Excess Amino Acid Polymorphism in Mitochondrial DNA: Contrasts Among Genes from *Drosophila*, Mice, and Humans

David M. Rand and Lisa M. Kann

Department of Ecology and Evolutionary Biology, Brown University

Recent studies of mitochondrial DNA (mtDNA) variation in mammals and *Drosophila* have shown an excess of amino acid variation within species (replacement polymorphism) relative to the number of silent and replacement differences fixed between species. To examine further this pattern of nonneutral mtDNA evolution, we present sequence data for the ND3 and ND5 genes from 59 lines of *Drosophila melanogaster* and 29 lines of *D. simulans.* Of interest are the frequency spectra of silent and replacement polymorphisms, and potential variation among genes and taxa in the departures from neutral expectations. The *Drosophila* ND3 and ND5 data show no significant excess of replacement polymorphism using the McDonald-Kreitman test. These data are in contrast to significant departures from neutrality for the ND3 gene in mammals and other genes in *Drosophila* mtDNA (cytochrome *b* and ATPase 6). Pooled across genes, however, both *Drosophila* and human mtDNA show very significant excesses of amino acid polymorphism. Silent polymorphisms at ND5 show a significantly higher variance in frequency than replacement polymorphisms, and the latter show a significant skew toward low frequencies (Tajima's $D = -1.954$). These patterns are interpreted in light of the nearly neutral theory where mildly deleterious amino acid haplotypes are observed as ephemeral variants within species but do not contribute to divergence. The patterns of polymorphism and divergence at charge-altering amino acid sites are presented for the *Drosophila* ND5 gene to examine the evolution of functionally distinct mutations. Excess charge-altering polymorphism is observed at the carboxyl terminal and excess charge-altering divergence is detected at the amino terminal. While the mildly deleterious model fits as a *net* effect in the evolution of nonrecombining mitochondrial genomes, these data suggest that opposing evolutionary pressures may act on different regions of mitochondrial genes and genomes.

Introduction

A number of recent reports have shown that the patterns of variation in mitochondrial DNA (mtDNA) are not consistent with several different predictions of the neutral theory of molecular evolution (Whittam et al. 1986; Excoffier 1990; Ballard and Kreitman 1994; Nachman, Boyer, and Aquadro 1994; Rand, Dorfsman, and Kann 1994; Nachman et al. 1996). These findings are of considerable importance to evolutionary biologists using mtDNA as a genetic marker, since most of the applications of, and inferences from, mtDNA assume neutrality of the marker. Unappreciated departures from neutrality will cloud our understanding of genetic variation within and among populations and compromise our interpretations of evolutionary events in the history of organisms (see fig. 4 in Rand, Dorfsman, and Kann 1994). Beyond simply providing evidence for or against selection, one hope is that tests of neutrality may shed light on general mechanisms of molecular evolution.

A prediction of the neutral theory is that the ratio of replacement (nonsynonymous) to silent (synonymous) differences observed between species should equal the replacement : silent ratio observed within species (McDonald and Kreitman 1991). In their original paper, McDonald and Kreitman (1991) observed an excess of replacement substitutions *between* species of *Drosophila* at the *Adh* locus, and interpreted this as evidence for adaptive fixations of new amino acid variants. A similar pattern of excess amino acid fixations has been reported by Eanes, Kirchner, and Yoon (1993) at the *G6pd* locus in *Drosophila melanogaster* and *D. simulans* (although other genes exhibit neutral evolution using the McDonald-Kreitman test; Brookfield and Sharp 1994). Both *Adh* and *G6pd* exhibit protein polymorphisms that appear to be maintained by some form balancing selection (Oakeshott et al. 1982; Kreitman and Aguadé 1986; Eanes 1994).

Quite a different pattern has been documented recently in the mitochondrial ND3 gene of mice. Nachman, Boyer, and Aquadro (1994) found that the replacement : silent ratio within *M. domesticus* is almost 10 times higher than the replacement : silent ratio observed fixed between *Mus domesticus* and *M. spretus.* Nachman, Boyer, and Aquadro interpret their data in light of the nearly neutral theory of molecular evolution (Ohta 1992*a*): amino acid variants in *M. domesticus* are weakly deleterious and may accumulate within species, but do not persist long enough to contribute to interspecific divergence. Comparable data have recently been reported for humans and African apes (Nachman et al. 1996). Similar but less striking patterns have been reported in *Drosophila* (ATPase 6 gene: Kaneko et al. 1993; ND5 gene: Rand, Dorfsman, and Kann 1994; cytochrome *b* gene: Ballard and Kreitman 1994).

While the departures from neutrality for the nuclear and mitochondrial examples cited above are opposite in direction, this has more to do with the mode of selection than any general distinction between nuclear and mitochondrial genes (Li and Sadler [1991] found an excess of amino acid polymorphism in human nuclear genes). Examples of neutral evolution (Schaeffer and Miller 1992), balancing selection (Kreitman and Hudson 1991; Eanes 1994) and selective sweeps (Aguadé et al. 1989;

Key words: mitochondrial DNA, neutral theory, natural selection, mildly deleterious mutations, genetic variation, *Drosophila,* human.

Address for correspondence and reprints: David M. Rand, Department of Ecology and Evolutionary Biology, Brown University, Box G, Providence, Rhode Island 02912. E-mail: david_rand@brown.edu.

Mol. Biol. Evol. 13(6):735–748. 1996

Begun and Aquadro 1991, 1993; Berry, Ajioka, and Kreitman 1991) have been reported for nuclear genes. These modes of evolution may be operating among, and even within, individual mitochondrial genes. However, the lack of recombination in animal mtDNA imposes constraints on how individual genes may exhibit the patterns of polymorphism that could indicate a history of distinct types of selection. An analogous problem has emerged from polymorphism studies in regions of low recombination in the *Drosophila* nuclear genome: current theory and data have not provided a clear means of discriminating between the effects of deleterious background selection (e.g., Charlesworth, Morgan, and Charlesworth 1993) and positive selective sweeps in the generation of reduced variation in regions of low recombination in *Drosophila* (e.g., Aguadé et al. 1989, Berry, Ajioka, and Kreitman 1991; Begun and Aquadro 1992; Hudson 1994; Braverman et al. 1995; but see Aquadro, Begun, and Kindahl 1994).

The purpose of this report is to study the patterns of polymorphism and divergence at silent and replacement sites in different mitochondrial genes. First, we present data for the ND3 gene in *Drosophila melanogaster* and *D. simulans* to provide a direct comparison to this gene in mammals. Next we present new data for the ND5 gene that more than double the data from an earlier study (Rand, Dorfsman, and Kann 1994). These data are compared to silent and replacement polymorphism data from the cytochrome *b* gene in *Drosophila* (Ballard and Kreitman 1994) and to recent data from complete mtDNA sequences in humans (Horai et al. 1995; Nachman et al. 1996). While individual genes generally show an excess of amino acid polymorphism, there is variation among genes, and between flies and mammals, in the significance of departures from neutrality. When polymorphism data are tabulated from all genes, both flies and hominids show a general pattern of a significant excess of amino acid polymorphism (or a reduction in the amount of amino acid fixation). The ND5 data show a significant excess of low-frequency amino acid polymorphisms. The data are consistent with a mildly deleterious model of molecular evolution for mtDNA (Ohta 1992*a*) which may be the "net" evolutionary effect on the nonrecombining mtDNA as a whole. However, the patterns of polymorphism and divergence for charge-altering amino acids in the ND5 gene suggest that some pressure for adaptive divergence may be acting in conflict with forces eliminating variation.

Materials and Methods

Fly Strains

Nine strains of *D. melanogaster* were chosen from diverse localities around the world. These lines are described in Rand, Dorfsman, and Kann (1994) and were obtained from Dr. Rama Singh and Dr. Andrew Clark. Five additional random samples of 10 lines each of *D. melanogaster* were obtained from Dr. C. F. Aquadro: Arvin, California (Arv2, Arv3, Arv8, Arv11, Arv12, Arv13, Arv15, Arv19, Arv21, Arv23); Australia (Aus1, Aus2, Aus3, Aus4, Aus5, Aus6, Aus7, Aus10, Aus11, Aus12); Beijing, China (Bei1, Bei2, Bei3, Bei4, Bei5, Bei6, Bei7, Bei8, Bei9, Bei10); Florida (Flo3, Flo4, Flo6, Flo9, Flo18, Flo20, Flo28, Flo32, Flo34, Flo35); and Zimbabwe, Africa (Zim3, Zim5, Zim6, Zim7, Zim8, Zim10, Zim11, Zim18, Zim22, Zim24).

Nine strains of *D. simulans* were obtained from the National Drosophila Species Resource Center (Bowling Green State University, Bowling Green, Ohio) and are described in Rand, Dorfsman, and Kann (1994). Two additional random samples of 10 lines each were obtained from Dr. C. F. Aquadro: Florida (FloSim1, FloSim2, FloSim3, FloSim4, FloSim5, FloSim6, FloSim8, FloSim10, FloSim12, FloSim13) and Zimbabwe (ZimSim22, ZimSim24, ZimSim25, ZimSim26, ZimSim27, ZimSim28, ZimSim29, ZimSim33, ZimSim34, ZimSim35).

DNA Preparation and Sequencing

Sequences for the ND5 gene were obtained for all of the above lines. For the ND3 gene, all of the *D. simulans* lines were sequenced, but the Australia lines, the California lines, Flo28, and Zim 24 of *D. melanogaster* were not sequenced. Hence the sample size for ND3 was 37 *D. melanogaster* and 29 *D. simulans*. Template for PCR amplification was prepared from single flies using a "squish prep" (Gloor and Engels 1991). Flies were homogenized with a sterile pipette tip holding 50 μL of 10 mM Tris (pH 8.2), 1 mM EDTA, and 25 mM NaCl and Proteinase K added to a concentration of 200 μg/mL. The resulting homogenate was incubated at 37°C for 20–30 min and the proteins were denatured by heat treatment at 95°C for 2–3 min.

Sequencing templates for the ND3 gene were prepared by PCR-amplification of a fragment that includes the entire ND3 gene. A 660-bp fragment was amplified with the 20-base primers 196R and 818L (see table 1). Two additional internal primers were used in the sequencing reactions: 252R and 797L. The primers used for amplification of the ND5 gene were 880R with 2230L and 1280R with 3017L as described Rand, Dorfsman, and Kann (1994). Additional sequencing primers spanning the ND5 gene are listed in table 1.

DNA was amplified in 50-μL reactions containing 1 μL of fly homogenate (see above), 5 μL of Promega Mg-free 10X buffer, 6 μL of 25 mM $MgCl_2$, 50 pmole of each primer, and 2 units of Promega *Taq* polymerase. Double-stranded products were cleaned with phenol: chloroform, precipitated with ammonium acetate and ethanol at room temperature, rinsed with cold 70% ethanol, dried, and resuspended in T.E. (10 mM Tris [pH 7.5], 0.1 mM EDTA). The PCR products were used in three sequencing reactions. An aliquot of double-stranded template with 50 pmol of sequencing primer was heated to 100°C for 3 min and transferred to a dry-ice–ethanol bath. These "snap-cooled" templates were melted in the presence of labeling mixture and transferred to the chain termination reactions (U.S. Biochemicals with Sequenase version 2.0). Sequencing gels were run "short" with sodium acetate in the lower buffer chamber (Sheen and Sneed 1988) resolving sequence from

Table 1
Primers Used for Amplifying and Sequencing the ND3 and ND5 Genes

Primer Name	Sequence (5′ → 3′)
ND3 primers:	
196R	GAAGCAGCTGCATGATATTG
252R	CACAATTTACTGATGAGGAG
797L	AACTAATTGCAATCAATCGC
818L	AATTACCTAAGATTAGGTCG
ND5 primers:	
880R	CCAAAAAGAGGCATATCACT
947R	ATCAAGTAAAAGCTGCTAAC
1109R	CTCCATAACATCTTCAATGTC
1280R	GACCTCCAAAATATTCTGAT
1261L(1280rev)	ATCAGAATATTTTGGAGGTC
1600R	AATCCTATTATACCACGGAG
1581L(1600rev)	CTCCGTGGTATAATAGGATT
1914L	TGTGCTGGGGCTATTATTCA
1932R(1913rev)	GTGAATAATAGCCCCAGCAC
2230L	AGCTATAGCTGCTCCTACAC
2299R	GCTCTTTTAGTTATAGCAGC
2509	AGGGTGAGATGGTTTAGGAC
2528R(2509rev)	GTCCTAAACCATCTCACCCT
3017L	TAGAAGAGGTAAAATTCGAG
3289L	CGTCTTGGAAGTCGAAGAAT

NOTE.—The number refers to the 3′ nucleotide of the primer based on the published sequence of Garesse (1988). The letter indicates the direction of elongation (Left, Right) with respect to figure 2 of Garesse (1988).

about 50 bp through 350 bp from the primer. "Long" gel runs without sodium acetate generally resolved sequence from 200 bp through 500 bp from the same primer. The sequences were determined in both directions.

Sequence Analysis

Sequences were aligned by eye with no ambiguity. Variable positions within and between species were tabulated as silent (S) or replacement (R). The amino acid (replacement) differences within and between species were further classified as nonpolar (NP) or uncharged polar (UP) (see Lehninger 1975, pp. 73–75). Amino acid changes within these categories were classified as "conservative" and changes between these categories were classified as "radical." Tests for departure from the neutral expectation were performed using a *G*-test with Williams' correction (Sokal and Rohlf 1981, pp. 735–742).

An index of neutrality is also presented ("Neutrality Index," N.I., in table 6). This index reflects the extent to which the levels of amino acid variation within species depart from the strict neutral model. It is calculated as a ratio of ratios:

$$\text{N.I.} = \frac{\dfrac{\text{no. of polymorphic replacement sites}}{\text{no. of fixed replacement sites}}}{\dfrac{\text{no. of polymorphic silent sites}}{\text{no. of fixed silent sites}}}.$$

Strict neutrality (equal ratios of polymorphism to divergence for both replacement and silent sites) has an index of 1.0. Values greater than one indicate an "excess" of amino acid variation within species, whereas values less than 1 indicate an "excess" of amino acid evolution between species (the term "excess" is relative since one's focus may change from silent vs. replacement to polymorphism vs. divergence). The index is intended to provide a qualitative indicator of direction and degree.

Results

ND3 Gene

The alignment of the most common haplotypes of the ND3 gene from *D. melanogaster* and *D. simulans* is presented in figure 1. In the 351-bp coding region (117 amino acids; the stop codon is not included in these analyses) there are 17 silent differences (4.8%) and 2

FIG. 1.—Sequence alignment of the ND3 gene in *Drosophila melanogaster* and *D. simulans*. The coding strand of *D. melanogaster* (Dm) sequence is shown, and numbering is after Garesse (1988). The amino acid sequence is shown below the nucleotide sequence with the single letter code aligned under the first codon position. Nucleotide differences fixed in *D. simulans* (Ds) are shown above the coding strand, and those substitutions that result in amino acid changes in *D. simulans* are indicated below the amino acid sequence for *D. melanogaster*.

replacement differences (0.56%) between the common haplotypes in each species (M1 and S1 in table 2). The various sequence haplotypes observed at the ND3 gene in our sample of 37 *D. melanogaster* and 29 *D. simulans* are presented in table 2. If one considers polymorphic sites, the number of unambiguously fixed differences between the species is 13 silent and 2 replacement changes (see tables 2 and 6). Taken together, these patterns of polymorphism and divergence do not depart from the neutral expectation (see table 6).

The data reveal a very slight excess of amino acid variation (N.I. = 1.3) when polymorphism from both species of *Drosophila* is considered. There are no replacement polymorphisms in *D. melanogaster* and a single one in *D. simulans* (N.I. = 0.0 and 4.0, respectively; both not significant). Data for polymorphism and divergence at the ND3 gene in mice (*Mus domesticus* vs. *M. spretus*; Nachman, Boyer, and Aquadro 1994) and hominids (humans vs. chimps; Nachman et al. 1996) are also presented in table 6. The data for mice show a strong and significant excess of amino acid polymorphism. The data for polymorphism within both humans and chimps, relative to divergence between these species, show a significant departure from neutrality, although a test using only human polymorphism is not significant (G_{adj} = 2.962, $0.05 < P < 0.10$; Nachman et al. 1996). These McDonald-Kreitman tests for the ND3 gene in mammalian species show considerably higher levels of excess amino acid variation (N.I. = 9.7, 6.2, and 4.43 for mice, hominids, and humans only, respectively) than do the comparable tests in flies. If corrections for multiple hits are allowed, the difference is even greater since the corrections have a greater effect on the number of silent differences between species (see Nachman, Boyer, and Aquadro 1994). Since the uncorrected tests are conservative, the raw data are presented.

ND5 Gene

Our earlier study found no overall departure from neutrality using a McDonald-Kreitman test, although a departure from neutrality was found in a localized region of the protein (Rand, Dorfsman, and Kann 1994, fig. 2). Fifty additional sequences have been added: three samples of 10 lines each of *D. melanogaster,* and two samples of 10 lines each of *D. simulans*. The haplotypes observed in these new samples, combined with the earlier data from Rand, Dorfsman, and Kann (1994), are reported in tables 3 and 4.

There are 21 polymorphic sites in the ND5 sequences of *D. melanogaster* (13 silent and 8 replacement; table 3) and 7 polymorphic sites in those of *D. simulans* (4 silent and 3 replacement; table 4). In *D. melanogaster,* the DNA haplotypes differ by up to five nucleotides (M1 and M3) while the protein haplotypes differ by at most three amino acids (MP8 vs. all others; table 3). In *D. simulans,* the haplotypes differ by at most six nucleotides and three amino acids. These data are consistent with earlier results (Rand, Dorfsman, and Kann 1994) indicating considerably more mtDNA polymorphism in *D. melanogaster* than in *D. simulans* (θ = 0.00298 ± 0.0064 for *D. melanogaster* and θ = 0.00118

Table 2
Polymorphic Nucleotide and Amino Acid Sites in the ND3 Gene of *Drosophila melanogaster* and *D. simulans*

D. MELANOGASTER

DNA Haplotype[b]	N	Locality[c]	Variable Nucleotide Positions[a] 377	550	676
S/R C/R[d] . . .			S	S	S
Consensus			C	G	G
M1	27	World, Z, Fl	•	•	•
M2	7	B	•	•	A
M3	2	Arg, Fl	•	A	•
M4	1	NA, SA	T	•	•

D. SIMULANS

DNA Haplotype[b]	N	Locality[c]	Variable Nucleotide Positions 392	433	544	Protein Haplotype[b]	N	Variable Amino Acid Position 392
S/R C/R[d]			R	S	S			C
Consensus			G	C	C			V
S1	13	World, Z, Fl	•	•	•	SP1	18	•
S2	11	Z	A	•	•	SP2	11	I
S4	5	NA, NG, SA	•	T	T			

NOTE.—N = sample size.
[a] Nucleotide and amino acid positions are numbered according to Garesse (1988) and figure 1.
[b] Haplotypes are identified with a letter (M for *melanogaster*, S for *simulans*) and a number. Protein haplotypes have a P added before the number.
[c] Arg = Argentina, B = Beijing, Fl = Florida, M = Mexico, NA = North America, NG = New Guinea, SA = South America, World = Diverse sample (see text), Z = Zimbabwe.
[d] For nucleotide sites, S = silent, R = replacement change; for amino acids, C = conservative, R = radical.

Table 3
Polymorphic Nucleotide and Amino Acid Sites in the ND5 Gene of *Drosophila melanogaster*

			Variable Nucleotide Positions[a]																					Variable Amino Acid Positions								
DNA Haplotype[b]	*N*	Locality[c]	1	2	2	2	5	6	6	6	8	8	9	1 0	1 0	1 1	1 1	1 1	1 2	1 3	1 3	1 4	1 4	Protein Haplotype[b]	*N*	1 6	2 1	5 2	6 8	1 8	3 0	4 6	4 7
			6	1	2	4	2	5	8	8	1	4	3	5	6	2	3	8	3	0	6	4	7										
			1	8	2	0	9	7	2	7	3	0	0	3	2	2	4	1	9	6	7	2	9			1	8	9	2	1	6	7	2
S/R C/R[d]			R	R	S	S	R	S	R	S	S	S	S	S	S	S	S	R	S	R	R	R	S			R	C	C	C	R	R	R	C
Consensus			G	T	A	A	T	A	T	G	T	A	A	G	C	A	C	G	G	G	A	T	T			S	N	S	I	T	P	I	N
M1	23	NA, A, B	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	MP1	46	•	•	•	•	•	•	•	•
M2	13	Z, E, A, B	•	•	•	G	•	•	•	•	•	•	•	•	•	G	•	•	•	•	•	•	•										
M3	7	B	•	•	•	G	•	•	•	•	C	•	•	A	•	G	•	•	•	•	•	C	•	MP2	7	•	•	•	•	•	•	•	S
M4	2	Fl	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	A	•	•	•	•										
M5	1	Z	•	•	•	G	•	•	•	•	•	G	•	•	•	G	•	•	•	•	•	•	•										
M6	1	Z	•	•	•	G	•	•	•	•	•	G	•	•	T	•	•	•	•	•	•	•	•										
M7	1	Z	•	C	•	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	MP3	1	•	S	•	•	•	•	•	•
M8	1	CAf	•	•	G	G	•	•	•	•	•	•	G	•	•	G	•	•	•	•	•	•	•										
M9	1	Fl	•	•	•	G	•	•	•	•	•	•	•	•	•	G	T	•	•	•	•	•	•										
M10	1	Fl	•	•	•	•	•	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•										
M11	1	Fl	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	MP4	1	F	•	•	•	•	•	•	•
M12	1	Fl	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	MP5	1	•	•	•	•	I	•	•	•
M13	1	Fl	•	•	•	•	•	•	•	A	•	•	•	•	•	•	•	•	•	•	•	•	•										
M14	1	Fl	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	G										
M15	1	Cal	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	A	•	•	•	MP6	1	•	•	•	•	•	T	•	•
M16	1	Arg	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	G	•	•	MP7	1	•	•	•	•	•	•	T	•
M17	1	J	•	•	•	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•										
M18	1	I	•	•	•	G	C	•	C	•	C	•	•	•	•	•	•	•	•	•	•	•	•	MP8	1	•	•	G	V	•	•	•	•

Note.—Data are for 59 sequences: 29 reported in Rand, Dorfsman, and Kann (1994) plus 30 new sequences from this study. *N* = sample size.

[a] Nucleotide and amino acid positions are numbered according to figure 1 of Rand, Dorfsman, and Kann (1994).

[b] DNA haplotypes are identified with an M (for *melanogaster*) and a number. Protein haplotypes are identified with an MP (for *melanogaster* protein) and a number.

[c] A = Australia, Arg = Argentina, B = Beijing, CAf = Central Africa, Cal = California, E = Europe, Fl = Florida, I = India, J = Japan, NA = North America, Z = Zimbabwe.

[d] For nucleotide sites, S = silent, R = replacement change; for amino acid sites, C = conservative, R = radical change.

Table 4
Polymorphic Nucleotide and Amino Acid Sites in the ND5 Gene of *Drosophila simulans*

DNA Haplotype[b]	N	Locality[c]	Variable Nucleotide Positions[a]							Protein Haplotype	N	Variable Amino Acid Positions		
					1	1	1	1	1				1	1
					3	3	3	4	5				3	5
				9	1	3	3	7	1			9	3	1
			3	1	4	2	5	9	5			1	5	5
S/R C/R[d] . . .			S	R	S	S	R	S	R			C	C	C
Consensus . .			A	C	A	A	A	A	C			L	I	L
S1	10	Z	•	•	•	•	•	•	A	SP1	10	•	•	F
S2	8	World	G	•	•	•	•	•	•	SP2	14	•	•	•
S3	6	Fl, M	•	•	•	•	•	•	•					
S4	4	NA, SA	G	•	G	G	T	G	•	SP3	4	•	M	•
S5	1	Fl	G	A	G	G	T	G	•	SP4	1	V	M	•

Note.—Data are from 29 sequences: 9 reported in Rand, Dorfsman, and Kann (1994) plus 20 new sequences from this study.

[a] Nucleotide and amino acid positions are numbered according to figure 1 of Ran, Dorfsman, and Kann (1994).

[b] DNA haplotypes are identified with an S (for *simulans*) and a number. Protein haplotypes are identified with an SP (for *simulans* protein) and a number.

[c] Fl = Florida, M = Mexico, NA = North America, SA = South America, Z = Zimbabwe.

[d] For nucleotide sites, S = silent, R = replacement change; for amino acid sites, C = conservative, R = radical change.

± 0.0003 for *D. simulans*), despite similar levels of maximal haplotypic divergence within each species (see tables 3–5).

While the indices of neutrality indicate an excess of amino acid polymorphism (N.I. > 2), no significant departures from neutrality are detected in the complete ND5 data set (see table 6). This result holds if one considers polymorphism in a single species relative to divergence from a single haplotype of the other species (data for 59 *D. melanogaster*: 15 fixed replacement (FR) and 52 fixed silent (FS) sites vs. 8 polymorphic replacement (PR) and 13 polymorphic silent (PS) sites (G_{adj} = 1.036, N.S.; N.I. = 2.13; data for 29 *D. simulans*: FR = 15; FS = 52; PR = 3; PS = 4; G_{adj} = 1.166, N.S.; N.I. = 2.60).

The different samples of flies from distinct geographic localities have different levels of nucleotide polymorphism and clear patterns of population differentiation (Rand, Dorfsman, and Kann 1994; unpublished data; see tables 3 and 4). If one excludes the sample of *D. melanogaster* from Florida, a significant excess of amino acid polymorphism is detected with the McDonald-Kreitman test. The remaining 49 *D. melanogaster* lines vs. 29 *D. simulans* lines show FR = 15, FS = 55, PR = 11, and PS = 13 (G_{adj} = 4.87, $P < 0.05$; N.I. = 3.10). However, a "balanced" comparison of 10 lines each from the Florida and Zimbabwe samples from both species revealed no departure from neutrality (FR = 15; FS = 52; PR = 6; PS = 12; G = 0.87; G_{adj} = 0.42, $P < 0.52$; N.I. = 1.73). If data from the ND3 and ND5 genes in *Drosophila* are combined into a single McDonald-Kreitman test, there is no significant departure from neutrality (FR = 17; FS = 65; PR = 12; PS = 21; G_{adj} = 2.850, N.S.; N.I. = 2.18).

ND5 sequences from three humans and the common chimpanzee are available from the complete sequences of three human mitochondrial genomes (Horai et al. 1995; an African, a Japanese, and the "Cambridge" sequence [Anderson et al. 1981]). There is an apparent excess of amino acid polymorphism at ND5 in humans (N.I. = 2.09), but the McDonald-Kreitman test is not significant (table 6).

Table 5
Nucleotide Diversity at the ND5 Gene in *D. melanogaster* and *D. simulans*

	No. of Sites	S[a]	Pn[b]	π[c]	θ[d]	Tajima's D[e]	H[f]
D. melanogaster (n = 59)							
All sites	1,515	21	0.01386	0.00149	0.00298	−1.564	0.7802
Silent sites	284	13	0.04577	0.00638	0.00985	−1.029	0.7234
Replacement sites	1,231	8	0.00650	0.00036	0.00140	−1.954*	0.3763
D. simulans (n = 29)							
All sites	1,515	7	0.00462	0.00149	0.00118	0.784	0.7420
Silent sites	284	4	0.01408	0.00493	0.00359	0.968	0.5898
Replacement sites	1,231	3	0.00244	0.00068	0.00062	0.212	0.6278

Note.—* indicates significance at the 0.05 level (Simonsen, Churchill, and Aquadro 1995).

[a] S = the number of variable sites.

[b] Pn = the proportion of variable sites.

[c] π = heterozygosity based on average pairwise proportion of differences between sequences.

[d] θ = heterozygosity based on variable sites (Watterson 1975).

[e] Tajima's D is from Tajima (1989).

[f] $H = 1 - \Sigma x_i^2$, where x_i is the sample frequency of the ith haplotype.

Table 6
Silent and Replacement Differences Within and Between Species

	Drosophila ND3[a]		Mammalian ND3[b]			
	Fixed	Polymorphic	Fixed	Polymorphic	Fixed	Polymorphic
	37 mel. vs. 29 *si*II sim.		*Mus domesticus* vs. *M. spretus*		61 Humans vs. 5 Chimps	
Replacement	2	1	2	11	4	8
Silent	13	5	23	13	31	10
Neutrality Index	1.30		9.73		6.20	
G_{adj}	0.031, N.S.		9.263, $P < 0.002$		6.747, $P < 0.02$	
	Drosophila ND5		Hominid ND5[c]			
	59 mel. vs. 29 *si*II sim.		3 Humans vs. Chimp			
Replacement	15	11	51	5		
Silent	52	17	143	7		
Neutrality Index	2.24		2.00			
G_{adj}	2.659, N.S.		1.179, N.S.			
	Drosophila Cyt b[d]		Hominid Cyt[c]			
	17 mel. vs. 16 *si*II sim.		3 Humans vs. Chimp			
Replacement	1	4	27	5		
Silent	47	7	102	5		
Neutrality Index	26.86		3.78			
G_{adj}	8.505, $P < 0.005$		3.494, N.S.			
	Drosophila ND3 + ND5 + Cyt b		Hominid ND5 + ND3 + Cyt b		All Mitochondrial Proteins[c]	
	59 (17) mel. vs. 29 (16) *si*II sim.		3 Humans vs. Chimp		3 Humans vs. Chimp	
Replacement	18	16	80	21	179	31
Silent	112	29	268	25	915	55
Neutrality Index	3.43		2.81		2.88	
G_{adj}	9.446, $P < 0.002$		9.716, $P < 0.01$		17.519, $P < 0.0005$	

[a] mel. = *D. melanogaster*; sim. = *D. simulans*.

[b] Mammalian ND3 data for mice (56 *Mus domesticus* vs. 1 *Mus spretus*) are from Machman, Boyer, and Aquadro (1994), and data for humans are from Nachman et al. (1996).

[c] Hominid ND3, ND5, Cyt b, and all mitochondrial protein data are from Horai et al. (1995) and Nachman et al. (1996).

[d] Drosophila Cyt b data are from Ballard and Kreitman (1994); values reported here reflect only the *D. melanogaster* haplotypes and the *si*II haplotypes for *D. simulans*.

Cytochrome *b* Gene

Ballard and Kreitman (1994) reported a significant excess of amino acid polymorphism in the *D. melanogaster* subgroup for cytochrome *b* (FR = 10; FS = 97; PR = 6; PS = 12; $G_{adj} = 5.846$, $P < 0.05$; N.I. = 4.85). This comparison, however, scored the differences between the *si*I, *si*II, and *si*III mtDNA haplotypes of *D. simulans* as polymorphisms. If one uses only the *si*II data from Ballard and Kreitman's study, the McDonald-Kreitman test shows a stronger departure from neutrality (table 6). Since this latter sample of flies from Ballard and Kreitman's (1994) study is most comparable to that from Rand, Dorfsman, and Kann (1994) and the current study, there does appear to be a difference between the ND5 and cytochrome *b* genes in the degree to which they depart from neutral expectations. Notably, in cytochrome *b* there is only a single amino acid difference fixed between *D. melanogaster* and *D. simulans si*II, but there are four polymorphic amino acid sites. This leads to a very high neutrality index (26.86). If the ND3, ND5, and cytochrome *b* data are pooled, the combined data show a significant excess of amino acid polymorphism (table 5). Using the published sequences of three complete human mtDNAs (Horai et al. 1995; Nachman et al. 1996), the cytochrome *b* gene does not depart from neutrality (table 6).

All Protein-Coding Genes in Human mtDNA

If one combines the ND3, ND5, and cytochrome *b* genes from humans, there is a significant departure from neutrality (table 6). Pooling all the protein-coding genes for three mtDNAs (African, Cambridge, and Japanese; Horai et al. 1995; Nachman et al. 1996), the McDonald-Kreitman test is highly significant (table 6). Figure 2 shows the patterns of silent and replacement polymorphism (per silent or replacement site, respectively) at each of the human mitochondrial genes. Also plotted is the neutrality index for each gene. Two general points should be noted: (1) Silent polymorphism, and the neutrality index, varies considerably across genes, and (2) the neutrality index is greater than 1.0 for 11 of the 12 genes showing replacement polymorphism (N.I. for ND4 = 0.93, the CO2 and ND4L genes show no replacement polymorphism, and only CO1 and ND6 show

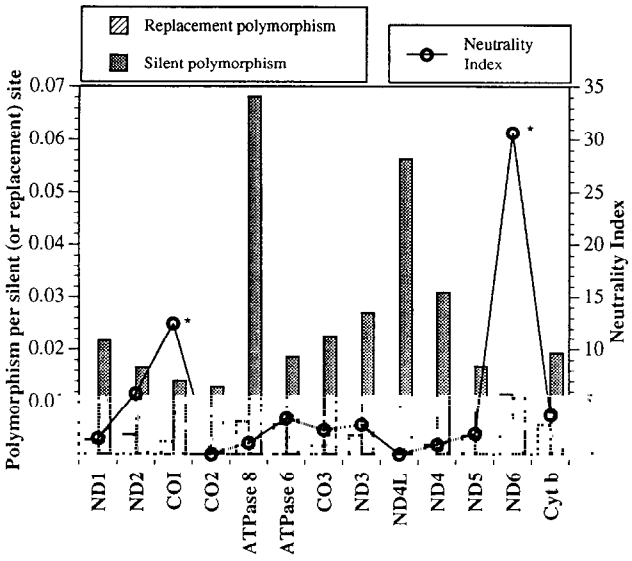

FIG. 2.—Silent, replacement and neutrality variation among genes in human mtDNA. Silent and replacement polymorphism is calculated as the number of variable silent (replacement) sites per effectively silent (replacement) site (see Li and Graur 1991). The neutrality index is described in the Materials and Methods. The asterisk (*) indicates a significant excess of amino acid polymorphism using the McDonald-Kreitman Test ($P < 0.05$). Sequence data are from Horai et al. (1995) and Nachman et al. (1996).

significant excesses of replacement polymorphism). Hence, in this sample of three complete sequences, excess amino acid variation is a general trend for human mtDNA, and the level of this excess varies among genes.

Nucleotide Site Frequencies and Diversity in *D. melanogaster*

One prediction of the nearly neutral theory (Ohta 1992*a*) is that mildly deleterious polymorphisms should not reach appreciable frequencies in the population, whereas neutral polymorphisms should range from rare to intermediate frequencies. As a result there should be a lower variance in frequency of mildly deleterious variants. The frequency spectra of silent and replacement nucleotide sites provide possible tests of these predictions. Figure 3 shows the frequencies of silent and replacement nucleotide positions in the ND5 gene of *D. melanogaster* and *D. simulans*. The frequency of the highest-frequency silent site is significantly greater than that of replacement sites (95% confidence limits are nonoverlapping; Rohlf and Sokal 1981, table 23). The silent sites also show significantly higher variance in frequency than replacement sites ($\text{Var(p)}_{\text{silent}} = 0.020$; $\text{Var(p)}_{\text{replacement}} = 0.0013$; $F_{13,8} = 16.023$, $P < 0.001$).

There is an excess of low-frequency polymorphisms at silent sites at the ND5 gene in *D. melanogaster*, but it is not significant (Tajima's (1989) $D = -1.0292$; see table 5). At replacement sites, however, there is a significant excess of low-frequency polymor-

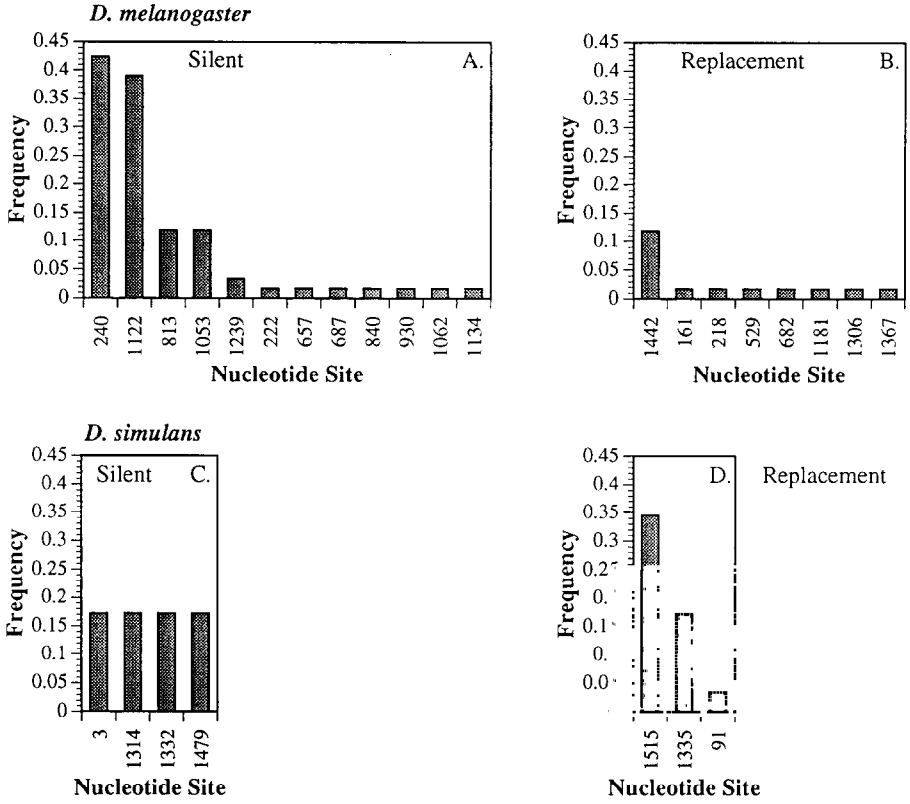

FIG. 3.—Frequency spectra for silent and replacement polymorphisms in the ND5 gene of *D. melanogaster* and *D. simulans*. (*A*) Silent sites, and (*B*) replacement sites in *D. melanogaster*. (*C*) Silent sites and (*D*) replacement sites in *D. simulans*.

phisms as revealed by Tajima's test ($D = -1.9549$, $P < 0.05$). The haplotypic diversity at ND5 in *D. melanogaster* is greater for silent haplotypes than for replacement haplotypes, consistent with stronger purifying selection at replacement sites (table 5). While the presence of fixed differences at both silent and replacement sites between geographic populations of *D. melanogaster* (unpublished data; see table 3) clouds the interpretation of these statistics, such patterns should *increase* Tajima's *D*. Hence the negative Tajima's *D*'s reported may be conservative.

Nucleotide Site Frequencies and Diversity in *D. simulans*

The patterns of nucleotide frequencies and diversity at the ND5 gene in *D. simulans* are very distinct from those in *D. melanogaster* (see fig. 3). Tajima's *D* is positive for both silent and replacement sites, and the haplotypic diversity is greater for replacement sites than for silent sites (table 5). While neither of these comparisons of silent and replacement sites are significant, these patterns are opposite in direction from those in *D. melanogaster*. Given the low level of nucleotide variability in the mtDNA of *D. simulans* (Baba-Aïsa et al. 1988; Hale and Singh 1991*a*, 1991*b*; Ballard and Kreitman 1994; Rand, Dorfsman, and Kann 1994), it is difficult to compare the frequency spectra in the two species. Moreover, a history of selection and the presence of two subtypes within the *si*II haplotype (Rand, Dorfsman, and Kann 1994) have left a more complex backdrop of nucleotide polymorphism upon which the footprint of mildly deleterious variation can be inferred.

Polymorphism and Divergence at Silent, Replacement, and Amino Acid Charge-Altering Sites

The patterns of polymorphism and divergence at silent (synonymous) and replacement (nonsynonymous) sites in the ND5 gene in *D. melanogaster* and *D. simulans* are plotted in a "sliding window" analysis in figure 4. The data were generated by sliding a window spanning approximately 300 nucleotide sites across the data set of 59 *D. melanogaster* and 29 *D. simulans* sequences. Figure 4*A* extends our earlier findings of variable levels of fixed differences between species across the ND5 gene for these two types of sites. One new replacement polymorphism has been uncovered near the carboxyl terminal of the *D. melanogaster* and *D. simulans* genes (positions 1442-Beijing and 1515-Zimbabwe, respectively). These polymorphisms strengthen the departure from neutrality in this region of the ND5 gene (left side of fig. 2 in Rand, Dorfsman, and Kann 1994; fig. 4 of this report).

If one considers only the amino acid replacement sites, these can be classified as "conservative" or "radical" changes with respect to amino acid charge or polarity (see Materials and Methods). Figure 4*B* presents the sliding window analysis for these two types of changes that are fixed between *D. melanogaster* and *D. simulans* at the ND5 gene. In the amino terminal region of the protein (right side of fig. 4), the majority of amino acid substitutions result in charge or polarity differences. This qualitative observation is not significant using a McDonald-Kreitman-type test (using the notation where the first letter [P or F] indicates polymorphic within or fixed between species and the second letter [R or C] indicates radical or conservative with respect to charge or polarity, from position 2105 to 2839, FR = 8, FC = 5, PR = 1, PC = 2, $G = 0.79$, N.I. = 0.31).

Figure 4*C* presents a sliding window analysis of the polymorphic nucleotide sites within *D. melanogaster*. Consistent with our earlier results, there is a slight excess of replacement (nonsynonymous) polymorphism and a slight excess of radical amino acid polymorphism in the carboxyl terminal end of the gene (left side of fig. 4). This excess is not significant (from position 1340 to 2090, FR = 1, FC = 2, PR = 3, PC = 1, $G = 1.24$, N.I. = 6.0). However, the amino and carboxyl halves of the ND5 gene show very different patterns of conservative and radical amino acid variation. From position 1340 to 2090 there are three polymorphisms and one fixed difference classified as radical; from position 2105 to 2839 there is one polymorphism and eight fixed differences classified as radical. This contrast is significant ($G_{adj} = 4.486$, $P < 0.05$). It appears as if the amino terminal half of the ND5 gene has experienced accelerated radical amino acid divergence, while the carboxyl terminal half exhibits excess radical amino acid polymorphism. It is difficult to account for these patterns by invoking a single evolutionary force.

Discussion

Polymorphism data for mitochondrial genes from both *Drosophila* and humans show a significant excess of amino acid polymorphism relative to that expected from the patterns of divergence at silent (synonymous) and replacement (nonsynonymous) sites (*cf.* McDonald and Kreitman 1991). This general statement for pooled data from multiple genes is supported qualitatively by polymorphism data at individual genes in *Drosophila* and human mtDNA. Of the 18 genes discussed in this study (4 in *Drosophila*, 1 in mice, and 13 in humans), 15 show an excess of amino acid polymorphism over that expected from a strictly neutral model, although significant excesses are not observed in all cases. In flies, the ND3 and ND5 data are not significant while the ATPase 6 and cytochrome *b* data are significant (Kaneko et al. 1993; Ballard and Kreitman 1994, respectively; see table 6). In humans only the ND6 and COI data are significant (the ND3 data are significant when chimpanzee polymorphism is included; Nachman et al. 1996). Thus, two general questions need to be addressed: (1) How does one account for an excess of amino acid polymorphism in mtDNA and (2) how does one account for apparent differences between genes and among taxa in the degree of departure from neutral models?

Excess amino acid variation could be due to some form of balancing selection acting on mtDNA. While theory suggests that it is difficult to maintain mtDNA variation by selection (Clark 1984; Gregorius and Ross 1984), some forms of frequency-dependent selection

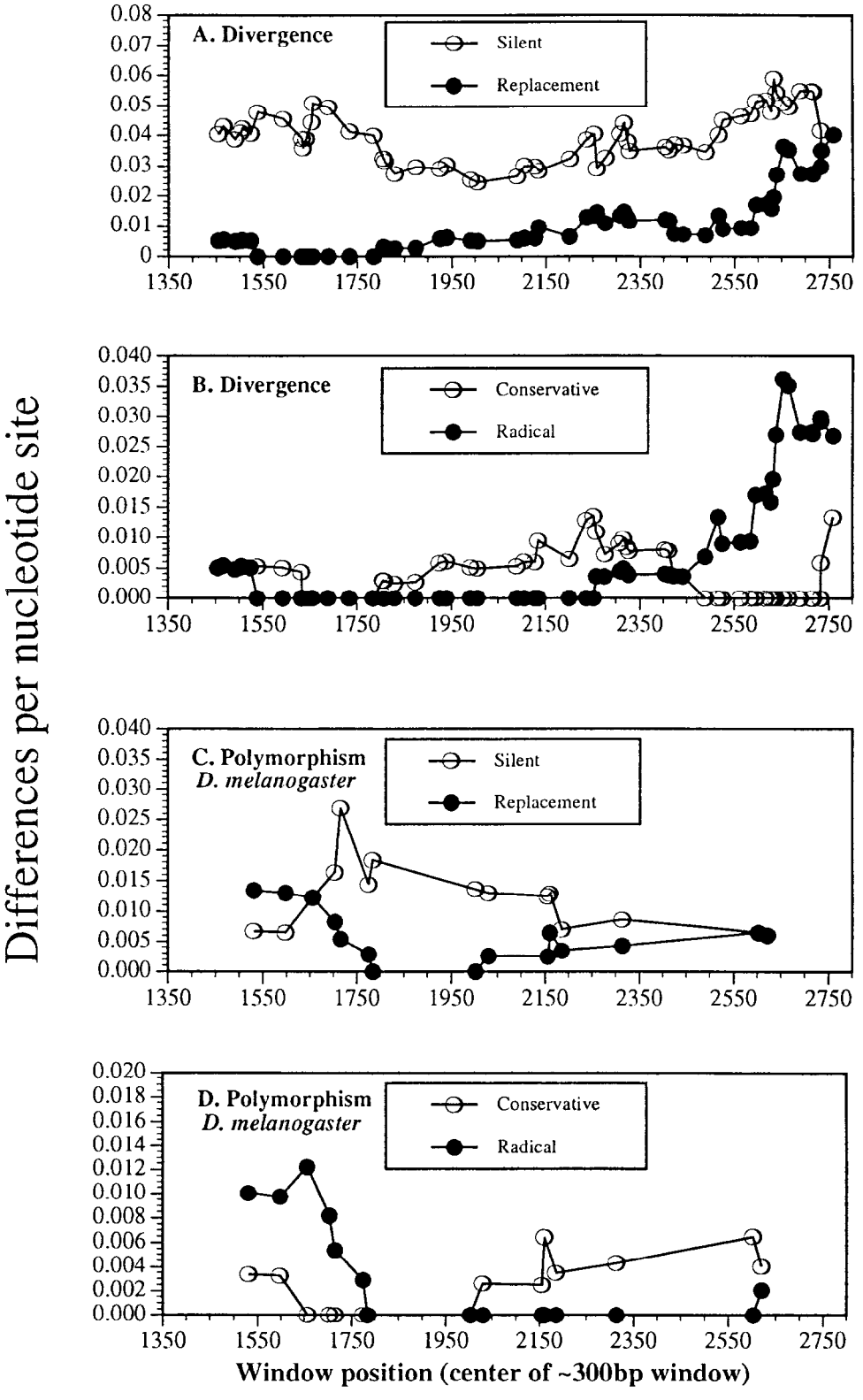

FIG. 4.—Sliding window analysis of nucleotide and amino acid changes in the ND5 gene. A window of approximately 300 bp was moved across the complete data set of 59 *D. melanogaster* sequences and 29 *D. simulans* sequences and nucleotide positions were scored as silent (synonymous) or replacement (nonsynonymous) with respect to the mitochondrial code. For those sites exhibiting amino acid changes, the sites were scored as either conservative with respect to charge (conservative) or charge-altering (radical). (*A*) nucleotide divergence between *D. melanogaster* and *D. simulans*; open dots represent silent divergence, solid dots represent replacement divergence. (*B*) Amino acid divergence between *D. melanogaster* and *D. simulans* with open dots indicating conservative, and solid dots representing radical changes. (*C*) Nucleotide polymorphism in *D. melanogaster* only, with open and solid dots as in *A*. (*D*) Amino acid polymorphism in *D. melanogaster* only with open and solid dots as in *C*. The numbering of the X-axis is according to Garesse (1988) and figure 2 of Rand, Dorfsman, and Kann (1994). See table 1 and figure 1 of Rand, Dorfsman, and Kann (1994) for conversion to a reversed order. Note the different scales on the vertical axes.

(Brooks and Curtsinger 1996) or marginal "overdominance" (e.g., Levene 1953) might maintain amino acid polymorphism in mtDNA. If such modes of selection are (or have been) operating, one should see an excess of high-frequency polymorphism in DNA sequence data. Moreover, one might expect some replacement polymorphism to be at high frequency. The data from *D. melanogaster* are not consistent with this prediction since replacement sites exhibit a significant excess of low-frequency polymorphisms (silent sites have a negative Tajima's *D,* but it is not significant; table 5). In *D. simulans,* Tajima's *D*'s are positive for silent and replacement sites, although neither is significant. The slight excess of high-frequency polymorphisms is due to the presence of two distinct subtypes within the *si*II mtDNA haplotype (Rand, Dorfsman, and Kann 1994). While this pattern suggests balancing selection in *D. simulans* mtDNA, there is a deficiency of mtDNA polymorphism in this species resulting from some history of selective sweeps or background selection, possibly mediated by an endosymbiont (Ballard and Kreitman 1994; Rand, Dorfsman, and Kann 1994). It is difficult to accommodate balancing selection in the face of reduced variation in a nonrecombining genome. Overall, it seems unlikely that balancing selection has played a significant role in generating the pattern of excess amino acid variation in the mtDNA *D. melanogaster* and *D. simulans.*

An alternative explanation is that amino acid polymorphisms are mildly deleterious and accumulate as short-lived polymorphism within species, but do not persist in the population long enough to become fixed and contribute to sequence divergence between species. This "nearly neutral" argument (Ohta 1992*a,* 1992*b*) has been invoked to account for excess amino acid variation in the mtDNA of mice and humans (Nachman, Boyer, and Aquadro 1994; Nachman et al. 1996). Under this mildly deleterious scenario, mtDNA polymorphism is governed by negative selection coefficients whose absolute values are approximately the reciprocal of the effective population size for mtDNA (N_{ef}; the effective breeding number of females). As such, the selection coefficients are neither sufficiently weak such that replacement mutations would drift as effectively neutral variants (i.e., $|s| \ll 1/N_{ef}$), nor sufficiently strong such that replacement mutations would be eliminated rapidly from the population (i.e., $|s| \gg 1/N_{ef}$). Notably, the variation in selection coefficients is a crucial factor in determining "nearly neutral" dynamics as this is indicative of the constancy of selection (Ohta 1992*a*). Tachida (1991) defined the nearly neutral range as $3 \geq 4N\sigma^2 \geq 0.2$ (for nuclear genes), where N is effective population size and σ^2 is the standard deviation of the selection coefficient.

Under the nearly neutral theory, effective population size will alter the fixation probability and levels of polymorphism at functionally constrained sites (DeSalle and Templeton 1988; Ohta 1992*a*; Lynch 1996). A potentially powerful test of the mildly deleterious model would be to compare levels of silent and replacement polymorphism in related species with distinctly different effective population sizes (N_e). Due to differences in the efficacy of selection, a negative relationship between N_e and the rate of amino acid divergence is predicted (Ohta 1995), and this should hold for amino acid polymorphisms as well (see Akashi [1995] for an example of presumably deleterious silent sites). It is suggestive that replacement (and total) polymorphism in *D. simulans* is lower than that of *D. melanogaster* at the ND5 gene (see table 5), given the lower N_e of the latter species (e.g., Aquadro, Lado, and Noon 1988). However, the presence of two distinct subtypes of mtDNA *and* reduced variation within the *si*II haplotype in *D. simulans* presents problems for the simple explanation invoking more effective background selection in *D. simulans.* Sequence surveys from other species in the *D. melanogaster* subgroup may help in testing the correlation between effective population size and silent vs. replacement polymorphism.

The frequencies of silent and replacement polymorphisms may also shed light on the mildly deleterious model of evolution. While certain "silent" sites are undoubtedly under some functional constraints (e.g., Akashi 1995; unpublished data), on average, silent polymorphisms are more likely to be governed by neutral forces than replacement polymorphisms. As such, silent polymorphisms should drift to higher frequencies, exhibit a wider range of frequencies, and show a less extreme Tajima's *D* than replacement sites. These patterns are observed in *D. melanogaster* (fig. 3 and table 5).

Despite the apparent fit of the data to the nearly neutral model, Nachman et al. (1996) note that it is difficult to attribute the excess of amino acid polymorphism to mildly deleterious evolution vs. a recent relaxation of selection. Given that *Drosophila* and mice are commensal with humans, a relaxation of selection seems plausible in the recent evolutionary pasts of these species. A recent expansion of population size, possibly associated with relaxed selection, could also generate a skew toward low-frequency polymorphisms (Rogers and Harpending 1992). In either case, it is unclear whether "low" frequencies in a sample of 59 sequences are consistent with the evolutionary effects implied under "deleterious" or "relaxed" selection. Alternative models, and combinations of models, may better account for the data (Gillespie 1994, 1995). Theoretical studies contrasting the values of Tajima's *D* for silent vs. replacement sites under various models of mildly deleterious evolution vs. relaxed selection would be very informative in light of the recent empirical data.

Variation Among Genes

Why should some genes show significant excess of amino acid polymorphism and others not? Some of the variation among genes in the significance of the McDonald-Kreitman tests may relate to the different effects of linkage on polymorphism vs. sequence divergence. The lack of recombination in mtDNA should impose constraints such that the patterns of polymorphism exhibited by individual genes are governed by the *net* effect of (for example) diversifying vs. purifying selection across mtDNA. Birky and Walsh (1988) observed, however, that linkage does not affect rates of divergence (see also Begun and Aquadro 1992). Some of the among-

gene variation (e.g., fig. 2) may be caused by these different consequences of linkage. This effect could also explain the contrasting patterns of charge-altering polymorphism and divergence within the ND5 gene that have the appearance of effects opposing selective forces (e.g., fig. 4). Additional sequence surveys of other genes are needed to determine whether comparable patterns at the among-gene level are common, whether there is a high sampling variance to the 2×2 tests, or whether excess amino acid polymorphism is the rule for all mitochondrial genes (see table 6).

Variation Among Taxa

Why should the same gene in different organisms (ND3 in flies and mammals) show such dramatically different levels of amino acid polymorphism? A possible explanation is a difference in mutation rate stemming from different thermal habits (e.g., Martin, Naylor, and Palumbi 1992; Rand 1994). Both the rate of ND3 evolution and the level of nucleotide polymorphism is higher in mice and humans than in flies (see above and Nachman, Boyer, and Aquadro 1994). If mutation were the primary cause, one would expect the contrast between flies and mammals to carry over to all mitochondrial genes. While the data are limited, this is not generally the case (compare the ND5 and cytochrome *b* genes in flies and humans; table 6). Mutation rate alone is by no means the primary effect of differences in metabolic rate (Martin and Palumbi 1993). Notably, Adachi, Cao, and Hasegawa (1993) provide support for changes in functional constraint, rather than mutation, as an explanation for faster mtDNA evolution in mammals. Thus, the contrast between flies and mammals at ND3 may reflect subtle changes in the function of homologous genes in different thermal contexts.

The different patterns at ND3 in flies and mammals may also relate to differences in effective population size. Amino acid variants that might drift as neutral or nearly neutral variants in small deme-structured populations of mice could be effectively eliminated from populations of *Drosophila* where N_e may be two orders of magnitude larger (e.g., Ohta 1992*a*, 1992*b*). While this explanation is appealing in light of the nearly neutral theory, this pattern should also carry over to all genes in mtDNA. Variation among genes in the levels of excess amino acid variation presents a problem for a simple explanation based on more effective selection. The present data are too limited to rigorously test the relationship between excess amino acid polymorphism and effective population size. However, a compilation of RFLP data from mtDNA does suggest a possible relationship. Avise (1992, fig. 9) shows for a variety of marine species that the discrepancy between organismal population size and N_e estimated from mtDNA polymorphism increases as organismal population sizes increase. More effective selection against mildly deleterious variation (or stronger selective sweeps) could account for this pattern.

In conclusion, it appears that levels of mtDNA variation within species cannot be predicted accurately from the patterns of divergence between species, and that selection is in part responsible for this discrepancy. Further contrasts of silent and replacement polymorphisms in other species promise to be very informative about mtDNA evolution.

Acknowledgments

We would like to thank Michael Nachman and Chip Aquadro for providing a copy of their manuscript prior to publication. Their comments, and those of an anonymous reviewer, improved the manuscript. Supported by NSF grant BSR 9120299 to D.M.R.

LITERATURE CITED

ADACHI, J., Y. CAO, and M. HASEGAWA. 1993. Tempo and mode of mitochondrial DNA evolution in vertebrates at the amino acid sequence level: rapid evolution in warm-blooded vertebrates. J. Mol. Evol. **36**:270–281.

AGUADE, M., N. MIYASHITA, and C. H. LANGLEY. 1989. Reduced variation in the *yellow-achete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122**: 607–615.

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

ANDERSON, S., A. T. BANKIER, B. G. BARELL, M. H. DE BRUIJN, A. R. COULSON, F. SANGER, P. H. SCHREIER, A. J. H. SMITH, R. STADEN, and G. YOUNG. 1981. Sequence and organization of human mitochondrial genome. Nature **290**: 457–465.

AQUADRO, C. F., D. BEGUN, and E. KINDAHL. 1994. Selection, recombination and DNA polymorphism in *Drosophila*. Pp. 46–65 *in* B. GOLDING, ed. Non-neutral evolution. Chapman and Hall, New York.

AQUADRO, C. F., K. M. LADO, and W. A. NOON. 1988. The *rosy* region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. Genetics **119**: 1165–1190.

AVISE, J. C. 1992. Molecular population genetic structure and the biogeographic history of a regional fauna: A case history with lessons for conservation biology. Oikos **63**:62–76.

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.

BALLARD, J. W. O., and M. E. KREITMAN. 1994. Unraveling selection in the mitochondrial genome of *Drosophila*. Genetics **138**:757–772.

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 **356**:519–520.

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.

BIRKY, C. W., and J. B. WALSH. 1988. Effects of linkage on rates of molecular evolution. Proc. Natl. Acad. Sci. USA **85**:6414–6418.

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.

BROOKFIELD, J. F. Y., and P. M. SHARP. 1994. Neutralism and selectionism face up to DNA data. Trends Genet. **10**:109–111.

BROOKS, L. D., and J. CURTSINGER. 1996. The amount of variation maintained by frequency-dependent selection in haploids. Genetics (in press).

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

CLARK, A. 1984. Natural selection with nuclear and cytoplasmic transmission. I. A deterministic model. Genetics **107**: 679–701.

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

EANES, W. 1994. Patterns of polymorphism and species divergence in the enzymes of central metabolism. Pp. 18–28 *in* B. GOLDING, ed. Non-neutral evolution. Chapman and Hall, New York.

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.

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.

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

GILLESPIE, J. 1994, Substitution processes in molecular evolution. III. Deleterious alleles. Genetics **138**:943–952.

———. 1995. On Ohta's hypothesis: most amino acid substitutions are deleterious. J. Mol. Evol. **40**:64–69.

GLOOR, G., and W. R. ENGELS. 1991. Single fly DNA preps for PCR. Drosophila Information Newsletter. Vol. 1. Available on the Internet at gopher://ftp.bio.indiana.edu/11/flybase/news.

GREGORIUS, H. R., and M. D. ROSS. 1984. Selection with gene-cytoplasm interaction. I. Maintenance of cytoplasm polymorphisms. Genetics **107**:165–178.

HALE, L. R., and R. S. SINGH. 1991*a*. A comprehensive study of genic 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.

———. 1991*b*. Contrasting patterns of genetic structure and evolutionary history as revealed by mitochondrial DNA and nuclear gene–enzyme variation between *Drosophila melanogaster* and *Drosophila simulans*. J. Genet. **70**:79–90.

HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE, and N. TAKAHATA. 1995. Recent African origin of humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA. **92**:532–536.

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

KANEKO, M., Y. SATTA, E. T. MATSUURA, and S. 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.

KREITMAN, M. E., and M. AGUADÉ. 1986. Excess polymorphism at the *Adh* locus in *Drosophila melanogaster*. Genetics **114**:93–110.

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.

LEHNINGER, A. H. 1975. Biochemistry, 2nd edition. Worth Publishers, New York.

LEVENE, H. 1953. Genetic equilibrium when more than one ecological niche is available. Am. Naturalist **87**:331–333.

LI, W.-H., and D. GRAUR. 1991. Fundamentals of molecular evolution. Sinauer Associates, Sunderland, Mass.

LI, W.-H., and L. A. SADLER. 1991. Low nucleotide diversity in man. Genetics **129**:513–123.

LYNCH, M. 1996. Mutation accumulation in transfer RNAs: molecular evidence for Muller's Ratchet in mitochondrial genomes. Mol. Biol. Evol. **13**:209–220.

MARTIN, A. P., G. J. P. NAYLOR, and S. R. PALUMBI. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature **357**:153–155.

MARTIN, A. P., and S. R. PALUMBI. 1993. Body size, metabolic rate, generation time and the molecular clock. Proc. Natl. Acad. Sci. USA **90**:4087–4091.

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

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

NACHMAN, M. W., S. N. BOYER, and C. F. AQUADRO. 1994. Non-neutral evolution at the mitochondrial ND3 gene in mice. Proc. Natl. Acad. Sci. USA **91**:6364–6368.

NACHMAN, M. W., W. M. BROWN, M. STONEKING, and C. F. AQUADRO. 1996. Excess amino acid polymorphism in mitochondrial DNA of humans and chimpanzees. Genetics (in press).

OAKESHOTT, J. G., J. B. GIBSON, P. R. ANDERSON, W. R. KNIBB, D. G ANDERSON, and G. K. CHAMBERS. 1982. Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. Evolution **36**:86–96.

OHTA, T. 1992*a*. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Systematics **23**:263–286.

———. 1992*b*. Theoretical study of near neutrality. II. Effect of subdivided population structure with local extinction and recolonization. Genetics **130**:917–923.

———. 1995. Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. **40**:56–63.

RAND, D. M. 1994. Thermal habit, metabolic rate and the evolution of mitochondrial DNA. Trends. Ecol. Evol. **9**:125–131.

RAND, D. M., M. DORFSMAN, and L. M. KANN. 1994. Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. Genetics **138**:741–756.

ROGERS, A. R., and H. HARPENDING. 1992. Population growth makes waves in the distribution of pairwise differences. Mol. Biol. Evol. **9**:552–569.

ROHLF, J. F., and R. SOKAL. 1981. Statistical tables, 2nd ed. W. H. Freeman, San Francisco.

SCHAEFFER, S. W., and E. L. MILLER. 1992. Molecular population genetics of an electrophoretically monomorphic protein in the alcohol dehydrogenase region of *Drosophila melanogaster*. Genetics **132**:163–178.

SHEEN, J., and B. SNEED. 1988. Electrolyte gradient gels for DNA sequencing. BioTechniques **6**:942–944.

SIMONSEN, K. L., G. A. CHURCHILL, and C. F. AQUADRO. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141**:413–429.

SOKAL, R., and J. F. ROHLF. 1981. Biometry, 2nd ed. W. H. Freeman, San Francisco.

TACHIDA, H. 1991. A study on a nearly neutral mutation model in finite populations. Genetics **128**:183–192.

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

WATTERSON, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7**:256–276.

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.

CHARLES F. AQUADRO, reviewing editor

Accepted February 20, 1996