Deleterious Mutations at the Mitochondrial ND3 Gene in South American Marsh Rats (Holochilus)

Patricia Kennedy and Michael W. Nachman

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721

Manuscript received October 31, 1997
Accepted for publication June 12, 1998

ABSTRACT

Statistical analyses of DNA sequences have revealed patterns of nonneutral evolution in mitochondrial DNA of mice, humans, and Drosophila. Here we report patterns of mitochondrial sequence evolution in South American marsh rats (genus Holochilus). We sequenced the complete mitochondrial ND3 gene in 82 *Holochilus brasiliensis* and 21 *H. vulpinus* to test the neutral prediction that the ratio of nonsynonymous to synonymous nucleotide changes is the same within and between species. Within *H. brasiliensis* we observed a greater number of amino acid polymorphisms than expected based on interspecific comparisons. This contingency table analysis suggests that many amino acid polymorphisms are mildly deleterious. Several tests of the frequency distribution also revealed departures from a neutral, equilibrium model, and these departures were observed for both nonsynonymous and synonymous sites. In general, an excess of rare sites was observed, consistent with either a recent selective sweep or with populations not at mutation-drift equilibrium.

A fundamental goal of population genetics is to understand the forces that give rise to and maintain genetic variation in natural populations. The relative ease of collecting DNA sequence data has facilitated measurement of genetic variation within species (for reviews, see Avise 1994; Kreitman and Akashi 1995; Moriyama and Powell 1996; Aquadro 1997), and recent theoretical advances allow us to analyze where and how selection is acting at the molecular level (*e.g.*, Hudson *et al.* 1987; Tajima 1989; McDonald and Kreitman 1991; Fu and Li 1993; Fu 1996, 1997; McDonald 1996, 1998; Templeton 1996).

Several statistical tests of the neutral model use data from a single locus and have been used to investigate evolutionary forces acting on mitochondrial DNA (mtDNA). For example, the neutral model predicts that the ratio of replacement to silent polymorphism within species is equal to the ratio of replacement to silent fixed differences between species (McDonald and Kreitman 1991). This prediction has been tested with mitochondrial sequences from several different species including humans (Nachman *et al.* 1996; Templeton 1996; Wise *et al.* 1998), mice (Nachman *et al.* 1994), and Drosophila (Ballard and Kreitman 1994; Rand *et al.* 1994; Rand and Kann 1996). Common to most of these studies is the observation of an excess of intraspecific amino acid polymorphism compared to the level of amino acid substitution. This pattern has been observed at ND3 in mice and humans (Nachman *et al.* 1994, 1996), COII in humans (Templeton 1996), ND2 in humans (Wise *et al.* 1998), cytochrome b in Drosophila (Ballard and Kreitman 1994), and portions of ND5 in Drosophila (Rand *et al.* 1994; Rand and Kann 1996). An excess of intraspecific amino acid polymorphism is also observed when all human mitochondrial genes are considered together (Nachman *et al.* 1996; Rand and Kann 1996); this suggests that the observed pattern is not specific to particular loci. However, some mitochondrial genes, such as ND3 in Drosophila, appear to fit the predictions of a neutral model (Rand and Kann 1996).

At least three hypotheses may explain the excess of intraspecific amino acid polymorphisms seen at mitochondrial genes. First, many amino acid mutations may be slightly deleterious (Ohta and Kimura 1971; Ohta 1992) and therefore contribute more to intraspecific heterozygosity than to interspecific divergence (Kimura 1983). Second, many amino acid mutations may be under positive selection that fluctuates in either space or time (Gillespie 1991). Third, mitochondrial genes in some species may have recently experienced a dramatic relaxation of selective constraint. Distinguishing between these hypotheses may be aided by extending these analyses to other species. For example, if the same pattern is found in several unrelated taxa that experience different selective pressures, the hypothesis of a recent relaxation of constraint becomes increasingly improbable.

We are interested in exploring the generality of the nonneutral patterns observed in mtDNA. Here we report mtDNA sequence variation within and between two species of South American marsh rats in the genus Holochilus. Marsh rats are semi-aquatic rodents that live

Corresponding author: Patricia Kennedy, Department of Ecology and Evolutionary Biology, Biosciences West Bldg., University of Arizona, Tucson, AZ 85721. E-mail: pkennedy@u.arizona.edu

in marshes and along stream banks throughout wet, lowland regions of South America (Hershkovitz 1955). The ND3 gene was sequenced so that direct comparisons could be made with the ND3 data collected in other taxa. As in previous studies of ND3 in mice and humans, we observed an excess of intraspecific amino acid variation. We argue that these results are best explained by mildly deleterious amino acid mutations.

MATERIALS AND METHODS

Samples and DNA preparation: A total of 103 marsh rats from 8 populations of *Holochilus brasiliensis* ($N = 82$) and 2 populations of *H. vulpinus* ($N = 21$) were collected along a river drainage extending from northern Paraguay to central Argentina (Figure 1), as previously described (Nachman 1992). Total genomic DNA was isolated from liver tissue using phenol-chloroform extractions according to standard protocols, and the DNA was resuspended in TE at pH 8.0 (Sambrook *et al.* 1989).

DNA amplification and sequencing: Sequencing templates were prepared using polymerase chain reaction (PCR) with primers that amplified a 488-bp fragment that included the entire ND3 gene. The primers PKND3-L9385 and PKND3-H9831 were used for both amplification and sequencing. Primer numbers refer to the position of the 3′ base in the complete mouse mitochondrial sequence of Bibb *et al.* (1981); L and H refer to the light and heavy strands, respectively.

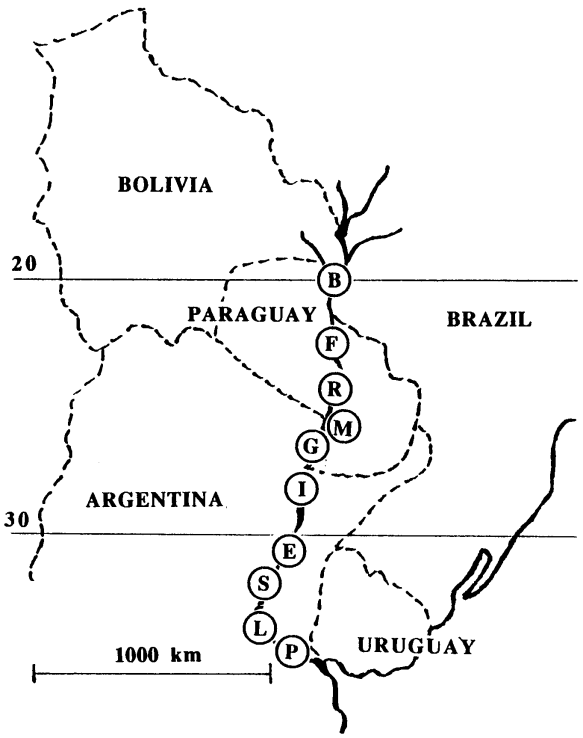

Figure 1.—Map showing collection localities for *H. brasiliensis* (eight northernmost populations) and *H. vulpinus* (two southernmost populations). The populations from north to south are Bahia Negra ($N = 9$), Fonciere ($N = 6$), Rosario ($N = 9$), Manduvira ($N = 6$), Golondrina ($N = 16$), Itati ($N = 23$), Esquina ($N = 7$), Santa Fe ($N = 6$), Las Cuevas ($N = 16$), and Puerto Ibicuy ($N = 5$).

Primer composition is as follows: PKND3-L9385, 5′-CGTYTC YATYTATTGATGAGG-3′; PKND3-H9831; 5′-CATAATCTAA TGAGTCGAAATC-3′. DNA was amplified in 50-μl reaction volumes with approximately 50 ng of template DNA using Amersham (Arlington Heights, IL) *Taq* polymerase with conditions as specified by the supplier. DNA was amplified in 40 cycles of 30 sec at 94°, 1 min at 40°, and 1 min at 72°. Amplified products were sequenced in both directions using Amersham's thermosequenase radiolabeled termination cycle sequencing kit labeled with ^{33}P. Thermocycling parameters for sequencing were identical to the amplification conditions above. Sequencing products were electrophoresed on 8.0% glycerol-tolerant acrylamide gels (Amersham). Sequences have been submitted to GenBank under accession numbers AF079374–AF079401.

Data analysis: Sequences were aligned by eye, and the numbers of replacement and silent polymorphisms and fixed differences were counted. The ND3 gene is 345 bp in length excluding the stop codon. Two different measures of nucleotide variation, π (Nei and Li 1979) and θ (Watterson 1975), were calculated from the sequence data for each species. Nucleotide diversity, π, is calculated from the average number of nucleotide differences between all pairs of sequences in a sample and θ is calculated from the number of segregating sites in a sample. Thus π takes into account the frequencies at which polymorphisms are present in the sample, while θ is based solely on the observed number of segregating sites. For mitochondrial sequences, both are estimators of the neutral parameter $2N_e\mu$, where N_e is the effective population size for females and μ is the neutral mutation rate.

To test the neutral prediction that the ratio of replacement to silent nucleotide changes is the same within and between species (McDonald and Kreitman 1991), we compared polymorphisms within *H. brasiliensis* and within *H. vulpinus* to the number of fixed differences between these species using *G* log-likelihood ratios. We compared the frequency distribution of segregating sites in our sample to those expected under a neutral model using Tajima's *D* (Tajima 1989), Fu and Li's *D* (Fu and Li 1993), and Fu's F_s (Fu 1996, 1997) statistics. Tajima's *D* is based on the difference between the number of segregating sites (θ) and the average number of pairwise nucleotide differences (π), while Fu and Li's *D* compares the distribution of mutations on internal (η_i) and external branches (η_e) of the gene tree. Both tests are based on the neutral prediction that these different estimators of $2N_e\mu$ will be the same; the expectation for both Tajima's *D* and Fu and Li's *D* under neutrality is zero. Simulations indicate that the power of these tests to reject the null model is not great unless sample sizes are quite large (*i.e.*, in excess of $N = 50$; Braverman *et al.* 1995; Simonsen *et al.* 1995); our sample of *H. brasiliensis* is large ($N = 82$), while our sample of *H. vulpinus* is considerably smaller ($N = 21$). Fu's F_s is based on the expected number of haplotypes in a sample for a given value of θ and may be more powerful for detecting population growth or hitchhiking events than Tajima's *D* or Fu and Li's *D* (Fu 1997).

RESULTS

Sequence variation: The aligned ND3 nucleotide and protein sequences are shown in Figure 2. For each species, the consensus sequence is shown and polymorphic nucleotide and amino acid sites are given in lower case.

Within *H. brasiliensis* ($N = 82$), 24 segregating sites and 25 different haplotypes were detected at the ND3 gene (Table 1). Within *H. vulpinus* ($N = 21$), 2 polymor-

```
Hv  M . . . . . L . T . T . . . L . . . . . . . . . . . . . . . . . . .
Hb  I N a L L I M L I N i L L A f L L A S V A F W L P Q P Y L Y T E K
Hb ATTAACgCCTTATTAATTATaCTAATCAACAtCCTACTGGCAtTTTTAtTAGCCTCtGTAGCATTCTGACTCCCCCAACCCTACtTATAcACAGAAAAA
Hv ..A......C.CC....CC..T...C...T.CT..T..T...C.GC..C....T.....................T..G.....TC.T.....T......
   1                                                                                                   99

Hv  . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Hb  a S P Y E C G F D P I N S A R L P F S M K F F L V G I T F L L F D
Hb gCCAGCCCCTATGAATGTGGgTTTGACCCAATCAACTCAGCCCGATTACCATTCTCAATAAAATTTTTCCTTGTAGGAATTACCTTCtTACTATTTGAC
Hv ........A...........T..C........T...........TC.T..C.................T........C..C......C.A..T..C...
   100                                                                                                 198

Hv  . . . . . . . . . . . . . . . . . . N . . Y . M . . . . . . . . . . . . . . .
Hb  L E I A L L L P I P W A M Q F T D T h T T I I T S f I L v S I L T
Hb CTCGAAATtGCTCTtCTcCTCCCAATCCCATGGGCtATACAATTCACAGACACACaCACAACtATCATTACTTCAtTcATcTTAgTCTCAATTCTAACC
Hv ..A.....C..C..A.....A...........A...........T.AT..T...T......TA........C..C..T..T........g...T.....
   199                                                                                                 297

Hv  . . . . . . . . m . . . . . . . . .
Hb  L G L A Y E W M N K G L E W t E *
Hb TTAGGCCTAGCCTATGAATGAATAAACAAGGGCtTAGAATGAaCAGAgTAG
Hv .....GT....T.........t...T..A..A...........C..A..A
   298                                          348
```

Figure 2.—Aligned nucleotide and amino acid sequences of the ND3 gene for *H. brasiliensis* (Hb) and *H. vulpinus* (Hv). The consensus sequences for each species are shown. Intraspecific polymorphic sites are indicated with lower case letters; the alternate states are presented in Tables 1 and 2. The positions of the nucleotides in the ND3 gene are numbered from 1 to 345, and the termination codon is indicated by an asterisk.

TABLE 1

Polymorphic sites at the ND3 gene for *H. brasiliensis* (N = 82)

Haplotype	Polymorphic sites 000000001112222222222333 023445890280113567778344 712397500077364414693105 * ** * * * * * GATTTTTCGGTTTCTATTCCGTAG	Distribution of haplotypes by locality B	F	R	M	G	I	E	S	Total
1	.........A.....	1	0	2	3	5	2	3	2	18
2	......C.....T....	0	0	0	0	0	16	0	0	16
3	..................T....	2	3	1	0	0	1	0	0	7
4	.............T.....T....	1	0	4	0	1	0	0	0	6
5	.........A.....A	0	0	0	0	4	0	0	1	5
6	..................A	0	0	0	0	1	2	1	0	4
7	...C.C...A.....	0	0	0	0	4	0	0	0	4
8	A........A.....	1	0	0	0	0	2	0	0	3
9	.G.......A.......C.....	0	0	0	1	0	0	0	1	2
10	.........A..C.....	0	1	1	0	0	0	0	0	2
11	.......T......C...TT...A	1	0	0	0	0	0	0	0	1
12	........A.....T....	1	0	0	0	0	0	0	0	1
13	..C......A.....T.C..	1	0	0	0	0	0	0	0	1
14	..................T...A	1	0	0	0	0	0	0	0	1
15	........A.....T...A	0	1	0	0	0	0	0	0	1
16	..................TA...	0	1	0	0	0	0	0	0	1
17	........AA.....	0	0	1	0	0	0	0	0	1
18	.............T.....T...A	0	0	0	1	0	0	0	0	1
19	....C....A.C.....	0	0	0	1	0	0	0	0	1
20	....C....A.....	0	0	0	0	1	0	0	0	1
21	.........A.....G.	0	0	0	0	0	0	1	0	1
22	..C......A.....	0	0	0	0	0	0	1	0	1
23	.G.......A.....	0	0	0	0	0	0	0	1	1
24	..........C....TC..T....	0	0	0	0	0	0	0	1	1
25	....	0	0	0	0	0	0	1	0	1

The consensus sequence for *H. brasiliensis* (N = 82) is shown. Replacement sites are indicated by an asterisk. Positions of polymorphic sites are given for *H. brasiliensis* ND3 gene; nucleotides in the ND3 gene are numbered from 1 to 348. The distribution of each haplotype, by locality, and the total number of individuals per haplotype are shown. Localities are as follows: B, Bahia Negra; F, Fonciere; R, Rosario; M, Manduvira; G, Golondrina; I, Itati; E, Esquina; and S, Santa Fe.

TABLE 2

Polymorphic sites at the ND3 gene for *H. vulpinus* ($N = 21$)

Haplotype	Polymorphic sites 23 82 80 * GT	Distribution of haplotypes by locality L	P	Total
1	..	13	0	13
2	A.	3	4	7
3	AC	0	1	1

The consensus sequence for *H. vulpinus* ($N = 21$) is shown. See legend to Table 1 for details. Localities are as follows: L, Las Cuevas; P, Puerto Ibicuy.

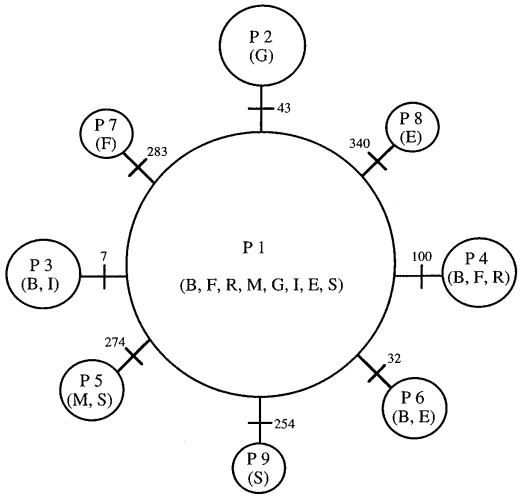

Figure 3.—Network linking the nine different proteins found among the *H. brasiliensis* populations. The protein variants are labeled P1–P9, and the localities of the variants are indicated in parentheses with abbreviations as in Table 1. The relative frequency of each protein is indicated by the size of the circle. Numbers associated with the hash marks indicate the position of the nucleotide change yielding each variant.

phic sites and 3 haplotypes were observed (Table 2). No insertion-deletion variation was observed within or between either species, and the length of the ND3 gene in Holochilus (348 bp) is the same as in humans (Anderson *et al.* 1981). Eight of the 24 (33%) polymorphisms in *H. brasiliensis* resulted in an amino acid change, creating nine different protein variants (Table 3). One of the two (50%) polymorphic mutations in *H. vulpinus* resulted in an amino acid change. Species-wide nucleotide diversity was 0.751% for *H. brasiliensis* and 0.171% for *H. vulpinus.*

Within *H. brasiliensis*, there is evidence for multiple mutations at the same site (violations of the infinite sites model). When parsimony trees are constructed using all the data, there are 617 equally parsimonious networks linking the 25 haplotypes; these trees have a consistency index of 0.8 (Swofford 1993). The length of each tree is 30 mutations, 6 of which are attributable to mutations at sites that already display a mutation. Over all 617 trees, these homoplasious mutations are observed at sites 100, 120, 216, 279, and 345. Site 100 is nonsynonymous and the other four are synonymous sites. The total number of silent and replacement mutations on these 617 trees ranges from 22 silent and 8 replacement to 20 silent and 10 replacement mutations. When only nonsynonymous variation is used to construct a tree, there is a single most parsimonious network that has no homoplasy (Figure 3). There is one common protein variant (P1) present in 79% of the individuals and eight other variants, each one mutational step removed from P1 and each present in fewer than 5% of the individuals.

There is some evidence of geographic structuring to the distribution of variation within *H. brasiliensis* (Tables 1 and 3). The most common protein variant was present in every population. However, of the eight rare variants,

TABLE 3

Protein Variants at the ND3 Gene for *H. brasiliensis* ($N = 82$)

Protein variants	Polymorphic sites 00012223 03405784 72304430 AIFAHFVT	Distribution of proteins by locality B	F	R	M	G	I	E	S	Total
P1	.....	6	4	10	3	12	21	5	4	65
P2	..L.....	0	0	0	0	4	0	0	0	4
P3	T.....	1	0	0	0	0	2	0	0	3
P4	...T....	1	1	1	0	0	0	0	0	3
P5	.....L..	0	0	0	1	0	0	0	1	2
P6	.T.....	1	0	0	0	0	0	1	0	2
P7	....L...	0	1	0	0	0	0	0	0	1
P8	.....A	0	0	0	0	0	0	1	0	1
P9	.....I.	0	0	0	0	0	0	0	1	1

Positions of polymorphic sites that yield protein variants (P1–P9), and the distribution of the protein variants among the populations are shown above for *H. brasiliensis* ND3 gene. See Table 1 legend for population names.

TABLE 4

Nucleotide diversity (π_w) among Holochilus populations

H. brasiliensis	*N*	π_w (%)	*H. vulpinus*	*N*	π_w (%)
Bahia Negra	9	0.885	Las Cuevas	16	0.094
Fonciere	6	0.580	Puerto Ibicuy	5	0.116
Rosario	9	0.612	Average	21	0.105
Rosario/Golondrina	6	0.773			
Golondrina	16	0.444			
Itati	23	0.428			
Esquina	7	0.386			
Santa Fe	6	0.831			
Average	82	0.617			

Percent nucleotide diversity (π_w) for the eight populations of *H. brasiliensis* and two populations of *H. vulpinus*. The sample size for each population is indicated with *N*.

four were restricted to single populations, three were observed in two populations, and one was observed in three populations (Figure 3). Nucleotide diversity among the eight *H. brasiliensis* populations ranged from a low of $\pi = 0.39\%$ to a high of $\pi = 0.89\%$ (Table 4). Average F_{ST} calculated among all *H. brasiliensis* populations was 0.178 and F_{ST} calculated between the two *H. vulpinus* populations was 0.386 (Table 5). To test for population subdivision, we performed a χ^2 test of haplotype frequencies in the different localities with rare haplotypes lumped such that the expected number of each haplotype in each locality was at least two (Nei 1987). This test rejected the null hypothesis of panmixia ($\chi^2 = 22.9$, d.f. = 7, $P = 0.0018$). We also performed permutation tests of subdivision (Hudson *et al.* 1992) between adjacent pairs of localities. These comparisons were significant only when the Itati sample was compared with either of its two neighboring localities, Golondrina and Esquina ($P < 0.001$ for each), suggesting that the Itati sample is distinct. Itati is the largest sample ($N = 23$) and contains one common haplotype and four relatively rare haplotypes. In contrast, the Bahia Negra sample ($N = 9$) contains eight haplotypes, all about equally common (Table 1).

The average uncorrected sequence divergence between *H. brasiliensis* and *H. vulpinus* was 18.9% for the entire ND3 gene. Divergence for the entire gene, corrected for multiple hits using Kimura's two-parameter model (Kimura 1980), was 22.0%. Uncorrected and corrected synonymous divergence estimates were 60.7% and 142.5%, respectively, and uncorrected and corrected nonsynonymous divergence estimates were 2.65% and 2.69%, respectively. The large difference between corrected and uncorrected synonymous divergence values implies that many more silent substitutions have occurred than are actually observed. This has important implications for interpreting the McDonald-Kreitman tests, as discussed below. Fifty-six fixed differences separate the ND3 sequences between *H. brasiliensis* and *H. vulpinus*; of these, seven (12.5%) result in amino acid changes.

TABLE 5

F_{ST} for *H. brasiliensis* and *H. vulpinus*

	$\overline{\pi}_w$ (%)	π_T (%)	F_{ST}
H. brasiliensis	0.617	0.751	0.178
H. vulpinus	0.105	0.171	0.386

$F_{ST} = \pi_T - \overline{\pi}_w/\pi_T$. π_T is the total nucleotide heterozygosity for all individuals sampled from each species, and $\overline{\pi}_w$ is the average nucleotide heterozygosity within populations.

Tests of neutrality: We observed 9 replacement and 17 silent polymorphisms within species (both species together), and 7 replacement and 49 silent fixed differences between species (Table 6). These ratios are significantly different from each other using a *G* log-likelihood ratio test ($P < 0.05$). The ratios are also significantly different from each other when polymorphism data from only *H. brasiliensis* are compared to fixed differences between the species ($P < 0.05$). There were too few segregating sites within *H. vulpinus* to construct a test with polymorphism data from that species alone. These comparisons are based on uncorrected levels of sequence divergence and do not account for multiple mutations at the same site, either within or between species. Corrected values do not represent independent observations and thus are inappropriate for use in a contingency table analysis (Sokal and Rohlf 1995). However, corrected values reflect more accurately the amount of evolutionary change that has occurred. When we use corrected values within and between species (Table 6), the resulting test is highly significant ($P < 0.001$) because silent site divergence is the most undercounted category when uncorrected values are used.

We investigated departures from a neutral frequency distribution by looking at the total number of segregating sites, the number of singletons, and the average number of pairwise differences within *H. brasiliensis* (Table 7). There is clear evidence for an excess of rare sites. Tajima's *D*, Fu and Li's *D*, and Fu's F_s are negative

TABLE 6

Silent and replacement differences within and between species at the ND3 gene

	Within			
	H. brasiliensis ($N = 82$)	*H. vulpinus* ($N = 21$)	Total ($N = 103$)	Between *H. brasiliensis* and *H. vulpinus*
Replacement	8 (8–10)	1	9	7 (7)
Silent	16 (20–22)	1	17	49 (100)

A G log-likelihood test was used to test the null hypothesis that the ratio of replacement to silent differences within and between species is equal. The G log-likelihood score ($G = 4.461$) was significant to reject the null hypothesis ($P < 0.03$). Numbers in parentheses are the corrected values for multiple mutations based on a parsimony analysis within species and a Jukes-Cantor (Jukes and Cantor 1969) model between species.

when all sites are considered together as well as when replacement or silent sites are considered alone. Tajima's D is significantly negative only for replacement sites, while Fu and Li's D is significantly negative for the entire data set and for silent sites alone. Fu's F_s is significantly negative for both replacement and silent sites. The frequency distribution of polymorphic sites is shown in Figure 4; 14 of 24 polymorphisms were present in just one or two individuals.

DISCUSSION

Excess amino acid polymorphism in natural populations: As in previous studies (Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Wise *et al.* 1998), we observe a greater number of intraspecific replacement polymorphisms than expected based on the interspecific comparison. This pattern is contrary to the predictions of the neutral theory of molecular evolution (Kimura 1983; McDonald and Kreitman 1991).

In principle, deviations from neutral expectations in a 2×2 contingency table analysis may be due to forces affecting the numbers in any or several of the four cells. For example, it is possible that the deviation we observe is due to the accumulation of adaptive synonymous substitutions between species. There is mounting evidence that selection on silent sites may play an important role in Drosophila (*e.g.*, Akashi 1994, 1995; Akashi and Schaeffer 1997). There are two reasons to doubt that this is the main cause of the observed deviation in our data. First, while selection on silent sites undoubtedly occurs, it is presumably weaker and less common than selection on changes affecting amino acid sequence and protein structure. Second, the ratio of replacement to silent changes seen between species in our data is in good general agreement with this ratio for a variety of other genes in interspecific comparisons (*e.g.*, Tucker and Lundrigan 1993). It seems reasonable, therefore, to consider explanations for the deviation that focus on the potential selective forces acting on nonsynonymous polymorphisms.

One possibility is that selection pressures have changed recently, allowing an accumulation of formerly deleterious, but currently neutral, amino acid polymorphisms (Figure 5). For example, Takahata (1993) has argued that human populations experienced a dramatic relaxation of selection since the Pleistocene, and that this has allowed for the accumulation of formerly deleterious mutations in human populations. There are several arguments against this hypothesis as a general explanation for the excess of mtDNA replacement polymorphisms observed in Holochilus and other taxa (humans, house mice, and Drosophila). First, the different ecologies of Holochilus, Homo, Mus, and Drosophila make it unlikely that each experienced a similar change in selection pressure on mitochondrial genes. Second, this hypothesis requires that a change in selection occurred at a specific point in time relative to our sampling (Figure 5). The likelihood that selection was relaxed independently in four species, and further that each of these species was sampled shortly after the relaxation of selection, seems remote. Third, the ND3 gene is one of the mitochondrial subunits of the NADH dehydrogenase complex; this complex functions in the creation of transmembrane proton gradients that are necessary for ATP synthesis (Weiss *et al.* 1991). In humans, several fatal and degenerative mitochondrial diseases have been associated with mutations that lead to NADH dehydrogenase deficiencies (Wallace 1994). The functional importance of genes associated with the mitochondrial electron transport pathway argues against relaxation of selection as a reasonable explanation for the observed patterns.

Another possibility is that multiple amino acid polymorphisms are being maintained in populations by some form of balancing selection. Although the uniparental inheritance of mtDNA precludes heterosis, it is possible that selection acting upon nuclear-cytoplasmic interactions is maintaining mitochondrial variants. Many of the mitochondrial genes for which excess amino acid variation has been documented (ATPase, cytochrome b, cytochrome oxidase, and NADH dehydrogenase complex) have subunits encoded by both the nucleus and mitochondrion. However, theoretical

TABLE 7

Nucleotide heterozygosity for *H. brasiliensis* ($N = 82$)

	S	k	θ (%)	π (%)	η (s)	Tajima's D	Fu and Li's D	Fu's F_s
All sites	24	2.593	1.397	0.751	10	−1.405	−2.062*	−15.275**
Replacement sites	8	0.406	0.587	0.148	3	−1.869*	−1.125	−8.289**
Silent sites	16	2.186	4.529	3.079	7	−0.919	−1.986*	−7.780*

S, Number of segregating sites; k, average number of pairwise differences between sequences; η(s), number of singletons; * $P < 0.05$; ** $P < 0.01$.

studies suggest that it is difficult to maintain multiple mitochondrial variants via cyto-nuclear interactions (Clark 1984). Another possibility is that variation is maintained by selection that varies over space or time (*e.g.*, Gillespie 1991, 1994). Balanced polymorphisms are expected to be maintained in populations, on average, longer than neutral polymorphisms, and are thus expected to be present at higher than average frequencies. However, some polymorphisms known to be targets of balancing selection are present at relatively low frequencies in some populations; *Adh* variants in Drosophila (Berry and Kreitman 1993) and *Hbb* variants in humans (Harding *et al.* 1997) are two such examples. The frequency distribution of polymorphisms in this study shows that replacement polymorphisms are only present at low frequencies (Figure 4). It is difficult to envision how multiple, linked replacement polymorphisms could be maintained at low frequencies by some form of positive selection; additional theoretical models with temporally or spatially varying selection coefficients would be useful for evaluating this hypothesis more carefully.

A third explanation for these data is that many amino acid mutations are weakly deleterious. First proposed by Ohta and Kimura (1971) and Ohta (1972), this hypothesis suggests that many replacement mutations fall within the neighborhood of $|s| = 1/N_e$ and, as such, are expected to contribute differentially to heterozygosity and to substitution (Kimura 1983). Specific models of weakly deleterious mutations have been strongly criticized, mainly on the grounds that they are improbable; *i.e.*, they exhibit their unique behavior only for a very restricted set of conditions (Gillespie 1994). In particular, if $|N_e s| < 1$, mutations will behave like neutral mutations, and if $|N_e s| \gg 1$, evolution will stop (Gillespie 1994, 1995). The difficulty with this explanation for the mitochondrial data in general is that marsh rats, house mice, fruit flies, and humans all presumably have different effective population sizes, yet all show an excess of intraspecific replacement polymorphism.

There are, however, other lines of evidence that support the view that deleterious mutations may be common in mtDNA. First, a number of mitochondrial missense mutations are known to cause disease in humans, and some disease phenotypes appear to be due to multiple mutations, each of small effect (Wallace 1994). Interestingly, many mitochondrial diseases show adult or late-onset (after reproductive maturity has been attained); this is consistent with their relatively mild effect on fitness. Second, several recent studies (*e.g.*, Parsons *et al.* 1997) have documented a much higher mitochondrial mutation rate in pedigrees than has been observed in phylogenetic comparisons. Such a difference is expected if many mutations are weakly deleterious and rarely fix in populations. In fact, it is possible that the per site mutation rate for mtDNA is sufficiently high that the excess replacement polymorphisms observed in population samples may be explained by a simple model of mutation-selection balance rather than by the more complicated models that include a stochastic component (*e.g.*, Ohta and Tachida 1990; Tachida 1991). Parsons *et al.* (1997) estimated that the mutation rate for the control region may be $\sim 10^{-5}$ per site. We can use this value and the average frequency of replacement polymorphisms in our sample to estimate the average selection coefficient on these polymorphisms assuming variation is solely a result of the balance between mutation and selection. The equilibrium allele frequency under mutation-selection balance for a haploid is given by μ/s, where μ is the mutation rate and s is the selection coefficient. The average frequency of replacement polymorphisms in our sample is 2.6%, leading to an average selection coefficient of $s = 0.038$. This calculation, while extremely rough, shows that mutations with moderate selection coefficients may still rise to appreciable frequencies if mutation rates are high.

The hypothesis that mitochondrial amino acid mutations are mildly deleterious makes at least one testable prediction. Ohta and Kimura (1971) pointed out that rates of evolution will be inversely proportional to effective population size under a slightly deleterious model. Ohta (1993, 1995) tested this prediction with nuclear gene sequences from flies and mammals and confirmed that nonsynonymous rates of evolution varied inversely with effective population size. Similar tests could be made with rates of mitochondrial protein evolution in species with different effective population sizes.

Frequency distribution of polymorphisms in *H. brasiliensis*: Tests of the frequency distribution were consistent in revealing an excess of rare sites (Table 7). These departures from neutral expectations were observed for

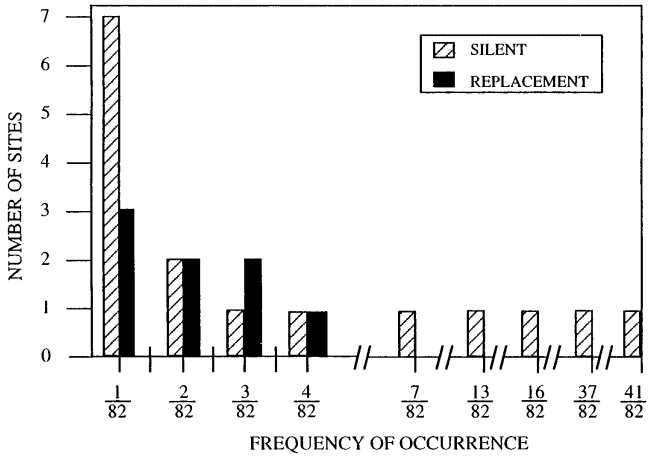

Figure 4.—Frequency distribution of polymorphic sites. Twenty-four polymorphic sites were observed among the 82 *H. brasiliensis* sampled. Silent sites are shown with hatch bars and replacement sites are shown with solid bars. Replacement sites are only present at low frequencies; however, silent sites are observed at both low and moderate frequencies.

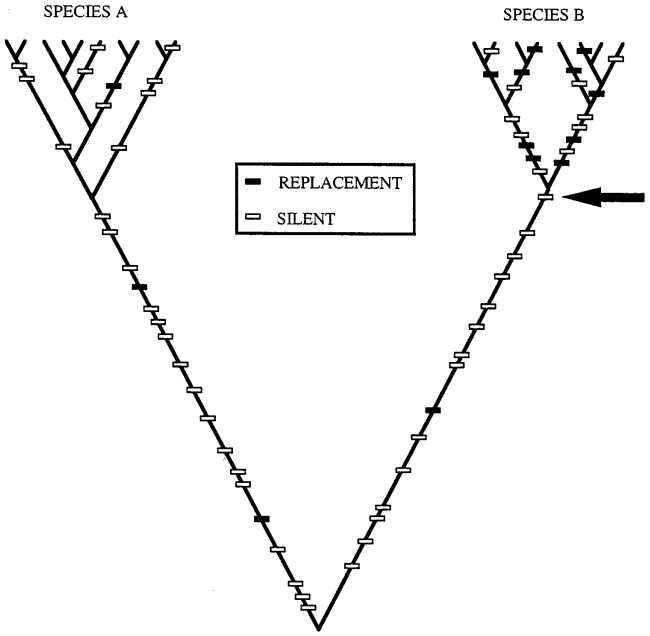

Figure 5.—Hypothetical gene tree depicting the effects of relaxed selection. The arrow indicates when a relaxation of selective constraint occurred. After selection is relaxed, replacement polymorphisms accumulate within species B but do not affect the ratio of replacement to silent fixed differences between species A and B. The hypothesis that relaxed selection accounts for the higher than expected levels of intraspecific polymorphisms requires that species are sampled at a specific point in time relative to the relaxation of constraint. If selection were relaxed earlier (*i.e.*, if the arrow were moved basally down the tree), the accumulation of replacement polymorphisms would affect both polymorphism and divergence.

both nonsynonymous and synonymous sites. These results could be due to a population expansion or a selective sweep on the mitochondrial genome (Tajima 1989). Population subdivision may also produce a skew in the frequency distribution if rare sites are restricted to single populations, as is the case for some of the polymorphisms in our data (Table 1 and Figure 3). Sampling additional loci would help distinguish between demographic and selective explanations for the observed frequency distribution.

Generality and implications of nonneutral mtDNA evolution: The results presented here show that the nonneutral patterns first documented for humans, house mice, and fruit flies are not unique to those species. This study, in conjunction with a recent analysis of published mtDNA datasets (Nachman 1998) raises the possibility that weakly deleterious amino acid polymorphisms may be a common feature of animal mtDNA. The pattern, however, is not ubiquitous, and some important questions remain. The ND3 gene has now been studied from this perspective in marsh rats (this study), humans (Nachman *et al.* 1996), Mus (Nachman *et al.* 1994), and Drosophila (Rand and Kann 1996). All three mammals show an excess of intraspecific replacement polymorphisms, while the patterns observed in both *Drosophila melanogaster* and *D. simulans* fit the predictions of a neutral model. Thus the same gene in different species may be under different selective pressures, or differences in effective population size may render mutations neutral in some species and visible to selection in other species. Further, different genes in the same species may show different patterns. In *D. melanogaster* and *D. simulans*, cytochrome b (Ballard and Kreitman 1994) and portions of ND5 (Rand *et al.* 1994; Rand and Kann 1996) show an excess of replacement polymorphism, while ND3 does not (Rand and Kann 1996). Thus, the patterns observed in this study, while common, are not characteristic of either all mitochondrial genes or of one gene in all species.

Another question raised by these observations is whether excess intraspecific amino acid polymorphisms are specific to mtDNA, or whether these observations may also extend to nuclear loci. Few data exist with which to address this issue, and most polymorphism and divergence data from nuclear genes come from Drosophila. In *D. melanogaster*, while some genes show an excess of intraspecific replacement polymorphism, most do not (*e.g.*, Brookfield and Sharp 1994; Moriyama and Powell 1996). For example, McDonald-Kreitman tests performed on 12 different loci using polymorphism data from *D. melanogaster* yielded only two significant results (Moriyama and Powell 1996). One locus, *Zw*, showed an excess of interspecific nonsynonymous substitution; conversely, another locus, *Pgi*, showed an excess of intraspecific nonsynonymous polymorphism. *Adh* data from some crop plants also show a slight excess of intraspecific replacement polymorphism (Gaut and Clegg 1993a,b). In addition, examination of Li and Sadler's (1991) human polymorphism data

revealed equal numbers of replacement and silent polymorphisms across 50 different loci (Takahata 1993), although divergence data for these genes are not available.

We thank A. S. Kondrashov for discussion, S. L. Crowell for help in the lab, and P. Andolfatto, R. R. Hudson, and one anonymous reviewer for comments on the manuscript. This work was supported by a National Science Foundation (NSF) Research Training Grant to P.K. and NSF grants to M.W.N.

LITERATURE CITED

Akashi, H., 1994 Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. Genetics **136:** 927–935.

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

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

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. De Brujin, A. R. Coulson *et al.*, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457-465.

Aquadro, C. F., 1997 Insights into the evolutionary process from patterns of DNA sequence variability. Curr. Opin. Genet. Dev. **7:** 835–840.

Avise, J. C., 1994 *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

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

Berry, A., and M. Kreitman, 1993 Molecular analysis of an allozyme cline in alcohol dehydrogenase in *Drosophila melanogaster* on the east coast of North America. Genetics **134:** 869–893.

Bibb, M. J., R. A. Van Etten, C. T. Wright, M. W. Walberg and D. A. Clayton, 1981 Sequence and gene organization of mouse mitochondrial DNA. Cell **26:** 167–180.

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.

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

Fu, Y. X., 1996 New statistical tests of neutrality for DNA samples from a population. Genetics **143:** 557–570.

Fu, Y. X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking, and background selection. Genetics **147:** 915–925.

Fu, Y. X., and W.-H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

Gaut, B. S., and M. T. Clegg, 1993a Nucleotide polymorphism in the *ADH1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics **135:** 1091–1097.

Gaut, B. S., and M. T. Clegg, 1993b Molecular evolution of the *AdhI* locus in the genus *Zea*. Proc. Natl. Acad. Sci. USA **90:** 5095–5099.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, Oxford.

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

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

Harding, R. M., S. M. Fullerton, R. C. Griffiths, J. Bond, M. J. Cox *et al.*, 1997 Archaic African and Asian lineages in the genetic ancestry of modern humans. Amer. J. Human Genet. **60:** 772–789.

Hershkovitz, P., 1955 South American marsh rats, genus Holochilus, with a summary of Sigmodont rodents. Fieldiana Zool. **37:** 639–673.

Hudson, R. R., M. Kreitman and M. Aquadé, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Hudson, R. R., D. D. Boos and N. L. Kaplan, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol. **9:** 138–151.

Jukes, T. H., and C. R. Cantor, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism*, edited by H. N. Munro. Academic Press, New York.

Kimura, M., 1980 A simple method for estimating evolutionary rate or base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. **16:** 111–120.

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

Kreitman, M., and H. Akashi, 1995 Molecular evidence for natural selection. Annu. Rev. Ecol. Syst. **26:** 403–422.

Li, W.-H., and L. A. Sadler, 1991 Low nucleotide diversity in man. Genetics **129:** 513–523.

McDonald, J. H., 1996 Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol. Biol. Evol. **13:** 1114–1118.

McDonald, J. H., 1998 Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. Mol. Biol. Evol. **15:** 377–384.

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

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13:** 261–277.

Nachman, M. W., 1992 Geographic patterns of chromosomal variation in South American marsh rats, *Holochilus brasiliensis* and *H. vulpinus*. Cytogenet. Cell Genet. **61:** 10–16.

Nachman, M. W., 1998 Deleterious mutations in animal mitochondrial DNA. Genetica **102/103:** 61–69.

Nachman, M. W., S. N. Boyer and C. F. Aquadro, 1994 Nonneutral evolution at the mitochondrial NADH dehydrogenase subunit 3 gene in mice. Proc. Natl. Acad. Sci. USA **91:** 6364–6368.

Nachman, M. W., W. M. Brown, M. Stoneking and C. F. Aquadro, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics **142:** 953–963.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei, M., and W.-H. Li, 1979 Mathematical models for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269–5273.

Ohta, T., 1972 Population size and rate of evolution. J. Mol. Evol. **1:** 305–314.

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

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.

Ohta, T., 1995 Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. Genetics **40:** 56–63.

Ohta, T., and M. Kimura, 1971 On the constancy of the evolutionary rate of cistrons. J. Mol. Evol. **1:** 18–25.

Ohta, T., and H. Tachida, 1990 Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. Genetics **126:** 219–229.

Parsons, T. J., D. S. Muniec, K. Sullivan, N. Woodyatt, R. Allison-Greiner *et al.*, 1997 A high observed substitution rate in the human mitochondrial DNA control region. Nature Genet. **15:** 363–368.

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

Rand, D. M., and L. M. Kann, 1996 Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Mol. Biol. Evol. **13:** 735–748.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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. R., and F. J. Rohlf, 1995 *Biometry*. W. H. Freeman and Company, New York.

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

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.
Takahata, N., 1993 Relaxed natural selection in human populations during the Pleistocene. Jpn. J. Genet. **68:** 539–547.
Templeton, A. R., 1996 Contingency tests of neutrality using intra/interspecific gene trees: the rejection of neutrality for the evolution of the mitochondrial cytochrome oxidase II gene in the Hominoid primates. Genetics **144:** 1263–1270.
Tucker, P. K., and B. L. Lundrigan, 1993 Rapid evolution of the sex determining locus in Old World mice and rats. Nature **364:** 715–717.
Wallace, D. C., 1994 Mitochondrial DNA mutations in diseases of energy metabolism. J. Bioenerg. Biomembr. **26:** 241–250.
Watterson, G. A., 1975 On the number of segregating sites in genetic models without recombination. Theor. Pop. Biol. **7:** 256–276.
Weiss, H., T. Friedrich, G. Hofhaus and D. Preis, 1991 The respiratory-chain NADH dehydrogenase (complex I) of mitochondria. Eur. J. Biochem. **197:** 563–576.
Wise, C. A., M. Sraml and S. Easteal, 1998 Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. Genetics **148:** 409–421.

Communicating editor: R. R. Hudson