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

Alan R. Templeton

Department of Biology, Washington University, St. Louis, Missouri 63130-4899

Manuscript received February 4, 1996
Accepted for publication July 24, 1996

ABSTRACT

Contingency tests of neutrality are performed using mitochondrial *cytochrome oxidase II* (*COII*) DNA sequences from hominoid primates, including humans. An intra-/interspecific haplotype tree is estimated, including a statistical assessment of ambiguities in tree topology and branch lengths. Four functional mutational categories are considered: silent and replacement substitutions in the transmembrane portion of the COII molecule, and silent and replacement substitutions in the cytosolic portion. Three tree topological mutational categories are used: intraspecific tips, intraspecific interiors, and interspecific fixed mutations. A full contingency analysis is performed, followed by nested contingency analyses. The analyses indicate that replacement mutations in the cytosolic portion are deleterious, and replacement mutations in the transmembrane portion and silent mutations throughout tend to be neutral. These conclusions are robust to ambiguities in tree topology and branch lengths. These inferences would have been impossible with an analysis that only contrasts silent and replacement *vs.* polymorphic and fixed. Also, intraspecific interior mutations have similar evolutionary dynamics to fixed mutations, so pooling tip and interior mutations into a single "polymorphic" class reduces power. Finally, the detected deleterious selection causes lowered inbreeding effective sizes, so arguments for small effective sizes in recent human evolutionary history based upon mitochondrial DNA may be invalid.

A fundamental prediction of the neutral theory is that the neutral mutation rate determines both the rate of interspecific divergence and influences the amount of intraspecific polymorphism (KIMURA 1968). This double impact of the neutral mutation rate arises because intraspecific polymorphism is simply a transient phase of the stochastic process that ultimately leads to fixation of alleles under the neutral theory. MAYNARD SMITH (1970) was the first to propose using this predicted relationship between interspecific divergence and intraspecific polymorphism to test the neutral hypothesis. Unfortunately, the level of polymorphism also depends upon the long-term inbreeding effective sizes of species and upon how close species are to mutation/drift equilibria. For example, polymorphism in the mitochondrial DNA (mtDNA) of the *Drosophila melanogaster* species group has undoubtedly been influenced by the effect of hitch-hiking of host mtDNA haplotypes on the rapid spread of the cytoplasmic microbe Wolbachia through host populations (KILPATRICK and RAND 1995). As a consequence, although rates of fixation and levels of polymorphism should be positively correlated under the neutral theory, there is no simple or universal relationship between these variables. This makes it difficult to test the neutral theory using this predicted relationship between divergence and polymorphism. One way of controlling for variation among species in long-term effective sizes or in departures from equilibria is to contrast the rates of fixation and polymorphism for different classes of mutation that occur in a single evolutionary tree of genetic variation that incorporates both within and between species differences. Although different classes of mutational change may have different neutral mutation rates, the different classes of mutations found in a common evolutionary tree (and thereby sharing common inbreeding effective sizes and degrees of departure from mutation/drift equilibrium) should have the same relative proportion of fixed *vs.* polymorphic mutational events if all mutational categories are neutral (and if all mutational categories are observed with equal resolution in both the intra- and interspecific portions of the data set, an important condition that will be addressed). Consequently, testing neutrality becomes equivalent to a simple contingency test of homogeneity in which one dimension consists of the mutational categories and the other dimension consists of the tree position categories of "fixed" (interspecific differences) *vs.* "polymorphic" (intraspecific mutations).

This straightforward contingency test of neutrality using a combined inter-/intraspecific evolutionary tree of genetic variation was first proposed and executed by TEMPLETON (1987). However, few data sets with both inter- and intraspecific mutational information on a

Address for correspondence: Alan R. Templeton, Department of Biology, Campus Box 1137, Washington University, St. Louis, MO 63130-4899. E-mail: temple_a@biodec.wustl.edu

common evolutionary tree existed at that time, so this contingency approach to testing neutrality did not gain much popularity until its rediscovery by McDONALD and KREITMAN (1991). By 1991, many DNA sequence data sets existed that allowed the estimation of haplotype trees having both an intra- and interspecific component. As sequencing has become more and more common, the number of data sets amenable to this contingency test of neutrality has grown accordingly. However, the full potential of this contingency approach has not yet been fully explored. For example, McDONALD and KREITMAN (1991) used only two mutational categories: amino acid replacement substitutions and silent substitutions. However, there is no inherent limit in the contingency approach to just two types of mutations, and indeed TEMPLETON (1987) used up to five mutational categories in the original use of the contingency test of neutrality. In this paper it will be shown how using more than two mutational classes can yield greater insight into the nature of selection on the *cytochrome oxidase II* (*COII*) gene in hominoid primate mtDNA when departures from neutrality are revealed.

Both TEMPLETON (1987) and McDONALD and KREITMAN (1991) limited the tree category dimension of their contingency tests to "fixed" *vs.* "polymorphic." HUDSON (1993) pointed out that finer distinctions are possible in principle but did not suggest any specific ones. However, recent work in coalescent theory by CASTELLOE and TEMPLETON (1994) suggests a meaningful finer categorization: splitting the intraspecific "polymorphic" class into those mutations falling on "tip" branches *vs.* "interior" branches. A tip haplotype is connected to only one other haplotype in the tree. An interior haplotype is connected to two or more other haplotypes in the tree, and hence represents an interior node in a topological sense. CASTELLOE and TEMPLETON (1994) showed that for an intraspecific haplotype tree, interior haplotypes strongly tend to be older than tip haplotypes and that interior haplotypes also tend to be more frequent in the gene pool than tip haplotypes. Hence, the topological contrast of tip *vs.* interior for the most part corresponds to a contrast of young *vs.* old and (to a lesser extent) rare *vs.* common. Given that the mutations that are ultimately fixed come preferentially from the old, common class and not the young, rare class under neutrality, then polymorphic mutations on interior branches may well have evolutionary properties more akin to fixed interspecific mutations than to polymorphic tip branches. This possibility is strengthened by the work of FU and LI (1993) who show that deviations from neutrality can be detected with intraspecific data by discriminating between "external" (*i.e.*, tip) *vs.* "internal" (*i.e.*, interior) mutations. This paper will provide an empirical investigation of the utility of discriminating among these two topological classes of mutations with respect to the evolution of the *COII* gene in the hominoid primates by creating a new mutational contrast: "young" mutations (those found on the tip branches, all of which are intraspecific) *vs.* "old" mutations (those found on the interior branches, both intra- and interspecifically).

MATERIALS AND METHODS

***COII* sequence data:** Sequence data were obtained from RUVOLO *et al.* (1993) and RUVOLO *et al.* (1994) for six humans (*Homo sapiens*) that defined four distinct haplotypes, five common chimpanzees (*Pan troglodytes*) that defined four distinct haplotypes, four pygmy chimpanzees (*P. paniscus*) that defined four distinct haplotypes, six gorillas (*Gorilla gorilla*) that defined six distinct haplotypes, and one orangutan (*Pongo pgymaeus*). The haplotype abbreviations used in this paper are the same as those used in the articles cited above.

Tree estimation and evaluation: The limits of parsimony and the probable extent of deviations from parsimony were estimated with the algorithm given in TEMPLETON *et al.* (1992) using the Mathematica (WOLFRAM 1991) package ParsimonyAnalysis (written by the author and available upon request). Because the rate dynamics of replacement and silent substitutions are so different, separate analyses were performed for these two categories of mutations. Moreover, because mtDNA has a strong transition bias, all calculations using ParsimonyAnalysis were performed assuming a strong transition bias (b = 1 in the parameterization given in TEMPLETON *et al.* 1992). To estimate these limits and probable deviations, it is also necessary to know the total number of sites at risk for a particular mutational type. Given that most silent mutations are third-position mutations, the number of sites at risk for silent mutation was set at one-third the total sequence length of 684 to yield 228. The number of sites at risk for replacement mutations was set to 456. These numbers are not precise because some third site substitutions are not silent, and some second site substitutions are not replacements. However, the expected deviations are minor, and given that these numbers are large relative to the number of differences to be analyzed, the resulting parsimony probabilities are not sensitive to small deviations from these numbers.

The haplotype sequence data were used to estimate a maximum parsimony tree using the estimation algorithm of TEMPLETON *et al.* (1992), which not only estimates a tree, but creates a 95% plausible set of trees that includes all connections that have a probability ≥ 0.95 under neutral coalescence. The program PAUP 3.1.1 (SWOFFORD 1993) with the branch and bound option was used as an aid to implement the algorithm of TEMPLETON *et al.* (1992). The organutan sequence is used as the outgroup.

Mutational categories: In addition to the standard categories of silent *vs.* replacement substitutions, an additional pair of categories was defined *a priori* on the basis of the known biochemical functioning of the COII polypeptide. On the N-terminal side of the central aromatic domain, the COII polypeptide is hydrophobic and is found in association with the transmembrane portion of the cytochrome oxidase complex. The C-terminal side of the molecule is hydrophilic and protrudes into the cytosol. It contains the Cu_A site, crucial for the transfer of electrons to O_2 and for the cytochrome c binding site (LARSSON *et al.* 1995; TSUKIHARA *et al.* 1995). Because of the extreme difference in the biochemical role played by these two regions, mutations are categorized as N-terminal mutations (the hydrophobic region plus aromatic divider, corresponding to nucleotide sites 1–330) or C-terminal mutations (nucleotide sites 331–684). Hence, there are a total of four mutational categories: N-terminal silent, N-terminal replacement, C-terminal silent, and C-terminal replacement.

Contingency analyses: Two-by-two contingency tables were analyzed using Fisher's exact test (FET). Larger tables were analyzed with an exact permutational test using the algorithm of ROFF and BENTZEN (1989) and using 1000 random permutations of the data to simulate the null hypothesis of homogeneity.

RESULTS

Tree estimation and evaluation: Resolving or iterating over ambiguities in the haplotype tree is a necessary first step in the contingency analysis because it determines the categories into which sequence differences are sorted. If the categories are not properly erected, then subsequent analyses are suspect. Figure 1 shows the 95% plausible set of estimated trees. The orangutan sequence was used to root the tree using the outgroup method, but that portion of the tree is not illustrated as it will not be used in any subsequent analyses. There is a loop of ambiguity in the portion linking the common chimpanzee haplotypes to the remainder of the tree. This loop can be broken in three equally parsimonious ways, but only the two branches of the loop shown with dashed lines are allowed to be broken by the algorithm of TEMPLETON *et al.* (1992). Phylogenetic resolution I will refer to the tree with the left-handed dashed line broken; phylogenetic resolution II will refer to the tree with the right-handed dashed line broken. Although only two alternative tree topologies are likely with these data, this ambiguity is important because the solid branch in the loop can either be an interior branch or part of a tip branch, depending upon which of the two resolutions is true. In all subsequent analyses, whenever this ambiguity is relevant, tests will be performed under both resolutions to ensure robustness of the test results to this phylogenetic uncertainty.

As can also be seen from Figure 1, the longest branch with respect to replacement substitutions has a length of 4. The probability of this branch being completely parsimonious (that is, no unobserved replacement substitutions due to multiple mutational hits) is 0.989, and all other branches have even higher probabilities for replacement substitutions. For all branches with six or fewer silent substitutions, the confidence of parsimony exceeds 95%. All but two of the intraspecific branches (the exceptions are the branches of parsimonious length 7 and 9 for silent substitutions found within the *G. gorilla* portion of the tree) but none of the interspecific branches are within this 95% limit. Table 1 shows the probable deviations from parsimony for all branches in the tree with seven or more silent substitutions. These ambiguities in branch lengths could affect the exact connections to the interspecific nodes shown in Figure 1. However, with respect to the categories in the contingency analyses to follow, these ambiguities do not affect the tree topology or the tree categories of the mutations on the branches. These ambiguities do affect the mutational counts in some of the contingency categories, so the results of the contingency analysis will be checked for robustness to ambiguities in branch lengths.

Contingency analyses: Contingency tables were constructed by counting the number of mutational events in the various categories and on the various types of branches for the maximum parsimony tree shown in Figure 1. The contingency tables and test results are shown in Table 2 for the full contrast of all four mutational categories (N-terminal silent, C-terminal silent, N-terminal replacement, C-terminal replacement) *vs.* all three haplotype tree topological categories (tip, interior, fixed). The more standard MCDONALD-KREITMAN (1991) contrast collapses the N- and C-terminal categories into the silent/replacement mutational categories and collapses the tip and interior categories into a single polymorphic category, yielding the 2 by 2 table given in Table 3.

The contingency test for the full model can be subdivided to investigate the evolutionary dynamics of the different mutational categories. First, the impact of the biochemical region of the molecule upon the evolutionary dynamics of silent and replacement mutations can be examined by a contingency test of the first and second columns of Table 2 (which contrasts the evolutionary dynamics of silent substitutions in the C- *vs.* the N-terminal regions) and a separate contingency analysis of the third and fourth columns (replacement substitutions across the two biochemical regions). The permutational probability values for the contingency analysis of silent mutations in the N *vs.* C regions are 0.590 and 0.670 for phylogenetic resolutions I and II, respectively. The permutational probability for the replacement mutations across biochemical regions is 0.119. None of these results are significant at the 5% level, and this may be due in part to the small numbers of observations in some of the categories in the contingency table. To enhance power, further pooling was done on the evolutionary tree axis: polymorphic (tip + interior) *vs.* fixed, yielding an FET probability of 0.170 for silent substitutions and 0.091 for replacement substitutions; and young (tip) *vs.* old (interior + fixed), yielding FET probabilities of 0.227 and 0.308 for trees I and II, respectively, for silent substitutions and 0.040 for replacement substitutions.

A second nested series of additional contingency tests examines the evolutionary dynamics of silent *vs.* replacement mutations within the N- and C-regions separately. Table 4 presents the relevant contingency tables extracted from Table 2. The C-terminal results were highly significant, but neither phylogenetic resolution yields significant results for the N-terminal region (Table 4). To enhance power, the pooling categories of polymorphic/fixed and young/old were used for the N-terminal region. The resulting FET probability for the polymorphic/fixed categories is 0.114, and the FET probabilities for the young/old categories are 0.358 and 0.408 for phylogenetic resolution I and II, respectively.

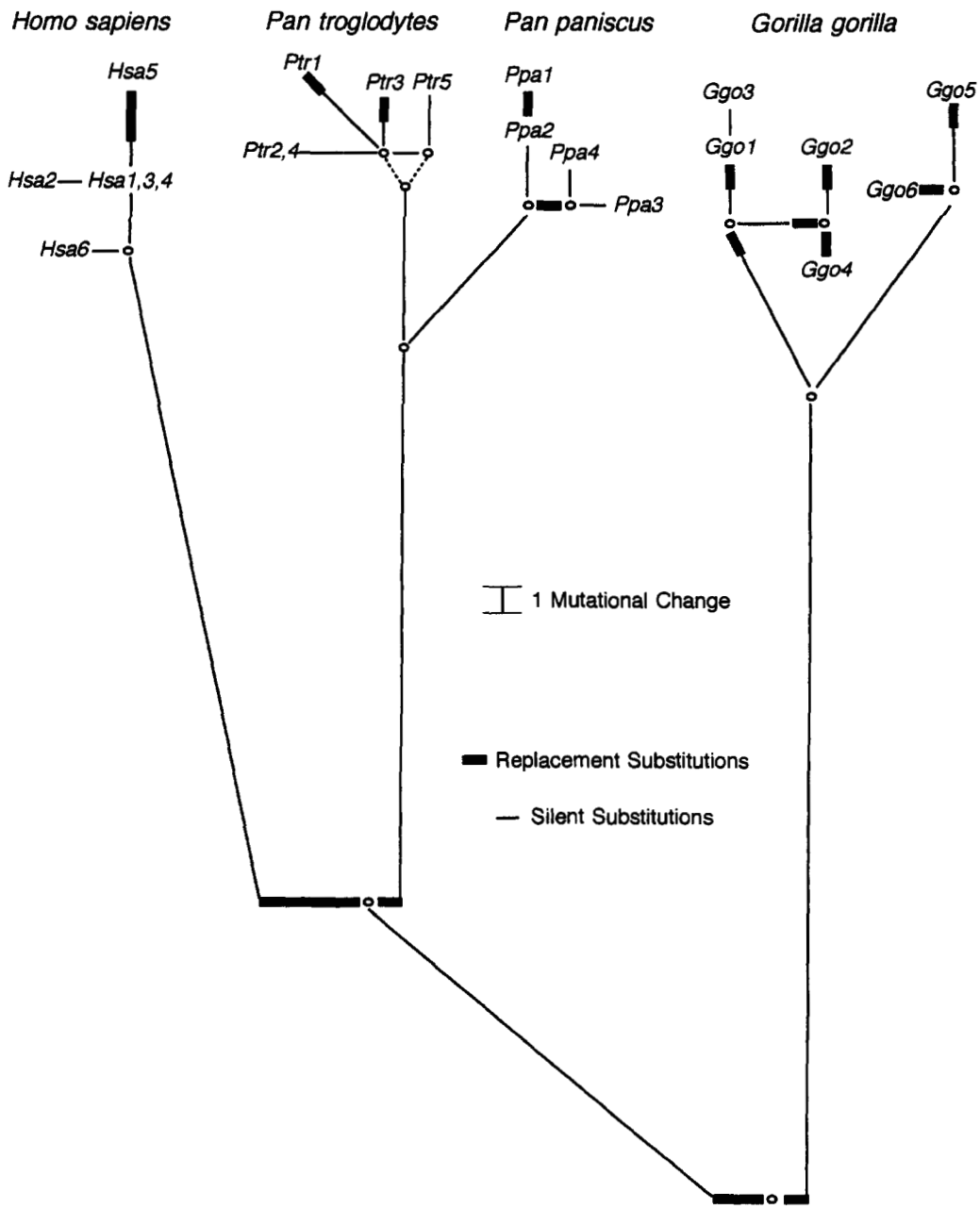

FIGURE 1.—Evolutionary tree of the *COII* haplotypes from hominoid primates as estimated by the algorithm of TEMPLETON *et al.* (1992). The designations of the haplotype sequences are those given in RUVOLO *et al.* (1994). Two alternative paths connecting the *Pan troglodytes* haplotypes to the remainder of the tree are indicated by dashed lines, which are mutually exclusive alternatives. Branch lengths are drawn proportional to their maximum parsimony branch length, with those portions of the branch due to replacement substitutions shown by thick lines and those portions due to silent substitutions shown by narrow lines. The orangutan sequence found in RUVOLO *et al.* (1994) is not shown but was used as an outgroup. The location of the root determined by this outgroup procedure under maximum parsimony is indicated by the internal node at the bottom of the figure.

Table 5 shows the comparable analyses when the C-terminal mutations are pooled into the polymorphic/fixed and young/old sets of categories.

DISCUSSION

The contingency approach to testing neutrality is simple, straightforward, and does not require statistical equilibrium under the neutral model (TEMPLETON 1987; HUDSON 1993). However, the statistical contrasts depend on accurate and unbiased counts of the numbers of mutations in various categories. This in turn requires that the topology and branch lengths of the tree be estimated in an accurate, unbiased fashion. Several recent tests of neutrality have used the contingency approach (*e.g.*, AKASHI 1995; KELLOGG and APPELS 1995; SHIBATA and YAMAZAKI 1995), but none of these addressed the effect of *estimation* accuracy of topologies

TABLE 1

Results of the analysis of the limits of parsimony for those branches that have seven or more silent substitutions under maximum parsimony

Parsimonious branch length	Probability of parsimony	Probability of one additional undetected mutation	Probability of two additional undetected mutations	Probability of three or more undetected mutations
7	0.940	0.058	0.001	0.000
9	0.905	0.091	0.004	0.000
20	0.621	0.300	0.068	0.011
22	0.561	0.330	0.091	0.018
26	0.446	0.367	0.144	0.043
32	0.291	0.369	0.224	0.116

and branch lengths upon the numbers in the contrasted categories.

The algorithm of TEMPLETON *et al.* (1992) estimates a 95% plausible set of trees, thereby explicitly documenting the extent of ambiguity that is likely in tree topology and branch lengths. For the *COII* data set, the 95% plausible set of tree topologies contained two alternatives that affected the numbers of mutations in some of the categories (Tables 2 and 4). By repeating the contingency analyses over all alternatives in the plausible set of tree topologies, a direct assessment can be made of robustness of conclusions to error in tree topology estimation. In this case, the impact of tree ambiguity was minor, and all subsequent conclusions about the evolution of the *COII* gene in the hominoid primates are robust to this topological ambiguity.

There is also phylogenetic ambiguity in the accuracy of the estimated branch lengths. This source of ambiguity could also have an impact on these contingency tests. For example, MAYNARD SMITH (1994) has shown how the observed number of mutational differences (assuming the tree topology is known) can lead to a rejection of neutrality even when the neutral model is true because some classes of mutations are more likely to experience unobserved multiple hits than other classes. MAYNARD SMITH (1994) pointed out that synonymous mutations tend to be transitions, and replacement mutations are more often transversions. When the genetic distance between species is large, there is a bias toward undercounting the number of transitions in a molecule such as mtDNA that has a strong transition bias. This in turn leads to an apparent excess of synonymous mutations, which can lead to as much as a twofold difference in the ratio of replacement to synonymous changes between species *vs.* within (MAYNARD SMITH 1994). This effect is not important if the between species differences are not close to saturation (MAYNARD SMITH 1994).

In light of MAYNARD SMITH's conclusions, an analysis of mutational saturation of silent and replacement substitutions should be done before interpreting a rejection of the null hypothesis as evidence for selection. Such an analysis is presented in Table 1. In the case of *COII,* although all of the replacement mutational branch lengths fall well within the 95% confidence bounds of parsimony, several branches involving silent mutations do not. In particular, all of the young tip branches fall within the bounds of parsimony for silent mutations, but two of the older interior branches do not. These two interior branches have a significant probability (>0.05) of exceeding parsimony by one additional mutation at most. All of the interspecific branches have a significant probability of deviating from parsimony for silent substitutions: two branches are likely to deviate by at most one additional mutation, three branches by two additional mutations, and the

TABLE 2

Contingency analysis of the full mutational *vs.* the full tree topological categories under both of the likely phylogenetic resolutions shown in Figure 1

	N-terminal silent	C-terminal silent	N-terminal replacement	C-terminal replacement
Tip	8 (9)	12	2	7
Interior	10 (9)	12	3	2
Fixed	60	53	6	2

Permutational probability under the null hypothesis of homogeneity: 0.000 (0.002)

The counts in only two categories were affected by phylogenetic ambiguity, as indicated by the categories with a number followed by a second number in parenthesis. The first number is the count under phylogenetic resolution I, and the number in parenthesis is the count under phylogenetic resolution II. The probability levels are presented in the same manner, with the first probability referring to phylogenetic resolution I, and the probability in parenthesis to II.

TABLE 3

Contingency analysis of the silent/replacement mutations *vs.* polymorphic/fixed tree topological categories

	Silent	Replacement
Polymorphic	42	14
Fixed	113	8

FET probability under the null hypothesis of homogeneity: 0.001

The probability under the null hypothesis of homogeneity is determined by Fisher's Exact Test (FET).

longest branch by up to three additional mutations. Hence, in the worst case scenario, there are two additional silent interior mutations, and 11 additional silent fixed mutations. Note that these likely deviations are nonrandomly distributed over the tree topology in the direction predicted by MAYNARD SMITH (1994) despite the fact that these hominoid species are not close to saturation.

To assess the maximum possible impact of silent substitution undercounting, 13 additional undercounted silent mutations (the worst case scenario outlined above) will be allocated to the N- and C-regions according to their relative lengths, rounded to the nearest integer. That is, the two possible undercounted interior silent mutations are allocated one each into the N- and C-terminal regions; and five of the 11 possible undercounted fixed silent mutations are allocated to the N-terminal region and six to the C-terminal region. This even allocation procedure is conservative because it favors the null hypothesis of homogeneity. A rejection of the null hypothesis under this worst case scenario should be of biological significance.

Table 6 presents a comparison of the original test results with those under this worst case scenario for all contingency tests affected by additional mutations.

TABLE 4

Contingency analyses of how mutations within the N-terminal region and within the C-terminal region are distributed across silent and replacement categories and across the full tree topological categories under both of the likely phylogenetic resolutions shown in Figure 1

Region	Tree position	Silent	Replacement
N-terminal	Tip	8 (9)	2
	Interior	10 (9)	3
	Fixed	60	6
	Permutational probability: 0.270 (0.230)		
C-Terminal	Tip	12	7
	Interior	12	2
	Fixed	53	2
	Permutational probability: 0.004		

TABLE 5

Contingency analysis of how mutations in the C-terminal region are distributed across silent and replacement categories and across the two alternative pooled tree topological categories of polymorphic *vs.* fixed and of young *vs.* old

Tree position	Silent	Replacement
Polymorphic	24	9
Fixed	53	2
	FET probability: 0.002	
Young	12	7
Old	65	4
	FET probability: 0.001	

Almost all of the test results are changed in only a minor fashion. Interestingly, the largest change in relative probabilities (with the P level going from 0.001 to 0.025) involves Table 3, the simple 2 by 2 test of polymorphic *vs.* fixed contrasted with silent *vs.* replacement. Overall, the test results are robust to undercounting of silent substitutions as well as to ambiguity in the haplotype tree topology. Given the robustness of the statistical conclusions, it is now time to explore their biological implications.

The full contingency analysis (Table 2) reveals a strong departure from neutrality. This departure is also seen in the MCDONALD and KREITMAN (1991) test (Table 3), although in this case the rejection is much weaker if undercounting of silent mutations had occurred (Table 6). One major advantage of doing a full contingency analysis as opposed to the simple 2 by 2 version, is that there is no possibility of refining the analysis once the null hypothesis has been rejected in the 2 by 2 version. However, in the full contingency analysis, further insights into the biological basis of the rejection of the null hypotheses are possible by performing additional contingency analyses nested within the original contingency table.

The first nested series explored the possibility that the null hypothesis was rejected because of nonrandomness in the distribution of mutations across the two regions of the *COII* gene. None of the tests involving silent substitutions resulted in a rejection of the null hypothesis, nor did the initial contingency test for replacement substitutions. However, there are very few replacement substitutions overall, so this lack of significance could be due to much lower statistical power in this case as compared to the silent substitution case. One way of regaining power in a contingency framework is to pool categories. However, power is only gained by pooling if the categories being pooled are truly homogeneous in their underlying properties. The standard pooling of "polymorphic *vs.* fixed" (TEMPLETON 1987) did not enhance statistical power, but pooling into "young *vs.* old" did indeed en-

TABLE 6

Test results (either exact permutational probabilities, or probabilities from Fisher's Exact Test) under the worst case scenario of undercounted silent substitutions compared to the original test results

Table or original contingency test	Original contingency test result	Worst case scenario contingency test result
Table 2	0.000 (0.002)	0.000 (0.000)
Table 3	0.001	0.025
Silent mutations across N and C	0.590 (0.670)	0.580 (0.610)
Silent mutations across N and C, polymorphic/fixed	0.170	0.190
Silent mutations across N and C, young/old	0.227 (0.308)	0.238 (0.321)
Table 4, N-region	0.270 (0.230)	0.250 (0.310)
Table 4, C-region	0.004	0.000
Silent/replacement in N, polymorphic/fixed	0.114	0.105
Silent/replacement in N, young/old	0.358 (0.408)	0.326 (0.373)
Table 5, polymorphic/fixed	0.0020	0.0013
Table 5, young/old	0.0015	0.0009

N refers to the N-terminal region of the *COII* gene, and C, the C-terminal region. Unless otherwise noted, the mutational tree positions include the following categories: tip, interior, and fixed. In cases where the two topological phylogenetic resolutions make a difference, the first number is the probability under resolution I, and the number in parentheses is the probability under II.

hance statistical power, leading to a significant rejection of the null hypothesis. Hence, from an evolutionary point of view, the classes young *vs.* old appear to be more relevant than the classes polymorphic *vs.* fixed. An examination of the deviation in this case indicates that there is an excess of young replacement substitutions in the C-terminal half of the *COII* gene relative to young replacement substitutions in the N-terminal half. This implies that either replacement mutations in the cytosolic portion of the molecule do not persist long in evolutionary time (that is, they tend to be deleterious) and/or replacement mutations in the transmembrane portion of the molecule tend to persist for long periods of time and become preferentially fixed (that is, they tend to be advantageous).

The second nested series examined the evolutionary dynamics of silent *vs.* replacement mutations within the N- and C-regions separately. No significant rejection of the null hypothesis occurs within the N-terminal region, but there is a strong rejection of neutrality in the C-terminal region, which is strongest for the contrast of young *vs.* old (Table 5). An examination of Table 5 reveals that there is an excess of replacement mutations in the young category, which implies either that replacement mutations in the cytosolic portion of the molecule are deleterious and/or silent mutations in the cytosolic portion are advantageous.

By combining the conclusions from these two nested series of analyses, only a single biological interpretation fits all of the test results: replacement mutations in the cytosolic portion of the COII molecule tend to be deleterious whereas replacement mutations in the transmembrane portion and silent mutations throughout the gene tend to be neutral. This precision of biological interpretation could not be achieved by performing a 2 by 2 test. In this paper, both dimensions of the contingency table were expanded to more than two categories, and these additional categories were critical in arriving at precise biological conclusions.

The conclusion that replacement mutations in the cytosolic portion of the COII molecule are deleterious also has implications for the out-of-Africa replacement hypothesis for modern human origins (popularly known as mitochondrial "Eve") (VIGILANT *et al.* 1991; TEMPLETON 1993). One frequent argument made in favor of the Eve hypothesis is that the low diversity levels in human mtDNA imply, under neutrality, a small inbreeding effective size during recent human evolutionary history (ROGERS and JORDE 1995). Because there is no recombination in mtDNA, fixation of a single selectively favored mutation in the recent evolutionary history of humans could also cause this depletion of variation and thus create the appearance of small effective size (TEMPLETON 1993, 1994). However, the type of deleterious mutations documented here for the *COII* gene would also reduce the apparent size of the inbreeding effective size. When there is no recombination, genomes carrying any deleterious mutations have a decreased chance of fixation (that is, they stay in the young category), and this "background selection" mimics the effects of a simple reduction in effective population size (CHARLESWORTH *et al.* 1995). Perhaps the method of FU (1994) for estimating effective size that takes into account the mutational distribu-

tion over the phylogenetic tree could be modified to correct for these effects, but at present even this method requires neutrality. Given that such deleterious mutations are strongly indicated by the present analysis in the *COII* gene, mtDNA will be biased to yield low effective size estimates, with the extent of the bias being accentuated if other loci in the mitochondrial genome also experience deleterious mutations. Hence, arguments for a low inbreeding effective size in recent human evolution based upon mtDNA should be regarded with skepticism because they are based upon a demonstrably false premise.

The author thanks two anonymous reviewers for their excellent suggestions on an earlier version of this paper. This work was supported by the National Institute of Heart, Lung, and Blood grant 1 R01 HL-39107.

LITERATURE CITED

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

CASTELLOE, J., and A. R. TEMPLETON, 1994 Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol. Phylogenet. Evol. **3:** 102–113.

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

FU, Y. X., 1994 Estimating effective population size or mutation rate using the frequencies of mutations of various classes in a sample of DNA sequences. Genetics **138:** 1375–1386.

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

HUDSON, R. R., 1993 Levels of DNA polymorphism and divergence yield important insights into evolutionary processes. Proc. Natl. Acad. Sci. USA **90:** 7425–7426.

KELLOGG, E. A., and R. APPELS, 1995 Intraspecific and interspecific variation in 5s RNA genes are decoupled in diploid wheat relatives. Genetics **140:** 325–343.

KILPATRICK, S. T., and D. M. RAND, 1995 Conditional hitchhiking of mitochondrial DNA: frequency shifts of *Drosophila melanogaster* mtDNA variants depend on nuclear genetic background. Genetics **141:** 1113–1124.

KIMURA, M., 1968 Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. **11:** 247–269.

LARSSON, S., B. KÄLLEBRING, P. WITTUNG and B. G. MALMSTRÖM, 1995 The Cu_A center of cytochrome-c oxidase: electronic structure and spectra of models compared to the properties of Cu_A domains. Proc. Natl. Acad. Sci. USA **92:** 7167–7171.

MAYNARD SMITH, J., 1970 Population size, polymorphism, and the rate of non-Darwinian evolution. Am. Nat. **104:** 231–236.

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

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

ROFF, D. A., and P. BENTZEN, 1989 The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. Mol. Biol. Evol. **6:** 539–545.

ROGERS, A. R., and L. B. JORDE, 1995 Genetic evidence on modern human origins. Human Biol. **67:** 1–36.

RUVOLO, M., S. ZEHR, M. N. DORNUM, D. PAN, B. CHANG *et al.*, 1993 Mitochondrial *COII* sequences and modern human origins. Mol. Biol. Evol. **10:** 1115–1135.

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.

SHIBATA, H., and T. YAMAZAKI, 1995 Molecular evolution of the duplicated *Amy* locus in the *Drosophila melanogaster* species subgroup—concerted evolution only in the coding region and an excess of nonsynonymous substitutions in speciation. Genetics **141:** 223–236.

SWOFFORD, D., 1993 PAUP 3.1 User's Manual. Illinois Natural History Survey.

TEMPLETON, A. R., 1987 Genetic systems and evolutionary rates, pp. 218–234 in *Rates of Evolution*, edited by K. S. W. CAMPBELL and M. F. DAY. Allen & Unwin, London.

TEMPLETON, A. R., 1993 The "Eve" hypothesis: a genetic critique and reanalysis. Am. Anthropol. **95:** 51–72.

TEMPLETON, A. R., 1994 "Eve": hypothesis compatibility *vs.* hypothesis testing. Am. Anthropol. **96:** 141–147.

TEMPLETON, A. R., K. A. CRANDALL and C. F. SING, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

TSUKIHARA, T., H. AOYAMA, E. YAMASHITA, T. TOMIZAKI, H. YAMAGUCHI *et al.*, 1995 Structures of metal sites of oxidized bovine heart cytochrome C oxidase at 2.8 Å. Science **269:** 1069–1074.

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

WOLFRAM, S., 1991 *Mathematica*. Addison-Wesley, Redwood City, CA.

Communicating editor: A. G. CLARK