Preponderance of Slightly Deleterious Polymorphism in Mitochondrial DNA: Nonsynonymous/Synonymous Rate Ratio is Much Higher Within Species than Between Species

Masami Hasegawa, Ying Cao,* and Ziheng Yang*†*

*The Institute of Statistical Mathematics, Tokyo; and †Department of Biology (Galton Laboratory), University College, London

We estimated synonymous (d_N) and nonsynonymous (d_S) substitution rates for protein-coding genes of the mitochondrial genome from two individuals each of the species human, chimpanzee, and gorilla. The genes were analyzed both separately and in a combined data set. Pairwise sequence comparisons suggest that the d_N/d_S rate ratios are about 5–10 times higher in within-species comparisons than in between-species comparisons. This result is confirmed by a more rigorous likelihood ratio test, which rejected the null hypothesis that the d_N/d_S rate ratios are identical within and between species. The likelihood models account for the genetic code structure, transition/transversion rate ratio, and codon usage bias and are expected to produce more reliable results than the commonly used contingency test. Separate analyses of different genes show that the d_N/d_S rate ratios are higher within species than between species for all 13 mitochondrial genes, with the difference being statistically significant for all except three small or slowly evolving genes. Furthermore, in conserved genes, nonsynonymous rates within species tend to be higher than the between-species rates by a greater proportion than in fast-changing genes. Our findings confirm and extend earlier results obtained from smaller data sets and suggest the operation of slightly deleterious mutations throughout the mitochondrial genome in the hominoids. Implications of the results for evolutionary studies and, in particular, for studies of the origin of modern humans, are discussed.

Introduction

Recently, Parsons et al. (1997) measured the mutation (substitution) rate in the mitochondrial DNA (mtDNA) control region using human pedigree data and found that the rate is much higher (about 20-fold higher) than those estimated in phylogenetic analyses. This result is consistent with the earlier observation of Howell, Kubacka, and Mackay (1996) based on a smaller number of generation events in humans, and with observations for mtDNA of other mammalian species (see Loewe and Scherer 1997 for a review). Parsons et al. (1997) attributed the disparity between the pedigree and phylogenetic estimates of the substitution rate to the presence of slightly deleterious mutations which are removed from the population over time.

We note that estimates obtained from the pedigree analysis largely reflect the mutation rate (although lethal mutations are probably not observed even in a pedigree analysis). However, phylogenetic analysis using multiple-species data measures the substitution rate or the rate of fixed mutations. While the two rates may be expected to be identical if the within-species polymorphism and the between-species divergence are both caused by neutral mutations, this will not be the case if evolution of the DNA is not neutral. An important prediction of the neutral theory is that the nonsynonymous/synonymous substitution rate ratios (d_N/d_S) will be identical within and between species (McDonald and Kreitman 1991). A contingency test was suggested to test this prediction of McDonald and Kreitman (1991; see also Akashi 1995; Templeton 1996). The test classifies observed (inferred) mutations into different classes (synonymous vs. nonsynonymous, within- vs. between-species) and tests the independence of the two groupings. The test has been used to examine the mitochondrial cytochrome *b* (Ballard and Kreitman 1994) and ND5 (Rand, Dorfsman, and Kann 1994) genes of *Drosophila*, mitochondrial ND3 (Nachman, Boyer, and Aquadro 1996) and COII (Templeton 1996) genes of humans, chimpanzees, and gorillas, and mitochondrial ND2 genes of humans and chimpanzees (Wise, Sraml, and Easteal 1998). Neutrality was rejected in almost all of those analyses, with a significantly higher d_N/d_S rate ratio within than between species.

Violation of the neutral assumption in mitochondrial DNA evolution has important implications for phylogenetic as well as population genetic analyses. For example, estimation of the time to the most recent common ancestor of all humans is highly important in the controversy concerning the origin of modern humans. A common practice has been to use a species divergence date (such as the date of human–chimpanzee separation) to calculate the substitution rate, and then to use this rate to estimate the age of the most recent common ancestor of humans (see, e.g., Vigilant et al. 1991). If the mutation (substitution) rate differs within and between species, such molecular-clock calibration will lead to seriously biased estimates. In particular, if the within-species rate is higher than the between-species rate with elevated polymorphisms, the age to the most recent common ancestor will be overestimated, while if the within-species rate is lower with reduced polymorphisms, the age will be underestimated (see, e.g., Rand, Dorfsman, and Kann 1994; Nachman et al. 1996).

It should be noted that reliable estimation of the synonymous and nonsynonymous substitution rates is not an easy task (Ina 1995; Yang and Nielsen 1998).

Key words: hominoids, mitochondrial DNA, slightly deleterious mutations, codon substitution, nonsynonymous/synonymous rate ratio, likelihood ratio test.

Address for correspondence and reprints: Masami Hasegawa, Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106–8569, Japan. E-mail: hasegawa@ism.ac.jp.

Mol. Biol. Evol. 15(11):1499–1505. 1998

Counting of sites and differences is always complicated by features of sequence evolution such as transition/transversion rate bias and codon usage bias. The contingency test (McDonald and Kreitman 1991) used in previous analyses is crude and does not correct for multiple substitutions at the same site (Graur and Li 1991; Whittam and Nei 1991; Maynard Smith 1994). Templeton (1996) suggested ways to guard against saturation of substitutions and biased estimates of phylogenetic trees and branch lengths, but these methods lack a rigorous statistical basis. Graur and Li (1991) and Whittam and Nei (1991) suggested tests based on pairwise estimates of synonymous and nonsynonymous substitution rates. However, these tests rely on unreliable normal approximation to the test statistics (Sawyer and Hartl 1992).

To test the hypothesis of neutral evolution in the mitochondrial DNA and to obtain more reliable estimates of synonymous and nonsynonymous rates, we analyze a large data set containing all of the protein-coding genes of the mitochondria from humans, chimpanzees, and gorillas. Whole mitochondrial genome sequences are available for two individuals of each of these species. We estimated the d_N/d_S rate ratios in within- and between-species comparisons of the gene sequences. Further, a recently developed likelihood ratio test (Yang and Nielsen 1998; Yang 1998) is used to test whether the d_N/d_S rate ratios are identical along the within- and between-species lineages. In contrast to the contingency test used by previous authors, the likelihood models explicitly take into account the genetic code structure, the transition/transversion rate bias, and codon usage biases (Goldman and Yang 1994; Yang and Nielsen 1998). The probabilistic model of codon substitution adopted in the likelihood calculation corrects properly for multiple hits (Goldman and Yang 1994; Muse and Gaut 1994). The likelihood ratio test thus provides a framework for testing the neutral hypothesis and can be expected to provide more reliable estimates of the synonymous and nonsynonymous substitution rates.

Materials and Methods

Complete mitochondrial genome data for two individuals from each of the hominoid species human (*Homo sapiens,* accession numbers D38112 and X93334), common chimpanzee (*Pan troglodytes,* accession numbers D38113 and X93335), and gorilla (*Gorilla gorilla,* accession numbers D38114 and X93347) were used. One set of sequences from each of the three species was determined by Horai et al. (1995), and another set was determined by Arnason, Xu, and Gullberg (1996) and Xu and Arnason (1996). The 12 protein genes encoded by the same strand of mtDNA were analyzed in a combined data set (with 3,569 codons, excluding initiation and termination codons and overlapping regions between ATP6 and ATP8 and between ND4 and ND4L). The ND6 gene is encoded by the other strand of mtDNA, with very different codon usage biases, and is not included in the combined data set. All 13 genes were then analyzed separately. Alignment of the sequences was simple, as no gaps needed to be introduced.

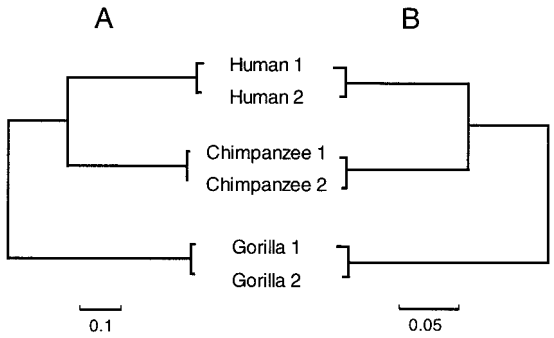

FIG. 1.—Phylogenetic trees constructed using maximum-likelihood estimates of (*A*) synonymous and (*B*) nonsynonymous substitution rates. Rates are estimated under the model that assumes two different nonsynonymous/synonymous ratios for the within- and between-species branches. Note that the two trees have different shapes; the tree of nonsynonymous rates has longer within-species branches relative to the tree of synonymous rates. Branch length is measured by the expected number of nucleotide substitutions per site.

As a preliminary analysis, the numbers of synonsymous (d_S) and nonsynonymous (d_N) substitutions per site were estimated from pairwise comparisons using the methods of Nei and Gojobori (1986) and Ina (1995). Both methods count the numbers of synonymous and nonsynonymous sites and synonymous and nonsynonymous differences between two sequences and apply approximate formulas to correct for multiple hits. Ina's (1995) method is based on Kimura's (1980) two-parameter model of nucleotide substitution and accounts for the transition/transversion bias while ignoring nucleotide or codon usage bias. Ina (1995) suggested two slightly different methods. Method I estimates the mutational transition/transversion rate ratio using data from the third codon positions, while method II uses a more sophisticated iterative algorithm. The two methods produced similar estimates for the mitochondrial genes, and only results from method II are presented. The method of Nei and Gojobori (1986) ignores both the transition/transversion bias and the codon usage bias. The maximum-likelihood method based on a Markov model of codon substitution was also used for pairwise comparison (Goldman and Yang 1994). The approach calculates d_S and d_N according to their definitions directly from maximum-likelihood estimates of model parameters. The complicated task of counting differences and correcting for multiple hits is taken care of by the probability theory. The likelihood model accounts for the transition/transversion rate bias and the codon usage bias (see Goldman and Yang [1994] and Yang and Nielsen [1998] for details). We use nucleotide frequencies at the three codon positions to calculate codon frequencies, with $3 \times (4 - 1) = 9$ parameters used. In a separate set of analyses, codon frequencies were used as free parameters; the results are very similar and thus are not presented.

The codon substitution models were also used to perform joint likelihood comparisons of all sequences in the data set (fig. 1). Two models concerning the non-

Table 1
Sequence Divergence Statistics for Pairwise Comparisons

	d_0	d_1	d_2	d_3	d_a
Within species					
Human	3,511	58	0	0	21
Common chimpanzee	3,553	16	0	0	9
Gorilla	3,535	33	1	0	18
Between species					
Human/chimpanzee	2,568.5	942.5	55	3	163.5
Human/gorilla	2,386	1,079	97	7	210
Chimpanzee/gorilla	2,433	1,036.5	93.5	6	214.5

NOTE.—d_k (k = 0, 1, 2, 3) is the number of codon sites at which k codon positions are different between the two sequences, and $d_0 + d_1 + d_2 + d_3 = 3{,}569$ being the number of codons in the sequence. d_a is the number of amino acid differences. Between-species counts are averages over all four comparisons.

synonymous/synonymous rate ratios ($\omega = d_N/d_S$) among lineages were used. Model 0 assumes the same ratio for all branches in the tree, while model 1 assumes that the ratio (ω_B) for the three between-species branches is different from the ratio (ω_W) for the six within-species branches (see fig. 1). The log-likelihood values under the two models can be compared with a χ^2 distribution with one degree of freedom to test whether the within- and between-species branches have the same ratio. This constitutes a likelihood ratio test of the neutral prediction that $\omega_B = \omega_W$. A model that assumes independent d_N/d_S rate ratios for all branches in the phylogenetic tree was also used to obtain estimates of synonymous and nonsynonymous rates along lineages without constraints.

Results

Combined Data Set

Some basic statistics of the sequence data are given in table 1. The test of McDonald and Kreitman (1991) requires unambiguous classification of observed differences into synonymous and nonsynonymous categories. This does not seem possible for the mitochondrial genes analyzed in this paper, as some codon sites differ at two or three codon positions between species, and it is not clear whether the differences should be counted as synonymous or nonsynonymous (table 1). Furthermore, a number of nucleotide sites are both different between species and polymorphic within species and pose difficulties in the procedure of McDonald and Kreitman (see Graur and Li 1991; Whittam and Nei 1991). The gene sequences are divergent enough that methods explicitly accounting for multiple hits are necessary.

Estimates of synonymous and nonsynonymous rates from pairwise comparisons of sequences are shown in table 2. The nonsynonymous rates are all very low, and different methods produced very similar estimates. Estimates of synonymous rates are quite high, and substantial differences exist among methods. The likelihood method may be expected to be more reliable, as it does not involve approximations of the other methods and is based on more realistic assumptions about the substitution process. Compared with the likelihood method, the method of Nei and Gojobori underestimates the synonymous rates, probably because it is based on a simple substitution model. The transition/transversion rate bias and codon usage bias appear to have opposite effects on estimates of synonymous and nonsynonymous rates (Yang and Nielsen 1998), so estimates obtained with Ina's method are more different from the likelihood estimates than are estimates obtained with Nei and Gojobori's method. The patterns found here are the same as those reported by Yang and Nielsen (1998) from analyses of nuclear genes.

While all estimates of d_N/d_S are much smaller than one, indicating the operation of negative selection, the within-species estimates are much greater than the between-species estimates. By the likelihood estimates, the within-species d_N/d_S ratios are 5–10 times higher than the between-species estimates (table 2). Although it is

Table 2
Pairwise Estimates of d_N and d_S by Different Methods

	Nei and Gojobori		Ina's Method II		Maximum Likelihood (κ = 20)		
	d_N	d_S	d_N	d_S	d_N	d_S	d_N/d_S + SE
Within species							
Human	0.0028	0.0143	0.0030	0.0099	0.0027	0.0138	0.198 ± 0.054
Common chimpanzee	0.0012	0.0027	0.0012	0.0023	0.0012	0.0025	0.478 ± 0.235
Gorilla	0.0024	0.0064	0.0026	0.0045	0.0024	0.0060	0.397 ± 0.017
Between species							
Human/chimpanzee	0.0234	0.4517	0.0258	0.2796	0.0234	0.5911	0.040 ± 0.004
Human/gorilla	0.0318	0.5826	0.0346	0.3597	0.0313	0.9465	0.033 ± 0.003
Chimpanzee/gorilla	0.0323	0.5371	0.0352	0.3350	0.0319	0.8020	0.040 ± 0.003

NOTE.—The methods used are Nei and Gojobori's (1986) methods, Ina's (1995) method II, and maximum likelihood (Goldman and Yang 1994). Estimates of d_N and d_S are shown for each method, while the ratios ($\omega = d_N/d_S$) and their standard errors are also shown for maximum likelihood.

Table 3
Separate Analysis of Different Genes

Order	Gene	No. of Codons	Tree Length (nonsynonymous)	Tree Length (synonymous)	ω_W	ω_B	ω_W/ω_B	$2\Delta\ell$
1	COI	512	0.013	0.979	0.368	0.010	38.4	17.1*
2	COII	226	0.019	1.136	0.187	0.015	12.6	2.75
3	ATP8	52	0.120	1.460	0.373	0.073	5.1	1.99
4	ATP6	210	0.062	0.812	0.468	0.068	6.9	4.54*
5	COIII	260	0.024	1.176	0.149	0.018	8.5	4.19*
6	ND3	114	0.065	1.040	∞	0.050	∞	13.32**
7	ND4L	95	0.016	0.876	0.140	0.013	11.5	2.41
8	ND4	456	0.046	1.113	0.167	0.038	4.4	5.54*
9	ND5	602	0.080	1.144	0.265	0.067	4.0	5.40*
10	ND6	173	0.034	0.987	∞	0.028	∞	10.47**
11	Cyt *b*	379	0.064	1.967	0.304	0.027	11.5	22.75**
12	ND1	317	0.043	0.976	0.211	0.038	5.6	6.58*
13	ND2	346	0.050	1.230	0.383	0.035	10.8	10.84**

NOTE.—The transition/transversion rate ratio is fixed at 20. Estimates of the tree length (sum of branch lengths) and synonymous and nonsynonymous substitution rates are obtained under the two-ratio model assuming different nonsynonymous/synonymous rate ratios for the within- (ω_W) and between-species (ω_B) branches. The likelihood ratio statistic, $2\Delta\ell$, is twice the log-likelihood difference between the two-ratio and one-ratio models. Order is the order of genes in the mitochondrial genome (Nedbal and Flynn 1998).

* Significant ($\chi^2_1 = 3.84$ at 5%).

** Extremely significant ($\chi^2_1 = 6.63$ at 1%).

possible to use pairwise estimates to construct a test of the null hypothesis that the d_N/d_S ratio is identical within and between species (Graur and Li 1991; Whittam and Nei 1991; but see Sawyer and Hartl 1992 for criticisms), this was not attempted here. The standard errors of the maximum-likelihood estimates of the d_N/d_S ratios in table 2 were obtained by the curvature method, that is, by inverting the matrices of the second deratives of the log-likelihood function with respect to model parameters. The standard errors for between-species estimates are simple averages over the four possible comparisons and are thus overestimated. The correct standard errors are not easy to calculate, due to the dependence among different pairwise estimates, but should be greater than half of the estimates shown in table 2.

A likelihood model that assumes independent d_N/d_S rate ratios for different lineages (Yang and Nielsen 1998) was used to estimate synonymous and nonsynonymous rates for branches in the phylogenetic tree. This model involves nine d_N/d_S ratio parameters (ω's) for the nine branches (fig. 1) and the log-likelihood value is $\ell = -20{,}440.52$. Estimates of the between-species ratios under this model are 0.033 ± 0.005 for the lineage ancestral to humans, 0.047 ± 0.007 for the lineage leading to chimpanzees, and 0.035 ± 0.004 for the lineage leading to gorillas (see fig. 1). Estimates of the within-species ratios are 0.199 and 0.209 for humans, 0.647 and 0.160 for chimpanzees, and 0.393 and 0.393 for gorillas. The rate ratios are much higher along the within-species branches than along the between-species branches.

The model that assumes the same d_N/d_S ratio for all branches in the tree gave a log-likelihood value $\ell_0 = -20{,}486.03$, with an estimated d_N/d_S ratio of 0.044 ± 0.003 and a transition/transversion rate ratio (α/β) of 20.7 ± 0.1. The model that assumes two different d_N/d_S rate ratios for the within- and between-species branches in the tree gave a log-likelihood value $\ell_1 = -20{,}444.10$, with estimates of $\hat{\omega}_W = 0.287 \pm 0.010$ for the within-species branches and $\hat{\omega}_B = 0.037 \pm 0.003$ for the between-species branches. The estimate of the transition/transversion rate ratio is 21.6 ± 0.2 under this model. The likelihood ratio statistic, $2\Delta\ell = 2(\ell_1 - \ell_0) = 83.86$, is much greater than the χ^2critical value with one degree of freedom, suggesting that the d_N/d_S rate ratios are significantly different among lineages. Indeed, the estimated ratio is about 7.8 (= 0.287/0.037) times higher along within-species lineages than along between-species lineages, in congruence with the analysis based on pairwise distances discussed above. Figure 1 shows the phylogenetic trees constructed using the synonymous and nonsynonymous substitution rates. It is obvious that the two trees have different shapes and the tree of nonsynonymous rates has much longer within-species branches than does the tree of synonymous rates.

Separate Analyses of Different Genes

Results obtained from separate analyses of all 13 protein-coding genes are shown in table 3. First, we note that the tree lengths (sums of branch lengths along the tree) calculated using synonymous rates are quite homogeneous among genes (about one synonymous substitution per synonymous site along the tree), although cytochrome *b* has a somewhat higher rate (almost two synonymous substitutions per site). The synonymous rates of genes are not correlated with their locations or replication times in the mitochondrial genome, in congruence with the conclusion of Nedbal and Flynn (1998). The nonsynonymous rates are much more variable among genes, and the tree lengths for nonsynonymous substitutions range from 0.013 for the conserved COI gene to 0.122 for the fast-changing ATP6 gene. It is noteworthy that for all 13 genes, the d_N/d_S ratio within species (ω_W) is much higher than the ratio between species (ω_B). This is also the case for the ND6 gene, which is encoded by the other strand of mitochondrial DNA. By the likelihood ratio test, the difference is statistically

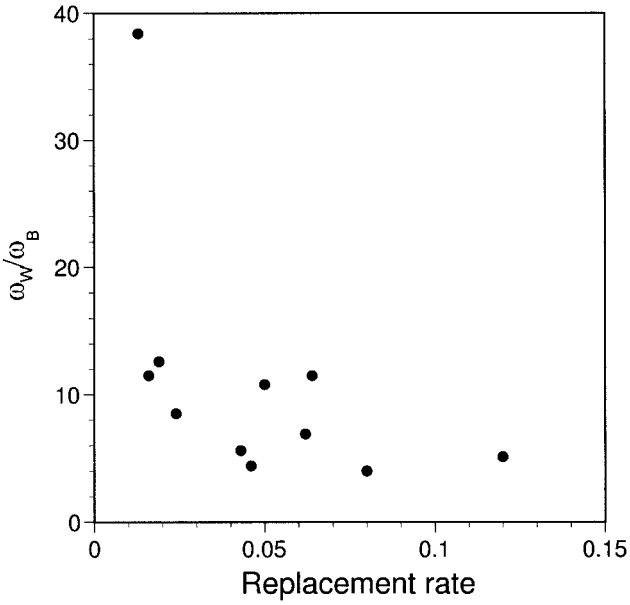

FIG. 2.—The ω_W/ω_B ratio plotted against the tree length for nonsynonymous substitutions for different genes (see table 3). Results for the ND3 and ND6 genes are not plotted.

significant for all but 3 of the 13 genes (table 3). The exceptions are the two very small genes (ATP8 and ND4L) and the conserved COII gene. These genes probably do not contain enough information to reject the null hypothesis of a single d_N/d_S rate ratio within and between species. We note that the χ^2_1 approximation of the likelihood ratio statistic relies on large-sample properties. Previous simulations under nucleotide-based models suggest that the χ^2 approximation is reliable (e.g., Yang, Goldman, and Friday 1995). At any rate, the approximation is expected to be reliable for large or fast-changing genes, for which significant differences are found.

Another interesting pattern in table 3 is the negative correlation between the variability of the gene, as reflected by the tree length of nonsynonymous substitutions, and the ω_W/ω_B ratio (fig. 2). While the ω_W/ω_B ratios are different within and between species for all genes, the difference is greater for conserved genes than for fast-changing genes. One possible explanation is that in conserved genes, a higher proportion of nonsynonymous mutations are deleterious and are tolerated in the population without contributing to between-species diversity, while in fast-changing genes, more nonsynonymous mutations are nearly neutral and contribute to both within-species polymorphisms and between-species divergences. Since synonymous rates are much more homogeneous within and between species than are nonsynonymous rates, the pattern of figure 2 can be explained by elevated nonsynonymous polymorphisms in conserved genes. It should be noted, however, that all genes in the mitochondrial genome are linked. Even neutral polymorphisms (such as synonymous substitutions) are affected by both background selection, eliminating deleterious mutations (Charlesworth, Charlesworth, and Morgan 1995), and hitchhiking effect, driving advantageous mutations to fixation (Maynard Smith and Haigh 1974). The population dynamics of synonymous and nonsynonymous mutations in the mitochondrial genes must be very complicated.

Discussion

Sawyer and Hartl (1992) pointed out that hypothesis testing using a 2×2 contingency table is well studied in statistics, and other tests, such as those suggested by Graur and Li (1991) and Whittam and Nei (1991), may be less powerful statistically. However, this argument assumes that entries in the contingency table are correctly recorded. This may not be the case for the test of McDonald and Kreitman (1991) except for nuclear genes of very closely related species. Indeed, Maynard Smith (1994) was able to construct a purely neutral model under which the ω_B/ω_W ratio, estimated using the approach of McDonald and Kreitman (1991) without correcting for multiple hits, approaches 2 at high levels of sequence divergence, irrespective of the true ratio. The mitochondrial genes analyzed in this paper, although from closely related species, involve considerable synonymous changes between species (>0.5 substitutions per site; see table 2), and the test of McDonald and Kreitman (1991) does not appear to be applicable. Sawyer and Hartl (1992) pointed out that the tests suggested by Graur and Li (1991) and Whittam and Nei (1991) are based on normality assumptions of test statistics, while the true distributions are unknown. These methods also suffer from the use of simplistic evolutionary models and limitations of pairwise comparisons. In contrast, the likelihood ratio test used in this paper has a sound statistical basis and is also based on more realistic substitution models. While there is reason to expect the likelihood ratio test to give more reliable results, the performances of those tests, when their assumptions are and are not met, need further investigation.

The neutral theory predicts that the neutral mutation rate determines both the divergence between species and the polymorphism within species (Kimura 1983). Although it is difficult to test this prediction directly, as the genetic diversity within a species depends on other factors besides the neutral mutation rate, such as the population size and departure from the equilibrium, the d_N/d_S rate ratios in within- and between-species comparisons should be the same if the neutral theory is correct (McDonald and Kreitman 1991). McDonald and Kreitman (1991) found, using the simple contingency test, that evolution at the *Adh* locus in *Drosophila* is not neutral, with the d_N/d_S ratios being much higher between species than within species. While our likelihood ratio test also rejected neutrality for mitochondrial protein-coding DNAs, our result is the opposite of theirs: the d_N/d_S rate ratio is much higher within species than between species. This is congruent with previous results obtained from analyzing a few mitochondrial genes of hominoids (Nachman et al. 1996; Templeton 1996; Wise, Sraml, and Easteal 1998), and

suggests that the pattern is common to all genes in the mitochondria.

A number of factors have been suggested to explain the elevated nonsynonymous rates within species, including balanced selection, relaxed selective constraints, and slightly deleterious mutations (Ballard and Kreitman 1995; Nachman, Boyer, and Aquadro 1994; Takahata 1993). Since mitochondrial genomes are haploid, heterozygous advantage is not possible. Although mitochondrial genomes may be potential targets of balanced selection due to environmental changes, population subdivision, or cytoplasmic–nuclear interactions, the predictions of this hypothesis are not generally consistent with data (Nachman, Boyer, and Aquadro 1994). The high d_N/d_S rate ratio within the human species relative to that between species has previously been attributed to relaxed selective constraints in the human lineage, possibly due to cultural evolution or improved environment (Takahata 1993; Adachi and Hasegawa 1996). However, results of this paper suggest that the pattern is not unique to humans, since it is also observed in chimpanzees and gorillas.

The most likely interpretation seems to be the hypothesis of slightly deleterious mutations (Ohta 1992). Deleterious mutations should ultimately be eliminated from the population and rarely contribute to between-species divergence. Since the synonymous rates are expected to be homogeneous within and between species, such a variation in nonsynonymous rates caused by deleterious mutations will lead to higher d_N/d_S rate ratios within species than between species. Our estimates of the d_N/d_S rate ratios are lower (0.2) for humans than for chimpanzees and gorillas (0.4). It is not clear whether the differences are due to different population sizes in those species. The hypothesis of slightly deleterious mutations predicts that higher polymorphisms will be maintained in smaller populations. The difference will be explicable if the human species has had a larger population size than the chimpanzee and gorilla species. While Takahata, Satta, and Klein (1995) obtained estimates of the human population size, estimates for chimpanzees and gorillas are lacking.

Our estimates suggest that the nonsynonymous rates in mitochondrial protein-coding genes were about 5–10 times higher within species than between species. This difference is still smaller than the difference in substitution (mutation) rate estimates obtained for the D-loop from human pedigree and phylogenetic analyses (see *Introduction*). Studies suggest that the D-loop mtDNAs are functional and under selective constraints. One likely explanation for the extremely high mutation rate found in the pedigree analysis is that more deleterious mutations are observed in a pedigree analysis than in analyses of samples of individuals, as performed in this paper. In any case, it is clear that the mitochondrial DNA is not a neutral marker (Ballard and Kreitman 1994, 1995; Rand, Dorfsman, and Kann 1994; Nachman et al. 1996; Hey 1997), and one should be cautious in using substitution rates obtained from phylogenetic analysis to date within-species events.

Acknowledgments

We thank Dr. N. Takahata for discussion and Dr. M. Uyenoyama and three anonymous referees for comments. This work was supported by grants from the Ministry of Education, Science, Sports and Culture of Japan.

LITERATURE CITED

ADACHI, J., and M. HASEGAWA. 1996. Tempo and mode of synonymous substitutions in mitochondrial DNA of primates. Mol. Biol. Evol. **13**:200–208.

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

ARNASON, U., X. XU, and A. GULLBERG. 1996. Comparison between the complete mitochondrial DNA sequences of *Homo* and the common chimpanzee based on nonchimeric sequences. J. Mol. Evol. **42**:145–152.

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

———. 1995. Is mitochondrial DNA a strictly neutral marker? Trends Ecol. Evol. **10**:485–488.

CHARLESWORTH, D., B. CHARLESWORTH, and M. MORGAN. 1995. The pattern of neutral molecular variation under the background selection model. Genetics **141**:1619–1632.

GOLDMAN, N., and Z. YANG. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.

GRAUR, D., and W.-H. LI. 1991. Neutral mutation hypothesis test. Nature **354**:114–115.

HEY, J. 1997. Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol. Biol. Evol. **14**:166–172.

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

HOWELL, N., I. KUBACKA, and D. MACKEY. 1996. How rapidly does the human mitochondrial genome evolve? Am. J. Hum. Genet. **59**:501–509.

INA, Y. 1995. New methods for estimating the numbers of synonymous and nonsynonymous substitutions. J. Mol. Evol. **40**:190–226.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

———. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England.

LOEWE, L., and S. SCHERER. 1997. Mitochondrial Eve: the plot thickens. Trends Ecol. Evol. **12**:422–423.

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

MAYNARD SMITH, J. 1994. Estimating selection by comparing synonymous and substitutional changes. J. Mol. Evol. **39**:123–128.

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

MUSE, S. V., and B. S. GAUT. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to chloroplast genome. Mol. Biol. Evol. **11**:715–724.

NACHMAN, M., S. BOYER, and C. AQUADRO. 1994. Non-neutral evolution at the mitochondrial NADH dehydrogenase sub-

unit 3 gene in mice. Proc. Natl. Acad. Sci. USA **91**:6364–6368.

NACHMAN, M., W. BROWN, M. STONEKING, and C. AQUADRO. 1996. Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics **142**:953–963.

NEDBAL, M. A., and J. J. FLYNN. 1998. Do the combined effects of the asymmetric process of replication and DNA damage from oxygen radicals produce a mutation-rate signature in the mitochondrial genome? Mol. Biol. Evol. **15**: 219–223.

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

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

PARSONS, T., D. MUNIEC, K. SULLIVAN et al. (11 co-authors). 1997. A high observed substitution rate in the human mitochondrial DNA control region. Nat. Genet. **15**:363–368.

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

SAWYER, S. A., and D. L. HARTL. 1992. Population genetics of polymorphism and divergence. Genetics **132**:1161–1176.

TAKAHATA, N. 1993. Relaxed natural selection in human populations during the Pleistocene. Jpn. J. Genet. **68**:539–547.

TAKAHATA, N., Y. SATTA, and J. KLEIN. 1995. Divergence time and population size in the lineage leading to modern humans. Theor. Popul. Biol. **48**:198–221.

TEMPLETON, A. 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.

VIGILANT, L., M. STONEKING, H. HARPENDING, K. HAWKES, and A. WILSON. 1991. African populations and the evolution of human mitochondrial DNA. Science **253**:1503–1507.

WHITTAM, T., and M. NEI. 1991. Neutral mutation hypothesis test. Nature **354**:115–116.

WISE, C., M. SRAML, and S. EASTEAL. 1998. Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 in humans, but not in chimpanzees. Genetics **148**: 409–421.

XU, X., and U. ARNASON. 1996. A complete sequence of the mitochondrial genome of the Western lowland gorilla. Mol. Biol. Evol. **13**:691–698.

YANG, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

YANG, Z., N. GOLDMAN, and A. E. FRIDAY. 1995. Maximum likelihood trees from DNA sequences: a peculiar statistical estimation problem. Syst. Biol. **44**:384–399.

YANG, Z., and R. NIELSEN. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. **46**:409–418.

MARCY K. UYENOYAMA, reviewing editor

Accepted July 31, 1998