh3 (Rhodopsin-3)	M17718	92D	40.9	0.391		0.00338
MtnB (Metallothionein-B)	M16251	92E		0.503	0.456	0.00342
cdc2c (cdc2c)	X57486	92F	52.7	0.322		0.00347
Atpa (Na-pump-a-subunit)	X14476	93B	42.8	0.511		0.00353
NK1 (NK1)	X55393	93E	48.6	0.373		0.00361
tin (tinman)	X55192	93E	44.2	0.465	0.288	0.00361
T-cp1 (mouse-T-cp1-like)	M21159	94B	38.1	0.581		0.00374
cnc (cap-n-collar)	M37495	94E	50.0	0.374		0.00380
unk (unkempt)	Z11527	94E	49.0	0.370		0.00380
HmG-CoAR (HMG-Coenzyme-A-reductase)	M21329	95A	49.2	0.370		0.00383
nau (nautilus)	M68897	95A	51.8	0.317		0.00383
Mst95E (male-specific-RNA-95E)	M32022	95E	51.9	0.175	0.293	0.00387
crb (crumbs)	M33753	95F	51.6	0.366		0.00388
Acr96Aa (Acetylcholine-receptor-B)	X07194	96A	44.4	0.442		0.00389
Acr96Ab (Acetylcholine-receptor-E)	X53583	96A	46.9	0.409		0.00389
tld (tolloid)	M76976	96B	52.2	0.334		0.00389
bam (bag-of-marbles)	X56202	96C	52.9	0.339	0.322	0.00389
E(spl) (Enhancer-of-split-m7-peptide)	X16553	96F	40.6	0.464		0.00388
E(spl) (Enhancer-of-split-m8-peptide)	X16553	96F	41.1	0.515		0.00388
HLHm5 (E(spl)-region-transcript-m5)	X16552	96F	38.5	0.534		0.00388
gro (groucho)	M20571	96F	44.8	0.379		0.00388
m4 (E(spl)-region-transcript-m4)	X16551	96F	31.0	0.695		0.00388
Elg (ets-like-gene)	X58481	97D	57.4	0.276		0.00385
Tl (Toll)	M19969	97D	49.7	0.328		0.00385
NepYr (Neuropeptide-Y-receptor-like)	M81490	97E	46.7	0.369		0.00384
Tubb97EF (tubulin-b97EF)	M20419	97E	33.5	0.664		0.00384
Bd (Beaded)	X56811	97F	50.1	0.340		0.00383

APPENDIX

Table A1 (*Continued*)

Locus Abbreviations (and Name)	AC[a]	Map[b]	ENC[c]	CAI[d]	G+C$_i$[e]	Rec[f]
Chromosome 3: (*Continued*)						
Mlc1 (Myosin-alkali-light-chain-1)	M10125	98A	39.4	0.555	0.407	0.00381
RpL1 (Ribosomal-protein-L1)	X13382	98A	34.0	0.678		0.00381
fkh (forkhead)	J03177	98D	42.9	0.386		0.00377
Pkc98E (Protein-C-kinase-98E)	J04848	98E	41.6	0.463		0.00376
Sry-c (serendipity-cognate)	M23391	98E	50.8	0.329	0.210	0.00376
yema (yema-gene-nuclein-a)	X63503	98F	49.4	0.355	0.344	0.00374
Ocd (Outcold)	X54794	99A	40.7	0.453		0.00371
Ptp99A (protein-tyrosine-phosphatase-99A)	M81795	99A	47.3	0.373		0.00371
stg (string)	M24909	99A	37.7	0.535		0.00371
Fra (Fos-related-antigen)	X54143	99B	39.9	0.416		0.00368
ncd (non-claret-disjunctional)	X52814	99B	45.4	0.407	0.382	0.00368
trp (transient-receptor-potential)	M34394	99C	48.1	0.420	0.379	0.00367
M(3)99D (Minute-(3)-99D)	X00848	99D	37.0	0.521	0.473	0.00365
Ser99Da (Serine-protease-1)	M24379	99D	29.8	0.716		0.00365
Ser99Dc (Serine-protease-3)	M24380	99D	39.5	0.493		0.00365
Sry-b (serendipity-b)	X03121	99D	35.5	0.540	0.531	0.00365
Sry-a (serendipity-a)	X03121	99D	46.2	0.388		0.00365
Sry-d (serendipity-d)	X03121	99D	34.1	0.606		0.00365
Takr99D (Tachykinin-like-receptor)	X62711	99D	41.8	0.394		0.00365
Tpi (Triose-phosphate-isomerase)	X57576	99D	28.4	0.781		0.00365
Anr (Andropin)	X16972	99E	60.0	0.278	0.241	0.00363
Mlc2 (Myosin-light-chain-2)	M28643	99E	37.9	0.612	0.450	0.00363
tll (tailless)	M34639	100A	43.1	0.435		0.00358
zfh1 (Zn-finger-homeodomain-1)	M63449	100A	41.2	0.461		0.00358
Gprk2 (G-protein-coupled-receptor-kinase-2)	M80494	100C	42.9	0.414		0.00352
awd (abnormal-wing-discs)	X13107	100C	27.9	0.776		0.00352
ttk (tramtrack)	Z11723	100D	49.1	0.359		0.00350
Ef1a100E (elongation-factor-1a100E)	X06869	100E	31.1	0.746	0.436	0.00349
Map205 (Microtubule-associated-protein-205)	X54061	100F	57.2	0.242		0.00347
mod (modulo)	X15702	100F	59.2	0.245		0.00347
Chromosome 4:						
ci (cubitus-interruptus)[g]	X54360	102A	51.8	0.165	0.147	0.00000
zfh2 (Zn-finger-homeodomain-2)	M63450	102C	59.1	0.210		0.00000
Cam (CAM-kinase)	M74583	102E	56.0	0.178		0.00000

NOTE.—Because the most recent sequences for several loci were derived from cDNAs, some intron sequences were taken from alternate GenBank sequences given in the Loci.text file of the Ashburner (1992) database. Genes located in regions near the centromeres are boxed.

[a] GenBank/EMBL accession no.

[b] Polytene-chromosome map position.

[c] Wright (1990).

[d] Sharp and Li (1987*b*).

[e] Intron G+C content. Does not include intron splice sequences, i.e., GT from the 5′ end or AG from the 3′ end.

[f] Recombination rate (see Material and Methods).

[g] Intron sequence provided by A. Berry (personal communication).

LITERATURE CITED

ASHBURNER, M. 1989. Drosophila: a laboratory handbook. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

———. 1992. Flybase: a Drosophila genetic database, version 9209 (available electronically from the ftp.bio.indiana.edu and EMBL file servers).

BEGUN, D. J., and C. F. AQUADRO. 1991. Molecular population genetics of the distal portion of the X chromosome in Drosophila: evidence for genetic hitchhiking of the *yellow-achaete* region. Genetics **129**:1147–1158.

———. 1992. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster.* Nature **356**:519–520.

BENNETZEN, J. L., and B. D. HALL. 1982. Codon selection in yeast. J. Biol. Chem. **257**:3026–3031.

BERRY, A. J., J. W. AJIOKA, and M. KREITMAN. 1991. Lack of polymorphism on the Drosophila fourth chromosome resulting from selection. Genetics **129**:1111–1117.

BRIDGES, C. B. 1935. Salivary chromosome maps with a key to the banding of the chromosomes of *Drosophila melanogaster.* J. Hered. **26**:60–64.

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

CHARLESWORTH, B., M. T. MORGAN, and D. CHARLESWORTH. 1993*a*. The effect of deleterious mutations on neutral molecular variation. Genetics. **134**:1289–1303.

CHARLESWORTH, D., M. T. MORGAN, and B. CHARLESWORTH. 1993*b*. Mutation accumulation in finite outbreeding and inbreeding population. Genet. Res. **61**:39–56.

DOBZHANSKY, T., F. J. AYALA, G. L. STEBBINS, and J. W. VALENTINE. 1977. Evolution. W. H. Freeman, San Francisco.

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

FITCH, D. H. A., and L. D. STRAUSBAUGH. 1993. Low codon bias and high rates of synonymous substitution in *Drosophila hydei* and *D. melanogaster* histone genes. Mol. Biol. Evol. **10**: 397–413.

GONZALES, F., S. ROMANI, P. CUBAS, J. MODOLELL, and S. CAMPUZANO. 1989. Molecular analysis of the *asense* gene, a member of the *achaete-scute* complex of *Drosophila melanogaster,* and its novel role in optic lobe development. EMBO J. **8**:3553–3562.

GOUY, M., and C. GAUTIER. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. **10**:7055–7074.

GRANTHAM, R., C. GAUTIER, M. GOUY, M. JACOBZONE, and R. MERCIER. 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9**:r43–r74.

GRANTHAM, R., C. GAUTIER, M. GOUY, R. MERCIER, and A. PAVE. 1980. Codon catalog usage and the genome hypothesis. Nucleic Acids Res. **8**:r49–r62.

HILL, W. G., and A. ROBERTSON. 1966. The effect of linkage on limits to artificial selection. Genet. Res. **8**:269–294.

HOCHMAN, B. 1976. The fourth chromosome of *Drosophila melanogaster.* Pp. 903–928 *in* M. ASHBURNER and E. NOVITSKI, eds. The genetics and biology of Drosophila. Academic Press, London.

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

LAI, Z., M. E. FORTINI, and G. M. RUBIN. 1991. The embryonic expression patterns of *zfh-1* and *zfh-2,* two *Drosophila* genes encoding novel zinc-finger homeodomain proteins. Mech. Dev. **34**:123–134.

LANGLEY, C. H., J. MACDONALD, N. MIYASHITA, and M. AGUADE. 1993. Lack of correlation between interspecific divergence and intraspecific polymorphism at the suppressor of forked region in *Drosophila melanogaster* and *Drosophila simulans.* Proc. Natl. Acad. Sci. USA **90**:1800–1803.

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

LINDSLEY, D. L., and L. SANDLER, 1977. The genetic analysis of meiosis in female *Drosophila melanogaster.* Philos. Trans. R. Soc. Lond. [B] **277**:295–312.

MAYNARD SMITH, J., and J. HAIGH. 1974. The hitch-hiking effect of a favorable gene. Genet. Res. **23**:23–35.

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

MULLER, H. J. 1964. The relation of recombination to mutational advance. Mutat. Res. **1**:2–9.

SHARP, P. M., and K. M. DEVINE. 1989. Codon usage and gene expression level in *Dictyostelium discoideum:* highly expressed genes do "prefer" optimal codons. Nucleic Acids Res. **17**: 5029–5039.

SHARP, P. M., and W.-H. LI. 1987*a*. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol. Biol. Evol. **4**:222–230.

———. 1987*b*. The codon adaptation index: a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15**:1281–1295.

SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS, and F. WRIGHT. 1988. "Silent" sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. **5**: 704–716.

SORSA, V. 1988. Chromosome maps of Drosophila. Vol. **2**. CRC, Boca Raton, Fla.

WHARTON, K. A., G. H. THOMSEN, and W. M. GELBART. 1991. *Drosophila* 60A gene, another transforming growth factor β family member, is closely related to human bone morphogenetic proteins. Proc. Natl. Acad. Sci. USA **88**:9214–9218.

WRIGHT, F. 1990. The "effective number of codons" used in a gene. Gene **87**:23–29.

PAUL M. SHARP, reviewing editor

Received February 17, 1993; revision received June 22, 1993

Accepted June 22, 1993