Unraveling Selection in the Mitochondrial Genome of Drosophila

J. William O. Ballard*,† and Martin Kreitman*

**Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, and †CSIRO Division of Entomology, Canberra, ACT 2601, Australia*

Manuscript received April 8, 1994
Accepted for publication July 25, 1994

ABSTRACT

We examine mitochondrial DNA variation at the cytochrome *b* locus within and between three species of *Drosophila* to determine whether patterns of variation conform to the predictions of neutral molecular evolution. The entire 1137-bp cytochrome *b* locus was sequenced in 16 lines of *Drosophila melanogaster*, 18 lines of *Drosophila simulans* and 13 lines of *Drosophila yakuba*. Patterns of variation depart from neutrality by several test criteria. Analysis of the evolutionary clock hypothesis shows unequal rates of change along *D. simulans* lineages. A comparison within and between species of the ratio of amino acid replacement change to synonymous change reveals a relative excess of amino acid replacement polymorphism compared to the neutral prediction, suggestive of slightly deleterious or diversifying selection. There is evidence for excess homozygosity in our world wide sample of *D. melanogaster* and *D. simulans* alleles, as well as a reduction in the number of segregating sites in *D. simulans*, indicative of selective sweeps. Furthermore, a test of neutrality for codon usage shows the direction of mutations at third positions differs among different topological regions of the gene tree. The analyses indicate that molecular variation and evolution of mtDNA are governed by many of the same selective forces that have been shown to govern nuclear genome evolution and suggest caution be taken in the use of mtDNA as a "neutral" molecular marker.

PATTERNS of genetic variation within and between species can provide insights into the processes influencing molecular evolution and this information can be used to distinguish between adaptive and neutral evolutionary change. Mitochondrial genes have been used extensively in evolutionary studies because of their uniparental mode of inheritance and high rate of evolution (BROWN *et al.* 1979, 1982; AQUADRO *et al.* 1984; IRWIN *et al.* 1991). Many studies rely on the assumption that the mitochondrial genome (mtDNA) evolves neutrally, but few studies have been carried out to test this assumption. The extent of hitchhiking of neutral mutations in response to selection on another part of the genome depends on the recombination distance from the site under selection (MAYNARD-SMITH and HAIGH 1974; KAPLAN *et al.* 1989). In the extreme case of no recombination, such as in mitochondrial DNA, the whole genome is a single completely linked entity. Thus the selective fixation of any mutation in the mtDNA will lead to the concomitant fixation of all variants in that genome, and the population can only regain polymorphism by the accumulation of new mutations. Similarly, selection acting to maintain multiple alleles in the population, such as diversifying selection, will lead to their divergence.

WHITTAM *et al.* (1986) examined human mtDNA restriction fragment length polymorphisms (RFLP) frequencies for the conformity to the theoretical equilibrium neutral distribution, under an infinite alleles model. In 35 comparisons, 71% of allele frequency distributions fell within the range predicted by the neutral model; Excesses in the frequencies of common alleles and in the number of singleton alleles were observed at 9 loci. Of these, the cytochrome *b* locus showed the greatest departure from neutrality. EXCOFFIER (1990) found that the mitochondrial RFLPs of African populations conformed well to the theoretical frequency distribution but several Oriental and Caucasoid populations had high frequencies of particular haplotypes, exceeding the neutral prediction. Predominance of one mitochondrial genotype might indicate selection, population expansion or population subdivision. ROGERS and HARPENDING (1992) studied the distribution of pairwise nucleotide differences for human mitochondrial data and found that the distribution does not conform to an equilibrium neutral model. Simulations of alternative models indicated that the results fit equally well with either a rapid expansion in population size or a population bottleneck. A bottleneck could well have been the result of a selective sweep of the mitochondrial genome rather than an actual population size reduction. In these studies, the lack of agreement with neutral equilibrium models can be explained by a variety of processes including selection. MARJORAM and DONNELLY (1994) have recently argued for the inclusion of selection in population genetics models of mtDNA.

Evidence for selection in mitochondria comes from a recent study in mice (NACHMAN *et al.* 1994). The authors sequenced the NADH dehydrogenase subunit 3 (*ND3*) gene from 56 wild individuals of *Mus domesticus* and

compared polymorphism within species to divergence between *M. domesticus*, *Mus musculus* and *Mus spretus*. Synonymous changes exceeded replacement changes between species, but were approximately equal within *M. domesticus*. The difference was statistically significant by Fisher's exact test, indicating incompatibility with a strictly neutral model of mtDNA evolution. The authors presented two alternative hypotheses, maintainance of protein variation by diversifying selection or reduction in the rate of protein evolution by selection against slightly deleterious mutations.

To investigate selection on mtDNA we present a population genetic analysis of sequence variation of cytochrome *b* genes from naturally derived lines of *Drosophila melanogaster*, *Drosophila simulans* and *Drosophila yakuba*. The cytochrome *b* gene was chosen because it has been used extensively to infer phylogenetic relationships (MEYER and WILSON 1990; IRWIN *et al.* 1991; MARTIN and PALUMBI 1993). Moreover, mutational and evolutionary studies of the cytochrome *b* gene have facilitated the development of a structure/function model and the identification of the sites of electron transfer and inhibitor action (HOWELL and GILBERT 1988; HOWELL 1989; DI RAGO *et al.* 1990). RFLP analysis of *D. melanogaster* and *D. yakuba* mtDNA suggests moderate variability (HALE and SINGH 1987, 1991; MONNEROT *et al.* 1990). Three mitochondrial haplotypes, *si*I, -II and -III (BABA-AÏSSA and SOLIGNAC 1984; SOLIGNAC *et al.* 1986; BABA-AÏSSA *et al.* 1988) have been identified in *D. simulans*. The *si*II haplotype has a worldwide distribution, whereas the *si*I haplotype is known to occur only in the Seychelles, New Caledonia, Polynesia, Hawaii and Madagascar. The *si*III haplotype has been found only in Madagascar (SOLIGNAC and MONNEROT 1986).

The assumption that mitochondrial DNA evolves as a neutral marker has been adopted more for convenience than as a logical deduction from experimental tests designed to test the hypothesis. Here, we test the neutral hypothesis by investigating explicit and testable predictions of the theory (KIMURA 1983). First, we investigate variability in the rate of change along mitochondrial lineages. We employ a simple statistical test of the molecular evolutionary clock hypothesis (TAJIMA 1993). Second, we investigate the patterns of variation within species and compare them to the patterns between species, utilizing tests of MCDONALD and KREITMAN (1991), HUDSON *et al.* (1987), WATTERSON (1978) and EWENS (1973). Third, we investigate the patterns of substitution across structural regions of the gene and across topological levels of the phylogenetic tree. CLAREY and WOLSTENHOLME (1984) noted a high bias for A- and T-ending codons in *D. yakuba*. We devise a test to determine whether this pattern has a selective basis. Our analysis is concluded by testing for variability of mutation rates among nonconserved sites.

TABLE 1

Primers for the amplification and sequencing of the cytochrome b and cytochrome oxidase I (COI) loci of Drosophila

Locus	5′ 3′
Cytochrome *b*	
10068+a	TCACCCATTAGCTTTAGGAT
10493+as	CTAAACTATTTAAAGGACCT
10493Y+as	.....T...
10780+s	TTACACGCTAACGGTGCATC
11143+s	ACAGGATCTAATAACCCTAT
11400+s	AGGAGGAGTTATTGCATTAG
11400S+s	.....C..C.....
10766−s	AAAAATGATGCACCGTTAGC
11125−s	TATAGGGTTATTAGATCCTG
11382−s	ACTAATGCAATAACTCCTCC
11382S−s	.....G..G.....
11664−as	CATACGCTTGTTCAAGCTCA
11842−a	TTACCTCGGTTTCGTTATGA
COI	
1377+a	GCAGTTTGATATCATTATTG
1653+s	TAATTGTTACTGCACATGCT
1945−s	GTAATAAAATTTACAGCTCC
2439−a	GAGTTCCATGTAAAGTAGCT

The numbering of primers corresponds to the 3′ position in the *D. yakuba* mtDNA (CLAREY and WOLSTENHOLME 1985). Y, *D. yakuba* specific primer; S, *D. simulans* specific primer; a, amplification primer; s, sequencing primer; and ±, mtDNA strands.

MATERIALS AND METHODS

Stocks: The 16 *D. melanogaster* lines represent a worldwide collection of isofemale lines consisting of three lines from Trinidad, West Indies (Tr); two from Kenya (Ke); two from Lawrenceville, New Jersey (Te); two from Luangwa Valley, Northern Zambia (La); one from Akavanga River, Botswana (O); two from Valparaiso, Indiana (Va); two from Belleville, Illinois (Be) and two from Zimbabwe, Africa (Zm). The published cytochrome *b* sequence of *D. melanogaster* was taken from GARESSE (1988). The initial sample of 16 *D. simulans* lines is also from a worldwide collection of isofemale lines consisting of two lines from Valparaiso, Indiana (Va); five from Mpala Ranch, Kenya (Mp); two from Northern Australia (Au); two from Lantana, Florida (La); two from Trinidad, West Indies (Tr) and three from California (one R and two W strains of HOFFMANN *et al.* 1986). The 13 *D. yakuba* lines were taken from one collection in Brazzaville, Congo. The published cytochrome *b* sequence of *D. yakuba* was taken from CLAREY and WOLSTENHOLME (1985).

Preliminary analysis indicated that our worldwide sample of 16 lines of *D. simulans* included only the *si*II haplotype. To further investigate mitochondrial variation in *D. simulans* we obtained a representative line of both the *si*I and -III haplotypes. The *si*I line was from Hawaii; and the *si*III line was from Mont d'Ambre (Madagascar).

DNA preparation and polymerase chain reaction (PCR): DNA from individual flies was prepared using Chelex 100 (WALSH *et al.* 1991). Fragments 1.2–1.6 kb and 0.8 kb in length were PCR-amplified from cytochrome *b* and cytochrome oxidase I (COI), respectively, using 20-mer oligonucleotide primers (Table 1). Amplification was carried out in 100-μl reaction volumes with two units of *Taq* polymerase under standard conditions (SAIKI *et al.* 1988) and a 2 mM magnesium concentration. Following amplification the product was precipitated with 1 volume of ethanol and 0.5 volume of 7.5 M NH_4OAc, washed with 1 ml of 70% ethanol and dried. The DNA was separated electrophoretically on a 1% TBE agarose gel and the desired product was purified using the Quiagen gel purifica-

tion kit. This product then served as a stock for reamplification. The DNA was reamplified, as described above, and purified with the Promega Magic Preps DNA purification system. The concentration of the amplicon was estimated by comparison with a DNA ladder of known concentration.

Sequencing and alignment: Both strands were sequenced using Taq-DyeDeoxy Terminator Cycle sequencing (Applied Biosystems) employing 150–200 ng of the amplicon (500–1200 bp) and 80 ng of primer. For cytochrome *b* three internal primers and the amplification primer were used in each direction (Table 1). For COI one internal primer and the amplification primers were used in each direction (Table 1). Following cycle sequencing the reactions were cleaned with two phenol/water/chloroform extractions, precipitated with 15 μl of 2 M NaOAc and 300 μl of 100% ethanol and then washed with 1 ml of 70% ethanol. The dried sample was resuspended in 3 μl of deionized formamide and 50 mM EDTA (5:1 ratio) and electrophoresed on an Applied Biosystems 373A DNA sequencer. Sequences were imported into the Applied Biosystems SeqEd v1.03 software program, the chromatograms investigated and the sequences aligned. Because of the possibility of heteroplasmy and variation in the chromatogram peak heights associated with Taq-DyeDeoxy Terminator Cycle sequencing, we confirmed any equivocal sites by direct PCR sequencing using the λ-exonuclease technique of HIGUCHI and OCHMAN (1989), as described by BERRY *et al.* (1991), or by cloning into pGEM (Promega). These additional sequences confirmed that none of the individuals employed in this study were heteroplasmic.

Phylogenetic analysis: The genealogical relationship of alleles was analyzed by parsimony using PAUP 3.1 (SWOFFORD 1993) and McClade (MADDISON and MADDISON 1992). Permutation tail probability (PTP) testing was employed to investigate phylogenetic structure (FAITH and CRANSTON 1991). Topological permutation tail probability (T-PTP) testing (FAITH 1991) and bootstrapping (EFRON 1982; FELSENSTEIN 1985) were used to test monophyly. PTP tests a null model in which the original number of characters and their character states are maintained, but for each character, the states are randomly reassigned to the taxa. The cladistic PTP is defined as the estimate of the proportion of times that a tree can be found as short or shorter than the original tree. Here the T-PTP is the estimate of the proportion of times that an *a priori* monophyletic assemblage can be found as short or shorter than the original assemblage. The PTP and T-PTP testing of assemblages was based on 999 randomized data sets, with the conventional 95% confidence level used as significant support for a given hypothesis (PTP and T-PTP less than or equal to 0.05). The bootstrap is a method of sampling the original dataset with replacement to construct a series of bootstrap estimates of the same size. Each of these is analyzed and the variation among these replicate estimates is taken as an indication of the error involved in making estimates from the original data. For this study 1,000 pseudosamples were generated to estimate the bootstrap proportions. The strict consensus was generated from all equally parsimonious trees.

To investigate the hypothesis of equal evolutionary rates among lineages we used the 2D method of TAJIMA (1993). The 2D test is appropriate when the outgroup is known. We employ *D. yakuba* as the outgroup to *D. melanogaster* and *D. simulans*. The method is based on the chi-square test and is applicable when the pattern of substitution rates is unknown and/or the substitution rates vary among different sites.

Tests of neutrality: To investigate protein evolution we compared the number of amino acid replacement substitutions to synonymous substitutions in cytochrome *b* (MCDONALD and KREITMAN 1991). The test of neutrality is based on the prediction that the ratio of replacement to synonymous fixed differences between species should be the same as the ratio of replacement to synonymous polymorphisms within species. We applied Williams' correction prior to calculation of the *G* statistic (SOKAL and ROHLF 1981). For *D. simulans* we use the 16 *si*II lines because they represent the initial worldwide "random" sample. Inclusion of the *si*I and the *si*III lines would bias our results because they were known, *a priori*, to be distinct mitochondrial haplotypes.

To compare the levels of variation of mitochondrial and autosomal genes we modified the segregating sites (Seg) HKA test, Equation 5 of HUDSON *et al.* (1987), and the pairwise difference test (Pwd) of KREITMAN and HUDSON (1991) so that the effective population size of mitochondrial genes was one-quarter that of autosomal genes. Similarly, the population size of sex-linked genes is assumed to be three-quarters that of autosomal genes. The cytochrome *b* data were compared with the following nuclear genes (relevant references are given in Table 5 legend): *D. melanogaster* cytochrome *b* was compared to 18 *Adh* sequences, 11 *Adh* 5′-flanking region sequences and six sequences of the *period* locus. All of the *D. melanogaster* loci used in the HKA tests, both mitochondrial and nuclear, consist of sequences obtained from worldwide collections of flies. The *D. simulans* nuclear gene data consist of six *Adh* and six *period* locus sequences from worldwide samples. None of these lines were collected from the Pacific Islands or Madagascar and, hence all the lines from the worldwide samples are likely to be of the *si*II mitochondrial haplotype. The 13 sequences of *D. yakuba Adh*, several of which were polymorphic, were resolved into 19 alleles by the method of CLARK (1990). The HKA test is based on the assumption that each sample is taken from a single random mating population. The departure of the samples from this assumption is not expected to bias the test toward finding greater (or lesser) variation at any one locus. Nevertheless, this assumption may affect the validity of the test results.

We also examined haplotype (allele) frequencies for departure from neutrality (WATTERSON 1978). The WATTERSON test calculates the probability of observing a sample homozygosity equal to or greater than the observed sample homozygosity under an infinite allele neutral model with no recombination, an appropriate model for mitochondrial DNA. The test would be inappropriate for samples taken from subdivided populations. The WATTERSON test for homozygosity can be applied to synonymous and/or replacement substitutions. We present combined probability values for synonymous and replacement substitutions and for synonymous substitutions alone. We also employ a "singletons" test which calculates the probability of observing as many or more singleton alleles compared to that expected under the infinite alleles, no recombination neutral model (EWENS 1973). In both tests, exact probabilities are calculated for each dataset using EWENS' (1973) sampling formula.

Distribution of mutations in cytochrome *b*: There is a strong bias toward A- and T-ending codons in the 13 protein coding genes of *D. yakuba* (CLAREY and WOLSTENHOLME 1984). We have constructed a statistical test of neutral molecular evolution to investigate whether the biased distribution is governed only by mutation and genetic drift, or whether selection favors A- and T-ending codons relative to C- and G-ending codons. This test is based on a method for investigating the type of selection maintaining codon bias in nuclear genes (H. AKASHI, personal communication). Using the strict consensus tree (see RESULTS) we parsimoniously polarized all transitional changes within and between *D. melanogaster* and *D. simulans*. Ambiguous sites were not included in the analysis. Synonymous changes were categorized as either mutating to C

TABLE 2

Variable nucleotides from the coding region of the cytochrome *b* locus of three species of Drosophila

		D. melanogaster		*D. simulans* *si* I		*D. simulans* *si* II	*D. yakuba* *si* III		*D. yakuba*		All data		Not including *si*I and *si*III	
	C	P Tr Ke Te La O Va Be Zm 1 2 4 4 5 9 1 2 4 3 1 2 4 2 3 5 3 0 7 5 5 9 3		Ha		Va Mp Au La Tr R W 1 8 1 2 3 4 5 1 2 6 7 1 2 1 1 2 7 3	Ma		P 1 2 3 4 5 6 7 8 9 1 1 1 1 0 1 2 3					
4	A	. . . .		.		. . . .	.		C C C C C C C C C C C C C C		r	fixed	.	.
24	C	. . . .		T		T T T T T T T T T T T T T T T T	T		. . . .		s	fixed	.	.
27	C	T T T T T T T T T T T T T T T T T		.		. . . .	.		. . . .		s	fixed	.	.
30	T	. . . .		C	s	. . . .	.		. . . .		s	poly.		
31	T	C C C C C C C C C C C C C C C C C		.		. . . .	C	s	. . . .		S	poly.	s	fixed
45	C	. . . .		T	s	. . . .	.		T T T T T T T T T T T T T T		s	poly	s	fixed
51	T	. . . .		C		C C C C C C C C C C C C C C C C	C		. . . .		s	fixed	.	.
60	A	. . . .		.		. . . .	.		T T T T T T T T T T T T T T		s	fixed	.	.
84	T	. . . .		C		C C C C C C C C C C C C C C C C	C		. . . .		s	fixed	.	.
105	A	. . . .		.		. . . .	.		. G . . . . . . G	s	s	poly.	.	.
115	G	. . . . A	s	.		. . . .	.		. . . .		s	poly.	.	.
117	A	. . . .		G		G G G G G G G G G G G G G G G G	G		. . . .		s	fixed	.	.
123	T	. . . .		C	s	. . . .	.		. . . .		s	poly.		
144	C	. . . .		.		. . . .	.		T T T T T T T T T T T T T T		s	fixed	.	.
165	T	. . . .		C	s	. . . .	.		C C C C C C C C C C C C C C		s	poly.	s	fixed
168	C	. . . .		T		T T T T T T T T T T T T T T T T	.	s	. . . .		s	poly.	s	fixed
171	A	. . . .		T		T T T T T T T T T T T T T T T T	.	s	. . . .		s	poly.	s	fixed
174	T	. . . .		.		. . . .	.		A A A A A A A A A A A A A A		s	fixed	.	.
178	A	. . . .		.		. . . .	.		G G G G G G G G G G G G G G		r	fixed	.	.
180	T	. . . .		C	s	. . . .	.		. . . .		s	poly.		
183	T	. . . .		.		. . . .	.		C C C C C C C C C C C C C C		s	fixed	.	.
184	T	C C C C C C C C C C C C C C C C C		.		. . . .	C	s	. . . .		s	poly.	s	fixed
192	T	C C C C C C C C C C C C C C C C C		.		. . . .	C	s	. . . .		s	poly.	s	fixed
210	T	. . . .		C	s	. . . .	.		. . . .		s	poly.		
213	T	. . . .		.		. . . .	.		C C C C C C C C C C C C C C		s	fixed	.	.
217	G	. . . .		.		. . . .	.		. A	r	r	poly.	.	.
219	T	C C C C C C C C C C C C C C C C C		.		. . . .	.		. . . .		s	fixed	.	.
222	T	. . . .		.		. . . .	.		A A A A A A A A A A A A A A		s	fixed	.	.
246	T	. . . .		.	s	C C C C C C C C C C C C C C C C	C		. . . .		s	poly.	s	fixed
252	C	T T T T T T T T T T T T T T T T T		.		. . . .	.		. . . .		s	fixed	.	.
273	T	. . . .		C	s	. . . .	.		. . . .		s	poly.		
288	C	. . . .		T		T T T T T T T T T T T T T T T T	T		. . . .		s	fixed	.	.
295	G	. . . .		.		. . . .	.		A	r	r	poly.	.	.
297	A	. . . .		.		. . . .	.		T T T T T T T T T T T T T T		s	fixed	.	.
300	A	. . . .		.		. . . .	.		T T T T T T T T T T T T T T		s	fixed	.	.
309	T	. . . .		C	s	. . . .	.		. . . .		s	poly.		
315	C	. . . .		T		T T T T T T T T T T T T T T T T	T		. . . .		s	fixed	.	.
318	T	. . . .		.		. . . .	.		A A A A A A A A A A A A A A		s	fixed	.	.
325	A	. . . .		.		. . . .	.		T T T T T T T T T T T T T T		r	fixed	.	.
326	T	A A A A A A A A A A A A A A A A A		.		. . . .	.		. . . .		r	fixed	.	.
330	T	. . . .		.	s	C C C C C C C C C C C C C C C C	C		. . . .		s	poly.	s	fixed
333	T	. . . .		.		. . . .	.		A A A A A A A A A A A A A A		s	fixed	.	.
346	A	. . . .		.		. . . .	.		G G G G G G G G G G G G G G		r	fixed	.	.

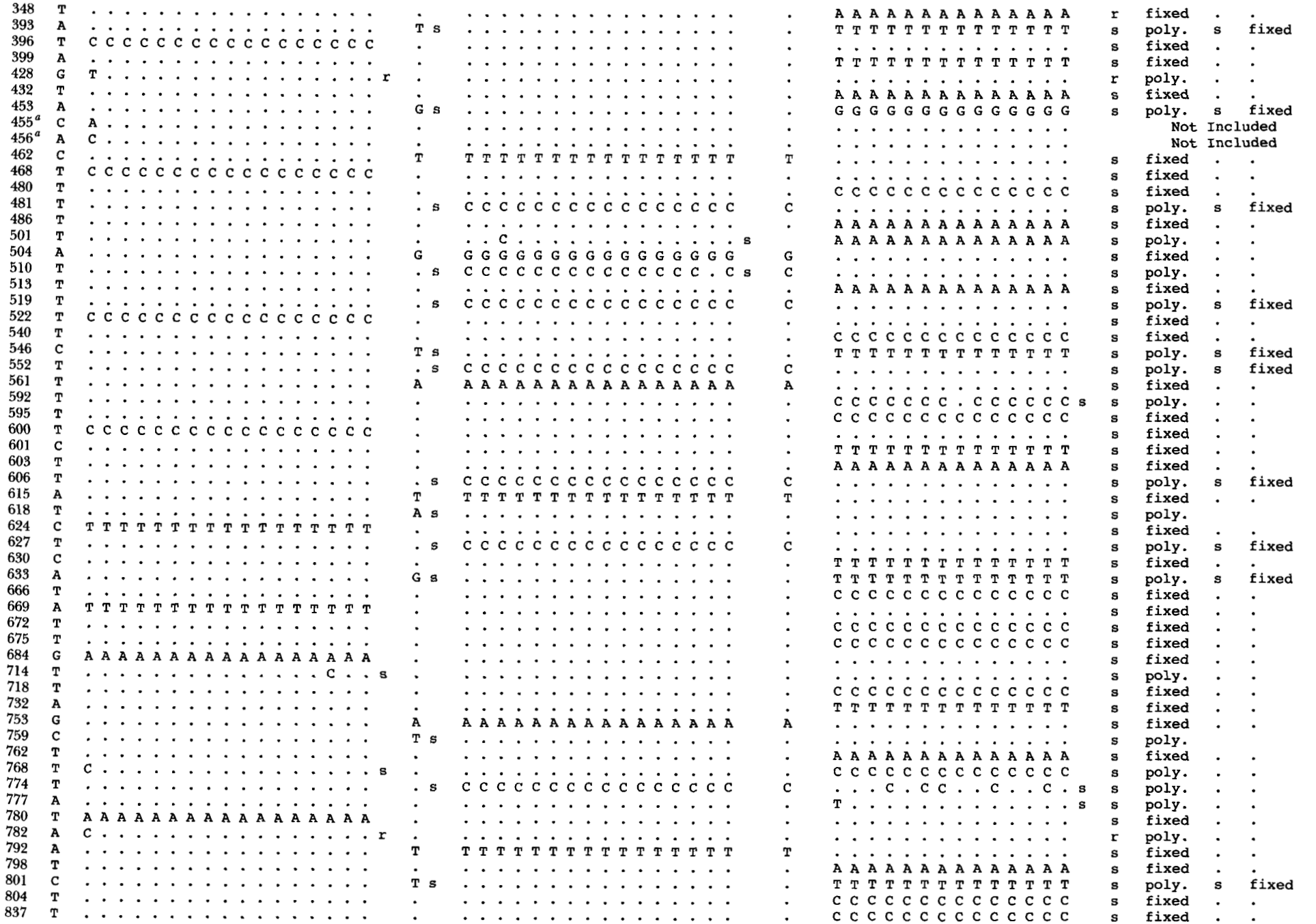
348 T
393 A
396 T C C C C C C C C C C C C C C C C
399 A
428 G T r
432 T
453 A G s
455[a] C A
456[a] A C
462 C T T T T T T T T T T T T T T T T T
468 T C C C C C C C C C C C C C C C C C
480 T
481 T
486 T
501 T C s
504 A G G G G G G G G G G G G G G G G G
510 T
513 T
519 T
522 T C C C C C C C C C C C C C C C C C
540 T
546 C
552 T
561 T A A A A A A A A A A A A A A A A A
592 T
595 T
600 T C C C C C C C C C C C C C C C C C
601 C
603 T
606 T
615 A
618 T
624 C T T T T T T T T T T T T T T T T T
627 T
630 C
633 A
666 T
669 A T T T T T T T T T T T T T T T T T
672 T
675 T
684 G A A A A A A A A A A A A A A A A A
714 T
718 T
732 A
753 G
759 C
762 T
768 T C
774 T
777 A
780 T A A A A A A A A A A A A A A A A A
782 A C
792 A
798 T
801 C
804 T
837 T
Not Included
Not Included

TABLE 2

Continued

		D. melanogaster																	
	C	P	Tr			Ke		Te		La		O	Va		Be		Zm		
			1	2	4	4	5	9	1	2	4	3	1	2	4	2	3	5	
									3	0	7	5				5	9	3	
840	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
844	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
846	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
849	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
852	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
858	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
865	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
870	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
876	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
879	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
897	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
904	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
912	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
913	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
915	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
918	T	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	s
939	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
945	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
948	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
954	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
960	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
966	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
970	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	A	.	r
972	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
973	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
978	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
993	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
1011	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1017	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1032	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1063	T	.	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	s
1069	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1072	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1074	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1080	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1086	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1093	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1095	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1098	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1102	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1105	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
1107	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	

	D. simulans																			
	si		si																	si
	I		II																	III
	Ha		Va		Mp					Au		La		Tr		R	W			Ma
			1	8	1	2	3	4	5	1	2	6	7	1	2	1	1	2		
										7	3									
840	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
844	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
846	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
849	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
852	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
858	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
865	C	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
870	T	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
876	.	s	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		C
879	C		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		C
897	C		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		C
904	.	s	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		C
912	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
913	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
915	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
918	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
939	T	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
945	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
948	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
954	C	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
960	A	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
966	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
970	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
972	T	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
973	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
978	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
993	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1011	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1017	.	s	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		C
1032	C	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1063	C	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1069	A	r	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	r	.
1072	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1074	G	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1080	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1086	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1093	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1095	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1098	T	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1102	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1105	.		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.
1107	C	s	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		.

	D. yakuba																		
																		Not including *si*I and *si*III	
	P	1	2	3	4	5	6	7	8	9	1	1	1	1		All data			
											0	1	2	3					
840	A	A	A	A	A	A	A	A	A	A	A	A	A	A		s	fixed	.	.
844	C	C	C	C	C	C	C	C	C	C	C	C	C	C		s	fixed	.	.
846	T	T	T	T	T	T	T	T	T	T	T	T	T	T		s	fixed	.	.
849	T	T	T	T	T	T	T	T	T	T	T	T	T	T		s	fixed	.	.
852	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	fixed	.	.
858	T	T	T	T	T	T	T	T	T	T	T	T	T	T		s	fixed	.	.
865	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.		
870	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.		
876	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.	s	fixed
879	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	fixed	.	.
897	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	fixed	.	.
904	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.	s	fixed
912	T	T	T	T	T	T	T	T	T	T	T	T	T	T		s	fixed	.	.
913	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	fixed	.	.
915	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	fixed	.	.
918	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.	.	.
939	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.		
945	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	fixed	.	.
948	C	C	C	C	C	C	C	C	C	C	C	C	C	C		s	fixed	.	.
954	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.		
960	A	A	A	A	A	A	A	A	A	A	A	A	A	A		s	poly.	s	fixed
966	C	C	C	C	C	C	C	C	C	C	C	C	C	C		s	fixed	.	.
970	A	A	A	A	A	A	A	A	A	A	A	A	A	A		r	poly.	.	.
972	T	T	T	T	T	T	T	T	T	T	T	T	T	T		s	poly.	s	fixed
973	T	T	T	T	T	T	T	T	T	T	T	T	T	T		r	fixed	.	.
978	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	fixed	.	.
993	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	fixed	.	.
1011	A	A	A	A	A	A	A	A	A	A	A	A	A	A		s	fixed	.	.
1017	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.	s	fixed
1032	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.		
1063	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.	.	.
1069	A	A	A	A	A	A	A	A	A	A	A	A	A	A		r	poly.	.	.
1072	A	A	A	A	A	A	A	A	A	A	A	A	A	A		r	fixed	.	.
1074	T	T	T	T	T	T	T	T	T	T	T	T	T	T		s	poly.	s	fixed
1080	T	T	T	T	T	T	T	T	T	T	T	T	T	T		s	fixed	.	.
1086	.	.	.	C	.	C	.	.	.	.	.	.	.	.	s	s	poly.	.	.
1093	A	A	A	A	A	A	A	A	A	A	A	A	A	A		r	fixed	.	.
1095	T	T	T	T	T	T	T	T	T	T	T	T	T	T		s	fixed	.	.
1098	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.		
1102	C	C	C	C	C	C	C	C	C	C	C	C	C	C		s	fixed	.	.
1105	G	G	G	G	G	G	G	G	G	G	G	G	G	G		r	fixed	.	.
1107	.	.	.	.	.	.	.	.	.	.	.	.	.	.		s	poly.		

C, consensus; s, silent; r, replacement. *D. melanogaster* isofemale lines: P, published (GARESSE 1988), West Indies (Tr), Africa (Ke, La, O, Zm), continental United States (Va, Be). *D. simulans* isofemale lines: *si*I, Hawaii (Ha); *si*II, West Indies (Tr), Africa (Mp), Continental USA (Va, La, R, W), Australia (Au); and *si*III, Madagascar (Ma). *D. yakuba* isofemale lines: P, published (CLARY and WOLSTENHOLME 1985), Africa (1–13).

[a] Not included in the analysis, as the published *D. melanogaster* sequence had these adjacent nucleotides reversed.

or G (T → C and A → G) or mutating to A or T (C → T and G → A). The neutral prediction is that the ratio of the two kinds of changes (to C or G and to A or T) should be the same within and between species. As an alternative, consider a model of weak selection favoring A- and T-endings over C- and G-endings. According to the nearly neutral theory, selection against slightly deleterious mutations will have a smaller effect on heterozygosity than on divergence compared to completely neutral mutations (KIMURA 1983, p. 44). Conversely, slightly advantageous mutations will have a relatively greater effect on divergence than on heterozygosity. The prediction, then, is that if mutations to C- and G-ending codons are slightly deleterious, they will be overrepresented within species compared to between species, relative to mutations to A- and T-ending codons.

A model of structure/function for mouse cytochrome *b* (HOWELL 1989) was used to derive a similar model for Drosophila cytochrome *b*. According to this model there are five internal eight transmembrane and four external domains. The number of silent and replacement substitutions was calculated for each domain and compared with the expected number of substitutions, assuming an equal probability of substitution over the length of the molecule. The segments of the protein hypothesized to constitute the Q_o and Q_i reaction centers and the hypothesized heme-ligating histidine residues (*H*) were not analyzed statistically because of the small size of the region.

Transitions exceed transversions in the mtDNA of Drosophila (SATTA *et al.* 1987; SATTA and TAKAHATA 1990). To test whether some sites are substituted more often than would be expected by chance we analyzed silent transitions and transversions separately. Synonymous changes (transitions or transversions) were mapped onto one of the most parsimonious trees and the number of sites that mutated once, twice and three times was calculated (no site changed more than three times). The number of sites without any mutations was estimated by subtracting the number of sites that mutated one or more times from the "effective" number of transition or transversion sites. The proportion of sites with no mutations was then used to estimate the Poisson parameter, λ. The number of observed substitutions was then compared with that expected under this Poisson distribution. The test is one-tailed because we are only interested in testing for a higher number of multiple substitutions than would be expected by chance. The test is conservative against this alternative because (1) parsimony by definition seeks to minimize the number of steps on a tree and (2) a site hit twice within the same lineage will be scored as either having no change or a single change. The underestimation of multiple hits will be a more severe problem when branches are long. As a consequence the test will be most powerful in a multibifurcating topology where each branch is short.

RESULTS

To investigate the potential for population subdivision in the *D. melanogaster* and *D. simulans* collections, at least two lines were sequenced from each locality. The only exceptions were the *D. melanogaster* line from Botswana (O) and the *D. simulans* R line. The *D. yakuba* lines are from a single collection. A previous study by MCDONALD and KREITMAN (1991) showed these same lines to be highly polymorphic at the *Adh* locus.

To link the 16 lines of *D. simulans* with the previously designated mitochondrial haplotypes (SOLIGNAC *et al.* 1986; SATTA and TAKAHATA 1990), we sequenced a small diagnostic region of the *COI* gene from the same individuals of *D. simulans* Mp4, Au17 and La6 (Table 2).

TABLE 3

Summary statistics of DNA variability in the cytochrome *b* locus of three species of Drosophila

	All sites (1137)	Silent sites (≈245)
D. melanogaster		
Sample size	17	17
No. segregating sites	8	5
π	0.0009	0.0005
$\hat{\theta}$ (per site)	0.0021	0.006
D. simulans siII		
Sample size	16	16
No. segregating sites	3	2
π	0.0003	0.0002
$\hat{\theta}$ (per site)	0.0011	0.0037
D. yakuba		
Sample size	14	14
No. segregating sites	7	5
π	0.0014	0.0012
$\hat{\theta}$ (per site)	0.0019	0.0064

This region contained 15 segregating sites diagnostic for the three haplotypes (SATTA and TAKAHATA 1990). Analysis of these sites demonstrated we had sampled only the *si*II haplotype. To extend our investigation of variation in the mtDNA of *D. simulans* we obtained two additional lines as described above.

Intraspecific variation: Polymorphism data is summarized in Table 3. The 16 *D. melanogaster* lines contain three synonymous polymorphisms and one replacement polymorphism. These 16 lines differed from the published sequence by two additional synonymous substitutions and two replacement substitutions (Table 2). We have excluded positions 455 and 456 of the published *D. melanogaster* sequence from all our analyses because all lines sequenced in this study inverted these adjacent positions. The 16 *si*II lines have two synonymous changes and one replacement change. Each mutation occurs once in the sample. The *si*I line differs from the *si*II lines by 37 synonymous substitutions. Inclusion of the *si*III line adds five more synonymous substitutions (Tables 2 and 3). The 13 lines of *D. yakuba* have four synonymous and one replacement polymorphism. Inclusion of the published *D. yakuba* cytochrome *b* sequence (CLAREY and WOLSTENHOLME 1985) adds one replacement and one synonymous change (Tables 2 and 3).

Phylogenetic analysis: The strict consensus of the four 162 step trees is presented in Figure 1. The topology of the trees differ only in the placement of *D. simulans* W1 and *D. simulans* Au23. Parsimony analysis with PTP testing indicates significant structure in these data (*PTP* = 0.001). When *D. yakuba* is employed as the outgroup, parsimony analyses with T-PTP testing and bootstrapping support monophyly of *D. melanogaster* (*T-PTP* = 0.001, 20 steps and 100%) and *D. simulans* (*T-PTP* = 0.001, 6 steps; 93%), relative to each other (Figure 1). In addition, the *si*II and -III haplotypes form a distinct monophyletic lineage (*T-PTP* = 0.001,

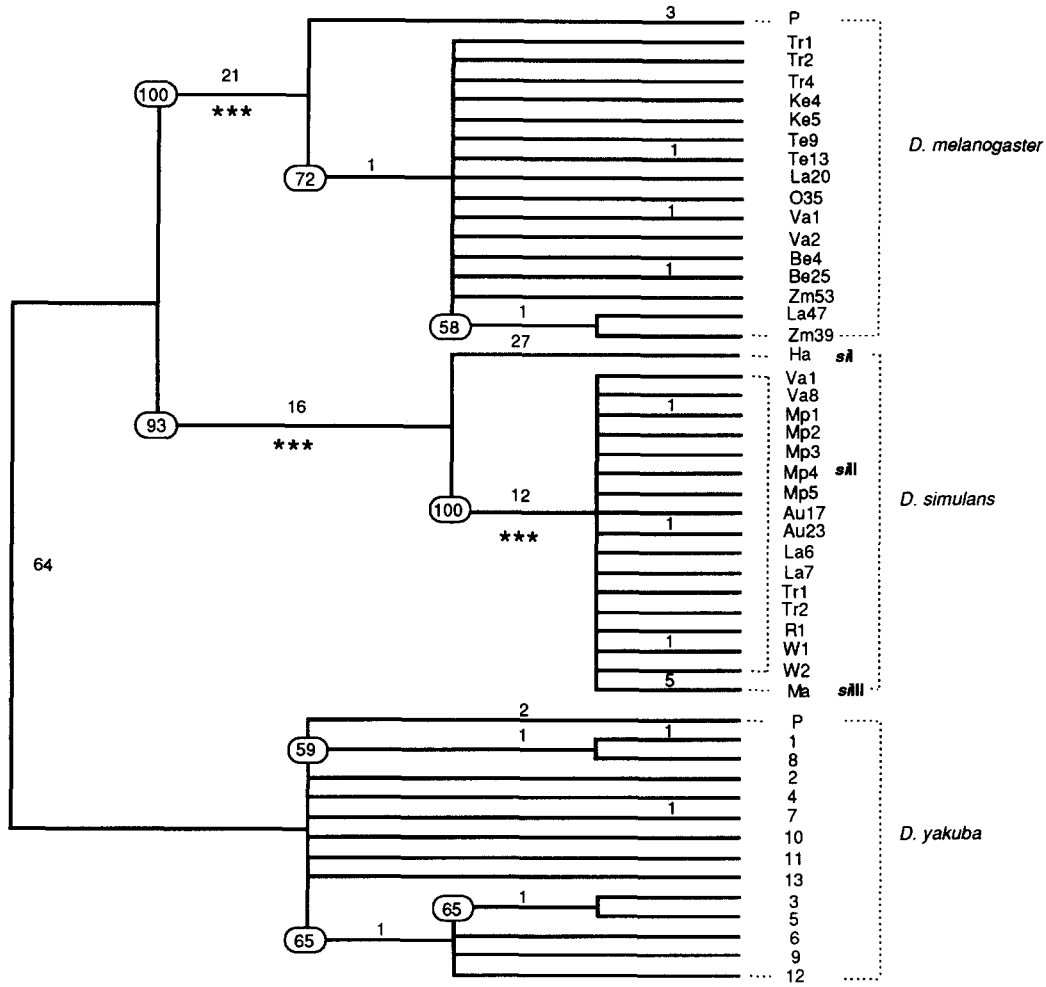

FIGURE 1.—Strict consensus of the four equally parsimonious topologies of length 162 steps (consistency index = 0.88, retention index = 0.99) generated from the 49 taxa 1137 cytochrome *b* sequence matrix (117 informative sites). The parsimony analyses with T-PTP testing based on 999 randomized data sets evaluated the support for each *a priori* monophyly hypothesis. Asterisks below the lines show significantly monophyletic groups ($P \leq 0.05$). Numbers above the lines refer to branch lengths and numbers in circles indicate bootstrap percentages from 1000 replicates. For clarity 0 branch lengths are not indicated on the tree. *D. melanogaster* isofemale lines: published (P) (GARESSE 1988); West Indies (Tr); Africa (Ke, La, O, Z); continental United States (Va, Be). *D. simulans* isofemale lines: *si*I, Hawaii (Ha); *si*II, West Indies (Tr), Africa (Mp), continental United States (Va, La, R, W), Australia (Au); *si*III, Madagascar (Ma). *D. yakuba* isofemale lines: published (P) (CLAREY and WOLSTENHOLME 1985), Africa (1–13).

10 steps; 100%), relative to *D. melanogaster* and the *si*I lineage (Figure 1). These data suggest that the *si*I haplotype is highly divergent from the *si*II and -III lineage.

We investigated the constancy of evolutionary rate utilizing the 2D method of TAJIMA (1993). With *D. yakuba* (CLAREY and WOLSTENHOLME 1985) as the outgroup, *D. melanogaster* (GARESSE 1988) is not evolving at a significantly different rate compared to the *si*I haplotype ($\chi^2_2 = 2.22$, $P = 0.33$) or the *si*II haplotype (Va1) ($\chi^2_2 = 3.83$, $P = 0.144$). However, with *D. melanogaster* (GARESSE 1988) as the outgroup, we can reject the hypothesis that the *si*I haplotype is evolving at the same rate as the *si*II haplotype (Au23) ($\chi^2_2 = 7.46$, $P = 0.023$) or the *si*III haplotype ($\chi^2_2 = 11.15$, $P = 0.004$). With 27 changes assigned to the *si*I lineage and 12 to the *si*II and -III lineage, the test result indicates a significantly higher rate of evolution along the *si*I branch. Possible explanations for this accelerated evolution are presented in the DISCUSSION.

Neutral tests: We compared the ratios of amino acid replacement polymorphism and divergence with synonymous polymorphism and divergence to test whether the patterns conform to the neutral prediction. The test result, shown in Table 4, yields a significant departure from neutrality. The departure is consistent with a relatively greater than expected number of within-species replacement polymorphisms or a relative paucity of between-species replacement substitutions compared to synonymous changes. These results do not include a multiple hit correction and therefore are likely to underestimate the number of fixed synonymous substitutions. Hence, this result is conservative.

We used only the silent variation for the HKA tests. Separate HKA tests were carried out for each species

TABLE 4

Number of replacement and synonymous substitutions within and between *D. melanogaster*, *D. simulans siII* and *D. yakuba*

	Fixed between species	Polymorphic within species	Total
Replacement	10	6	16
Synonymous	97	12	109
	107	18	$G_{adj.}{}^{a} = 6.22$
			$P = 0.01$

[a] We applied WILLIAMS' correction prior to calculation of the *G* statistic (SOKAL and ROHLF 1981).

using polymorphism data from only one species. This procedure assumes the population size of the common ancestor is the same as the species from which the polymorphism estimate is derived (KREITMAN and HUDSON 1991).

The cytochrome *b* polymorphism level in each species, estimated either by the number of segregating sites or by the average pairwise difference, was compared to the corresponding level in a nuclear gene. Divergence was estimated by comparing a sequence of each gene with a sequence from one of the other species. The species and strains and the test results are presented in Table 5. The pairwise difference test (Pwd) yielded uniformly lower X^2 values than the segregating sites test, even though cytochrome *b* polymorphism frequencies were low in *D. melanogaster* and the worldwide *D. simulans* sample. With only a small number of polymorphic sites, the Pwd test may not be as powerful as the segregating sites test at detecting departures from neutrality.

In *D. melanogaster*, cytochrome *b* was tested against three nuclear gene regions: the *Adh* coding region and introns 2 and 3, the 5′-flanking region of *Adh* and the coding region and three introns of *period*. As expected, the comparison to *Adh* is statistically significant (Table 5). *Adh* in this species has a high level of silent polymorphism, thought to be the result of a balanced polymorphism at the locus (HUDSON *et al.* 1987; KREITMAN and HUDSON 1991). The cytochrome *b* polymorphism and divergence is not significantly different from either the *Adh* 5′ flanking region or *period* (Table 5), nor is the *Adh* 5′ flanking region different from *period* (X^2 =0.76, P = 0.383). Considering all the data and tests, we cannot reject a neutral model for the polymorphism and divergence of cytochrome *b* in *D. melanogaster* or *D. yakuba* (Table 5).

The *si*II haplotype is segregating at only two synonymous sites. HKA segregating sites tests comparing cytochrome *b* against either *Adh* or *period* are significant, indicating a departure from neutrality (Table 5). That *Adh* and *period* are not significantly different from each other (X^2 = 0.12, P = 0.729), allows us to conclude that the silent polymorphism level is lower than expected for the cytochrome *b* locus, given its rate of evolution. Consequently, the worldwide *si*II haplotype is a candidate for a recent selective sweep.

We used the WATTERSON (1978) test for homozygosity to ask whether there are too many rare alleles in the samples (Table 6). The WATTERSON test is significant for haplotype frequencies that include all and silent *D. melanogaster* changes (F = 0.446, P = 0.049 and F = 0.599, P = 0.024, respectively). Such a pattern might be expected after a bottleneck and subsequent expansion or following a selective sweep. To further investigate the distribution of alleles in *D. melanogaster*, we reanalyzed HALE and SINGH's (1991) RFLP study of a worldwide collection of 144 isofemale lines. Although there is little population subdivision in this species for mtDNA haplotypes, HALE and SINGH divided the populations into three geographic regions: Western Hemisphere, Euro-African and Far East. The Euro-African sample (12 haplotypes, N = 58) and the Far East sample (eight haplotypes, N = 50) show no departure from neutrality (F = 0.171, P = 0.29 and F = 0.188, P = 0.652, respectively), whereas the Western Hemisphere sample (eight haplotypes, N = 58) does show a highly significant departure (F = 0.665, P = 0.003). Because of the large computational requirements we did not perform the WATTERSON test on the combined dataset, but there is no evidence for an excess of singletons as determined by the EWENS (1973) test (ten singletons among 23 haplotypes, N = 144; P = 0.118).

Three of the four *si*II haplotypes are singletons (N = 16) of which one of the singleton alleles is an amino acid replacement. The combined data (silent and replacement) is significant at the 0.06 level (F = 0.672, P = 0.053); but it is not significant for silent sites alone (F = 0.773, P = 0.12). These results are suggestive of an excess of rare mitochondrial alleles in this species.

The WATTERSON (1978) test for homozygosity is not significant for *D. yakuba* (F = 0.214, P = 0.430 and F = 0.225, P = 0.767 for all and silent sites, respectively). Furthermore, there is no evidence for non-neutrality in the eight African and Madagascan lines of MONNEROT *et al.* (1990) (F = 0.781, P = 0.44).

Codon selection: Approximately 90% of cytochrome *b* codons end in A or T. In nuclear genes of Drosphila, SHIELDS *et al.* (1988) have shown that biased codon usage (always favoring G- or C-ending codons) is the result of natural selection rather than mutation. To our knowledge this problem has not been investigated for mtDNA genes. Restricting our attention to the third "wobble" position we polarized all silent transition changes as either mutating to C or G (T → C and A → G) or mutating to A or T (C → T and G → A) within *D. melanogaster*, *D. simulans* and *D. yakuba* and between *D. melanogaster* and *D. simulans* using the strict consensus tree.

A simple test of neutrality is to compare the ratio of the two types of changes when they are polymorphic

TABLE 5

Single species HKA test of the silent site variability in three species of Drosophila

Species	Gene	Region	Sample size	Polymorphic sites	Divergence between Line 1 and Line 2	Line 1 (species)[a]	Line 2 (species)[a]	Average pairwise difference	Reference[b]	χ^2 (P) Seg[c]	χ^2 (P) Pwd[d]
D. melanogaster	Cytochrome *b*	Coding	17	5	49	Tr4 (m)	Im5 (s)	0.59			
	Adh	Coding + 2 introns	18	27	16	Af-sx (m)	Drsadha (s)	7.85	A	4.72 (0.03)	3.52 (0.06)
		5′ flanking	11	30	78	Af-s (m)	Drsadha (s)	11.56	B	0.13 (0.72)	1.01 (0.31)
	period	Coding + 3 introns	6	28	67	L12 (m)	Ca1 (s)	11.26	C	1.66 (0.20)	1.56 (0.21)
D. simulans	Cytochrome *b*	Coding	16	2	49	Im5 (s)	Tr4 (m)	0.25			
	Adh	Coding	6	11	10	Drsadha (s)	Af-s (m)	4.60	D	8.03 (0.005)	4.27 (0.04)
	period	Coding + 3 introns	6	46	67	Ca1 (s)	L12 (m)	18.40	C	6.70 (0.01)	2.84 (0.09)
D. yakuba	Cytochrome *b*	Coding	14	5	82	4 (y)	Im5 (s)	1.31			
	Adh	Coding	19	17	23	a (y)	g (s)	3.39	D	2.84 (0.09)	0.81 (0.37)

Each test compares the number of polymorphic sites of one species to divergence with another at the cytochrome *b* locus and an autosomal locus.

[a] m = *D. melanogaster*, s = *D. simulans*, y = *D. yakuba*.

[b] Reference A = compiled from GeneBank by M. WAYNE (personal communication), B = KREITMAN and HUDSON (1991), C = KLIMAN and HEY (1993), D = MCDONALD and KREITMAN (1991).

[c] Seg = HKA test based on segregating sites.

[d] Pwd = pairwise difference HKA test.

TABLE 6

WATTERSON (1978) test for homozygosity

	All sites (1137)	Silent sites (245)
D. melanogaster		
Sample size (alleles)	17 (6)	17 (5)
Allele (freq)	4 (1) + 1 (2) + 1 (11)	4 (1) + 1 (13)
Observed *F* value	0.446	0.599
Probability	0.049	0.024
*D. simulans sil*I		
Sample size (alleles)	16 (4)	16 (3)
Allele (freq)	3 (1) + 1 (13)	2 (1) + 1 (14)
Observed *F* value	0.672	0.773
Probability	0.053	0.12
D. yakuba		
Sample size (alleles)	14 (7)	14 (5)
Allele (freq)	4 (1) + 1 (2) + 1 (3) + 1 (5)	2 (1) + 2 (2) + 1 (3) + 1 (5)
Observed *F* value	0.214	0.225
Probability	0.430	0.767

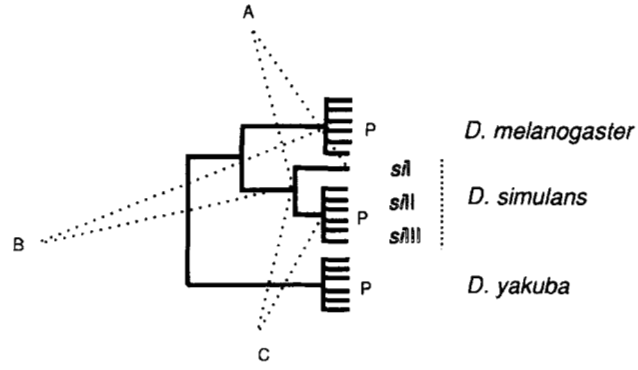

FIGURE 2.—For the analysis presented in Table 7, we divided the tree into four topological regions. These were the differences between *D. melanogaster* and *D. simulans* (B), changes in the *sil* lineage (A), changes in the *sil*I and -III lineage (C), and polymorphisms within *D. yakuba*, *D. melanogaster* and the *sil*I and -III haplotypes (P) (Table 7).

within a species or are fixed between them. However, for this analysis we are faced with the difficulty of deciding how to treat changes within *D. simulans* (Figure 1). Phylogenetic analysis shows that the *sil*I and -III haplotypes form a monophyletic group that is highly divergent from the *sil* haplotype. For this analysis, we divided the tree into four topological regions, as shown in Figure 2: differences between *D. melanogaster* and *D. simulans* (B), changes in the *sil* lineage (A), changes in the *sil*I and -III lineage (C), and polymorphisms within *D. yakuba*, *D. melanogaster* and the *sil*I and -III haplotypes (P) (Table 7). Previous tests of neutrality have characterized variation as polymorphic or fixed. However, it is equally valid to compare the ratios of the two types of changes within and between the four topological regions–the prediction being that the proportions should be the same in each region.

The pattern of third position synonymous transition changes in the four topological regions are not the same. There is no preference between species for changes to either A and T or C and G, as would be expected if the overall usage of A and T *vs.* C and G is remaining constant. Between species, there are 11 changes to C or G and 10 to A or T. Within *D. melanogaster*, *D. yakuba* and the *sil*I and -III haplotypes there is a distinct bias favoring mutations that change codons to C and G: there are nine codon changes to C or G but only one change to A or T.

The substitution pattern within the *sil* lineage is similar to that between species. There is no distinct bias to either A and T or C and G (12 codons change to C or G and seven change to A or T). In contrast, the substi-

TABLE 7

Third position synonymous transition changes within *D. melanogaster*, *D. simulans* and *D. yakuba* and between *D. melanogaster* and *D. simulans*

Transition type		*D. simulans sil* lineage A	Between *D. melanogaster* and *D. simulans* B	Fixed on *D. simulans silI* and -III C	Polymorphic P	Total
C → T	G → A	7	10	0	1	18
T → C	A → G	12	11	10	9	42
Total		19	21	10	10	$\chi^2_3 = 9.72$ $P = 0.02$

tution pattern in the *si*II and -III lineage, like canonical polymorphisms, is biased highly toward C and G: there are 10 changes to C or G and none to A or T. These data strongly suggest the action of natural selection on codon usage in Drosophila, but the nature of the selection is equivocal.

Are substitutions random?: Silent and replacement changes within and between species among the five internal, eight transmembrane and four external segments should be randomly distributed if there are no selective constraints, or if constraints are the same for all regions. The data are presented in Table 8 and Figure 3. A larger-than-expected number of amino acid replacements occur in the second internal and small eighth transmembrane segments (3/18 and 5/18, respectively) but there are not enough changes within transmembrane domains to carry out a useful statistical test of heterogeneity. There is no evidence for heterogeneity of amino acid replacement changes among the three types of domains ($\chi^2_2 = 0.639$, $P = 0.72$). There is no evidence for heterogeneity of silent changes within the internal domains ($\chi^2_3 = 1.03$, $P = 0.79$ and $\chi^2_3 = 3.10$, $P = 0.38$, for all substitutions and fixed substitutions, respectively), transmembrane domains ($\chi^2_7 = 6.13$, $P = 0.52$ and $\chi^2_7 = 3.883$, $P = 0.79$, all and fixed substitutions, respectively) and external domains ($\chi^2_2 = 0.856$, $P = 0.65$ and $\chi^2_2 = 1.85$, $P = 0.39$, all and fixed substitutions, respectively), nor is there evidence for heterogeneity of silent changes between the three types of domains ($\chi^2_2 = 3.068$, $P = 0.22$ and $\chi^2_2 = 3.202$, $P = 0.2$, all and fixed substitutions, respectively) (Table 8). Furthermore, there is no evidence of heterogeneity in the number of substitutions within and between species for the three domains ($\chi^2_2 = 0.865$, $P = 0.65$).

Heterogeneity in the rate of substitution among silent sites may occur if certain sites mutate (or substitute) more frequently than others. We compared the observed distribution of the numbers of substitutions at each variable site in a most parsimonious tree with that expected if every site had the same probability of change. The Poisson parameter, λ, was estimated separately for transitions and transversions. The effective number of silent sites (246.67) was subdivided into transversion (101.00) and transition (145.67) effective sites. Poisson parameters were estimated from the proportion of sites having no change. For transversions, the numbers of sites with zero, one and two substitutions are 59, 38 and 4, respectively. The expected number under the Poisson distribution with the same mean as the observed are 64.05, 29.17 and 6.64. The Poisson model can explain the observed distribution of transversion substitutions ($\chi^2_2 = 2.02$, $P = 0.37$). For transitions, the number of sites with zero, one and two substitutions are 59.66, 75 and 11, respectively. The expected number under the Poisson distribution with the same mean as the observed are 74.84, 49.84 and 16.59. The Poisson model does not explain the observed distribution of transition substitutions ($\chi^2_2 = 7.84$, $P = 0.02$). However, rather than an excess of sites being hit more than once there is an excess of sites with a single substitution. We cannot envisage any reasonable biological mechanism leading to an overdispersion of silent changes. As a consequence, we suspect this nonconformity to the Poisson distribution is an artifact caused by not being able to detect multiple substitutions along the long branches (such as the one in the *si*I lineage, Figure 1). Alternatively, the tree topology may be incorrect. As a consequence we do not reject the hypothesis that transitions are evolving at an equal rate in cytochrome *b*.

TABLE 8

Synonymous mutations in the three functional regions of cytochrome *b*

Internal	Transmembrane	External	Total
Fixed substitutions			
28	30	35	93
"Effective" number of silent sites			
58.33	111.66	78.33	248.32
Expected number of fixed substitutions			
21.85	41.82	29.34	$\chi^2_2 = 3.202$ $P = 0.20$

DISCUSSION

Data presented here indicate that molecular variation and evolution of mtDNA are governed by many of the same selective forces that have been shown to govern

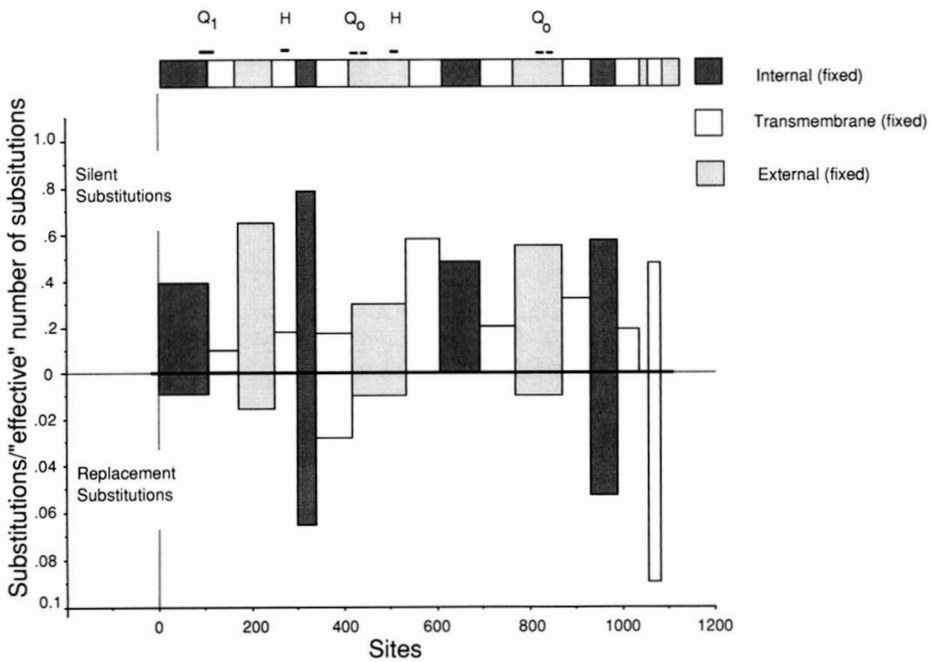

FIGURE 3.—Distribution of fixed silent and replacement substitutions between *Drosophila melanogaster*, *D. simulans* and *D. yakuba* divided by the effective number of sites (silent above the line and replacement below the line) within each functional domain. The schematic map of the locus is based on the model by HOWELL (1989). The segments of the protein hypothesized to constitute the Q_o and Q_i reaction centers and the hypothesized heme-ligating histidine residues (H) are plotted onto the figure.

nuclear genome evolution. This finding has important implications for anyone who uses mtDNA as a "neutral" molecular marker. We do not find a global explanation for the departure from neutral expectations. This comes as something of a surprise. We predicted, by analogy to the general lack of variation on the very small nonrecombining fourth chromosome of Drosophila, that positive selection acting on mtDNA would also lead to the reduction of variation (BERRY *et al.* 1991). Such a crisp pattern of variation is not observed in all three species. Rather, we are forced to propose a series of rather tentative explanations for the complex mosaic of selection pressures that may be acting on Drosophila mtDNA. A summary of significant departures from the neutral theory predictions are presented in Table 9.

Population substructure: We designed our initial sampling to allow the detection of strong population subdivision within *D. melanogaster* and *D. simulans* (specifically the *si*II mitochondrial haplotype). Although the number of intraspecific substitutions is low, there is no evidence of population differentiation. No two taxa from any one location (with the exception of *D. yakuba* where all lines were from the one locality) form a monophyletic group (Figure 1, Table 2). Note that all of the singleton alleles have a second representative sampled from the same location. BABA-AÏSSA *et al.* (1988) similarly do not note any population subdivision within 118 lines of the *si*II haplotype from around the globe. Within 144 isofemale lines of *D. melanogaster* from 18 geographic locations HALE and SINGH (1991) identify 23 haplotypes. Contrary to their earlier observation of geographic clustering (HALE and SINGH 1987), the larger study finds only limited geographical clustering of mitochondrial haplotypes. There is extensive nuclear gene flow within and between most populations of *D. melanogaster* (SINGH and RHOMBERG 1987a,b) with the notable exception of some lines from Zimbabwe. BEGUN and AQUADRO (1993) reported that a nuclear DNA restriction site analysis shows a population of *D. melanogaster* from Zimbabwe to be more than twice as variable as North American populations and most variants are not shared. However, *D. melanogaster* from Zimbabwe may be in the process of forming a distinct species, as there is evidence for assortative mating between them and other lines of *D. melanogaster* (C.-I WU, personal communication). In this study the two Zimbabwe lines did not form a monophyletic group.

Intraspecific variation: The estimates of nucleotide diversity (π) within the cytochrome *b* gene of *D. melanogaster* and *D. yakuba* are low: *D. melanogaster* π_t = 0.0009 (silent + replacement π), π_s = 0.0005 (synonymous π); *D. yakuba* π_t = 0.0014, π_s = 0.0012. The diversity of synonymous substitutions is similar to that calculated by HALE and SINGH (1987) for restriction site polymorphism of whole *D. melanogaster* mtDNA (π_r = 0.002 restriction site π). Within the cytochrome *b* locus of the *si*II haplotype we estimate the genetic variability to be π_t = 0.0003, π_s = 0.0002. The synonymous diversity within this haplotype is similar to that calculated from RFLP data for the coding region of mtDNA (π_r = 0.006) and an order of magnitude less than that calculated for the complete mtDNA including the (A + T)-rich region (π_r = 0.046) (BABA-AÏSSA *et al.* 1988).

Interestingly, the level of variation is unrelated to geographic size. The *si*I haplotype has the highest genetic diversity of the three *D. simulans* haplotypes (BABA-AÏSSA *et al.* 1988). Also, *D. heteroneura* and *D. silvestris* are about five times more variable than *D. melanogaster* and the worldwide *si*II haplotype (from data of DESALLE *et al.* 1986).

TABLE 9

Summary of statistically significant tests of neutrality

Question	Test	Conclusion	Reference
Rate heterogeneity	*D. simulans sil vs. D. simulans sil*I and -III	Higher substitution rate in *D. simulans sil*	TAJIMA (1993)
Protein evolution	Within and between species variation at linked sites	Excess intraspecific replacements or interspecific synonymous substitutions: slightly deleterious or diversifying selection	MCDONALD and KREITMAN (1991)
Hitchhiking	*D. simulans sil*I (p)[a] *vs. D. melanogaster*	Reduced polymorphism within the *sil*I haplotype: possibly caused by an external agent, *Wolbachia*	HUDSON *et al.* (1987)
Selective sweep	Homozygosity test: *D. melanogaster* (silent + replacement) and *D. simulans sil*I (silent)	Suggestion of too many rare alleles: selective sweep, purifying selection or population subdivision	WATTERSON (1978)
Codon selection	Substitution bias in different topological regions	Heterogeneity for C/G → A/T substitutions at silent sites	

[a] Polymorphism.

Evolutionary rates: We observed a significantly higher rate of substitution along the *sil* lineage compared to the *sil*I and -III lineage. The rates of substitution may be related to the population structure and the evolutionary histories of these lineages. While the *sil*I haplotype has a worldwide distribution the *sil* haplotype has been found only on Pacific islands, and it may have undergone repeated founder events leading to the fixation of deleterious alleles. Alternatively, the generation time of the Pacific island populations may be shorter than the other populations. DESALLE and TEMPLETON (1988) compared rates of mtDNA evolution, as determined by restriction-site changes, in two closely related lineages of Hawaiian Drosophila. The β lineage, which included species that are thought to have undergone repeated founder events, evolved at three times the rate of the α lineage, the latter having probably arisen from a large ancestral population. This result assumes that the α and β lineages form a monophyletic group: if either is outside the designated outgroup, the result would be invalid. Although they attribute the increased rate of molecular evolution to the founder events, they provide no mechanistic explanation. That explanation can only be the increased rate of fixation of slightly deleterious mutations in small populations (OHTA 1973).

Neutral tests: The *sil*I haplotype sample and the *D. melanogaster* sample show low levels of cytochrome *b* variation, both for the total numbers of polymorphic sites and for the frequency of each polymorphism in the samples. KANEKO *et al.* (1993) observed one synonymous polymorphism at the ATPase 6 locus in four lines of *D. melanogaster*. The segregating sites HKA test indicates a reduction of polymorphisms in the *sil*I haplotype sample. If selective sweeps are responsible for the low level of variation, the target of selection can be a mutation anywhere in the mitochondrial genome. However, selection need not even be acting on the mitochondrial genome itself. Wolbachia is a maternally inherited rickettsia-like microorganism infecting *D. melanogaster* and *D. simulans*. Infected females can mate with any male and produce progeny, but uninfected females produce fewer progeny when they mate with infected males (HOFFMANN *et al.* 1986; TURELLI and HOFFMANN 1991). Here the mtDNA is carried passively as the microorganism sweeps through the population (TURELLI *et al.* 1992). The parasite is occasionally lost from infected flies, so there is not a complete barrier to gene flow once a population becomes infected (HOFFMANN and TURELLI 1988; HOFFMANN *et al.* 1990).

In Australia there is weak incompatibility between infected and uninfected strains of *D. melanogaster* (HOFFMANN 1988). Crosses between infected males and uninfected females exhibit a 15–30% reduction in egg hatch (HOFFMANN *et al.* 1994). Progeny tests indicate that Wolbachia infections are widespread in Australian *D. melanogaster* populations and that populations are polymorphic for the presence of the infection (HOFFMANN *et al.* 1994). SOLIGNAC *et al.* (1994) have extended this observation noting that many populations in the world are polymorphic for Wolbachia infection.

There is evidence for restricted gene flow between the three *D. simulans* haplotypes. The *sil*II haplotype has the M-strain of Wolbachia. This strain exhibits partial incompatibility with the type R-strain associated with the *sil*I haplotype and unidirectional incompatibility with the S-strain associated with the *sil* haplotype (ROUSSET *et al.* 1992). The *sil* haplotype appears intimately associated with the S-strain of Wolbachia (MONTCHAMP-MOREAU *et al.* 1991; ROUSSET *et al.* 1992). It is bidirectionally incompatible with the *sil*I associated type R-strain (O'NEILL and KARR 1990) and unidirectionally incompatible with the *sil*I associated type W-strain (MONTCHAMP-MOREAU *et al.* 1991). Wolbachia infections may also have a significant influence on the effective population size. The *sil*I haplotype has a worldwide distribution. However, if there has been a series of Wolbachia-induced selective sweeps, as recorded in California (TURELLI *et al.* 1992), the effective population size

of the *si*II haplotype may, be lower than that of the geographically restricted *si*I haplotype.

About 10% of the fixed differences that are fixed between species are replacement substitutions, but more than 35% of the polymorphisms are replacements. The substitution pattern in the *si*I lineage is similar to that between species whereas the pattern in the *si*II and -III lineage is more like that within species. This suggests that the former has evolved as a distinct biological entity. We believe selection is acting on replacement changes rather than silent changes. Selection for biased codon usage will lead to a relative increase in the level of synonymous polymorphism compared to the number of fixed differences among species. This is in the opposite direction to the observed pattern for synonymous and replacement changes. Thus, we do not believe that codon selection can explain this departure from neutrality.

We are left with two viable hypotheses: selection against slightly deleterious amino acid replacement changes and diversifying selection. It is difficult to distinguish between them because the number of replacement polymorphisms are low. Slightly deleterious mutations are expected to be present in populations parimarily and will rarely fix as low frequency polymorphisms. The *si*I lineage, with its accelerated evolution provides a possible test of this prediction. If the population size of the geographically restricted *si*I haplotype has been small, then a relatively higher proportion of amino acid replacements might have been expected on this lineage compared to synonymous changes. That it is not observed–only one of 27 changes in the *si*I lineage is an amino acid replacement–suggests either the population size is not small compared to the other haplotypes (as we have argued above) or replacement changes are not, in general, deleterious. OHTA (1993) showed a higher rate of *Adh* protein evolution relative to silent evolution for Hawaiian *Drosophila* species compared to mainland species. We cannot invoke a similar explanation for the *si*I haplotype.

These data do not lend support to the hypothesis that the synonymous changes are slightly deleterious but have escaped detection in the *si*I populations. Our reasoning is that if the *si*I haplotype has escaped selection on synonymous changes, then there should have been an excess accumulation of C and G ending codons, assuming they are deleterious (see *Codon selection* below). We also do not find support for diversifying selection. If diversifying selection is coarse-grained with regard to the environment, we would expect amino acid replacement polymorphisms to be present as fixed differences among our population samples. This is not observed. We are left, by default, with the possibility of other mechanisms maintaining protein variation, such as frequency dependent selection.

Codon selection: In Drosophila, nuclear codon usage is biased toward mostly C- and G-ending codons (SHIELDS *et al.* 1988). For nuclear genes in *D. simulans*, H. AKASHI (personal communication) finds a significantly higher ratio of polymorphism to divergence for synonymous changes from preferred to unpreferred codons than for changes in the reverse direction. Comparison of preferred and unpreferred synonymous changes in the *D. melanogaster* and *D. simulans* lineages show an excess of unpreferred fixations in *D. melanogaster*.

The mitochondrial genome is strongly biased toward A- and T-ending codons. Indeed, there is a higher A and T content in the silent sites of coding regions than in the presumably neutral A + T-rich region in *D. yakuba*.To investigate the hypothesis that selection is acting on synonymous changes we divided the strict consensus tree into four topological regions. If there is no selection acting, the ratio of the two types of changes (to A and T, and to C and G) within and between the four to pological regions (Figure 2) should be the same. The third position synonymous transition changes in the four topological regions are not the same (Table 7).

The direction of changes in the two canonical topological regions (within *D. melanogaster*, *D. yakuba* and the *si*II and *si*III haplotypes and between *D. melanogaster* and *D. simulans*) is consistent with the hypothesis that changes to C or G are slightly deleterious and have a lower probability of reaching fixation than mutations to A or T. The substitution pattern in the two other topological regions cannot be easily explained. We have argued above that the *si*I haplotype evolved as a distinct biological entity and that changes to C or G and A or T are proportionally similar to those between species. However, there is no such obvious explanation for the dramatic bias toward C and G in the *si*II and -III lineage. We have argued above that a reduction in the polymorphism of the *si*II haplotype is consistent with the notion of a microorganism induced selective sweep. Moreover, it is possible that there is microorganism induced restriction to gene flow between the *si*II and *si*III haplotypes. If the *si*II -III lineage has undergone recurrent Wolbachial sweeps, then deleterious mutations to C- or G-ending codons, otherwise present at low frequencies, may have repeatedly swept to fixation.

Conclusion: Systematists have known for some time that it is important to have multiple representatives of each clade to accurately infer phylogeny. This study underscores this importance at the species level. Selective sweeps of completely neutral mutations are unlikely to dramatically alter phylogenetic inference, particularly at the higher level. However, data presented here suggest there is both a codon bias toward A- and T-ending codons and a non-neutral pattern of amino acid replacements. These results may affect the rate of evolution and, hence, phylogenetic inference. As a consequence, phylogenetic studies of closely related species

employing mtDNA should include multiple representatives of some terminal taxa and consider the potential for heterogeneity in evolutionary rate.

The non-neutrality of mtDNA has severe implications for interpreting population parameters estimated from polymorphism data, such as the effective population size and migration rate. We present evidence that a variety of evolutionary forces including slightly deleterious or diversifying selection, parasite induced selective sweeps and/or reduced gene flow, may lead to an incorrect interpretation of heterozygosity. A large amount of variation may indicate diversifying selection. A low level may indicate the presence of selection, not small population size. We propose that it is essential to consider the action of selection in all future studies employing mtDNA as a population genetic marker.

We would like to thank MICHEL SOLIGNAC for supplying the *D. simulans si*I and *si*III lines and MICHAEL TURELLI for supplying the *D. simulans si*II R and W lines. Thanks to HIROSHI AKASHI for discussions leading to the test of codon usage. BRIAN CHARLESWORTH programmed the EWENS (1973) and WATTERSON (1978) tests. We also recognize comments made by KIRRIE BALLARD and two anonymous reviewers. Financial support to J.W.O.B. was supplied by a C. J. Martin Postdoctoral Fellowship from the Australian National Health and Medical Research Council. The work is supported by NIH grant GM39355 to MK.

LITERATURE CITED

AQUADRO, C. F., N. KAPLAN and N. RISKO, 1984 An analysis of the dynamics of mammalian mitochondrial DNA sequence evolution. Mol. Biol. Evol. **5:** 423–434.

BABA-AÏSSA, F., and M. SOLIGNAC, 1984 La plupart des populations de *Drosophila simulans* ont problement pour ancêtre une femelle unique dans un passé récent. C. R. Acad. Sci. Paris **299:** 289–292.

BABA-AÏSSA,F., M. SOLIGNAC, N. DENNEBOUY and J. R. DAVID, 1988 Mitochondrial DNA variability in *Drosophila simulans*: quasi absence of polymorphism within each of the three cytoplasmic races. Heredity **61:** 419–426.

BEGUN, D. J., and C. F. AQUADRO, 1993 African and North American populations of *Drosophila*: are very different at the DNA level. Nature **365:** 548–550.

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.

BROWN, W. M., M. GEORGE and A. C. WILSON, 1979 Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. USA **76:** 1967–1971.

BROWN, W. M., E. M. PRAGER, E. M. WANG and A. C. WILSON, 1982 Mitochondrial DNA sequence of primates: Tempo and mode of evolution. J. Mol. Evol **18:** 225–239.

CLAREY, D. O., and D. R. WOLSTENHOLME, 1984 A cluster of six tRNA genes in *Drosophila* mitochondrial DNA that includes a gene for an unusual tRNA ser AGY. Nucleic Acids Res. **12:** 2367–2379.

CLAREY, D. O., and D. R. WOLSTENHOLME, 1985 The mitochondrial molecule of *Drosophila yakuba*: nucleotide sequence, gene organization and genetic code. J. Mol. Evol. **22:** 252–271.

CLARK, A. G., 1990 Inference of haplotypes from PCR-amplified samples of diploid populations. Mol. Biol. Evol. **7:** 111–122.

DESALLE, R., and A. R. TEMPLETON, 1988 Founder effects and the rate of mitochondrial DNA evolution in Hawaiian *Drosophila*. Evolution **42:** 1076–1084.

DESALLE, R., L. VAL GIDDINGSAND and A. R. TEMPLETON, 1986 Mitochondrial DNA variability in natural populations of Hawaiian *Drosophila*. J. Mol. Evol. **26:** 157–164.

DI RAGO, J.-P., P. NETTER and P. P. SLONIMSKI, 1990 Pseudo-wild type revertants from inactive apocytochrome *b* mutants as a tool for the analysis of the structure/function relationships of the mitochondrial ubiquinol-cytochrome *c* reductase of *Saccharomyces cerevisiae*. J. Biol. Chem. **265:** 3332–3339.

EFRON, B., 1982 The jackknife, the bootstrap, and other resampling plans. Conf. Board Math. Sci. Soc. Ind. Appl. Math. **38:** 1–92.

EWENS, W. J., 1973 Testing for increased mutation rate for neutral alleles. Theor. Popul. Biol. **4:** 251–258.

EXCOFFIER, L., 1990 Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. J. Mol. Evol. **30:** 125–139.

FAITH, D. P., 1991 Cladistic permutation tests for monophyly and nonmonophyly. Syst. Zool. **40:** 366–375.

FAITH, D. P., and P. CRANSTON, 1991 Could a data set this short have arisen by chance alone? On permutation tests for cladistic structure. Cladistics **7:** 1–28.

FELSENSTEIN, J., 1985 Confidence limits on phylogenies: an approach using bootstrap. Evolution **39:** 783–791.

GARESSE, R., 1988 *Drosophila melanogaster* mitochondrial DNA: gene organization and evolutionary implications. Genetics **118:** 649–663.

HALE, L. R., and R. S. SINGH, 1987 Mitochondrial DNA and genetic structure in populations of *Drosophila melanogaster*. Mol. Biol. Evol. **4:** 622–637.

HALE, L. R., and R. S. SINGH, 1991 A comprehensive study of genetic variation in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation and the role of history *vs.* selection in the genetic structure of geographic populations. Genetics **129:** 103–117.

HIGUCHI, R. G., and H. OCHMAN, 1989 Production of single stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. **17:** 5865.

HOFFMANN, A. A., 1988 Partial cytoplasmic incompatibility between two Australian populations of *Drosophila melanogaster*. Entomol. Exp. Appl. **48:** 61–67.

HOFFMANN, A. A., and M. TURELLI, 1988 Unidirectional incompatibility in *Drosophila simulans*: inheritance, geographic variation and fitness effects. Genetics **119:** 435–444.

HOFFMANN, A. A., M. TURELLI and G. M. SIMMONS, 1986 Unidirectional incompatibility between populations of *Drosophila simulans*. Evolution **40:** 692–701.

HOFFMANN, A. A., M. TURELLI and L. G. HARSHMAN, 1990 Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*. Genetics **126:** 933–948.

HOFFMANN, A. A., D. J. CLANCY and E. MORTON, 1994 Cytoplasmic incompatibility in Australian populations of *Drosophila melanogaster*. Genetics **136:** 993–999.

HOWELL, N., 1989 Evolutionary conservation of protein regions in the proton-motive cytochrome *b* and their possible roles in redox catalysis. J. Mol. Evol. **29:** 157–169.

HOWELL, N., and K. GILBERT, 1988 Mutational analysis of the mitochondrial cytochrome *b* gene. J. Mol. Biol. **29:** 157–169.

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

IRWIN, D. M., T. D. KOCHER and A. C. WILSON, 1991 Evolution of the cytochrome *b* gene of mammals. J. Mol. Evol. **32:** 128–144.

KANEKO, M., Y. SATTA, E. T. MATSUURA and S. I. CHIGUSA, 1993 Evolution of the mitochondrial ATPase 6 gene in *Drosophila*: Unusually high level of polymorphism in *D. melanogaster*. Genet. Res., Camb. **61:** 195–204.

KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

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

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

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.

MADDISON, W. P., and D. R. MADDISON, 1992 MacClade: analysis of phylogeny and character evolution, Version 3.0. Sinauer Associates, Sunderland, Mass.

MARJORAM, P., and P. DONNELLY, 1994 Pairwise comparisons of mitochondrial DNA sequences in subdivided populations and implications for early human evolution. Genetics **136:** 673–683.

MARTIN, A. P., and S. R. PALUMBI, 1993 Protein evolution in different cellular environments: cytochrome *b* in sharks and mammals. Mol. Biol. Evol. **10:** 873–891.

MAYNARD-SMITH, J., and J. HAIGH 1974 The hitchhiking effect of a favorable gene. Genet. Res. **23:** 23–25.

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

MEYER, A., and A. C. WILSON, 1990 Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. J. Mol. Evol. **31:** 359–364.

MONNEROT, M., M. SOLIGNAC and D. WOLSTENHOLME, 1990 Discrepancy in divergence of the mitochondrial and nuclear genomes of *Drosophila teissieri* and *Drosophila yakuba*. J. Mol. Evol. **30:** 500–508.

MONTCHAMP-MOREAU, C., J.-F. FERVEUR and M. JACQUES, 1991 Geographic distribution of three cytoplasmic incompatibility types in *Drosophila simulans*. Genetics **129:** 399–407.

NACHMAN, M. W., S. N. BOYER and C. F. AQUADRO, 1994 Contrasting levels of amino acid polymorphism and divergence at the mitochondrial ND3 gene in mice. Evolution (in press).

OHTA, T., 1973 Slightly deleterious mutant substitutions in evolution. Nature **246:** 96–98.

OHTA, T., 1993 Amino acid substitution at the *Adh* locus of *Drosophila* is facilitated by small population size. Proc. Natl. Acad. Sci. USA. **90:** 4548–4551.

O'NEILL, S. L., and T. L. KARR, 1990 Bidirectional incompatibility between conspecific populations of *Drosophila simulans*. Nature **384:** 178–180.

ROGERS, A. R., and H. HARPENDING, 1992 Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **93:** 552–559.

ROUSSET, F., D. VAUTRIN and M. SOLIGNAC, 1992 Molecular identification of *Wolbachia*, the agent of cytoplasmic incompatibility in *Drosophila simulans*, and variability in relation with host mitochondrial types. Proc. R. Soc. Lond. B **247:** 163–168.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. G. HIGUCHI *et al.*, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SATTA Y., and N. TAKAHATA, 1990 Evolution of *Drosophila* mitochondrial DNA and the history of the *melanogaster* subgroup. Proc. Natl. Acad. Sci. **87:** 9558–9562.

SATTA, Y., H. ISHIWA and S. I. CHIGUSA, 1987 Analysis of nucleotide substitutions of mitochondrial DNAs in *Drosophila melanogaster* and its sibling species. Mol Biol. Evol. **4:** 638–650.

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.

SINGH, R. S., and L. R. RHOMBERG 1987a A comprehensive study of genetic variation in natural populations of *Drosophila melanogaster*. I. Estimates of gene flow from rare alleles. Genetics **115:** 313–322.

SINGH, R. S., and L. R. RHOMBERG 1987b A comprehensive study of genetic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and causes of geographic differentiation. Genetics **115:** 313–322.

SOKAL, R. R., and F. J. ROHLF, 1981 *Biometry: The Principles and Practice of Statistics in Biological Research*, Ed. 2. W. H. Freeman & Co., New York.

SOLIGNAC, M., and M. MONNEROT, 1986 Race formation, speciation, and introgression within *Drosophila simulans, D. mauritiana*, and *D. sechellia* inferred from mitochondrial DNA analysis. Evolution **40:** 531–539.

SOLIGNAC, M., V. VAUTRIN and F. ROUSSET, 1994 Widespread occurrence of the proteobacteria *Wolbachia* and partial cytoplasmic incompatibility in *Drosophila melanogaster*. C. R. Acad. Sci. Paris **317:** 461–470.

SOLIGNAC, M., M. MONNEROT and J.-C. MOUNOLOU, 1986 Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Evol. **23:** 31–41.

SWOFFORD, D. L., 1993 PAUP: phylogenetic analysis using parsimony, Version 3.1. Computer program distributed by the Illinois Natural History Survey, Champaign, Ill.

TAJIMA, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis. Genetics **135:** 599–607.

TURELLI, M., and A. A. HOFFMANN, 1991 Rapid spread of an inherited incompatibility factor in California *Drosophila*. Nature **353:** 440–442.

TURELLI, M., A. A. HOFFMANN and S. W. MCKECHNIE, 1992 Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. Genetics **132:** 713–723.

WALSH, S. P., D. A. METZGER and R. HIGUCHI, 1991 Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques **10:** 506–513.

WATTERSON, G. A., 1978 The homozygosity test of neutrality. Genetics **88:** 405–417.

WHITTAM, T. S., A. G. CLARK, M. STONEKING, R. L. CANN and A. WILSON, 1986 Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. Proc. Natl. Acad. Sci. USA **83:** 9611–9615.

Communicating editor: A. G. CLARK