The Age of Nonsynonymous and Synonymous Mutations in Animal mtDNA and Implications for the Mildly Deleterious Theory

Rasmus Nielsen* and Daniel M. Weinreich[†,1]

*Department of Organismic and Evolutionary Biology and †Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT

McDonald/Kreitman tests performed on animal mtDNA consistently reveal significant deviations from strict neutrality in the direction of an excess number of polymorphic nonsynonymous sites, which is consistent with purifying selection acting on nonsynonymous sites. We show that under models of recurrent neutral and deleterious mutations, the mean age of segregating neutral mutations is greater than the mean age of segregating selected mutations, even in the absence of recombination. We develop a test of the hypothesis that the mean age of segregating synonymous mutations equals the mean age of segregating nonsynonymous mutations in a sample of DNA sequences. The power of this age-of-mutation test and the power of the McDonald/Kreitman test are explored by computer simulations. We apply the new test to 25 previously published mitochondrial data sets and find weak evidence for selection against nonsynonymous mutations.

ONE of the greatest standing controversies in evolutionary biology is the debate over the causes of molecular variation in natural populations (*e.g.*, Kimura 1983; Gillespie 1991). The neutral theory asserts that the vast majority of variation is selectively neutral and when two or more alleles coexist in a population, natural selection cannot distinguish between the alleles. Under the neutral theory, allelic frequencies are determined by genetic drift. Similarly the theory holds that most allelic fixation events are the consequence of genetic drift and not of natural selection. In contrast, selective theories, including the mildly deleterious model (Ohta 1973), claim that a significant amount of the allelic variation and fixations in natural populations is determined by natural selection.

The McDonald/Kreitman test (McDonald and Kreitman 1991) analyzes the neutral assumption that the same processes determine allelic polymorphism and allelic fixation by comparing the number of synonymous and nonsynonymous sites segregating within and fixed between species in a 2×2 contingency table. Recently, neutrality has been rejected by the McDonald/Kreitman test in a number of mitochondrial DNA (mtDNA) data sets (Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Rand *et al.* 1994; Rand and Kann 1996, 1998; Nachman 1998; Hasegawa *et al.* 1998; Wise *et al.* 1998). The direction of deviation from neutrality in

each of these data sets is the same: an excess number of polymorphic nonsynonymous sites segregating within species relative to the number of nonsynonymous sites fixed between species. Mildly deleterious alleles may segregate in a population in low or intermediate frequencies but are not as likely to become fixed (Kimura 1983), and thus if some nonsynonymous mutations are mildly deleterious, we expect that they will segregate within populations but rarely reach fixation. This is the interpretation commonly given to the significant McDonald/Kreitman tests observed for animal mtDNA (Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Nachman 1998; Rand and Kann 1996, 1998; Rand *et al.* 1994; Hasegawa *et al.* 1998; Wise *et al.* 1998).

We present a new method for testing the hypothesis that segregating nonsynonymous sites in animal mtDNA are mildly deleterious, based on a comparison of the age of segregating nonsynonymous and synonymous mutations. Throughout we assume that there is no recombination between adjacent nucleotide sites. This implies that nonsynonymous and synonymous sites share the same underlying gene genealogy. We first show that the mean age of a segregating deleterious mutation is less than the mean age of a segregating neutral mutation, even in the absence of recombination. Thus under the hypothesis that nonsynonymous mutations are mildly deleterious, the mean age of nonsynonymous segregating sites should be less than the mean age of synonymous segregating sites despite the fact that the underlying gene genealogy is identical for the two types of sites. We present a test for comparing the mean ages of two arbitrary classes of segregating sites (*e.g.*, synonymous and nonsynonymous) in a sample and de-

Corresponding author: Rasmus Nielsen, Department of Organismic and Evolutionary Biology, Harvard University, 288 Biol. Labs., Cambridge, MA 02138. E-mail: rnielsen@oeb.harvard.edu

[1] *Present address:* Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912.

termine its power as a function of the strength of purifying selection. We then determine the power of the McDonald/Kreitman test to detect purifying selection. Finally we apply our test to a set of 25 previously published mtDNA data sets.

METHODS

Computer simulations: Throughout, computer simulations were used to test theoretical predictions and to examine the properties of the suggested statistical tests. In these simulations a haploid population of size N chromosomes was simulated under the Wright-Fisher model. To model the mutational process an infinite sites model was used (Watterson 1975). The mutational makeup of each individual chromosome and the generation in which each segregating mutation arose were stored in computer memory. In each generation the population of chromosomes underwent mutation, selection, and genetic drift. The probability of a new mutation occurring in a given generation on a chromosome is $\mu = \mu_s + \mu_n$, where μ_s is the mutation rate of selected mutations and μ_n is the mutation rate of neutral mutations. Simple multiplicative fitness was employed; thus the relative fitness of the ith chromosome is given by

$$w_i = (1 + s)^{n_i} / \sum_j (1 + s)^{n_j}, \qquad (1)$$

where s is the selection coefficient of selected mutations and n_i is the number of selected mutations carried by the ith chromosome. For the sake of simplicity we assume that s does not vary in time or among selected mutations and that there are no epistatic interactions. Selection together with genetic drift is simulated by sampling multinomially between generations among the k segregating chromosomes according to the probabilities specified by the w_i, $i = 1, 2, \ldots, k$ given by (1). In general, samples were taken from the population every $10N$ generations after the first $100N$ generations. For each sample, some function of the sample was evaluated, such as the mean age of selected or neutral segregating sites or the P value under one of the suggested tests. Denote the obtained value of the evaluated function in the ith sample by f_i. Assuming that stationary distributions exist for these functions, the theorem of ergodicity ensures that

$$\frac{1}{n} \sum_{i=1}^{n} f_i \to E(f) \quad \text{as } n \to \infty,$$

where $E(f)$ is the expectation of the function f and n is the number of samples. To examine the power of the McDonald/Kreitman test, we also scored the number of fixed differences between the sequences in each sample and a single sequence sampled t generations in the past. Population sizes of $N = 1000$ chromosomes were used in the simulations throughout. Except where otherwise noted, the parameter values used in the simulations

were chosen to fit the estimated values based on the analyzed data sets (see *Data analysis*); *i.e.*, the number of sequences in a sample was set to $n = 25$, $\theta = 10$ ($\theta = 4N\mu$), and the level of intraspecific divergence was set to $\mu t = 180$. The ratio of nonsynonymous to synonymous mutations was set to $\mu_n / \mu_s = 1$, corresponding to assuming that $\sim^3/_4$ of all new nonsynonymous mutations are immediately lethal to the organism.

Test statistics: We explored several tests of the hypothesis that the mean ages of segregating sites in two distinct classes of sites (*e.g.*, neutral and selected) in a sample are equal. The basics of these tests are the same: a score is calculated for each segregating site. Denote the value of this score for the ith site by ω_i. Then, we define the test statistic

$$\Omega = \frac{\Sigma_i (1 - I_i) \cdot \Sigma_i I_i \omega_i}{\Sigma_i I_i \cdot \Sigma_i (1 - I_i) \omega_i}, \qquad (2)$$

where I_i is an indicator function that returns 0 if site i is a neutral site and 1 if site i is selected. Thus Ω is the ratio of the average of ω for selected sites to the average of ω for neutral sites. The null distribution of Ω under the hypothesis that the value of ω_i is independent of the class of site i can be approximated by repeatedly recalculating Ω after random relabeling of sites such that the total number of selected sites remains constant and the total number of neutral sites remains constant. A total of 100,000 replicates of this resampling scheme were used to establish critical values of Ω. If we denote the ith value of Ω by Ω_i and the observed value of Ω by Ω_{obs}, then a test based on k resampled samples rejects at the 5% level if $(1/k)\Sigma_{i=1}^{k} I_{(\Omega_{obs} \geq \Omega_i)} < 0.05$, where $I_{()}$ is an indicator function that returns 1 if the Boolean condition is true and 0 otherwise. We refer to the statistic Ω as the age-of-mutation (AOM) statistic.

The power of such a test to detect differences in mean age between classes is obviously dependent on how strongly ω_i is correlated with the age of the mutation responsible for polymorphic site i. The power of the test using the different estimators for ω was evaluated as described in *Computer simulations.* One obvious candidate for ω is an estimator of the age of a mutation under a neutral equilibrium infinite sites model suggested by Griffiths and Tavaré (1999). This estimator uses $E(T \mid X)$ (where T is the age of a mutation and X is the observed data) as an estimator of T. $E(T \mid X)$ is calculated by integrating over all possible gene genealogies (branch lengths and topologies) while weighting each genealogy by its likelihood. The integral is evaluated by the Monte Carlo method described in Griffiths and Tavaré (1995, 1999). This estimator may have desirable statistical properties because all the information in the data regarding T is applied in the estimation of T. However, application of this estimator is computationally very time consuming. Furthermore, in cases where the data do not conform to the infinite sites model, it is

necessary to remove conflicting sites or conflicting sequences. For this reason, considerable effort was devoted to developing additional estimators that could be calculated fast and did not require elimination of any data.

Given only b, the observed number of chromosomes bearing a mutation in a sample of n chromosomes, the expected age (a) of the mutation under strict neutrality is given by

$$E(a|b) = 2 \binom{n-1}{b}^{-1} \sum_{j=2}^{n} \binom{n-j}{b-1} \frac{n-j+1}{n(j-1)} \qquad (3)$$

(Griffiths and Tavaré 1998). Results such as this, based on the marginal frequency of a mutation in a single site, can obviously be applied to estimate ages of individual mutations. However, they were in general found to perform much worse than estimators employing more of the data. Also, estimators based on neutrality such as $E(T \mid X)$ may not perform well when the assumption of selective neutrality is violated. Instead, a heuristic estimator was devised on the basis of estimation of the structure of the underlying genealogy. In this method, an estimate of the underlying genealogy was obtained by the unweighted pair group method using arithmetic averages (UPGMA; Sokal and Michener 1958). Mutations were subsequently assigned to branches (edges) using Fitch and Wagner parsimony (*e.g.*, Fitch 1971), and the expected age of a mutation was estimated as the average age of the branch on which the mutation occurred, *i.e.*, the average of the time of the start of the branch and the time of the end of the branch. In cases where two or more mutations were needed to account for the observed site pattern, the average of the estimates obtained for the different assignments of the mutation was used as an estimator. We denote this estimator by $\hat{a}$. Note that under the infinite sites model, $\hat{a}$ is the expected age of a mutation conditional on the estimated genealogy.

This estimator ($\hat{a}$) is applied in addition to $E(T \mid X)$ because it can be calculated quickly, making it suitable for simulation purposes, and because it does not require the elimination of data when the assumptions of the infinite sites model are violated. However, we do not suggest that this heuristically derived estimator should be used for more general purposes. It was chosen here solely because it provided more power in the test than estimators based on the marginal site pattern such as estimators derived from Equation 3.

Power analysis: The power of the test statistic given in Equation 2 and of the McDonald/Kreitman test to detect the presence of deleterious mutations was evaluated by computer simulations as described in *Computer simulations* for different values of the selection coefficient. The proportion of simulated data sets that gave significant test statistic values was tabulated for a range of values of Ns for $\theta = 10$ and for a range of values of θ for $Ns = -3$.

McDonald/Kreitman tests (McDonald and Kreitman 1991) of simulated data were performed as a G-test in a 2×2 contingency table using the Williams correction for small sample size (see, for example, Sokal and Rohlf 1981). The numbers entering the contingency table were the number of fixed selected and neutral mutations and the number of segregating selected and neutral mutations in the sample. Empty cell entries were ignored when calculating the test statistic. This mimics a common method for performing the McDonald/Kreitman test and should not inflate the chosen significance level markedly under biologically reasonable parameter values.

Data analysis: The null hypothesis that we wish to test is that the average age of nonsynonymous mutations is the same as the average age of synonymous mutations. To analyze the data, all polymorphisms were partitioned into nonsynonymous and synonymous polymorphisms. To do this, the following heuristic approach was applied: the most common sequence was chosen as a reference sequence representing a hypothesized ancestral state. This was done only to establish an unambiguous criterion to determine if a mutation was synonymous or nonsynonymous. The identity of a variable site in a codon was determined by comparison to the homologous codon in the reference sequence. In this manner, all polymorphisms could be scored unambiguously as nonsynonymous or synonymous even when there are multiple segregating sites in a single codon. If three nucleotides were segregating in the same site, two mutations were inferred and represented as two variable sites. In most cases only one mutation was observed in each codon, suggesting that the effect of this heuristic approach does not seriously bias the analysis. Also, because the approach was applied blindly with regard to the class of the mutation, it should not inflate the chosen significance level in the resampling test.

For each data set, two resampling tests were performed, one using $E(T \mid X)$ and one using the heuristically derived $\hat{a}$ as estimators of the age of a mutation. In each case, the data were resampled 100,000 times to provide a P value. When $E(T \mid X)$ was applied, the Watterson (1975) estimator of θ was used to drive the simulations, and 100,000 genealogies (runs of the Markov chain) were sampled to provide estimates of the age of each mutation (see Griffiths and Tavaré 1995). Calculation of $E(T \mid X)$ requires that the data strictly conform to the infinite sites model. For example, Harding *et al.* (1997) chose to eliminate apparent recombinant sequences. In the present analysis we instead eliminated sites to obtain a subset of sites that are consistent with the infinite sites model. The following heuristic algorithm was implemented for this purpose. Two diallelic sites are in conflict if all of the patterns in Table 1 are observed, where the labeling of state (0, 1) is arbitrary. Let C_i be the number of sites that conflict with site i. Then site i is eliminated from the data set when

TABLE 1

Conflicting site patterns

Site 1	Site 2
0	0
0	1
1	0
1	1

$C_i = \max\{C_1, C_2, \ldots, C_h\}$, where h is the number of sites. This elimination scheme is repeated until $C_1 = C_2 = \ldots = C_h = 0$. The test based on $E(T \mid X)$ was performed on these reduced data sets, whereas the test based on â was performed on the full data sets. Data sets so small that a significant result was not theoretically possible were not included in the analysis. Four data sets were eliminated for this reason.

Amino acid saturation dynamics: To examine the relative degree of mutational saturation in nonsynonymous and synonymous substitutions in mtDNA, the number of nonsynonymous and synonymous nucleotide differences between pairs of species was plotted for 13 mtDNA genes (NADH1, NADH2, NADH3, NADH4, NADH4L, NADH5, NADH6, ATPase 6, ATPase 8, COI, COII, COIII, Cyt. b). The species included in this analysis were *Homo sapiens* (Anderson *et al.* 1981); *Pan paniscus*, *Pongo pygmaeus* (Horai *et al.* 1995); *Mus musculus* (Bibb *et al.* 1981); *Gallus gallus* (Desjardins and Morais 1990); and *Drosophila melanogaster* (Garesse 1988). Sequences were aligned by eye and nonsynonymous and synonymous nucleotide differences were counted as described in *Data analysis.*

RESULTS

Expected ages of selected and linked neutral segregating mutations: Several results regarding the ages of neutral mutations are known. For example, the age of a neutral mutation with frequency b in a sample of size n is given by Equation 3. However, selection will affect the underlying gene genealogy and thus the distribution of coalescence times in those genealogies (*e.g.,* Kaplan *et al.* 1988; Hudson and Kaplan 1994) and will therefore affect the expected age of a sampled selected mutation and of linked neutral mutations. In addition, the age of a selected mutation may differ from the neutral expectation because selected mutations may not be distributed on the genealogy according to a Poisson process.

We are interested in the ratio of the age of selected to the age of neutral mutations in models without recombination. We show that in the models of interest, selection tends to reduce this ratio below unity and that this property of models with selection can be used to test neutrality. First, consider a haploid model in which

there is only one selected site in the population and assume that there are only two possible states for this site (A and a). A haplotype carrying state A has fitness 1 and a haplotype carrying state a has fitness $1 + s$. Furthermore, assume that mutations occur from state A to state a at rate υ and from state a to state A at rate υ. Assume that neutral mutations occur according to an infinite sites model at rate μ and that there is no recombination. Following Hudson and Kaplan (1994), we analyze this model in the limit of large population sizes so that the frequency of A is maintained at the equilibrium value. Under this assumption the mutation/selection model can be analyzed as a model of migration between two populations. Mutation between allele A and a is equivalent to migration between two isolated subpopulations with population sizes $Nf(A)$ and $Nf(a)$, respectively, where N is the population size and $f(A)$ and $f(a)$ are the frequencies of A and a, respectively, in the population. The equilibrium frequency of a for biologically plausible values of s and υ is given by

$$f(a) = \left(s + s\upsilon + 2\upsilon \right.$$

$$\left. - \sqrt{(s + s\upsilon + 2\upsilon)^2 - 4s\upsilon(1 + s)} \right) \Big/ (2s) \quad (4)$$

(see, for example, Nagylaki 1992, pp. 9–13). The probability that an A allele arose by mutation in the previous generation is given by

$$P_{aA} = \frac{f(a)\upsilon}{f(a)\upsilon + f(A)(1 - \upsilon)} \quad (5)$$

and the probability that an a allele arose by mutation in the previous generation is given by

$$P_{Aa} = \frac{f(A)\upsilon}{f(A)\upsilon + f(a)(1 - \upsilon)}. \quad (6)$$

Equations 5 and 6 correspond to Equation 4 of Hudson and Kaplan (1994). Let $T(A, a)$ be the expected coalescence time of a sample of size 2 given that one A allele and one a allele were obtained in the sample, $T(2a)$ be the expected coalescence time of a sample of size 2 given that two a alleles were obtained in the sample, and $T(2A)$ be the expected coalescence time of a sample of size 2 given that two A alleles were obtained in the sample. Assuming that only one mutation or coalescence event can happen in each generation, we can establish the following system of recursive equations for $T(A, a)$, $T(2A)$, and $T(2a)$,

$$T(2A) = \frac{1 + T(A, a)2P_{aA}}{2P_{aA} + 1/(Nf(A))}$$

$$T(2a) = \frac{1 + T(A, a)2P_{Aa}}{2P_{Aa} + 1/(Nf(a))}$$

$$T(A, a) = \frac{1 + T(2a)P_{Aa} + T(2A)P_{aA}}{P_{Aa} + P_{aA}}, \quad (7)$$
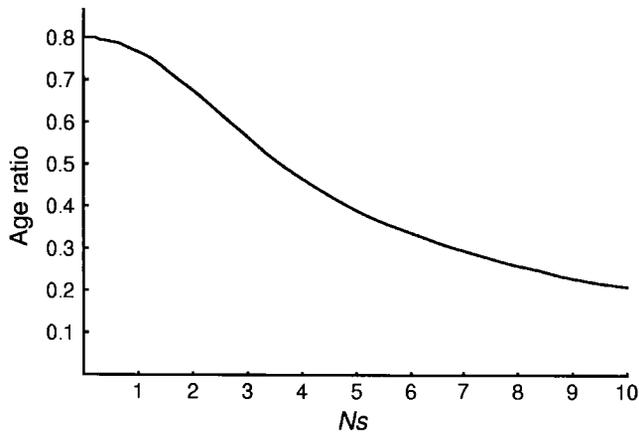
Figure 1.—The ratio of the expected age of a detectable selected mutation to the age of a neutral mutation (age ratio) in the model described in the text with $Nv = 1.0$.

which can easily be solved for $T(A, a)$, $T(2A)$, and $T(2a)$. Then the mean coalescence time can be found as

$$T_2 = f(a)^2 T(2a) + f(A)^2 T(2A) + 2f(A)f(a) T(2a) \qquad (8)$$

(Hudson and Kaplan 1994). Using this result, the expected age of a segregating neutral mutation in a sample of size 2 is then simply given by $T_2/2$. The expected age of a detectable selected mutation in a sample of size 2 is given by

$$\frac{1}{P_{aA} + P_{Aa}}. \qquad (9)$$

We can therefore evaluate the ratio of the expected age of a detectable selected mutation to the expected age of a neutral mutation in a sample of size 2 (Figure 1). The absolute ratio of the ages in the figure does not approach 1.0 as Ns approaches 0 because it is a function of v. However, note that the ratio is a monotonically decreasing function of s. The reason is that selection has only a marginal effect on the distribution of coalescence times but has a strong effect on the age of the selected mutation. That selection has only little effect on the expected coalescence time in a diallelic model at equilibrium has previously been noted by Neuhauser and Krone (1997). The model described above does not easily pertain to real data because of the assumption that the frequency of A is maintained at the equilibrium value and because only one selected site is considered. A more realistic model would assume an infinite sites model for both selected and neutral mutations and would allow the frequency of an allele to be a random variable through time. However, such a model does not lend itself to analytical treatment. Instead, we examined this model using computer simulations as described in methods (Figure 2). Note that the ratio of the mean age of a selected segregating mutation to the age of a segregating neutral mutation is a monotonically increasing function of the selection coefficient in the examined
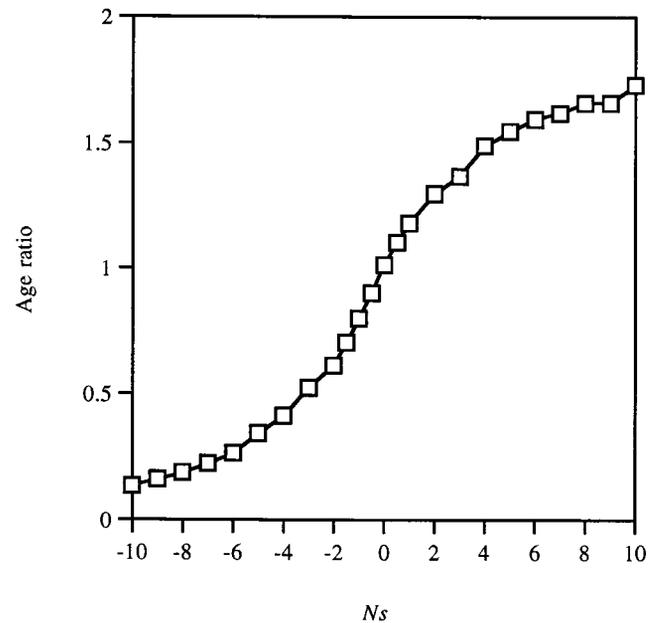
Figure 2.—The ratio of the average age of nonsynonymous mutations to synonymous mutations in a sample when nonsynonymous mutations are under constant positive or negative selection. Parameter values were $n = 25$ (sample size), $\theta = 10$, $N = 1000$, $\mu_s/\mu_n = 1$. A total of 1000 data sets were obtained to evaluate each point. Simulations were performed as described in methods.

range of Ns and that this relationship extends to positive values of Ns. These results confirm the notion that the differences in the average age of nonsynonymous and synonymous mutations may be used to detect selection against slightly deleterious mutations. If a large fraction of nonsynonymous mutations is slightly deleterious, their average age should be less than the average age of synonymous mutations. However, if the excess of nonsynonymous mutations segregating in mtDNA within populations is caused by other forces such as balancing selection, one would not necessarily expect the age of nonsynonymous mutations to be less than the age of synonymous mutations. Indeed, balancing selection is expected to increase the age of the selected mutation relative to the neutral expectation.

Power analyses: The power of the McDonald/Kreitman test (McDonald and Kreitman 1991) has not received much attention in the literature, although Akashi (1999) recently explored its power under the assumption of free recombination between adjacent sites. We were interested in examining its power against the hypothesis of purifying selection in the absence of recombination. This is the case applicable to mitochondrial DNA. Such analysis may provide a crude estimate of the size of the selection coefficients required to explain the repeated observations of departures from neutrality in animal mtDNA (Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Rand *et al.* 1994; Rand and Kann 1996, 1998; Nachman 1998; Wise *et al.* 1998). We examined the power of the test as
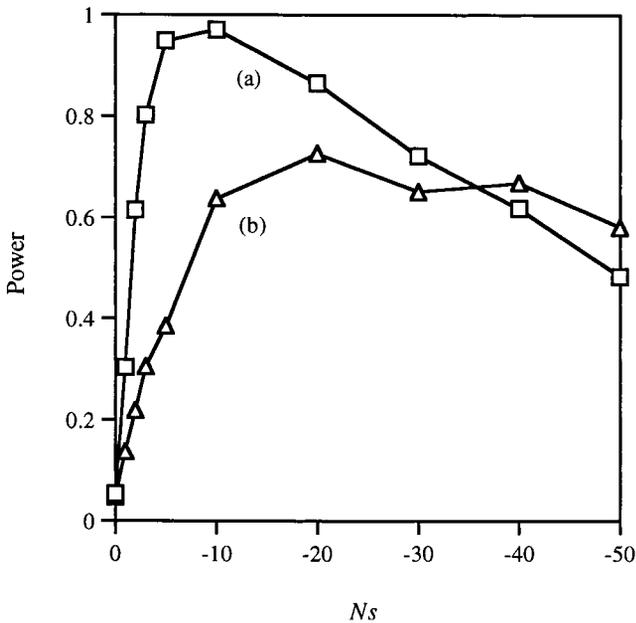
Figure 3.—The power (the proportion of tests significant at the 5% level) of the McDonald/Kreitman test (a) and the test of the ratio of mean ages of selected to neutral mutations (b) under negative selection. The parameter values are as in Figure 2; $\mu t = 180$.

described in methods (Figure 3) and found that, although it has a maximum at $\sim Ns = -10$, the test has power >50% over the range $-2 > Ns > -40$. Thus selection coefficients acting on nonsynonymous mutations in mtDNA should fall in this range if they are to account for the many observed deviations from neutrality.

Having demonstrated that under purifying selection the mean age of segregating selected sites is less than the mean age of segregating neutral sites in a population (Figures 1 and 2), we next turned to the problem of comparing mean ages of selected and neutral sites segregating in a sample of chromosomes drawn from a population. We examined the power of the AOM test (Equation 2) when using $\hat{a}$ as an estimator of the age of mutations by computer simulations as described in methods and found it to increase as Ns is reduced from 0 to -10 with a maximum in the range $-10 > Ns > -50$ (Figure 3a). Power analysis of the AOM test using $E(T \mid X)$ as an estimator of age of mutations by simulation is computationally too expensive to perform.

Data analysis: The hypothesis that the ages of nonsynonymous and synonymous polymorphisms are equal was tested for 25 previously published animal mtDNA data sets (Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Rand and Kann 1996; Wise *et al.* 1998; and references in Nachman 1998 and Rand and Kann 1998). The results of the tests are described in Table 2. Note that in only 6 cases using $\hat{a}$ as the estimator of relative age and in only 4 cases using $E(T \mid X)$ as the estimator of age was the hypothesis of an equal average age of nonsynonymous and synonymous mutations rejected by the AOM test,

whereas the McDonald/Kreitman test rejected neutrality in 14 of 25 cases.

DISCUSSION

Application of the McDonald/Kreitman test (McDonald and Kreitman 1991) to animal mtDNA has repeatedly demonstrated that the evolution of this genetic element is not consistent with strict neutrality (Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Rand *et al.* 1994; Rand and Kann 1996, 1998; Nachman 1998; Hasegawa *et al.* 1998; Wise *et al.* 1998). Relative to the number of nonsynonymous sites fixed between species, these data sets all show an excess number of segregating nonsynonymous sites within species. Kimura (1983) has shown that deleterious alleles contribute more to (within-species) heterozygosity than to (between-species) divergence and thus, these observations may be consistent with the notion that most segregating synonymous mutations are roughly neutral but that most segregating nonsynonymous mutations are mildly deleterious. We set out to explicitly test this hypothesis. On the basis of theoretical predictions (Figures 1 and 2) we chose the ratio of the estimated mean age of nonsynonymous mutations to the estimated mean age of synonymous mutations as a test statistic. Two estimators of the age of a mutation were used to analyze published mtDNA data sets: $E(T \mid X)$ and $\hat{a}$ (an estimator based on estimation of the underlying gene genealogy using classical phylogenetic methods). $E(T \mid X)$ has the advantage of using all the information in a sample, but assumes that the data fit the infinite sites model. As shown in Table 2, most of the published data sets (18/25) do not fit this model. Furthermore, the calculation of $E(T \mid X)$ is very time consuming, making it unsuitable for simulation purposes. This is the motivation for using $\hat{a}$ as an age estimator of the age of a mutation in addition to $E(T \mid X)$.

However, note that the selective removal of sites when $E(T \mid X)$ is used will not lead to falsely significant results. To realize this, note that under the null hypothesis of strict neutrality, both nonsynonymous and synonymous mutations should be distributed according to a Poisson process on the underlying shared gene genealogy. Because the distribution of the times of events in a Poisson process is given by the order statistics of a uniform random variable, the expected age of a mutation in a site is independent of the number of mutations occurring at the site. Thus the expected age of mutations in a data set will not change by the removal of sites with many mutations nor will the ratio of the expected ages of nonsynonymous to synonymous mutations change.

On the basis of an analysis of the power of the McDonald/Kreitman test (Figure 3a), we conclude that if purifying selection alone is responsible for the observed deviations from neutrality in animal mtDNA, selection coefficients must be in the range $-2 > Ns > -40$. It

TABLE 2

The results of the test of equal age of nonsynonymous and synonymous segregating mutations

Data set	N	Sites[a]	P value ($\hat{a}$)	P value ($E(T \mid X)$)	P value (McD/Kr)
Gadus [b,c]	41	25/4	0.699	0.670	0.25
Isothrix [b,c]	10	118/10	0.474	0.474	0.26
D. melanogaster [c,d]	17	8/0	0.714	0.891	0.005
D. melanogaster [e,f]	59	21/1	0.400	0.562	<0.1
Mesomys [b,c]	10	142/43	0.000	0.000	0.075
M. domesticus [g,h]	56	24/0	0.824	0.718	0.05/0.005[i]
P. troglodytes [j,k]	19	48/4	0.170	0.338	0.02
Scurius [b,c]	20	164/3	0.017	0.032	0.025
D. simulans [e,f]	29	7/0	0.142	0.057	0.5
Ursus [b,c]	28	96/6	0.932	0.966	<0.5
D. yakuba [c,d]	14	7/0	0.475	0.234	0.01
H. sapiens [j,k]	21	20/2	0.699	0.520	0.02
H. sapiens [l,m]	147	56/11	0.230	0.210	<0.00005
Phyllobates [c,n]	7	74/5	0.968	0.824	0.060
Passarella [c,o]	15	15/2	0.467	0.508	0.743
Microtus rossiaemeridionaiis [c,p]	15	20/9	0.676	0.788	<0.001
Microtus arvalis [c,p]	15	34/4	0.049	0.063	0.008
Melospiza melodia [c,o]	11	7/0	0.571	0.285	0.039
Grus antigone [c,q]	9	17/2	0.362	0.478	0.002
Ensatina eschscholtzii [c,r]	23	239/104	0.044	0.065	<0.001
Emoia imopara [c,s]	9	78/1	0.040	0.032	0.067
Emoia cyanura [c,t]	10	35/0	0.143	0.124	0.067
G. gorilla [t,u]	6	30/3	0.010	0.009	0.001
H. sapiens [t,u]	6	7/0	0.428	0.905	0.001
P. troglodytes [t,u]	5	14/1	0.671	0.439	0.001

[a] Number of segregating sites/number of sites removed in order to fit the infinite sites model.
[b] Rand and Kann (1998).
[c] Cytochrome B.
[d] Ballard and Kreitman (1994).
[e] Rand and Kann (1996).
[f] NADH dehydrogenase 5.
[g] Nachman *et al.* (1994).
[h] NADH dehydrogenase 3.
[i] Relative to *M. domesticus* and *M. musculus*, respectively.
[j] Wise *et al.* (1998).
[k] NADH dehydrogenase 2.
[l] Nachman *et al.* (1996).
[m] RFLP survey of entire mtDNA.
[n] Summers *et al.* (1997).
[o] Zink and Blackwell (1996).
[p] Baker *et al.* (1996).
[q] Wood and Krajewski (1996).
[r] Moritz *et al.* (1992).
[s] Bruna *et al.* (1996).
[t] Ruvolo *et al.* (1994).
[u] Cytochrome *C* oxidase.

was surprising to us that the McDonald/Kreitman test retains power over such a wide range and that such large (negative) values of Ns may be responsible for the observations. Akashi (1999) found that little power remained for values of $Ns < -10$ and that essentially no power remained for values of $Ns < -20$. The discrepancy between the results can presumably be explained by the effect of linkage, which is known to reduce the efficacy of selection (*e.g.*, Hill and Robertson 1966).

The AOM test based on $\hat{a}$ has less power to reject the hypothesis of neutrality across most of the ranges of Ns examined (Figure 3b) than the McDonald/Kreitman test. Power rises as Ns is reduced from -1 to -10 after which it appears to reach a plateau of $\sim$65%. It is not surprising that the McDonald/Kreitman test has more power in most of the parameter space because this test employs more data (interspecific comparisons in addition to intraspecific comparisons) than the test based on $\hat{a}$. However, the new test based on $\hat{a}$ uses information in the data not employed by the McDonald/Kreitman

test, and it is able to distinguish between hypotheses of balancing selection and selection on deleterious mutations. It should therefore provide additional information regarding selection coefficients when applied in addition to the McDonald/Kreitman test. Another advantage of the new test is that it is quite robust to the model assumptions. Under strict neutrality, the distribution of site patterns in nonsynonymous sites and in synonymous sites should be identical. If the weights assigned to different site patterns are poorly chosen, this will lead to a reduction in power but not to an inflation of the chosen significance level. This is also true even if the population is not at stationarity because of the decoupling of the mutational and the genealogical processes under neutrality.

Six out of 25 tests were significant using $\hat{a}$ as an estimator of the age of an allele (Table 2). The probability of randomly obtaining 6 or more significant results in 25 tests at the 5% significance level is $P = 0.0012$. One of the data sets (Mesomys) has a P value so low ($P < 0.001$) that the null hypothesis of no difference in age of synonymous and nonsynonymous mutations for all data sets can be rejected by a Bonferroni test. It is curious to note that all six significant data sets contain a relatively large number of segregating sites; indeed three of the six are the three largest data sets shown in Table 2 with respect to number of segregating sites. This may suggest that the small number of significant tests is caused by a lack of statistical power when only few polymorphic sites are available. In contrast there appears to be no bias among significant data sets with respect to number of chromosomes examined. Also, in 17/25 cases, the estimated mean age of nonsynonymous mutations is less than the estimated mean age of synonymous substitution (not shown). Together, these results provide weak support for the hypothesis that the average age of nonsynonymous mutations is less than the average age of synonymous mutations.

To test the hypothesis that the small number of significant AOM results in Table 2 could be the consequence of small sample size (polymorphic sites), additional power analyses of both the McDonald/Kreitman and the AOM tests were performed in which Ns was held constant and θ varied (Figure 4). Indeed the reduction in power when θ is small is somewhat more severe for the AOM test than for the McDonald/Kreitman test.

Supposing that purifying selection is acting on amino acid replacement mutations, the next question is which range of values of Ns are consistent with the observations from both the McDonald/Kreitman test and the AOM test presented here. Figure 5 (solid bars) shows the expected frequency distributions of P values for the test based on $\hat{a}$ derived from simulation with $Ns = 3$. Note that $Ns = 3$ is near the bottom of the power curve of the AOM test (Figure 3b); however, if Ns is larger, the expected distribution would be even more skewed toward low P values. In contrast the frequency distribu-
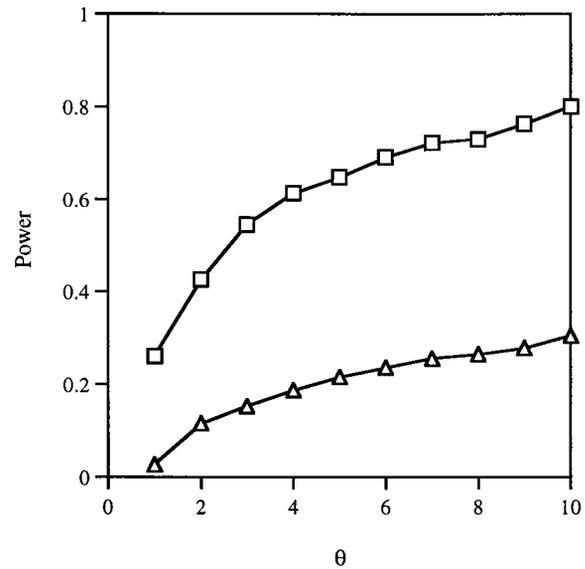
Figure 4.—The power (the proportion of tests significant at the 5% level) of the two tests (□, McDonald/Kreitman test; △, AOM test) when $Ns = 3$ and θ is varied. The remaining parameters are as in Figure 2.

tion of P values observed in the 25 data sets is nearly uniform (Figure 5, hatched bars). It appears that Ns must be <3 to account for the distribution of P values obtained in the test based on $\hat{a}$. However, if selection against deleterious mutations is to explain the many significant McDonald/Kreitman tests, Ns should be >2. There is therefore only a small window of values of Ns that are consistent with the findings in this article. Given the variability of population sizes in naturally occurring
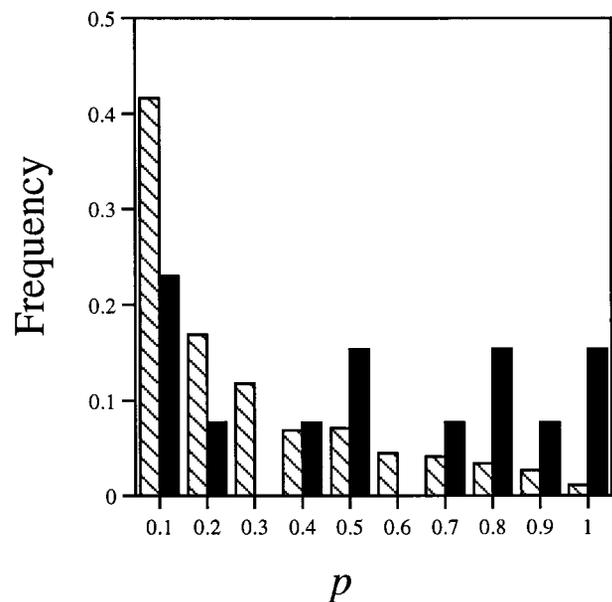
Figure 5.—The observed distribution of P values (solid bars) for the 25 data sets and the expected distribution (hatched bars) for $Ns = 3$ and the same parameter values as in Figure 2.
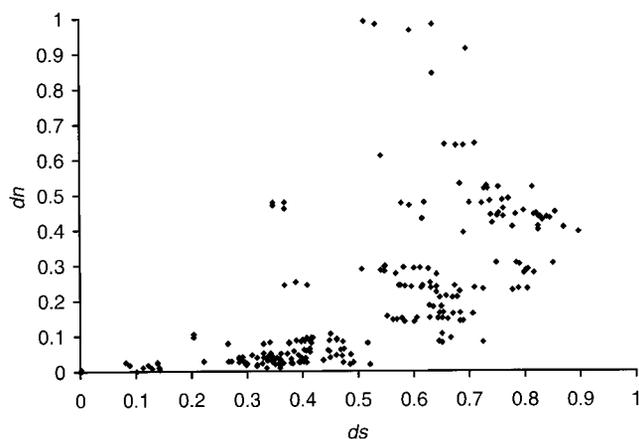
Figure 6.—Observed nonsynonymous differences per site (*dn*) and synonymous differences per site (*ds*) in pairwise comparisons for multiple mtDNA-encoded proteins.

populations, it is difficult to explain why all values of *Ns* should fall in this range.

The model applied to describe selection against deleterious mutations is very simple in that it assumes that all nonlethal nonsynonymous mutations have the same selection coefficient and that this selection coefficient is constant in time. Likewise, it is assumed in the simulations that the populations are in stationarity, *i.e.*, a constant population size and no population subdivision. Although there is no reason to assume that a more complex demographic model or a more complex distribution of negative selection coefficients would change the relative power of the McDonald/Kreitman test *vs.* the AOM test, we cannot rule out this possibility.

If purifying selection alone is not responsible for the repeated observations of deviation from neutrality, what other processes are consistent with the data? One obvious explanation is that the significant results partly are artifacts caused by the simple methods used to estimate the number of nonsynonymous and synonymous mutations (*e.g.*, Hasegawa *et al.* 1998). Such an explanation would require that saturation occur faster in nonsynonymous than in synonymous sites, *i.e.*, the number of nonsynonymous nucleotide differences should be a convex function of the number of synonymous nucleotide differences (Nielsen 1997). To examine this hypothesis, we plotted the number of nonsynonymous nucleotide differences between pairs of species against the number of synonymous nucleotide differences for several mitochondrial genes (Figure 6). As expected, saturation occurs faster in synonymous sites than in nonsynonymous sites. Thus the use of nucleotide differences instead of the estimated number of substitutions in the McDonald/Kreitman test does not appear to be responsible for the apparent pattern of deviation from neutrality in mtDNA evolution. In fact, saturation by multiple hits should make the McDonald/Kreitman test more conservative against the alternative of segregating deleterious mutations.

Another problem is sequencing errors. Such errors could inflate the significance level of the McDonald/Kreitman test because they would tend preferentially to elevate the number of intraspecific nonsynonymous mutations. Likewise, sequencing errors could be of importance in explaining the significant results from the AOM test presented in this article. However, sequencing errors should also elevate the ratio of nonsynonymous mutations to synonymous mutations for small levels of divergence in interspecific data. Because no such pattern is observed in Figure 6, we do not believe that this explanation is likely. Moreover, elevated intraspecific nonsynonymous mutations have not been observed in nuclear data sets (McDonald and Kreitman 1991; D. M. Weinreich and D. Rand, unpublished results). However, we cannot conclusively rule out that sequencing errors may be of some importance in explaining the large number of significant McDonald/Kreitman tests.

Theoretically, balancing selection acting on mtDNA could maintain levels of nonsynonymous polymorphism within species above that seen between species (Nachman *et al.* 1994; Rand and Kann 1996). Balancing selection could act through cytoplasmic-nuclear interaction (McRae and Anderson 1988; Babcock and Asmussen 1996; Goodisman and Asmussen 1997). However, it is difficult to imagine that this type of selection can act on sufficiently many nucleotide sites simultaneously to account for the observed results. Indeed, selection against slightly deleterious mutations appears at the moment to be the most plausible explanation for the significant results of the McDonald/Kreitman test in animal mtDNA. However, a combination of forces cannot be excluded.

LITERATURE CITED

Akashi, H., 1999 Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. Genetics **151:** 221–238.

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

Babcock, D. S., and M. A. Asmussen, 1996 Effects of differential selection in the sexes on cytonuclear polymorphism and disequilibria. Genetics **144:** 839–853.

Baker, R. J., R. A. VanDenBussche, A. J. Wright, L. E. Wiggins, M. J. Hamilton *et al.*, 1996 High levels of genetic change on rodents of Chernobyl. Nature **380:** 707–708.

Ballard, J. W. O., and M. Kreitman, 1994 Unraveling selection in the mitochondrial genome of *Drosophila.* Genetics **138:** 875–888.

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.

Bruna, E. M., R. N. Fisher and T. J. Case, 1996 Morphological and genetic evolution is decoupled in Pacific skinks (Squamata: Scinidae: Emoia). Proc. R. Soc. Lond. Ser. B. **263:** 681–688.

Desjardins, P., and R. Morais, 1990 Sequence and gene organiza-

tion of the chicken mitochondrial genome. J. Mol. Biol. **212:** 599–634.

Fitch, W. M., 1971   Toward defining the course of evolution: minimum change for a specific tree topology. Syst. Zool. **20:** 406–416.

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

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

Goodisman, M. A. D., and M. A. Asmussen, 1997   Cytonuclear theory for haplodiploid species and *X*-linked genes. I. Hardy-Weinberg dynamics and continent-island, hybrid zone models. Genetics **147:** 321–338.

Griffiths, R. C., and S. Tavaré, 1995   Unrooted genealogical tree probabilities in the infinitely-many-sites model. Math. Biosci. **127:** 77–98.

Griffiths, R. C., and S. Tavaré, 1998   The age of a mutation in a general coalescent tree. Stoch. Model. **14:** 273–295.

Griffiths, R. C., and S. Tavaré, 1999   The ages of mutations in gene trees. Ann. Appl. Prob. (in press).

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. Am. J. Hum. Genet. **60:** 772–789.

Hasegawa, M., Y. Cao and Z. Yang, 1998   Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. Mol. Biol. Evol. **15:** 1499–1505.

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

Horai, S., K. Hayasaka, R. Kondo, K. Tsugane and N. Takahata, 1995   Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA **92:** 532–536.

Hudson, R. R., and N. L. Kaplan, 1994   Gene trees with background selection, pp. 140–153 in *Non-Neutral Evolution*, edited by B. Golding. Chapman and Hall, New York.

Kaplan, N. L., T. Darden and R. R. Hudson, 1988   The coalescent in models with selection. Genetics **120:** 819–829.

Kimura, M., 1983   *The Neutral Theory of Molecular Revolution.* Cambridge University Press, Cambridge, United Kingdom.

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

McRae, A. F., and W. W. Anderson, 1988   Evidence for non-neutrality of mitochondrial DNA haplotypes in *Drosophila pseudoobscura.* Genetics **120:** 485–494.

Moritz, C., C. J. Schneider and D. B. Wake, 1992   Evolutionary relationships within the *Ensatina eschscholtzii* complex confirm the ring species interpretation. Syst. Biol. **41:** 273–291.

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   Non-neutral 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.

Nagylaki, T., 1992   *Introduction to Theoretical Population Genetics.* Springer-Verlag, Berlin.

Neuhauser, C., and S. M. Krone, 1997   The genealogy of samples in models with selection. Genetics **145:** 519–534.

Nielsen, R., 1997   The ratio of replacement to silent divergence and tests of neutrality. J. Evol. Biol. **10:** 217–231.

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

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

Rand, D. M., and L. Kann, 1998   Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA. Genetica **102/103:** 393–407.

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

Ruvolo, M., D. Pan, S. Zehr, T. Goldberg, T. R. Disotell *et al.*, 1994   Gene trees and hominoid phylogeny. Proc. Natl. Acad. Sci. USA **91:** 8900–8904.

Sokal, R. R., and C. D. Michener, 1958   A statistical method for evaluating systematic relationship. Univ. Kansas Sci. Bull. **28:** 1409–1438.

Sokal, R. R., and F. J. Rohlf, 1981   *Biometry.* W. H. Freeman and Company, New York.

Summers, K., E. Bermingham, L. Weigt, S. McCafferty and L. Dahlstrom, 1997   Phenotypic and genetic divergence in three species of dart-poison frogs with contrasting parental behavior. J. Hered. **88:** 8–13.

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

Wise, C. A., S. Micheala 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.

Wood, T. C., and C. Krajewski, 1996   Mitochondrial DNA sequence variation among the subspecies of Sarus Crane (*Grus antigone).* Auk **113:** 655–663.

Zink, R. M., and R. C. Blackwell, 1996   Patterns of allozyme, mitochondrial DNA, and morphometric variation in four sparrow genera. Auk **113:** 59–67.