Intraspecific Nuclear DNA Variation in *Drosophila*

Etsuko N. Moriyama and Jeffrey R. Powell
Department of Biology, Yale University

We have summarized and analyzed all available nuclear DNA sequence polymorphism studies for three species of *Drosophila, D. melanogaster* (24 loci), *D. simulans* (12 loci), and *D. pseudoobscura* (5 loci). Our major findings are: (1) The average nucleotide heterozygosity ranges from about 0.4% to 2% depending upon species and function of the region, i.e., coding or noncoding. (2) Compared to *D. simulans* and *D. pseudoobscura* (which are about equally variable), *D. melanogaster* displays a low degree of DNA polymorphism. (3) Noncoding introns and 3′ and 5′ flanking DNA shows less polymorphism than silent sites within coding DNA. (4) X-linked genes are less variable than autosomal genes. (5) Transition (Ts) and transversion (Tv) polymorphisms are about equally frequent in noncoding DNA and at fourfold degenerate sites in coding DNA while Ts polymorphisms outnumber Tv polymorphisms by about 2:1 in total coding DNA. The increased Ts polymorphism in coding regions is likely due to the structure of the genetic code: silent changes are more often Ts's than are replacement substitutions. (6) The proportion of replacement polymorphisms is significantly higher in *D. melanogaster* than in *D. simulans.* (7) The level of variation in coding DNA and the adjacent noncoding DNA is significantly correlated indicating regional effects, most notably recombination. (8) Surprisingly, the level of polymorphism at silent coding sites in *D. melanogaster* is positively correlated with degree of codon usage bias. (9) Three proposed tests of the neutral theory of DNA polymorphisms have been performed on the data: Tajima's test, the HKA test, and the McDonald–Kreitman test. About half of the loci fail to conform to the expectations of neutral theory by one of the tests. We conclude that many variables are affecting levels of DNA polymorphism in *Drosophila,* from properties of nucleotides to population history and, perhaps, mating structure. No simple, all encompassing explanation satisfactorily accounts for the data.

Introduction

The "struggle" to measure genetic variation in populations has a long history, with *Drosophila* playing a leading role as a model system for study (reviewed in Lewontin 1974; Powell 1994). The ultimate view of genetic variation must be at the level of the DNA sequences, which are now accumulating at a rapid pace. Sufficient numbers of studies have been done on intraspecific nucleotide variation in *Drosophila* that we can begin to see the patterns emerging. Here we review and analyze the available data to highlight these patterns. Inasmuch as pattern infers process, we also discuss processes or mechanisms of molecular evolution.

The Data

While many studies using restriction fragment length polymorphisms have been performed on *Drosophila,* we confine ourselves here to nucleotide sequence data. Likewise we confine ourselves to studies of nuclear genes as relatively few DNA sequencing studies have been done on *Drosophila* mtDNA. Three species have been studied in sufficient detail to be useful: *Drosophila melanogaster, D. simulans,* and *D. pseudoobscura.* The Appendix lists the genes, their map positions, accession numbers to GenBank database, and references to the original publications. All sequences analyzed were taken from GenBank 86.0, December 1994 or directly from the original references. Gene names and map positions were from "FlyBase (1994)," The *Drosophila* Genetics Data Base. Because *D. pseudoobscura* has not been extensively mapped at the DNA level, only chromosomal locations are given for this species.

Tables 1, 2, and 3 summarize the intraspecific variation for *D. melanogaster, D. simulans,* and *D. pseudoobscura,* respectively. *Drosophila melanogaster* is the best studied species with 24 genes or gene regions which have been characterized for intraspecific DNA sequence polymorphism. Half the number of genes, 12, have been studied in *D. simulans. Drosophila pseudoobscura* has been studied for only five genes, but the sample sizes are very large for two of the genes.

For most analyses we broke the nucleotide data into three classes:

Total or total coding: Silent (synonymous) and replacement sites in protein-coding DNA
Silent or silent coding: Only silent substitutions in protein-coding DNA
Noncoding: Introns and 5′ and 3′ sequences adjacent to coding DNA

Because not all genes have been studied for all classes of DNA, the sample size varies from case to case.

Key words: *Drosophila,* DNA polymorphism, neutral theory, recombination, codon usage bias.

Address for correspondence and reprints: Jeffrey R. Powell, Department of Biology, Yale University, New Haven, Connecticut 06520-8104. E-mail: jeffrey.powell@yale.edu.

Mol. Biol. Evol. 13(1):261–277. 1996

Table 1
Polymorphisms of Nuclear Genes in *D. melanogaster*

		Coding Region										Noncoding Region (5′, Introns, 3′)					
								Nucleotide Diversity[c]									
			Polymorphic Sites					Total		Silent			Polymorphic Sites			Nucleotide Diversity[c]	
Gene	Allele #[a]	Length (bp)	Ts[b]	Tv[b]	Syn	Rep	Total	π	$\hat{\theta}$	π	$\hat{\theta}$	Length (bp)	Ts	Tv	Total	π	$\hat{\theta}$
ase	6	1,068[h]	5 (1)	1 (0)	3	3	6	2.06	2.46	5.07	5.57						
su(s)[d]	(50)											3,214	3	7	10	1.00	0.70
su(w^a)[d]	(50)											1,985	8	9	17	1.75	1.96
pn	8 (6)	1,173[h]	1 (0)	0 (0)	0	1	1	0.22	0.33	0.00	0.00	345	1	0	1	0.97	1.27
Pgd	13	1,443	3 (2)	1 (0)	3	1	4	1.33	0.89	4.18	2.87	2,900	8	5	13	1.59	1.47
z	6	804[h]	5 (1)	0 (0)	5	0	5	2.09	2.74	10.11	13.27	183	0	0	0	0.00	0.00
per	6	1,682[h]	14 (6)	6 (5)	19	1	20	4.95	5.21	20.64	21.50	186	5	4	9	17.63	21.54
Yp2	6	1,046[h]	6 (1)	3 (1)	7	2	9	4.46	3.77	15.17	12.94	68	2	0	2	15.92	13.07
Zw	33	1,558[h]	16 (7)	8 (4)	22	2	24	3.84	3.80	15.81	15.90	147	3	6	9	20.65	17.06
Total (X-linked)			50 (18)	19 (10)	59	10	69						30	31	61		
Average (X-linked)								2.71	2.74	10.14	10.29					7.44	7.13
Acp26Aa	10	792	8 (1)	8 (2)	5	11	16	7.35	7.14	10.71	10.34	584[i]	7	6	13	8.76	7.98
Acp26Ab	10	270	4 (0)	1 (0)	3	2	5	8.81	6.55	24.27	20.39						
Adh	15	768	11 (5)	8 (5)	16	3	19	8.11	7.61	28.51	25.90	1,210	17	42	60[j]	19.10	17.02
Adhr	11	816	4 (1)	2 (1)	3	3	6	1.34	2.51	2.08	5.89	845	6	7	13	4.43	5.46
Lcp1Psi[e]	10	384	0 (0)	1 (1)	0	1	1	1.22	0.93	0.00	0.00	184	0	1	1	2.67	2.02
Pgi	11	1,674	3 (1)	1 (1)	2	2	4	0.78	0.82	1.62	1.79	889	4	2	6	1.89	2.30
Amy-d	8 (5)	1,482	21 (6)	16 (10)	28	9	37	8.82	9.63	27.29	31.86	503	20	9	29	26.48	27.95
Amy-p	10 (6)	1,482	25 (7)	9 (6)	26	8	34	9.75	8.11	33.51	27.19	437	4	15	19	24.50	20.00
Sod	11	441[h]	5 (1)	2 (1)	5	2	7	4.37	5.42	14.99	16.41	969	29	21	52[j]	17.81	18.83
(CRS)[g]	25	441[h]	5 (1)	2 (1)	5	2	7	3.58	4.20	12.27	12.73	969	29	21	52[j]	11.56	14.60
Est-6	13 (12)	1,632	32 (11)	13 (4)	28	17	45	7.21	8.89	22.05	25.06	119	1	6	7	17.19	19.48
tra	11	588	1 (1)	0 (0)	1	0	1	0.74	0.58	3.19	2.51	342	0	1	1	0.54	1.01
Rh3	5	1,149	2 (1)	0 (0)	2	0	2	0.70	0.84	2.97	3.57						
boss	5	1,566[h]	11 (2)	5 (3)	14	2	16	4.86	4.90	19.09	18.88						
Mlc1	16	314[h]	0 (0)	0 (0)	0	0	0	0.00	0.00	0.00	0.00	656	10	9	19	11.79	9.00
ci	10	958[h]	0 (0)	0 (0)	0	0	0	0.00	0.00	0.00	0.00						
Total (autosomal)			127 (37)	65 (33)	133	59	192						98	119	220		
Average (autosomal)								4.43	4.41	13.40	13.29					11.72	11.53
Total (all)			177 (55)	84 (43)	192	69	261						128	150	281		
Average (all)[f]								4.02	4.03	13.46	13.51					10.82	10.52

[a] If numbers of alleles differ between coding and noncoding regions, the number for noncoding region is shown in parentheses.
[b] Transitional and transversional polymorphisms at fourfold degenerate sites are in parentheses.
[c] Nucleotide diversities were multiplied by 10^3. In/del sites were excluded from estimation ("Length" is the total length of alignment and includes in/del sites). No multiple-hit correction was done. The number of silent sites for each coding region was estimated by Nei and Gojobori (1986)'s method.
[d] From direct sequencing and single-strand conformation polymorphism analysis.
[e] Pseudogene, excluded from calculations of "Total" and "Average" of coding regions.
[f] Nucleotide diversity for X-linked genes are multiplied by ⅔ and added to those for autosomal genes.
[g] Constructed Random Sample, which includes 22 Fast alleles and 3 Slow alleles (Hudson et al. 1994).
[h] Partial sequence.
[i] Noncoding regions for both of *Acp26Aa* and *Acp26Ab* are included.
[j] One site includes both of Ts (transition) and Tv (transversion) polymorphisms.

Analyses

Measures of Variation

Two measures of nucleotide diversity have been commonly used in the literature (Tajima 1993). The first is designated π (or k), which is the observed average proportion of nucleotide differences between alleles sequenced. This was calculated as:

$$\pi = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_{ij} \Big/ \binom{n}{2} \qquad \binom{n}{2} = n(n-1)/2$$

where n is the number of alleles sampled, and k_{ij} is the number of nucleotide differences per site between alleles i and j. The second measure is designated $\hat{\theta}$ (or S/a_1) and is the average number of nucleotides segregating per site based on the expected distribution of neutral

Table 2
Polymorphisms of Nuclear Genes in *D. simulans*

		Coding Region										Noncoding Region (5′, Introns, 3′)					
								Nucleotide Diversity[c]									
			Polymorphic Sites					Total		Silent			Polymorphic Sites			Nucleotide Diversity[c]	
Gene	Allele #[a]	Length (bp)	Ts[b]	Tv[b]	Syn	Rep	Total	π	$\hat{\theta}$	π	$\hat{\theta}$	Length (bp)	Ts	Tv	Total	π	$\hat{\theta}$
ase	6	1,068[h]	0 (0)	0 (0)	0	0	0	0.00	0.00	0.00	0.00						
pn	4	1,170[h]	9 (2)	5 (1)	8	6	14	6.13	6.53	15.00	15.75	367	4	0	4	5.61	6.36
z	6	804[h]	11 (3)	0 (0)	11	0	11	6.09	6.04	29.51	29.38	183	5	2	7	16.09	16.84
per	6	1,679[h]	27 (11)	19 (12)	40	6	46	10.72	12.00	40.48	45.50	193	5	2	7	15.35	16.05
Yp2	6	1,046[h]	1 (0)	1 (1)	2	0	2	0.83	0.84	3.66	3.71	66	0	1	1	5.24	6.95
Zw	12	1,558[h]	9 (5)	4 (4)	13	0	13	3.47	2.76	15.65	12.48	147	2	1	3	13.74	8.42
Total (X-linked)			57 (21)	29 (18)	74	12	86						16	6	22		
Average (X-linked)								4.54	4.70	17.38	17.80					11.21	10.99
Adh	5	768	8 (6)	3 (3)	11	0	11	6.77	6.88	27.24	27.79	206	6	4	11[j]	27.07	27.22
Pgi	6	1,674	11 (7)	7 (5)	16	2	18	3.95	4.71	15.53	18.44	945	9	5	14	5.82	6.61
Est-6[d]	4	1,626	41 (14)	21 (13)	52	12	64[j]	22.24	21.47	79.65	79.45	1,362	27	31	60[j]	24.98	24.22
(5′ far)[d]	(3)											500	10	15	25	33.81	33.81
Rh3	5	1,149	21 (8)	8 (7)	30	0	30[j]	12.03	12.53	51.55	53.93						
boss	5	1,566[h]	27 (13)	12 (9)	38	2	40[j]	12.45	12.26	52.13	50.95						
ci	9	958[h]	0 (0)	1 (0)	0	1	1	0.23	0.38	0.00	0.00						
Total (autosomal)			108 (48)	52 (37)	147	17	164						52	55	110		
Average (autosomal)								9.61	9.71	37.68	38.43					22.92	22.99
Total (all)			165 (69)	81 (55)	221	29	250						68	61	132		
Average (all)[f]								7.83	7.99	30.43	31.08					18.93	18.77

Note.—Footnotes same as table 1 except for d below.

[d] Noncoding region of *Est-6* was divided into "5′ far distal segment" and the remainder (Karotam, Boyce, and Oakeshott 1995).

variants in a panmictic population at equilibrium. In practice $\hat{\theta}$ is estimated as:

$$\hat{\theta} = S \Big/ \sum_{i=1}^{n-1} (1/i)$$

where S is the number of segregating sites per site. If all the nucleotide variants are neutral and assuming panmixia and equilibrium, then: $\pi = \hat{\theta} = 4N_e\mu$ for autosomal genes, $3N_e\mu$ for sex-linked genes. In all the tables, both π and $\hat{\theta}$ have been multiplied by 10^3.

X-linked Versus Autosomal

In order to test whether X-linked and autosomal genes differ in level of variation we performed *t*-tests based on the variance estimates of π and $\hat{\theta}$ as given in Tajima (1993). Tajima provides variance estimates for both the assumption of free recombination and no recombination. We tested all possible combinations: assuming free recombination, assuming no recombination, removing two loci with particularly low recombination (*ase* and *ci*) and assuming free recombination for the rest of the genes. We tested whether $\pi_a - \pi_x$ is significantly different from zero, where π_a and π_x are the mean π for autosomal and X-linked genes respectively. We then multiplied π_x by ⅓, subtracted this from the autosomal diversity, and tested whether this was significantly different from zero. This latter test was to see if X-linked genes conformed to the predicted ¾ the amount of variation predicted if variation was a direct function of population size and there is a 1:1 sex ratio. We repeated all tests in a like manner with $\hat{\theta}$.

Codon Usage Bias

Various measures of codon usage bias have been proposed. Here we use ENC, the "effective number of codons," which is based on the evenness of use of alternative codons somewhat analogous to effective number of alleles (Wright 1990). This measure has the advantage of not requiring a database from which to infer the "optimal" codons as does the codon adaptation index, CAI (Sharp and Li 1987). For the *D. melanogaster* genes analyzed here, ENC and CAI have a correlation of -0.94 ($P < 0.001$). A third measure is χ^2/L (Shields et al. 1988); this measure has a correlation with ENC of -0.95 ($P < 0.001$) for the present data set. Thus all three statistics are measuring essentially the same thing, so the use of a single measure, ENC, captures the information in the data. It is important to bear in mind that the level of codon bias is *negatively* correlated to

Table 3
Polymorphisms of Nuclear Genes in *D. pseudoobscura*

		Coding Region										Noncoding Region (5′, Introns, 3′)					
								Nucleotide Diversity[c]									
			Polymorphic Sites					Total		Silent			Polymorphic Sites			Nucleotide Diversity[c]	
Gene	Allele #[a]	Length (bp)	Ts[b]	Tv[b]	Syn	Rep	Total	π	$\hat{\theta}$	π	$\hat{\theta}$	Length (bp)	Ts	Tv	Total	π	$\hat{\theta}$
Adh	107	762	26 (10)	13 (13)	38	1	39	4.03	9.76	16.18	38.54	1,280	81	91	187[j]	20.21	31.49
(N. America)	99	762	26 (10)	13 (13)	38	1	39	4.22	9.91	16.92	39.12	1,279	79	91	183[j]	19.93	31.26
(Bogotá)	8	762	3 (1)	0 (0)	3	0	3	1.22	1.52	4.92	6.15	1,213	25	28	54[j]	14.37	17.89
Adhr	107	834	50 (20)	29 (19)	73	10	83[j]	10.87	18.97	40.88	74.82	737	24	26	55[j]	10.00	19.82
(N. America)	99	834	50 (20)	28 (18)	71	10	81[j]	10.71	18.80	40.06	73.87	737	23	27	54[j]	10.12	19.76
(Bogotá)	8	834	6 (2)	10 (7)	13	3	16	8.31	7.40	30.05	26.96	685	9	4	13	7.76	7.56
Amy-1	(6)											1,058	22	25	49[j]	20.92	20.28
ry	7	4,026	65 (25)	45 (25)	84	30	114[j]	11.01	11.56	36.52	35.60	1,454	33	28	62[j]	17.02	17.86
Rh3	3	1,081[h]	3 (1)	4 (3)	7	0	7	4.32	4.32	18.90	18.97						
Total			144 (56)	91 (60)	202	41	243						160	170	353		
Average								7.56	11.15	28.12	41.98					17.04	22.36

Note.—Footnotes same as table 1.

ENC. ENC for the *D. melanogaster* and *D. simulans* genes is given in the Appendix.

Recombination Rate

We used the estimates of recombination in *D. melanogaster* given by Kliman and Hey (1993*b*; R. Kliman, personal communication, for genes not included in their publication). This method plots the recombination map position against the cumulative amount of DNA along the length of a chromosome. A least-squares polynomial curve-fitting procedures is used to generate a curve. The recombination rate is calculated as the derivative of the curve at the map position of the gene in question. This makes the assumption that recombination rate is approximately constant per length of DNA. This recombination rate is given in the Appendix; we have assumed equal recombination rates for homologous genes in *D. simulans.*

Tests of Neutrality

Three tests of the neutrality of molecular polymorphisms have been proposed for the type of data presented here. Tajima's (1989) test asks the question of whether $\hat{\theta}$ and π are significantly different. The statistic we present is *D,* which, under the assumption of a beta distribution, has a mean of 0 and variance of 1; whether *D* is significantly different from zero (the expectation if $\hat{\theta} = \pi$) was determined from the confidence intervals given in Table 2 of Tajima (1989).

The second test is known as the HKA test (Hudson, Kreitman, and Aguadé 1987). It is based on the prediction of neutral theory that the same forces and dynamics that control polymorphism on the intraspecific level should also control levels of interspecific divergence (Kimura 1983). This test requires polymorphism data from two or more genes or gene regions and an interspecific comparison for these genes. One of the genes or gene regions is used as a reference of the expectation of neutrality; usually a noncoding region is used, for example Hudson, Kreitman, and Aguadé (1987) used the 5′ flanking region of *Adh.* The sequence tested against this presumed neutral region is usually coding DNA. We used the modifications of Berry, Ajioka, and Kreitman (1991) to take into account differences in sample sizes, and those of Begun and Aquadro (1991) to compare autosomal and X-linked genes. The third test is the McDonald and Kreitman (1991) test. This test requires polymorphism data for only one locus in two species. The test asks whether the ratio of silent to replacement intraspecific polymorphisms is the same as the ratio for fixed differences between species. The test statistic is a 2 × 2 *G*-test.

Results

1. Levels of Variation

Species Averages

The overall nucleotide diversity varies among species (table 4). In calculating the overall averages (X-linked plus autosomal), the X-linked measures were multiplied by ⅔. As will be clear later, genes in regions of particularly low recombination have low levels of nucleotide variation, so in table 4 the overall averages and averages excluding low recombination genes are shown for *D. melanogaster* and *D. simulans. Drosophila pseudoobscura* is less well studied in this regard, and as far

Table 4
Mean Diversity Measures

	Coding Region					
	Total		Silent		Noncoding	
Species	π	$\hat{\theta}$	π	$\hat{\theta}$	π	$\hat{\theta}$
melanogaster (all) . .	4.02	4.03	13.46	13.51	10.82	10.52
Minus low rec . . .	4.27	4.24	14.54	14.55	12.17	11.81
simulans (all)	7.83	7.99	30.43	31.08	18.93	18.77
Minus low rec . . .	9.38	9.54	36.52	37.30	18.93	18.77
pseudoobscura	7.56	11.15	28.12	41.98	17.04	22.36

Note.—Both π and $\hat{\theta}$ are multiplied by 10^3. "Minus low rec" refers to excluding genes in regions of low recombination. For *D. melanogaster* and *D. simulans,* coding, this excludes *ase* and *ci*; for *melanogaster* noncoding, it also excludes *su(s)* and *su(w^a)*.

as is known, none of the genes studied are in regions of particularly low recombination. We speculate in the Discussion why there may be differences among species in regard to level of polymorphism.

One pattern evident in table 4 is that noncoding variation is always lower than silent variation in coding regions. This indicates that noncoding regions (here defined as introns and 5′ and 3′ regions adjacent to coding sequences) have more selective constraints than silent sites within coding regions. This is not surprising given the evidence that these noncoding regions are often involved in gene regulation.

Sex-linked Versus Autosomal Genes

If nucleotide polymorphism is entirely or predominantly neutral, then the level of diversity predicted is directly proportional to effective population size. X-linked genes have an effective population size ¾ that of autosomal genes (assuming a 1:1 sex ratio), so sex-linked genes should exhibit ¾ the level of neutral variation as for autosomal genes. In *D. melanogaster* and *D. simulans,* enough genes of each type have been studied to make the comparison. In a qualitative sense, the data from both species conform to predictions in that sex-linked genes are less variable on average; this is true for total coding region, silent sites, and for noncoding DNA (table 5). The data for *D. melanogaster* are close to the predicted 75%, although the *t*-tests (data not shown) indicated that the difference between autosomal and X-linked genes in this species is not statistically significant in this species. In *D. simulans,* X-linked genes average about half the variation of autosomal genes. In this species $\pi_a - \pi_x$ and $\pi_a - \frac{4}{3}\pi_x$ are significantly different from zero using total coding sequences and assuming no recombination ($t = 4.33$, $P < 0.001$; $t = 2.70$, $P < 0.01$, for the two cases, respectively); no other tests were significant. These assumptions are surely not realistic, and the lack of statistical significance with more realistic assumptions emphasizes the large variances associated with these estimates. Nevertheless, there is still good reason to believe that in *D. simulans* X-linked genes may be less variable than strict neutral theory predicts.

Table 5
Ratio of Nucleotide Diversity of X-linked Gene Regions to Autosomal

Region	*melanogaster*	*simulans*
Total coding	0.617 (0.589)	0.478 (0.481)
Silent coding	0.766 (0.768)	0.462 (0.462)
Noncoding	0.627 (0.797)	0.482 (0.482)

Note.—The mean of the ratio of π and $\hat{\theta}$ is shown. First number is for all loci; number in parentheses excludes genes in regions of particularly low recombination as identified in footnote to table 4.

2. Types of Polymorphisms

Transitions versus transversions

Table 6 presents the frequencies of transition (Ts) and transversion (Tv) polymorphisms for the three *Dro-*

Table 6
Summary of Transition (Ts) and Transversion (Tv) Polymorphisms for Different Species and for Different Classes of DNA

	Coding									Chi-squared (Fisher's Exact *P*)		
	Total			Four-fold			Non-coding					
	Ts	Tv	Tv's	Ts	Tv	Tv's	Ts	Tv	Tv's	T/4f	T/non	4f/non
melanogaster	177	84	32.2%	55	43	43.9%	128	150	54.0%	4.26 (0.05)	25.98 (<0.001)	2.95 (0.10)
simulans	165	81	32.9%	69	55	44.4%	68	61	47.3%	4.63 (0.04)	7.42 (0.007)	0.22 (0.71)
pseudoobscura . . .	144	91	38.7%	56	60	51.7%	160	170	51.5%	5.36 (0.02)	9.04 (0.003)	0.002 (0.99)

Note.—"Four-fold" indicates fourfold degenerate, i.e., any nucleotide can be in the third position. Contingency chi-squares are presented for total versus fourfold (T/4f), total versus noncoding (T/non), and fourfold versus noncoding (4f/non). Chi-squared analyses indicate that the species are homogeneous for all three classes.

Table 7
Silent and Replacement Polymorphisms

Species	No. Silent	No. Replacement	% Replacement ± SD
melanogaster			
Total	192	69	26.4 ± 2.7
In common with *simulans*	118	33	21.9 ± 3.4
In common with *pseudoobscura*	21	6	22.2 ± 8.0
simulans	221	29	11.6 ± 2.0
pseudoobscura	202	41	16.9 ± 2.4

NOTE.—SD assumes a binomial distribution.

sophila species, the sum of the individual genes in tables 1–3. In the total coding region of all three species, Ts's outnumber Tv's about 2:1, exactly the opposite ratio of that predicted by random mutation. This pattern is homogeneous across the three species. However, in the noncoding regions for two of the three species (*D. melanogaster* and *D. pseudoobscura*), Tv's outnumber Ts's, but only slightly, not near the 2:1 ratio expected for random changes. In *D. simulans* the ratio is in favor of Ts's in the noncoding region, although the sample size for this species is smaller than for the other two and is not statistically significantly different from them.

One possible cause of the preponderance of Ts's over Tv's in the coding region is that it is due to the structure of the genetic code; synonymous or silent substitutions are more often Ts's than Tv's. Stronger selection against replacement substitutions than against silent substitutions raises the frequency of Ts's relative to Tv's. For fourfold degenerate codons, selection based on amino acids should be blind. Table 6 presents the frequencies of Ts and Tv polymorphisms at these sites. As predicted, Tv polymorphisms are more common at fourfold degenerate sites for all three species and not significantly different from the ratio for noncoding DNA.

Silent Versus Replacement

The ratio of silent to replacement substitutions varies considerably among genes and species (table 7). Overall totals for the species indicate that *D. melanogaster* tends to have the highest proportion of replacement polymorphisms, 26.4%, and *D. simulans* the least, only 11.6%; these differences are significant ($\chi^2 = 18.13$, $P < 0.001$). However, as we just noted, there is extreme heterogeneity among genes, so a fairer comparison among species would be for the same genes studied in different species. When this is done for *D. melanogaster* and *D. simulans,* the proportion of replacement polymorphism in *D. melanogaster* drops to 21.4%, although still significantly different from *D. simulans* ($\chi = 7.09$ $P < 0.01$. Relatively few genes in common with the other species have been studied in *D. pseudoobscura,* so it is difficult to make any firm comparisons. For genes studied in common, *D. melanogaster* has a higher proportion of replacement polymorphisms than does *D. pseudoobscura,* although not statistically significant ($\chi^2 = 0.489$, $P > 0.10$).

3. Variables Correlated with Nucleotide Diversity

In trying to explain levels of nucleotide polymorphism, especially variation among loci, several factors have been considered. In this section we present correlations for several of these for *D. melanogaster* and *D. simulans,* the sample of genes in *D. pseudoobscura* being too small for any meaningful tests. The variables are listed in the Appendix where the value for each locus is given. In carrying out the correlations, corrections were made. For comparisons of diversity (π and $\hat{\theta}$), X-linked estimates were multiplied by $\frac{4}{3}$. Taking into account the lack of recombination in *Drosophila* males, the recombination rates of X-linked genes were multiplied by $\frac{2}{3}$ ($\frac{2}{3}$ of all X chromosomes are in recombining females, assuming a 1:1 sex ratio) and autosomal recombination rates were multiplied by ½ as only half are in females (Begun and Aquadro 1992).

Coding Versus Noncoding

The first entries in table 8 indicate the correlations in levels of polymorphism in coding and adjacent noncoding DNA. These are the strongest correlations found in these data. For both species the correlation is positive, being highly so for *D. melanogaster,* with the smaller sample of *D. simulans* genes precluding statistical significance for this latter species. Thus we can conclude there is good evidence in the best studied species that the level of polymorphism in coding and adjacent noncoding DNA is positively correlated, with similar patterns exhibited by *D. simulans.* This is evidence for regional factors controlling levels of polymorphism regardless of the functions of the DNA.

Codon Usage

Genes vary considerably in their degree of codon usage bias, presumably due to variation in some kind of selective constraint on codon usage. If variation in level of nucleotide polymorphism is due to variation in selective constraints, one would predict there to be a negative correlation between codon usage bias and the level of silent polymorphism in a given gene. Correlations of level of nucleotide diversity with codon usage bias present conflicting results (middle part of table 8). For *D. melanogaster,* the results are opposite that just predicted, namely that the level of nucleotide polymorphism is

Table 8
Summary of Correlations between Variables Relevant to Intraspecific Variation

	D. melanogaster			*D. simulans*		
VARIABLES	*n*	π	$\hat{\theta}$	*n*	π	$\hat{\theta}$
Total vs. noncoding	17	+0.804**	+0.808**	8	+0.582	+0.595
Silent vs. noncoding	17	+0.858**	+0.864**	8	+0.689	+0.695
Total vs. ENC	21	−0.307	−0.257	12	+0.150	+0.136
Silent vs. ENC	21	−0.491*	−0.450*	12	+0.068	+0.057
Noncoding vs. ENC	16	−0.600*	−0.530*	9	+0.240	+0.305
Total vs. recomb.	22	+0.542**	+0.499*	12	+0.539	+0.538
Silent vs. recomb.	22	+0.499*	+0.486*	12	+0.592*	+0.582*
Noncoding vs. recomb.	19	+0.536*	+0.529*	9	+0.118	+0.129
ENC vs. recomb. all data	20	−0.174		12	−0.319	
Remove two loci	18	−0.476*				

NOTE.—Numbers shown are the correlation coefficient, *r*. Significance of this is indicated at the 0.05 and 0.01 level with one and two asterisks. "*n*" is the number of pairwise comparisons. "Recomb" is estimated recombination rate. In the last entry for *D. melanogaster* the two anomalous loci, *Acp26Aa* and *Acp26Ab*, have been removed. See text for details.

greater in genes with greater codon usage bias! (Recall that codon usage bias is negatively correlated with ENC.) The effect is especially notable in silent site polymorphism, again opposite that predicted by the selective constraint theory. But perhaps most revealing, the noncoding nucleotide diversity is also significantly positively correlated with codon usage bias in adjacent coding regions. There is no reason to think that nonrandom selection of codons in coding regions should have any effect on polymorphism in noncoding adjacent DNA. However, this observation is consistent with that in the first part of table 8 showing a strong correlation between level of nucleotide variation in coding regions and the adjacent noncoding regions. This previous result indicated there was a regional factor(s) at work. The implication here is that whatever the regional factor(s) is, it is simultaneously affecting codon usage bias and level of variation in all classes of DNA, so the correlation of the latter two variables is due to each being correlated to a third variable.

In *D. simulans* the pattern is exactly opposite that of *D. melanogaster* and consistent with selective constraints affecting both codon usage and level of polymorphism. ENC is positively correlated with level of nucleotide variation (table 8). But again, if one wants to reach any causal relationship, it need be kept in mind that the highest correlations exist for noncoding DNA. However, none of the correlations between codon bias and nucleotide polymorphisms are statistically significant at the traditional 5% level, so there may be nothing in the *D. simulans* data that requires "explaining away."

Recombination

Recombination is known to affect levels of DNA polymorphisms in *Drosophila* (e.g., Berry, Ajioka, and Kreitman 1991; Begun and Aquadro 1992; Aguadé and Langley 1994). Genes residing in regions of the genome exhibiting low levels of recombination have less polymorphism than genes in regions with higher levels of recombination. In *D. melanogaster,* recombination is low near centromeres and telomeres and is absent in the fourth dot chromosome. The Appendix presents the recombination rates for the genes studied for polymorphisms using the method of Kliman and Hey (1993b; see Analyses section). Rows 6–8 in table 8 present the correlations of this variable with levels of nucleotide polymorphism. As detected by others, the data show a positive correlation between recombination rate and level of polymorphism. This holds whether one considers total coding, silent, or noncoding polymorphism in *D. melanogaster.* Thus, as expected, recombination rate seems to have a regional effect that affects level of polymorphism regardless of the function of the DNA in the region. Perhaps due to a smaller sample of genes, in *D. simulans* the statistically significant positive correlation only holds for silent polymorphisms, although all classes of DNA have positive correlations.

Recombination and Codon Usage

Given the correlations of both codon usage and recombination with levels of polymorphism in *D. melanogaster,* it is of interest to determine if the former two variables are correlated with one another. The bottom of table 8 shows the results. For *D. melanogaster,* when all data are used, there is a negative correlation between level of codon bias and recombination but it is not significant at the 5% level. There are two data points that appear to be outliers in that they have very high levels of recombination and exhibit very low levels of codon usage bias; these are the closely linked *Acp26Aa* and

Table 9
Tajima's (1989) Test for Significant Differences between π and $\hat{\theta}$

	D. melanogaster			*D. simulans*		
	Coding			Coding		
GENE	Total	Silent	Noncoding	Total	Silent	Noncoding
ase	−0.93	−0.43		0	0	
su(s)			1.22			
su(w^a)			−0.35			
pn	−1.04	0	−0.93	−0.62	−0.48	1.11**
Pgd	1.67	1.44	0.34			
z	−1.34	−1.33	0	0.06	0.03	−0.260
per	−0.30	−0.25	−1.08	−0.68	−0.70	−0.25
Yp2	1.10	1.00	1.03	−0.05	−0.07	−0.96
Zw	0.04	−0.02	0.64	1.09	1.08	2.07*
Acp26Aa	0.14	0.14	0.45			
Acp26Ab	1.38	0.68				
Adh	0.27	0.40	0.53	−0.11	−0.14	−0.04
Adhr	−1.85*	−2.21**	−0.83			
Lcp1Psi	0.82	0	0.82			
Pgi	−0.15	−0.30	−0.71	−1.01	−0.98	−0.73
Amy-d	−0.45	−0.76	−0.39			
Amy-p	0.97	1.11	1.40			
Sod (CRS)	−0.46	−0.10	−0.80			
Est-6	−0.84	−0.52	−0.46	0.38	0.03	0.32
tra	0.67	0.66	−1.13			
Rh3	−0.97	−0.98		−0.30	−0.33	
boss	−0.07	0.08		0.12	0.17	
Mlc	0	0	1.24			
ci	0	0		−1.09	0	

NOTE.—Value shown is the D with the associated P * <0.05, ** <0.01.

Acp26Ab (see Appendix). When these two points are removed, the negative correlation becomes significant. For *D. simulans,* the correlation is likewise negative, but the small sample size does not allow any strong statistical conclusion, except to note the direction in correlation is the same as for *D. melanogaster.* We can conclude there is good evidence that recombination and codon usage are correlated, an observation made previously by Kliman and Hey (1993*b*).

Table 10
Tajima's (1989) Test on *D. pseudoobscura* Genes

		CODING REGION		
GENE		Total	Silent	NONCODING
Adh	All data	−1.82*	−1.79*	−1.19
	N. America	−1.79*	−1.77*	−1.21
	Bogotá	−0.81	−0.82	−1.06
Adhr	All data	−1.39	−1.46	−1.57
	N. America	−1.41	−1.49	−1.56
	Bogotá	0.63	−0.58	0.13
Amy-1				0.20
ry		−0.28	0.15	−0.27
Rh3		0	NA[a]	

[a] NA: Not applicable due to small sample size.
* P < 0.05.

4. Tests of Neutrality

Tajima's Test

Tables 9 and 10 present the results of Tajima's tests for all available data for the three species. In the case of *D. melanogaster,* it can only be concluded that the data are remarkably consistent in indicating no significant difference between π and $\hat{\theta}$ and thus, by this criterion, the data are consistent with neutral theory. Only two tests (out of 58) are significant, and both are for the coding region of the *Adhr* gene and thus the two data sets are not independent. *Adhr* is the *Adh*-related gene tightly linked to *Adh,* sometimes also called *Adh(dup).* Why this gene stands out is not clear. One might postulate it is because it is a duplication of *Adh* that has not yet reached equilibrium; however, this duplication is present in several other Sophophoran species so it is unlikely to be a young duplication. It may simply be chance that one out of 24 genes deviates at a significance level less than 5%. Similarly for *D. simulans,* 2 tests out of 30 indicate a significant difference, both occurring in noncoding regions of X-linked genes.

Drosophila pseudoobscura, on the other hand, exhibits a more interesting pattern with regard to Tajima's test (table 10). *Adh* and *Adhr* in this species are the best-

Table 11
HKA Tests on *D. melanogaster* Genes

	Noncoding Reference Locus																
Test Locus	5′	2	3	4	5	6	7	8	10	11	12	13	14	15	16	17	20
1. *ase*	—	—	*	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2. *pn*	—	—	—	—	*	—	**	*	*	—	—	*	*	*	*	—	*
	—	—	—	NA	*	*	**	*	*	—	—	*	*	*	*	—	*
3. *Pgd*	—	—	—	—	—	—	***	*	**	—	—	**	***	*	***	—	*
	—	—	—	—	—	—	***	*	*	—	—	**	***	*	***	—	*
4. *z*	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	*	—	—	—	—	—	—	—	—	—	—	—	—	—	—
5. *per*	—	—	***	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	***	—	—	—	—	—	—	—	—	—	—	—	—	—	—
6. *Yp2*	—	—	****	—	—	—	—	—	—	—	—	—	—	—	—	*	—
	—	—	****	—	—	—	—	—	—	—	*	—	—	—	—	*	—
7. *Zw*	—	—	**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	****	—	—	—	—	—	—	—	—	—	—	—	—	*	—
8. *Acp26Aa* ..	—	—	—	—	***	—	***	*	*	—	—	***	****	*	***	—	*
	—	—	—	—	—	—	—	—	—	—	—	—	*	—	—	—	—
9. *Acp26Ab* ..	—	—	***	—	—	—	—	—	—	—	—	—	—	—	—	*	—
	—	—	***	*	—	—	—	—	—	—	*	—	—	—	—	*	—
10. *Adh*	*	—	****	*	—	—	—	—	—	—	***	—	—	—	—	**	—
	*	—	****	*	—	—	—	—	—	—	***	—	—	—	—	**	—
11. *Adhr*	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	—	—	*	—	*	—	*	—	—	*	*	—	*	—	—
12. *Pgi*	—	—	—	—	*	—	*	—	*	—	—	*	*	—	*	—	—
	—	—	—	—	*	—	**	*	*	—	—	*	**	*	*	—	*
13. *Amy-d* ...	—	—	****	—	—	—	—	—	—	—	*	—	—	—	—	*	—
	—	—	****	—	—	—	—	—	—	—	*	—	—	—	—	*	—
14. *Amy-p* ...	—	—	****	—	—	—	—	—	—	—	*	—	—	—	—	*	—
	—	—	****	—	—	—	—	—	—	—	*	—	—	—	—	*	—
15. *Sod*	—	—	**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	*	—	—	—	—	—	—	—	—	—	—	—	—	—	—
16. *Est-6*	—	—	***	—	—	—	—	—	—	—	—	—	—	—	—	*	—
	—	—	***	—	—	—	—	—	—	—	—	—	—	—	—	—	—
17. *tra*	*	—	—	—	****	**	****	***	***	*	—	***	****	**	****	—	***
	—	—	—	—	*	—	**	*	*	—	—	*	**	*	**	—	*
18. *Rh3*	—	—	—	—	*	—	*	—	—	—	—	—	*	—	*	—	—
	—	—	—	—	—	—	*	—	—	—	—	—	*	—	—	—	—
19. *boss*	—	—	**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	—	—	**	—	—	—	—	—	—	—	—	—	—	—	—	—	—
20. *Mlc1*	—	—	—	NA	—	—	*	—	*	—	—	*	*	—	—	—	—
	—	—	—	NA	—	—	—	—	—	—	—	—	*	—	—	—	—
21. *ci*	*	*	—	NA	****	****	****	***	***	**	*	***	****	***	****	—	***
	*	—	—	NA	***	*	***	*	**	*	—	**	***	*	***	—	*

Note.—Upper entry in each cell is for total coding, and lower entry for silent only. The 5′ column is for the *Adh* 5′ region, the original region used by Hudson, Kreitman, and Aguadé (1987). Other numbers across the top refer to numbers next to genes in the first column. In some cases when there were no polymorphisms, the test could not be performed; these are indicated by the NA entries. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.005$, **** $P < 0.001$.

studied genes for intraspecific nucleotide polymorphism for all *Drosophila*. *Adh* exhibits a significant difference between $\hat{\theta}$ and π, the former being larger producing a negative *D*. This is consistent with a pattern of there being too many rare nucleotide polymorphisms with respect to predictions of the neutral theory (see e.g., Braverman et al. 1995). The difference is only evident for North American populations and does not appear to hold for the isolated Bogotá population. The closely linked *Adhr* and the genes on the other chromosomes do not indicate any differences between $\hat{\theta}$ and π, so it is unlikely that the cause of the deviation for N. American

Adh is due to some overall populational factors, which would be expected to affect all genes equally. Furthermore, unlike the case with *D. melanogaster* and *D. simulans,* many fewer genes have been studied in *D. pseudoobscura,* so many fewer tests are possible; therefore the significant deviation in one of the data sets is less likely to be a statistical artifact.

The HKA Test

This test combines information on interspecific divergence with level of intraspecific polymorphism. Given the relatively large numbers of genes studied for polymorphism and the large number of potential reference sequences against which to test them by the HKA test, the number of possible tests is very large. We have performed about 1,000 HKA tests on the data in tables 1 and 2. Rather than present the full tables of all tests, for many purposes it is sufficient to see the patterns in the results.

First we need to make some comments on the reference loci, the region assumed to be neutral against which one tests other regions. Hudson, Kreitman, and Aguadé (1987) first used the 5′ *Adh* noncoding region; this is the standard reference. We tested all 16 other noncoding regions from *D. melanogaster* using this standard and, with the single exception of *Pgd,* none displayed a statistically significant difference with the HKA test (data not shown). For *D. simulans,* we performed all possible pairwise HKA test for the eight noncoding regions studied in this species; none was significant (data not shown). With the single exception noted (*Pgd* in *D. melanogaster*), these results indicate it is reasonable to use the noncoding regions as references for HKA tests; they do seem to be behaving homogeneously as expected if they were all neutral. (However, as we noted in table 4, these noncoding regions are less variable than silent sites in coding regions, so there may be some selective constraints, although they would appear not to be affected sufficiently differently as to be detected by the HKA test.)

The significance of the HKA tests for *D. melanogaster* are presented in table 11. Because there are more than 700 tests, it is difficult to determine if any one asterisk in a cell has any biological significance. Some patterns, however, are very clear. For example, *Pgd* stands out as an unusual locus that does not conform to the neutral expectations of the HKA test. Whether used as the reference locus or test locus (i.e., both the column and row for *Pgd),* there are many significant deviations from neutrality; but note that the coding region for this locus does not fail the test when its own flanking DNA is used as the standard. This result is consistent with the fact that this was the only noncoding region that gave a

Table 12
HKA Test on *D. simulans* Genes

Gene	Noncoding Reference							
	2	3	4	5	6	7	8	9
1. *ase* ...	*	*	*	—	*	**	—	*
	—	—	—	—	—	*	—	—
2. *pn*	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
3. *z*	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
4. *per* ...	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
5. *Yp2* ...	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
6. *Zw*	—	—	—	—	—	*	—	—
	—	—	—	—	—	—	—	—
7. *Adh* ...	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
8. *Pgi* ...	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
9. *Est-6* ..	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
10. *Rh3* ..	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
11. *boss* ..	—	—	—	—	—	—	—	—
	—	—	—	—	—	—	—	—
12. *ci*	****	****	***	—	***	****	*	**
	***	***	**	*	**	***	*	**

Note.—Constructed the same as table 11.

significant HKA test when compared against the standard 5′ *Adh* noncoding region. Another locus clearly failing the HKA expectations of neutrality is the nonrecombining *ci* locus, the last row (no noncoding sequence is available to use this region as a reference). Other loci that tend to stand out are *pn, Acp26Aa,* and *tra. Adh* also tends to show significant deviations as already detected by Hudson, Kreitman, and Aguadé (1987); however, at that time they used only the 5′ noncoding *Adh* region as a reference. Here we can see that *Adh* deviates significantly from expectations when 5 other references are used, but not when the remaining 12 are used. Finally we note for this species that the one locus which stood out as significantly deviating from neutral expectations in the Tajima (1989) test, *Adhr* (see table 9), conforms reasonably well to neutral expectations with the HKA test, although several significant deviations are noted for silent sites for some reference loci.

The results of the HKA test on *D. simulans* are more clear (table 12). Only two loci stand out and account for almost all the significant deviations from neutral expectations: *ci* and *ase,* the first and last rows. Both of these genes are in regions of unusually low recombination. This is almost certainly related to the obser-

Table 13
Results of McDonald and Kreitman Test for *D. melanogaster* and *D. simulans*

	Fixed[a]		Polymorphic[a]								
			melanogaster			*simulans*			Total		
Gene	Silent	Rep.	Silent	Rep.	*G*	Silent	Rep.	*G*	Silent	Rep.	*G*
ase	13	11	3	3	0.03	0	0	FE	3	3	0.03
pn	22	5	0	1	FE	8	6	2.69	8	7	3.65
z	15	2	5	0	FE	11	0	FE	16	0	FE
per	25	1	19	1	0.04	40	6	1.83	58[b]	7	1.28
Yp2	11	5	8	1	1.40	2	0	FE	10	1	2.03
Zw	23	25	22	2	15.04***	12	1	9.72**	33[b]	3	19.82***
Adh	2	2	16	3	1.97	11	0	FE	27	3	3.35
Pgi	23	1	2	2	5.21*	16	2	0.74	18	4	2.45
Est-6 ...	19	16	30	15	1.27	50	14	5.95*	78[b]	27[b]	4.73*
Rh3	14	3	2	0	FE	30	0	FE*	32	0	FE*
boss	16	3	14	2	0.08	38	2	1.79	52	4	1.13
ci	23	21	0	0	FE	0	1	FE	0	1	FE

Note.—A *G*-test is used except when a zero appears in a cell; those cases are denoted FE for Fisher's exact *P* tests. * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0001$.

[a] Rep. = replacement polymorphisms.

[b] Both species have polymorphisms in the same site.

vations made by Begun and Aquadro (1992) that regions of low recombination show much less intraspecific polymorphism but "normal" degrees of interspecific divergence. However, *ase* in *D. melanogaster* does not show any signs of deviation from neutral expectations by the HKA test (table 11).

The McDonald–Kreitman Test

Table 13 presents the results of all the McDonald–Kreitman tests. Three loci produce *G*-tests with associated *P* values less than 5%. *Zw* deviates most from neutral expectation as noted by Eanes, Kirchner, and Yoons (1993). *Est-6* and *Rh3* deviate sufficiently to produce *P*'s just under 5%; *Pgi* deviates if only *D. melanogaster* polymorphisms are used. Interestingly, *Adh* does not significantly deviate when *D. melanogaster* and *D. simulans* are the test species, whereas McDonald and Kreitman's (1991) original use of this test was between *D. melanogaster* and *D. yakuba* for which this locus did significantly deviate from neutral expectation. It should be noted that *D. melanogaster* is more closely related to *D. simulans* than to *D. yakuba* (Caccone, Amato, and Powell 1988; Lachaise et al. 1988) so problems of saturation are less likely for the two closer related species used for the tests in table 13.

Discussion

Overall Variation

The overall estimated level of per nucleotide heterozygosity is 0.4%–2% depending upon region and species (table 4); silent site polymorphism can reach 4%. This amount of variation is sufficiently large that virtually *every diploid individual is heterozygous at every locus,* with the exception of genes in regions of unusually low recombination and in closely inbred individuals. This high level of variation exceeds that predicted by even the strongest proponents of the "balance" view of populations with the possible exception of B. Wallace who concluded in 1959 that "we feel that the proportion of heterozygosis among gene loci of representative individuals of a population tends toward 100 per cent." Of course whether this degree of variation, or what proportion of it, has any significance with regard to adaptive evolution remains an open question.

It appears from table 4 that, among these three species, *D. melanogaster* is the least variable and *D. pseudoobscura* the most variable. This is not unexpected given the history of these species. *Drosophila melanogaster* and *D. simulans* are cosmopolitan species associated with human habitats. Almost certainly they originated in sub-Saharan Africa, became adapted to manmade environments, and were spread around the world by human transport (David and Capy 1988). Nearly all the populations sampled for genetic diversity in these species have been human-associated and outside Africa. Thus it is likely these populations have some degree of founder effect and have not reached population genetic equilibrium. (For empirical evidence see Begun and Aquadro [1993], who did sample native *D. melanogaster* populations from Zimbabwe and found them to be more variable on the nucleotide level than those sampled in the New World.) *Drosophila pseudoobscura,* on the

other hand, is native to oak-pine forests of the western one-third of North America and can still be found only in this region. Thus all the samples of this species came from its native habitat, or very close to it, and such populations would be expected to have little founder effect evident and are likely closer to equilibrium, i.e., they have been stable large populations for a long time. If population size and stability are positively correlated to level of polymorphism (as population genetics theory predicts), then the level of nucleotide diversity difference between *D. melanogaster* and *D. pseudoobscura* is understandable. *Drosophila simulans* is thought to have a history similar to that of *D. melanogaster,* yet its nucleotide diversity is on a par with that of *D. pseudoobscura.* This implies *D. simulans* may have been subject to less severe founder effects and perhaps maintains larger, more stable populations compared to its sibling, *D. melanogaster.* Others (e.g., Aquadro 1992; Akashi 1995) have reached similar conclusions.

However, there is one anomalous finding at odds with this conclusion. The Tajima test did not detect much difference between π and $\hat{\theta}$ in *D. melanogaster* populations (table 9). The latter is expected to equal the former only in a panmictic equilibrium population. Populations that recently went through a bottleneck would be expected to have an excess of rare variant, which would produce significantly negative *D*'s by this test (Tajima 1989; Braverman et al. 1995). Negative and positive *D*'s in table 9 are about equally frequent, 30 and 26, respectively.

Sex-linked Versus Autosomal

Sex-linked genes are less variable than autosomal genes (table 5) although in the case of *D. melanogaster,* the differences are not statistically significant. In *D. simulans* the ratio of diversity in sex-linked genes to autosomal genes is even less than the predicted 0.75, being somewhat less than 0.5 for all classes of DNA. This has one important implication for understanding the forces controlling nucleotide variation. One of the major hypotheses to account for the reduction of polymorphism in regions of low recombination is the "background selection" model of Charlesworth, Morgan, and Charlesworth (1993; Charlesworth 1994). This model is based on the selective removal of deleterious recessive alleles along with linked neutral variation; the strength or distance over which such selective elimination will act is dependent upon recombination. Under such a model the effective population becomes essentially the frequency of chromosomes free of deleterious mutations linked sufficiently close to the studied locus as to affect it (Hudson 1994; Stephan 1994). Because of the hemizygosity of X chromosomes in males, one would expect, on average, more efficient selection against deleterious recessives and thus the fraction of X chromosomes free of deleterious loci to be higher than for autosomal chromosomes. The pattern observed, especially in *D. simulans,* is opposite the prediction of the background selection model.

An alternative model is the positive selective sweep model (Berry, Ajioka, and Kreitman 1991; Begun and Aquadro 1992; Hudson 1994). In this model, an advantageous mutation is hypothesized to sweep to fixation taking with it linked neutral variants. If most new advantageous mutants are recessive, then the positive selection should again be more effective for X-linked genes due to hemizygosity in males. This could account for the reduced heterozygosity for X-linked genes. One argument against the selective sweep model is the observed frequency distribution of variants; following a selective sweep, most new mutations would be expected to be in low frequency relative to the expected distribution of a population at equilibrium. Braverman et al. (1995) show that the data so far collected do not conform to this expectation.

What then could account for the lower diversity of X-linked genes in *D. simulans*? One possibility is the mating structure of the species (see, e.g., Charlesworth 1994). Specifically, if males have a higher effective population than females, this could cause the effective population size of X-linked genes to be less than the 75% of autosomal genes predicted from a 1:1 sex ratio. This could come about in *Drosophila* due to the prevalence of multiple insemination in many species. Males could be inseminating many females, but not all females may find a suitable larval substrate on which to deposit eggs. Because each female is carrying more than one male's sperm, her eggs represent contributions to the next generation of only a single female but multiple males. On the other hand, it must be noted that in the only direct measurements of sex-specific effective population sizes in laboratory populations of *Drosophila,* the general finding is that females have a larger effective population size than do males (e.g., Crow and Morton 1955).

Types of Polymorphisms

With regard to Ts's and Tv's, species and genes appear to be quite homogeneous. Generally in noncoding regions Tv's outnumber Ts's by a small amount (table 6), whereas in the coding region Ts's outnumber Tv's by about 2:1. The fact that fourfold degenerate sites have a ratio of Tv:Ts very similar to noncoding DNA indicates that the structure of the genetic code and selection against replacement polymorphisms can account for the preponderance of Ts's in the coding region. This assumes that the Tv:Ts patterns at fourfold degenerate sites and noncoding DNA represent the neutral mutation process. However, there is evidence of more selective

constraints on noncoding regions compared to silent sites (table 4); evidently these constraints do not change the ratio of Tv:Ts in these regions.

With regard to silent and replacement polymorphisms, perhaps the most notable aspect of the data is the great variation from locus to locus in their relative occurrence, although this could have been anticipated from allozyme data, which indicated considerable variation among loci with regard to electrophoretically detectable variation that must be due to replacement polymorphisms. While silent polymorphisms generally outnumber replacements (table 7), there are exceptions, e.g., *Acp26Aa* in *D. melanogaster* (table 1). *Adh* in *D. pseudoobscura* is the alternative extreme case with a single replacement polymorphism detected in more than 85 kb of sequence (107 alleles × 762 bp/allele; table 3). Whether this variation is due to differences in selective constraints or positive selection for polymorphism is not clear. The one replacement polymorphism in *D. melanogaster Adh,* which causes the fast/slow allozyme polymorphism, is very likely a case of positive selection (Kreitman 1991).

One surprise in considering the relative frequencies of replacement polymorphism is the difference between *D. melanogaster* and *D. simulans.* The former species has a significantly greater proportion of replacement polymorphism even when only genes studied in both species are considered (table 4). This is consistent with allozyme data which indicate *D. melanogaster* and *D. simulans* are equally variable (Choudhary and Singh 1987); since *D. melanogaster* has less overall nucleotide polymorphisms, a greater proportion must be replacements for allozyme variation to be equal. Why *D. melanogaster* should be more tolerant of replacement polymorphisms than is *D. simulans* remains unknown. It could be related to the observation that *D. simulans* is overall more variable at the nucleotide level than is *D. melanogaster* (table 4). If *D. simulans* has larger more stable populations than *D. melanogaster,* then it is expected that natural selection will be more effective in *D. simulans.* This implies that much of the replacement polymorphism in *D. melanogaster* is slightly deleterious but that selection is relatively ineffective in eliminating the mutations due to small and unstable populations. Akashi (1995) also detected reduced efficacy of selection against synonymous substitutions in *D. melanogaster* relative to *D. simulans.* This is consistent with a general populational cause (such as small size), which would affect all types of substitutions.

Recombination Effects

It is clear from table 8 that there are regional factors operating which similarly affect coding and adjacent noncoding DNA. Recombination rate was explored as possibly being a regional factor involved and there is indication it is positively correlated to level of variation in both *D. melanogaster* and *D. simulans.* Whether this is the only regional factor affecting variation is not clear. The correlations between nucleotide diversity and recombination have r values around 0.5, which means the amount of variance accounted for (r^2) is about 25%. Either recombination is not the only factor or the error in estimating both diversity and recombination rate account for the lack of a stronger correlation. (Another potential regional effect is mutation; however, there is no evidence that mutation rates or processes vary along *Drosophila* genomes [e.g., Woodruff, Slatko, and Thompson 1983; Moriyama and Hartl 1993].)

It is especially interesting to note the lack of a negative correlation between codon usage bias and level of nucleotide variation at silent sites; in fact, for *D. melanogaster* there is a positive correlation (table 8). Again, recombination may be the key in understanding this. The level of codon usage bias is positively correlated with recombination (table 8; see also Kliman and Hey [1993*b*]) presumably because selection can be more effective at single nucleotide positions when recombination is high, the so-called Hill–Robertson effect (Hill and Robertson 1966). Higher recombination also increases variation by reducing the hitchhiking effect of selective elimination of linked loci. So recombination positively affects level of variation and codon usage bias. Evidently, the selection for codon usage bias is not strong enough to offset the increase in variation due to reduction in hitchhiking. This is in strong contrast to what is seen in interspecific comparisons: more strongly biased genes show less divergence between species (Sharp and Li 1989; Moriyama and Powell, in preparation) but recombination rate does not correlate with interspecific divergence (Begun and Aquadro 1992). This seemingly contradictory pattern for intraspecific and interspecific nucleotide variation is lacking an adequate explanation.

Neutrality or Selection?

Are the patterns of DNA polymorphisms in *Drosophila* more consistent with selection playing a major role or does the hypothesis of selective neutrality adequately account for the data? The data are not decisive on this question.

Tajima's (1989) test rarely rejects neutrality. However, it may be that this test is sensitive to the sample size (number of alleles) as the test is strongest in rejecting neutrality for the largest data set, *Adh* from *D. pseudoobscura.* Perhaps the inability of Tajima's test to detect much if any selection in the data from *D. melanogaster* and *D. simulans* is due to the smaller sample sizes: average number of alleles studied per locus is 11

and 6, respectively, in contrast to the 99 alleles of *Adh* and *Adhr* for N. American *D. pseudoobscura.* It is clearly of interest to investigate theoretically the effect of sample size on the sensitivity of the Tajima test. The HKA test more often detects deviation from neutral expectation. For *D. melanogaster* (table 11), 7/21 loci clearly deviate from neutrality: *pn, Pgd, Acp26Aa, Adh, Pgi, tra,* and *ci.* For these loci too many asterisks appear in their rows than would be expected by chance. For *D. simulans* (table 12) only 2/12 loci show clear deviations with both loci (*ase* and *ci*) in regions of particularly low recombination. However, the genes failing the HKA test in *D. melanogaster* do not particularly cluster in regions of low recombination and one gene, *ase,* in such a region seems to conform well to neutral expectation as detected by the HKA test (table 11). The McDonald and Kreitman (1991) test detects deviations from expectation as does the HKA test. One quarter (3/12) of the loci tested in table 13 gave indications of deviation from neutral expectations. Of particular interest is the fact that these three loci, *Zw, Est-6,* and *Rh3,* gave no indication of deviations in either the HKA or Tajima tests. The set of loci failing the HKA test is mutually exclusive of the set that fails the McDonald/Kreitman test.

For *D. melanogaster* then, we can conclude that 11/24 loci failed one of the tests for neutrality. Seven failed the HKA test, one failed the Tajima test, and three contributed to the significant deviations in the McDonald–Kreitman test. *Drosophila simulans* had 5/12 loci fail one test. Can we then conclude that just less than half the genes in these species are evolving neutrally and half are subject to selection? This would be a very naive conclusion to draw from this analysis. The analyses presented in this paper emphasize the complexity of interpreting intraspecific DNA polymorphism data. The evidence for regional effects, at least partly mediated through recombination, is very clear. The fact that the selective constraints on codon usage seem not to affect the level of polymorphism at silent sites is perhaps the best example of how misleading it can be to consider too few variables in understanding these data. The correlation of level of codon bias in a coding region is highest with adjacent noncoding DNA. If we had not considered recombination as another variable, this pattern would have been completely mysterious.

Conclusions

We are particularly struck by two patterns in the data analyzed here. First is the clear evidence of regional factors affecting levels of nucleotide polymorphisms. Such forces seem to override other forces such as selection for codon usage. Second, the contrast between *D. melanogaster* and *D. simulans* in their patterns of nucleotide diversity is striking. They differ in level of overall diversity, in the proportion of replacement versus silent polymorphisms, and in the ratio of diversity of X-linked genes to autosomal genes. These two species are very closely related, differing by a single paracentric inversion; furthermore, ecologically the two species are very similar. One would think that any genetic processes such as mutation, recombination, etc., would be very similar in the two species. So explanations for the contrast in nucleotide diversity must be sought at a different level. We have speculated that size and stability of populations and the mating structure of the species could account for the differences. These observations emphasize the fact that in evaluating patterns of nucleotide diversity we need to consider several levels of organization: single nucleotides, codons, chromosomal regions, populations, and mating behavior. No single simple explanation seems adequate.

Acknowledgments

We thank Richard Kliman for supplying the missing recombination data, John Hartigan for statistical advice, and Gail Simmons for supplying details of her data. Brian Charlesworth made useful comments on an early version of this manuscript. This work was supported by NSF grant DEB 9318836.

This article is dedicated to the memory of Motoo Kimura without whom the field of evolutionary genetics would have been much less interesting for the last 30 years.

APPENDIX
Genes Used in This Paper[a]

GENE[b]	SPECIES	MAP[b]	ENC[c]	REC. RATE[d]	GENBANK ACC#	REFERENCE	DIVERGENCE BETWEEN *melanogaster* AND *simulans*[e]		
							Total	Silent	Non-coding
ase	*melanogaster*	1–0.0	57.4	0.00		Hilton, Kliman, and Hey 1995	0.025	0.059	
	simulans		57.8			Hilton, Kliman, and Hey 1995			
su(s)[f]	*melanogaster*	1–0.0		0.00		Aguadé et al. 1994			
su(w^a*)*[f]	*melanogaster*	1–0.1		0.00		Aguadé et al. 1994			
pn	*melanogaster*	1–[0.5]	54.4	2.02		Simmons et al. 1994	0.026	0.089	0.032
						G. Simmons, personal communicaton			
	simulans		53.5			Simmons et al. 1994			
						G. Simmons, personal communicaton			
Pgd	*melanogaster*	1–0.6	37.2	1.87		Begun and Aquadro 1994	0.038	0.136	0.060
z	*melanogaster*	1–1.0	42.8	2.74	L13043–8	Hey and Kliman 1993	0.032	0.142	0.077
	simulans		42.1		L13049–53, 55	Hey and Kliman 1993			
per	*melanogaster*	1–1.4	37.6	2.94	L07817–9, 21, 23, 25	Kliman and Hey 1993*a*	0.031	0.129	0.088
	simulans		36.5		L07826, 28–32	Kliman and Hey 1993*a*			
Yp2	*melanogaster*	1–30	33.7	2.88	L14421–3	Hey and Kliman 1993	0.020	0.061	0.121
	simulans		33.7		L14426–8	Hey and Kliman 1993			
Zw	*melanogaster*	1–62.9	32.7	2.31	L13880, 85–90, 95–920	Eanes, Kirchner, and Yoon 1993	0.037	0.096	0.068
	simulans		31.4		L13876–9, 81–4, 91–4	Eanes, Kirchner, and Yoon 1993			
Acp26Aa	*melanogaster*	2–[20]	61.0	4.97	X70888–97	Aguadé, Miyashita, and Langley 1992	0.140	0.157	0.036
Acp26Ab	*melanogaster*	2–[20]	52.2	4.97	X70888–97	Aguadé, Miyashita, and Langley 1992	0.030	0.058	
Adh	*melanogaster*	2–50.1	31.4	1.98	M17827, 28, 30–7	Lauric, Bridgham, and Choudhary 1991	0.014	0.047	0.059
	simulans		30.3		M19263, X57361–4	McDonald and Kreitman 1991			
	pseudoobscura	4			M60979–96, X68159–66, X62181–238, X64468–89, Y00602	Schaeffer and Miller 1992, 1993			
Adhr	*melanogaster*	2–50.1	59.6	1.98		Kreitman and Hudson 1991	0.033	0.132	0.039
	pseudoobscura	4			M60979–96, X68159–66, X62181–238, X64468–89, Y00602	Schaeffer and Miller 1992, 1993			
Lcp1Psi[g]	*melanogaster*	2–[58]		0.94	U17196–205	Pritchard and Schaeffer[h]			
Pgi	*melanogaster*	2–58.6	36.7	0.92	L27539–46, 53–5	McDonald and Kreitman[h]	0.016	0.068	0.044
	simulans		35.2		L27547–52	McDonald and Kreitman[h]			
Amy-d	*melanogaster*	2–77.9	28.4	3.38	L22717, 18, 20, 24, 28, 30, 32, 34	Inomata et al. 1995	0.032	0.097	0.074
Amy-p	*melanogaster*	2–77.9	28.4	3.38	L22716, 19, 21, 25–7, 29, 31, 33	Inomata et al. 1995	0.028	0.086	0.046
Amy-1	*pseudoobscura*	3			U09746–57	Popadic and Anderson 1994			
Sod	*melanogaster*	3–32.5	36.3	3.19		Hudson et al. 1994	0.027	0.106	0.065
Est-6	*melanogaster*	3–35.9	54.6	2.72	J04167	Cooke and Oakeshott 1989	0.044	0.129	0.076
	simulans		52.9		L34263–5, 70	Karotam, Boyce, and Oakeshott 1995			
tra	*melanogaster*	3–45	55.5	1.06	M17478, M19618–20, M19464–70	Walthour and Schaeffer 1994	0.069	0.162	0.054
Rh3	*melanogaster*	3–[67]	42.0	3.38		Ayala, Chang, and Hartl 1993	0.024	0.093	
	simulans		42.7			Ayala, Chang, and Hartl 1993			
	pseudoobscura	2				Ayala, Chang, and Hartl 1993			
hoss	*melanogaster*	3–90.5	49.2	3.88		Ayala and Hartl 1993	0.027	0.106	
	simulans		48.6			Ayala and Hartl 1993			
Mlc1	*melanogaster*	3–[98]	39.2	3.81	L37312–27	Leicht et al. 1995	0.019	0.082	0.041
ci	*melanogaster*	4–0.0	50.1	0.00		Berry, Ajioka, and Kreitman 1991	0.046	0.111	
	simulans		49.9			Berry, Ajioka, and Kreitman 1991			
ry	*pseudoobscura*	2				Riley, Kaplan, and Veuille 1992			

[a] Sequences were taken from GenBank 86.0 (December 1994), the daily update, or original references.

[b] Gene names and maps were referred from "FlyBase (1994). The *Drosophila* Genetic Database. Available from the ftp.bio.indiana.edu network server and Gopher site."

[c] Effective number of codons by Wright (1990).

[d] Recombination rate $\times 10^{-3}$ (Kliman and Hey [1993*b*] and personal communication from R. Kliman). For *Pgi*, the value is the average for the bands 44A–F.

[e] Numbers of nucleotide substitutions per site between *D. melanogaster* and *D. simulans*. When there are multiple sequences, one of each sequence is chosen arbitrarily. Numbers of silent substitutions are estimated by Nei and Gojobori's (1986) method without multiple-hit correction.

[f] Noncoding regions only from direct sequencing and single-strand conformation polymorphism analysis.

[g] Pseudogene.

[h] Names of submitters to GenBank, but no publication found.

LITERATURE CITED

AGUADÉ, M., and C. H. LANGLEY. 1994. Polymorphism and divergence in regions of low recombination in *Drosophila*. Pp. 67–76 *in* B. GOLDING, ed. Non-neutral evolution: theories and molecular data. Chapman and Hall, New York.

AGUADÉ, M., W. MEYERS, A. D. LONG, and C. H. LANGLEY. 1994. Single-strand conformation polymorphism analysis coupled with stratified DNA sequencing reveals reduced sequence variation in the *su(s)* and *su(w^a)* regions of the *Drosophila melanogaster* X chromosome. Proc. Natl. Acad. Sci. USA **91**:4658–4662.

AGUADÉ, M., N. MIYASHITA, and C. H. LANGLEY. 1992. Polymorphism and divergence in the *Mst26A* male accessory gland gene region in *Drosophila*. Genetics **132**:755–770.

AQUADRO, C. F. 1992 Why is the genome variable? Insights from *Drosophila*. Trends Genet. **8**:355–362.

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

AYALA, F. J., B. S. W. CHANG, and D. L. HARTL. 1993. Molecular evolution of the *Rh3* gene in *Drosophila*. Genetica **92**:23–32.

AYALA, F. J., and D. L. HARTL. 1993. Molecular drift of the *bride-of-sevenless* (*boss*) gene in *Drosophila*. Mol. Biol. Evol. **10**:1030–1040.

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.

———. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365**:548–550.

———. 1994. Evolutionary inferences from DNA variation at the 6-phosphogluconate dehydrogenase locus in natural populations of *Drosophila*: selection and geographic differentiation. Genetics **136**:155–171.

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.

BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY, and W. STEPHAN. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140**:783–796.

CACCONE, A., G. D. AMATO, and J. R. POWELL. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics **118**:671–683.

CHARLESWORTH, B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. **63**:213–227.

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

CHOUDHARY, M., and R. S. SINGH. 1987. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. III. Variations in genetic structure and their causes between *Drosophila melanogaster* and its sibling species *Drosophila simulans*. Genetics **117**:697–710.

COOKE, P. H., and J. G. OAKESHOTT. 1989. Amino acid polymorphisms for esterase-6 in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **86**:1426–1430.

CROW, J. F., and N. E. MORTON. 1955. Measurement of gene frequency drift in small populations. Evolution **9**:202–214.

DAVID, J. R., and P. CAPY. 1988. Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4**: 106–111.

EANES, W. F., M. KIRCHNER, and J. YOON. 1993. Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. Proc. Natl. Acad. Sci. USA **90**:7475–7479.

HEY, J., and R. M. KLIMAN. 1993. Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. **10**:804–822.

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

HILTON, H., R. M. KLIMAN, and J. HEY. 1995. Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. Evolution **48**:1900–1913.

HUDSON, R. R. 1994. How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? Proc. Natl. Acad. Sci. USA **91**:6815–6818.

HUDSON, R. R., K. BAILEY, D. SKARECKY, J. KWIATOWSKI, and F. J. AYALA. 1994. Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. Genetics **136**:1329–1340.

HUDSON, R. R., M. KREITMAN, and M. AGUADÉ. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153–159.

INOMATA, N., H. SHIBARA, E. OKUYAMA, and T. YAMAZAKI. 1995. Evolutionary relationships and sequence variation of α-amylase variants encoded by duplicated genes in the *Amy* locus of *Drosophila melanogaster*. Genetics **141**:237–244.

KAROTAM, J., T. M. BOYCE, and J. G. OAKESHOTT. 1995. Nucleotide variation at the hypervariable esterase 6 isozyme locus of *Drosophila simulans*. Mol. Biol. Evol. **12**:113–122.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KLIMAN, R. M., and J. HEY. 1993*a*. DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. Genetics **133**:375–387.

———. 1993*b*. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10**:1239–1258.

KREITMAN, M. 1991. Detecting selection at the level of DNA. Pp. 204–221 *in* R. SELANDER, A. CLARK, and T. WHITTAM, eds. Evolution at the molecular level. Sinauer, Sunderland, Mass.

KREITMAN, M., and R. R. HUDSON. 1991. Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127**:565–582.

LACHAISE, D., M.-L. CARIOU, J. R. DAVID, F. LEMEUNIER, L. TSACAS, and M. ASHBURNER. 1988. Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22**:159–225.

LAURIE, C. C., J. T. BRIDGHAM, and M. CHOUDHARY. 1991. Association between DNA sequence variation and variation in expression of the *Adh* gene in natural populations of *Drosophila melanogaster.* Genetics **129**:489–499.

LEWONTIN, R. C. 1974. The genetic basis of evolutionary change. Columbia University Press, New York.

LEICHT, B. G., S. V. MUSE, M. HANCZYC, and A. G. CLARK. 1995. Constraints on intron evolution in the gene encoding the myosin alkali light chain in *Drosophila.* Genetics **139**: 299–308.

MCDONALD, J. H., and M. KREITMAN. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila.* Nature **351**:652–654.

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

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

POPADIC, A., and W. W. ANDERSON. 1994. The history of a genetic system. Proc. Natl. Acad. Sci. USA **91**:6819–6823.

POWELL, J. R. 1994. Molecular techniques in population genetics: a brief history. Pp. 31–156 *in* B. SHIERWATER, B. STREIT, G. WAGNER, and R. DESALLE, eds. Molecular ecology and evolution: approaches and applications. Birkhauser Verlag, Basel.

RILEY, M. A., S. R. KAPLAN, and M. VEUILLE. 1992. Nucleotide polymorphism at the xanthine dehydrogenase locus in *Drosophila pseudoobscura.* Mol. Biol. Evol. **9**:56–69.

SCHAEFFER, S. W., and E. L. MILLER. 1992. Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. Genetics **132**:471–480.

———. 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.

SHARP, P. M., and W.-H. LI. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias and its potential applications. Nucl. Acids Res. **15**: 1281–1295.

———. 1989. On the rate of DNA sequence evolution in *Drosophila.* J. Mol. Evol. **28**:398–402.

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.

SIMMONS, G. M., W. KWOK, P. MATULONIS, and T. VENKATESH. 1994 Polymorphism and divergence at the *prune* locus in *Drosophila melanogaster* and *D. simulans.* Mol. Biol. Evol. **11**:666–671.

STEPHAN, W. 1994. Effects of genetic recombination and population subdivision on nucleotide sequence variation in *Drosophila ananassae.* Pp. 57–66 *in* B. GOLDING, ed. Non-neutral evolution, Chapman Hall, New York.

TAJIMA, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**: 585–595.

———. 1993. Measurement of DNA polymorphism. Pp. 37–59 *in* N. TAKAHATA and A. G. CLARK, eds. Mechanisms of molecular evolution. Sinauer, Sunderland, Mass.

WALLACE, B. 1959. The role of heterozygosity in *Drosophila* populations. Proc. 10th Int. Congr. Genetics **1**:408–419.

WALTHOUR, C. S., and S. W. SCHAEFFER. 1994. Molecular population genetics of sex determination genes: The *transformer* gene of *Drosophila melanogaster.* Genetics **136**:1367–1372.

WOODRUFF, R. C., B. E. SLATKO, and J. N. THOMPSON, JR. 1983. Factors affecting mutation rates in natural populations. Pp. 37–124 *in* M. ASHBURNER, H. L. CARSON, and J. N. THOMPSON, Jr, eds. The genetics and biology of *Drosophila,* Vol. 3c. Academic Press, New York.

WRIGHT, F. 1990. The 'effective number of codons' used in a gene. Gene **87**:23–39.

CHARLES F. AQUADRO, reviewing editor

Accepted August 18, 1995