Departure from Neutrality at the Mitochondrial NADH Dehydrogenase Subunit 2 Gene in Humans, but Not in Chimpanzees

Cheryl A. Wise, Michaela Sraml[1] and Simon Easteal

The John Curtin School of Medical Research, The Australian National University, Canberra, ACT 0200, Australia

Manuscript received April 8, 1997
Accepted for publication October 6, 1997

ABSTRACT

To test whether patterns of mitochondrial DNA (mtDNA) variation are consistent with a neutral model of molecular evolution, nucleotide sequences were determined for the 1041 bp of the NADH dehydrogenase subunit 2 (*ND2*) gene in 20 geographically diverse humans and 20 common chimpanzees. Contingency tests of neutrality were performed using four mutational categories for the ND2 molecule: synonymous and nonsynonymous mutations in the transmembrane regions, and synonymous and nonsynonymous mutations in the surface regions. The following three topological mutational categories were also used: intraspecific tips, intraspecific interiors, and interspecific fixed differences. The analyses reveal a significantly greater number of nonsynonymous polymorphisms within human transmembrane regions than expected based on interspecific comparisons, and they are inconsistent with a neutral equilibrium model. This pattern of excess nonsynonymous polymorphism is not seen within chimpanzees. Statistical tests of neutrality, such as Tajima's *D* test, and the *D* and *F* tests proposed by Fu and Li, indicate an excess of low frequency polymorphisms in the human data, but not in the chimpanzee data. This is consistent with recent directional selection, a population bottleneck or background selection of slightly deleterious mutations in human mtDNA samples. The analyses further support the idea that mitochondrial genome evolution is governed by selective forces that have the potential to affect its use as a "neutral" marker in evolutionary and population genetic studies.

THE neutral theory of molecular evolution asserts that most mutations are deleterious and are quickly removed from the population, thereby contributing little, if anything, to the levels of polymorphism detected within a species (Kimura 1983). Genetic variation within a species is largely caused by random genetic drift of mutations that are selectively equivalent (*i.e.*, neutral). One of the appealing features of the neutral theory is that it provides a number of straightforward predictions, and thus serves as a useful null hypothesis for studies of genetic variation within and between species. One such prediction is that the amount of nucleotide polymorphism within a species will be correlated with the amount of divergence between species (*e.g.*, Hudson *et al.* 1987). An additional prediction of the neutral theory, formulated into a test by McDonald and Kreitman (1991), is that the ratio of amino acid replacement (nonsynonymous) to silent (synonymous) nucleotide differences will be the same within and between species.

Mitochondrial DNA (mtDNA) is widely used as a marker in population and evolutionary studies, and it is generally assumed to evolve according to a neutral model of molecular evolution. This assumption may be important for such things as measuring gene flow (Slatkin 1985), estimating effective population size (Wilson *et al.* 1985), detecting population subdivision (Avise *et al.* 1987), and dating times of species divergence or historical events within a species using a molecular clock (Brown 1980; Cann *et al.* 1987). The apparent lack of genetic recombination in mammalian mtDNA (Thyagarajan *et al.* 1996) means that the whole genome is a single completely linked entity. Any selective force acting at one site will equally affect the history of the whole molecule. Thus, the selective fixation of an advantageous mutation, for example, will lead to the concomitant fixation of all other polymorphisms through the process of genetic hitchhiking. Analyses of the simple hitchhiking model predict a reduction in heterozygosity (Maynard-Smith and Haigh 1974; Stephan *et al.* 1992) and in the number of segregating (or polymorphic) sites (Kaplan *et al.* 1989) near the selected site.

A number of studies have compared patterns of RFLP variation in human mtDNA to neutral predictions (*e.g.*, Johnson *et al.* 1983; Whittam *et al.* 1986; Excoffier 1990; Merriwether *et al.* 1991) and found fewer intermediate frequency polymorphisms than expected using Watterson's (1978) test of homozygosity and/or Tajima's (1989) test. While these findings are

Corresponding author: Cheryl Wise, Human Genetics Group, John Curtin School of Medical Research, The Australian National University, Canberra, ACT 0200, Australia. Email: cheryl.wise@anu.edu.au
[1]*Present address:* Michaela Sraml, Woden Valley Hospital, Yamba Drive, Garran, ACT 2606, Australia.

inconsistent with neutral expectations, it is unclear whether the deviations arise from recent natural selection or changing population sizes. Rogers and Harpending (1992) studied the distribution of pairwise nucleotide differences for human mitochondrial data and found that the distribution does not conform to a neutral equilibrium model. They suggest that the results fit well with a rapid population expansion, but they do not rule out the possibility of a population bottleneck. A bottleneck could well have been the result of a selective sweep of a mtDNA type rather than an actual population size reduction. In these studies, departures from the neutral equilibrium model can be explained by a variety of processes, including selection.

Evidence for selection in mtDNA comes from more recent studies utilizing the McDonald and Kreitman (1991) approach (Ballard and Kreitman 1994; Nachman *et al.* 1994, 1996; Rand *et al.* 1994; Templeton 1996). Common to all of these studies was the finding of higher ratios of nonsynonymous to synonymous nucleotide differences within species than between species either for all or part of the genes in question.

In humans and chimpanzees, previous studies have involved DNA sequence data from the 345-bp NADH dehydrogenase subunit 3 (*ND3*) gene (Nachman *et al.* 1996) and the 783-bp cytochrome *c* oxidase subunit II (*COII*) gene (Templeton 1996) using limited sample sizes (particularly chimpanzees). To investigate neutral predictions further, we have sequenced the 1041-bp NADH dehydrogenase subunit 2 (*ND2*) gene in 20 geographically diverse humans and 20 common chimpanzees. The patterns of variation within these species were compared to the patterns between species by using a simple contingency test of neutrality (McDonald and Kreitman 1991; Templeton 1987, 1996). Departures from neutrality were also investigated using Tajima's (1989) *D* test (hereafter referred to as D_T), as well as the *D* and *F* tests proposed by Fu and Li (1993).

MATERIALS AND METHODS

Samples: Twenty human blood samples were obtained from indigenous populations of Africa (Bantu from Durban, South Africa), Asia (Cantonese from Hong Kong), Europe (Anglo-Celts from Canberra, Australia), and Australia (Aboriginal Australians from the Kimberley region of Western Australia).

Common chimpanzee (*Pan troglodytes*) blood samples were obtained from animals held under long-term observation in one of several primate colonies at the Laboratory of Slow, Latent and Temperate Virus Infection of the National Institutes of Health (NIH, Bethesda, MD). They were supplied by D. C. Gajdusek and C. J. Gibbs Jr. Twenty individuals were included in the study. These were drawn from a larger sample of 102 individuals, the majority of which were wild caught (Board *et al.* 1981). Three major subspecies of *Pan troglodytes* are currently recognized: *P. t. troglodytes*, *P. t. schweinfurthii*, and *P. t. verus*. In the wild, these subspecies are geographically isolated yet very similar morphologically. The geographic origin of most captive chimpanzees in the United States, including those in our sample, is unknown. Previous analysis of mitochondrial control region sequences (Wise *et al.* 1997) indicates, however, that all but two of the individuals (Pt175 and Pt176) included in the present study are from the west African subspecies *P. t. verus*. Furthermore, individual Pt281 may belong to the newly recognized subclade of western chimpanzees in Nigeria (Gonder *et al.* 1997).

DNA amplification and sequencing: Sequencing templates were prepared by polymerase chain reaction (PCR) amplification of three overlapping fragments (I–III) encompassing the 1041-bp *ND2* gene in all individuals. Amplification primers were designed using the sequences reported by Horai *et al.* (1992) and are given in Table 1. PCR cycling conditions for fragments I and III were 30 cycles of 96° for 1 min, 56° for 1 min, and 73° for 1 min. Fragment II required the lower annealing temperature of 52°. For each fragment, two separate PCR reactions were performed with the primers (L′ *biotinylated* and H′ *–21M13*) and (L′ *M13 reverse* and H′ *biotinylated*). For each individual, both H (heavy) and L (light) strands were sequenced on a DNA sequencer (model 373A; Applied Biosystems, Foster City, CA). Consensus sequences were obtained by aligning forward- and reverse-complement sequences from the same individual in the SeqEd 675 DNA Sequence Editor program (Applied Biosystems). Fragments I–III were concatenated for each individual, and sequence alignment was performed manually using Genetic Data Environment (GDE) 2.2 (Smith *et al.* 1994). Previously published human (Anderson *et al.* 1981) and chimpanzee (Horai *et al.* 1992) *ND2* sequences were used in the comparative analyses. The 20 human and 20 chimpanzee mitochondrial *ND2* sequences reported here have been deposited in the DDBJ/EMBL/GenBank International Nucleotide Sequence Database under accession numbers AF014882–AF014921.

Intraspecific variation: The amount of genetic variation within a species was estimated from the number of segregating (polymorphic) sites (S) using the following relation:

$$\theta = S / \sum_{i=1}^{n-1} 1/i \qquad (1)$$

(Watterson 1975; Equation 10.3 in Nei 1987). The variance of θ was calculated using Equation 4 in Tajima (1993). An alternative measure of genetic variation (π) was also estimated from the average number of pairwise nucleotide differences per site between sequences or nucleotide diversity (Equation 10.6 in Nei 1987). The total variance of π, which incorporates

TABLE 1

Primers for the amplification of the NADH dehydrogenase subunit 2 (*ND2*) gene of humans and chimpanzees

Primer	Sequence (5′ to 3′)
Fragment I	
L4448	CCCATACCCCGAAAATGTT
H4784	AAAGGGGGCTATTCCTAGT
Fragment II	
L4753	CTACCAATCAATACTCATC
H5166	ATGTTAGCTTGTTTCAGGT
Fragment III	
L5162	CACGACCCTACTACTATCT
H5529	TTGAAGGCTCTTGGTCTGT

The L or H in the primer name refer to the light or heavy strand, respectively, while the number identifies the base at the 3′ end according to the numbering system of Anderson *et al.* (1981).

the sampling variance as well as other stochastic factors, was calculated using Equation 10.9 in Nei (1987). The number of nonsynonymous and synonymous sites was estimated using the method of Li (1993).

Median network analysis: A median network approach (Bandelt *et al.* 1995) was used to portray the human and chimpanzee mitochondrial *ND2* sequence relationships. Median networks are generated by partitioning the groups of sequence types character by character. An unmodified network contains almost parsimonious solutions and displays graphically the full information content of the sequence data. This approach highlights any incompatibility between pairs of characters, which enables identification of homoplasy (parallel mutation events or reversals), and can assist in identifying sequencing errors.

Tests of neutrality: The data were tested for departures from the neutral expectation that the ratio of nonsynonymous to synonymous polymorphisms within species should equal the ratio of nonsynonymous to synonymous fixed differences between species (McDonald and Kreitman 1991). Within and between species, nucleotide differences were counted as the number of mutational events occuring along the various branches of a network connecting humans, chimpanzees, and gorilla. All of these nucleotide differences were classified as either nonsynonymous or synonymous.

In addition to the standard categories of nonsynonymous *vs.* synonymous mutations, an additional pair of categories was defined based on the predicted secondary structure of the human ND2 protein (Persson and Argos 1994). According to this model, there are 10 transmembrane domains of approximately equal length (21–29 codons) and 11 surface domains of various lengths (1–24 codons). Hence, there are a total of four mutational categories: transmembrane nonsynonymous, transmembrane synonymous, surface nonsynonymous, and surface synonymous.

Contingency tables were constructed in which one dimension consists of the structural mutational categories and the other dimension consists of the "fixed" *vs.* "polymorphic" categories. The intraspecific "polymorphic" class was further split into those mutations falling on "tip" branches *vs.* "interior" branches (Castelloe and Templeton 1994; Templeton 1996). 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 Zaykin and Pudovkin (1993) and using 1000 random permutations of the data to simulate the null hypothesis of homogeneity. Uncorrected values were used in the statistical tests. This ensures that all observations are independent and results in conservative tests when there is an excess of polymorphic nonsynonymous differences since the number of fixed synonymous differences may be underestimated (Maynard-Smith 1994).

For n nucleotide sequences, quantities such as π, the average number of pairwise nucleotide differences between sequences (Nei 1987), S, the number of segregating (or polymorphic) sites (Watterson 1975), and the total number (η) of mutations and the number (η_e) of mutations in the external branches (Fu and Li 1993) may be calculated. These quantities were used to investigate departures from neutrality using the D_T test [Equation 38 in Tajima (1989)], and the D and F tests proposed by Fu and Li (1993). This was done for all nucleotide sites, nonsynonymous sites, and synonymous sites in humans and chimpanzees.

RESULTS

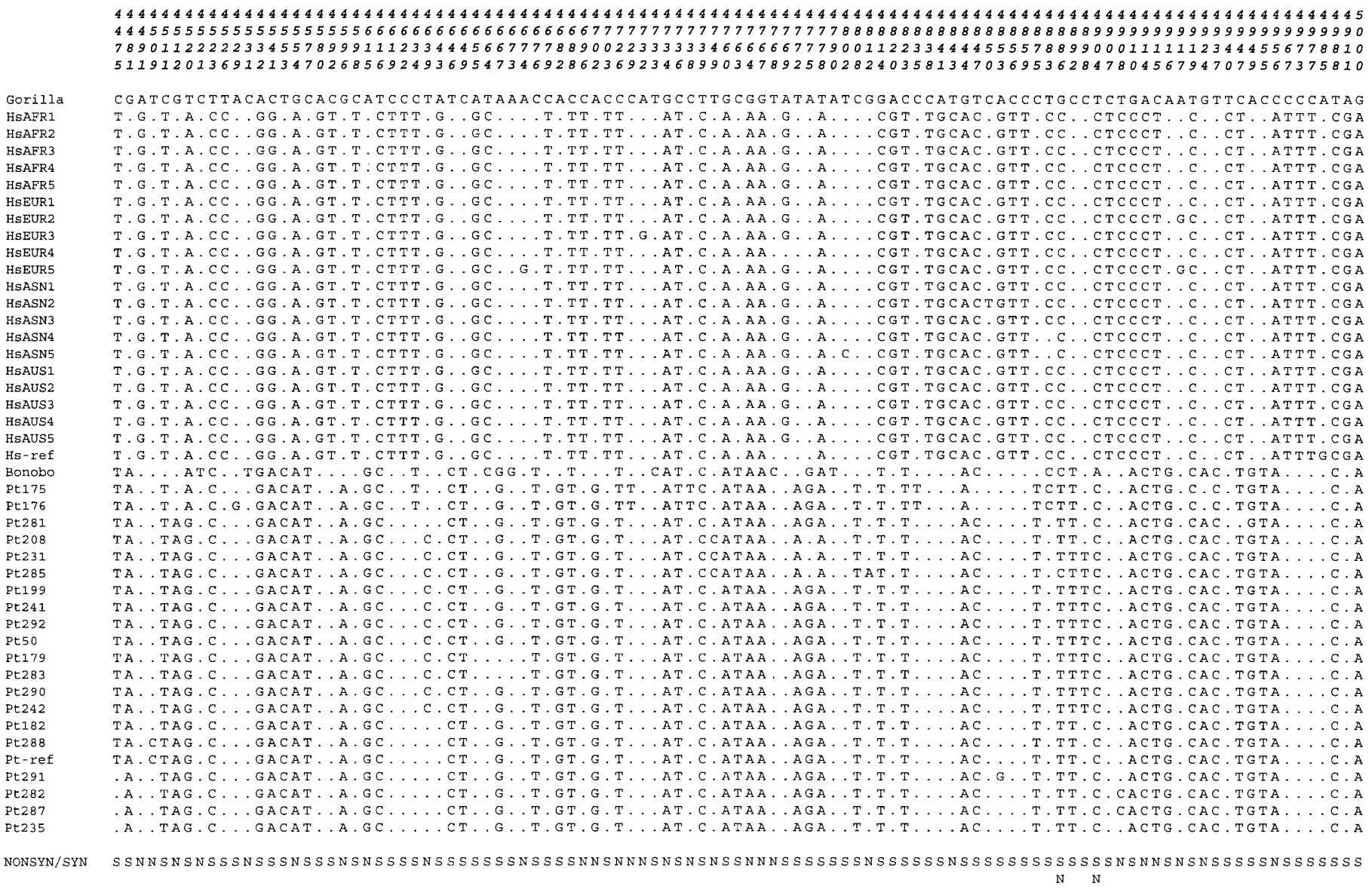
Intraspecific variation: We determined the nucleotide sequence of the 1041-bp coding region of the *ND2* gene (positions 4470–5510 according to the numbering system of Anderson *et al.* 1981) for 20 human and 20 common chimpanzee individuals. These sequences were compared with a previously published human (Anderson *et al.* 1981), chimpanzee, bonobo, and gorilla (Horai *et al.* 1992) sequence. All polymorphic and fixed nucleotide sites are shown in Figure 1. Polymorphism data within humans and chimpanzees are summarized in Table 2.

Within humans, there were 17 sequence types among 21 individuals, and two sequence types were shared among two or four individuals. Most differences between the sequences (92.7%) result from transition-type mutations, which is consistent with the general patterns of mtDNA sequence variation in humans (*e.g.*, Aquadro and Greenberg 1983; Greenberg *et al.* 1983; Vigilant *et al.* 1989; Horai and Hayasaka 1990; Horai *et al.* 1993; Watson *et al.* 1996). The uncorrected nucleotide diversity (per site) for the entire human sample was $\pi = 0.24\% \pm 0.15\%$, $\pi_N = 0.17\% \pm 0.12\%$ per nonsynonymous site, and $\pi_S = 0.46\% \pm 0.34\%$ per synonymous site (Table 2).

Within chimpanzees, there were 16 sequence types among 21 individuals, and four sequence types were shared among two or three individuals. Again, differences between the sequences (99.1%) result almost exclusively from transition-type mutations. The bias toward transitions has been noted in previous sequence comparisons of mtDNA in chimpanzees (Morin *et al.* 1994; Wise *et al.* 1997). The uncorrected nucleotide diversity (per site) for the entire chimpanzee sample was $\pi = 1.02\% \pm 0.54\%$, $\pi_N = 0.31\% \pm 0.20\%$ per nonsynonymous site, and $\pi_S = 3.12\% \pm 1.69\%$ per synonymous site (Table 2). This high level of diversity derives partly from the presence of a few very divergent sequence types (Pt175 and Pt176) that differ at 28 out of 1041 sites (2.69%). This is considerably greater than the most divergent human sequence types, which differ at six out of 1041 sites (0.58%). It is, however, similar to the level of divergence reported for a small section of the mitochondrial cytochrome *b* gene (2.8%; Morin *et al.* 1994) and the *ND3* gene (2.03%; Nachman *et al.* 1996) between *P. t. verus* and either *P. t. troglodytes* or *P. t. schweinfurthii*. Divergence between *P. t. troglodytes* and *P. t. schweinfurthii* at cytochrome *b* is $< 0.5\%$ (Morin *et al.* 1994). It is therefore likely that our sample includes *P. t. verus* and at least one of the two closely related subspecies, *P. t. troglodytes* and *P. t. schweinfurthii*, as noted previously (Wise *et al.* 1997).

To ensure that the chimpanzee sample represents a single interbreeding group, individuals Pt175 and Pt176 were excluded from all analyses. The uncorrected nucleotide diversity (per site) for *P. t. verus* is $\pi = 0.73\% \pm 0.40\%$, $\pi_N = 0.18\% \pm 0.13\%$ per nonsynonymous site, and $\pi_S = 2.35\% \pm 1.31\%$ per synonymous site (Table 2).

Network evaluation: Before the contingency analysis


4444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444444445
4445555555555555555555555566666666666666666666677777777777777777777778888888888888888888888999999999999999999999990
7890112222233455789991112334445667777889002233333334666666677790011223344445555578899000111111234445556778810
5119120136912134702685692493695473469286236923468990347892580282403581347036953628478045679470795673758105
Gorilla CGATCGTCTTACACTGCACGCATCCCTATCATAAACCACCACCCATGCCTTGCGGTATATATCGGACCCATGTCACCCTGCCTCTGACAATGTTCACCCCCATAG
HsAFR1 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsAFR2 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsAFR3 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsAFR4 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsAFR5 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsEUR1 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsEUR2 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT.GC..CT..ATTT.CGA
HsEUR3 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT.G.AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsEUR4 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA....A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsEUR5 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC..G.T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT.GC..CT..ATTT.CGA
HsASN1 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsASN2 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCACTGTT.CC..CTCCCT..C..CT..ATTT.CGA
HsASN3 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsASN4 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT..C..CTCCCT..C..CT..ATTT.CGA
HsASN5 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A.C..CGT.TGCAC.GTT..C..CTCCCT..C..CT..ATTT.CGA
HsAUS1 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsAUS2 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsAUS3 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsAUS4 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
HsAUS5 T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA.G..A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTT.CGA
Hs-ref T.G.T.A.CC..GG.A.GT.T.CTTT.G..GC....T.TT.TT...AT.C.A.AA....A....CGT.TGCAC.GTT.CC..CTCCCT..C..CT..ATTTGCGA
Bonobo TA....ATC..TGACAT....GC..T..CT.CGG.T..T...T..CAT.C.ATAAC..GAT...T.T....AC.....CCT.A..ACTG.CAC.TGTA....C.A
Pt175 TA..T.A.C...GACAT..A.GC..T..CT..G..T.GT.G.TT..ATTC.ATAA..AGA..T.T.TT...A.....TCTT.C..ACTG.C.C.TGTA....C.A
Pt176 TA..T.A.C.G.GACAT..A.GC..T..CT..G..T.GT.G.TT..ATTC.ATAA..AGA..T.T.TT...A.....TCTT.C..ACTG.C.C.TGTA....C.A
Pt281 TA..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC..GTA....C.A
Pt208 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.CCATAA..A.A..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
Pt231 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.CCATAA..A.A..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
Pt285 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.CCATAA..A.A..TAT.T....AC....T.CTTC..ACTG.CAC.TGTA....C.A
Pt199 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
Pt241 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
Pt292 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
Pt50 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
Pt179 TA..TAG.C...GACAT..A.GC...C.CT.....T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
Pt283 TA..TAG.C...GACAT..A.GC...C.CT.....T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
Pt290 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
Pt242 TA..TAG.C...GACAT..A.GC...C.CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TTTC..ACTG.CAC.TGTA....C.A
Pt182 TA..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
Pt288 TA.CTAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
Pt-ref TA.CTAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
Pt291 .A..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC.G..T.TT.C..ACTG.CAC.TGTA....C.A
Pt282 .A..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C.CACTG.CAC.TGTA....C.A
Pt287 .A..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C.CACTG.CAC.TGTA....C.A
Pt235 .A..TAG.C...GACAT..A.GC.....CT..G..T.GT.G.T...AT.C.ATAA..AGA..T.T.T....AC....T.TT.C..ACTG.CAC.TGTA....C.A
NONSYN/SYN SSNNSNSNSSSSNSSSNSSSNSNSSSSSNSSSSSSSSSNSSSSNNSNNNSNSNSNSSNNNSSSSSSSSNSSSSSSSNSSSSSSSSSSSSSSNSNNSNSNSSSSSSNSSSSSSS
N N

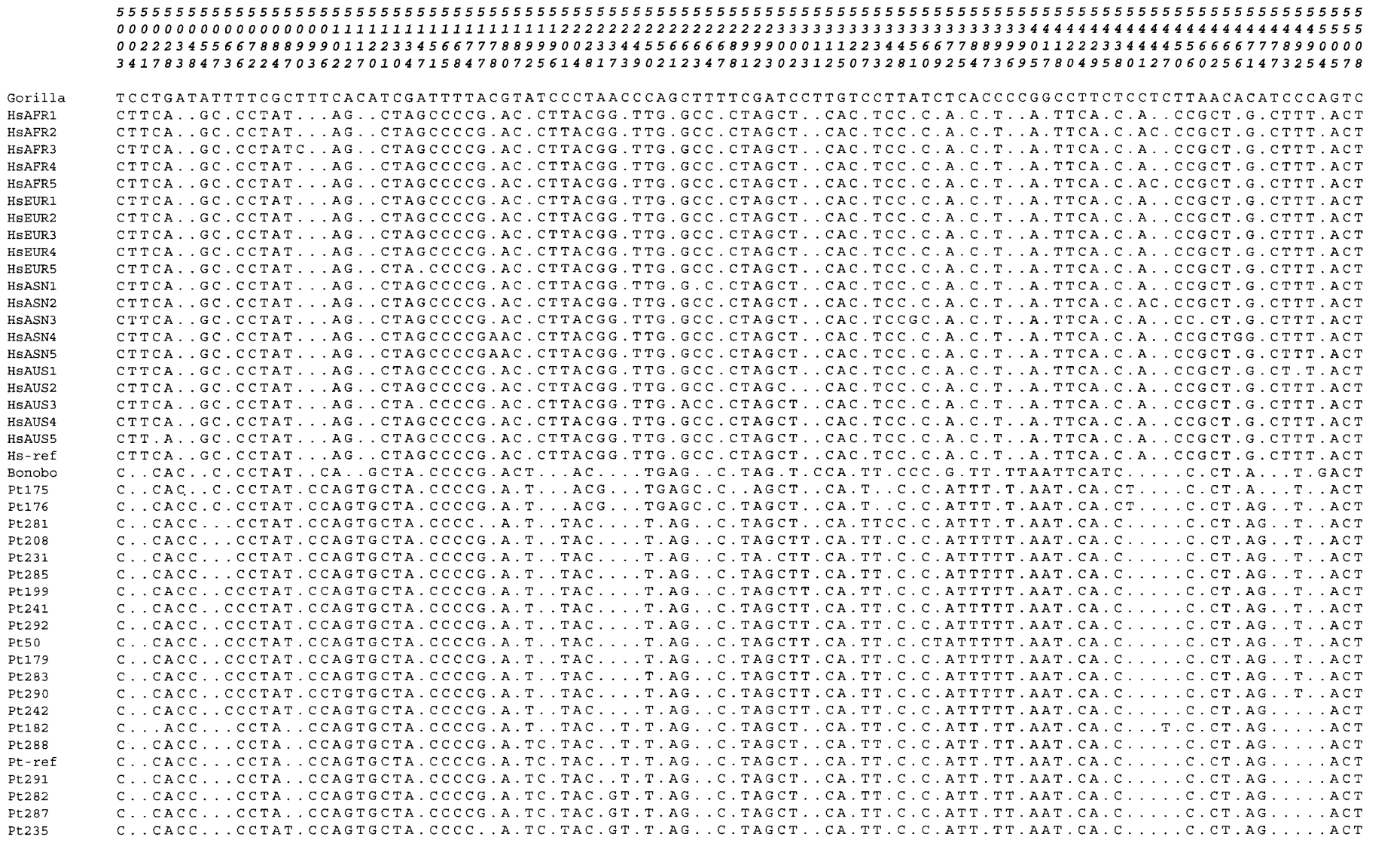


Figure 1.—Variable nucleotide sites in the coding region of the *ND2* gene from humans, chimpanzees, bonobo (Horai *et al.* 1992), and gorilla (Horai *et al.* 1992). Nucleotide positions are numbered according to Anderson *et al.* (1981). A dot indicates identity to the gorilla sequence, and differences are classified as either nonsynonymous (N) or synonymous (S). Human sequences are from Africa (AFR), Europe (EUR), Asia (ASN), and Australia (AUS). Hs-ref is a previously published human sequence (Anderson *et al.* 1981). The majority of the chimpanzee sequences are from the west African subspecies *P. t. verus*, with the exception of Pt175 and Pt176, which are from *P. t. troglodytes* and/or *P. t. schweinfurthii* (Wise *et al.* 1997). Pt-ref is a previously published chimpanzee sequence (Horai *et al.* 1992).

TABLE 2

Summary of *ND2* variation in humans and chimpanzees

	ND2				
	S	η	η_e	π ± SD (per site)	θ ± SD (per site)
Humans ($n = 21$)					
All sites	20	21	15	0.0024 ± 0.0015	0.0053 ± 0.0021
Nonsynonymous	10	10	7	0.0017 ± 0.0012	0.0035 ± 0.0016
Synonymous	10	11	8	0.0046 ± 0.0034	0.0109 ± 0.0049
All chimpanzees ($n = 21$)					
All sites	48	52	21	0.0102 ± 0.0054	0.0128 ± 0.0046
Nonsynonymous	12	12	4	0.0031 ± 0.0020	0.0042 ± 0.0018
Synonymous	36	40	17	0.0312 ± 0.0169	0.0394 ± 0.0145
P. t. verus ($n = 19$[a])					
All sites	28	39	17	0.0073 ± 0.0040	0.0077 ± 0.0030
Nonsynonymous	6	7	4	0.0018 ± 0.0013	0.0022 ± 0.0011
Synonymous	22	32	13	0.0235 ± 0.0131	0.0248 ± 0.0098

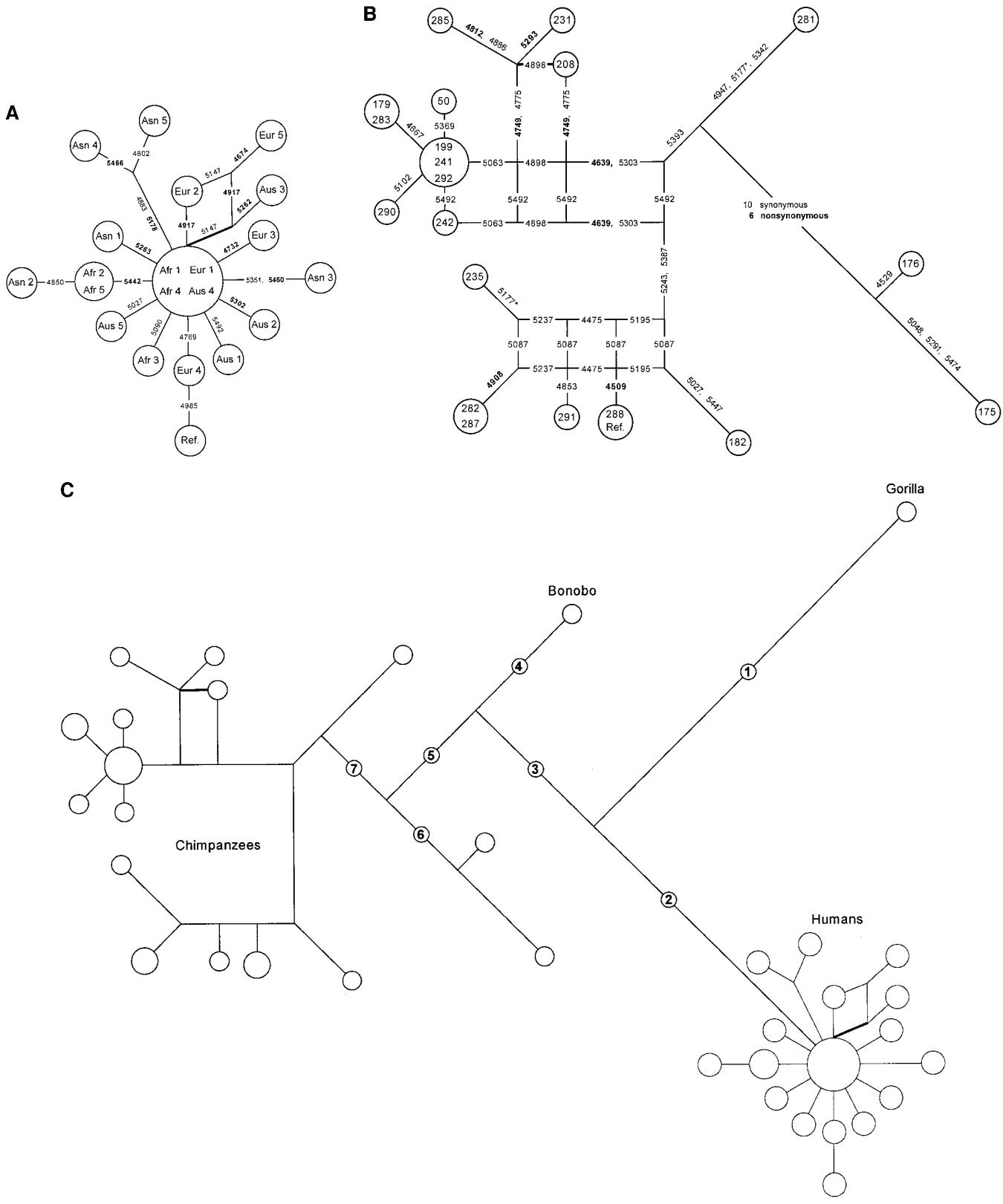
S, number of polymorphic sites; η, total number of mutations; η_e, number of unambiguous mutations in the external branches (inferred from Figure 2); π, average number of pairwise nucleotide differences per site between sequences; θ, average number of nucleotides segregating per site. SD, standard deviation. Values of π and θ and their standard deviations were calculated as outlined in materials and methods. No multiple-hit correction was made. The total number of sites compared for the *ND2* gene was 1041. The numbers of nonsynonymous (787) and synonymous (254) sites were estimated using the method of Li (1993).

[a] Excluding individuals Pt175 and Pt176 because they are not from the *P. t. verus* subspecies.

can be performed, it is necessary to resolve any ambiguities in the networks, since this will determine the topological categories into which sequence differences are sorted. Figure 2A shows the unrooted median network for 17 human *ND2* sequence types. There is a box of ambiguity involving nucleotide positions 4917 and 5147 (numbered according to Anderson *et al.* 1981). Since position 4917 is a nonsynonymous change, it is more parsimonious to assume that two mutational events have occurred at position 5147. This could involve either a reversal or two parallel mutations, which will be referred to as resolutions Ia and IIa, respectively. Figure 2B shows the unrooted median network for 16 chimpanzee *ND2* sequence types. Homoplasies occur at nucleotide position 5177 (asterisks) and probably positions 5087 and 5492. The box of ambiguity involving position 4898 can be broken in two equally parsimonious ways involving either two parallel mutations or a reversal, which will be referred to as resolutions Ib and IIb, respectively. These ambiguities are important because the thick lines in the boxes can either be an interior branch or part of a tip branch, depending on which resolution is chosen. In all analyses, whenever these ambiguities are relevant, tests are performed under all resolutions to ensure robustness of the test results to this network uncertainty. Figure 2C shows a reduced *ND2* network using the gorilla sequence found in Horai *et al.* (1992) as an outgroup. There are 11 fixed nonsynonymous nucleotide differences between humans and chimpanzees. The number of fixed synonymous differences varies between 84 and 91, depending on the branch placement of some mutations. In the contingency analyses to follow, the minimum number of fixed synonymous differences is used to ensure that the tests remain conservative.

Contingency tests of neutrality: Contingency tables were constructed by counting the number of mutational events in the various categories and on the various types of branches for the networks shown in Figure 2. The contingency tables and test results are shown in Table 3 for the full contrast of all four mutational categories (transmembrane nonsynonymous, surface nonsynonymous, transmembrane synonymous, and surface synonymous) *vs.* all three topological categories (tip, interior, and fixed). The null hypothesis of homogeneity is rejected when polymorphism data from humans are compared with fixed differences between species, but not when chimpanzee polymorphisms are compared with interspecific differences (Table 3). The more standard McDonald and Kreitman (1991) test collapses the transmembrane and surface categories into the nonsynonymous/synonymous mutational cate-

Figure 2.—(A) Unrooted median network of 17 human *ND2* sequence types. (B) Unrooted median network of 16 chimpanzee *ND2* sequence types. (C) Reduced human and chimpanzee *ND2* network (see results) using a gorilla sequence as the outgroup. Circles denote sequence types, and individuals are identified as described in Figure 1. Differences between sequence types are numbered according to Anderson *et al.* (1981). Amino acid replacement (nonsynonymous) mutations are shown in bold, and



likely parallel mutations (or reversals) are marked by asterisks. Unresolved homoplasies in the networks are indicated by thick lines. ①, 58 synonymous and 19 nonsynonymous; ②, 42 synonymous and 7 nonsynonymous; ③, 24 synonymous and 2 nonsynonymous; ④, 9 synonymous and 6 nonsynonsymous; ⑤, 10 synonymous and 2 nonsynonymous; ⑥, 4 synonymous and 5 nonsynonymous; ⑦, 6 synonymous and 1 nonsynonymous unambiguous nucleotide differences. The following sites are ambiguous with respect to their branch placement: 4511, branches 1 and 4, or 2 and 5; 4664, 5187, and 5420, branches 1 and 5, or 2 and 4; 4541, 4814, 4910, 5384, and 5471, branches 1 and 2, or 1 and 3, or 2 and 3.

TABLE 3

Contingency analysis of the full mutational categories *vs.* the full network topological categories within and between species at the *ND2* gene

			Humans		Chimps	
Mutation	Region	Fixed	Tip	Interior	Tip	Interior
Nonsynonymous	Transmembrane	9	7	3	4	2
	Surface	1	0	0	0	1
Synonymous	Transmembrane	54	4	1	9 (10)	15 (14)
	Surface	28	4 (5)	2 (1)	4	4
	Permutational probability:		0.049 (0.045)		0.476 (0.577)	

Exact permutational tests were used to test the null hypothesis of homogeneity. Categories that were affected by ambiguities in the networks are indicated by a number followed by a second number in parentheses. The first number is the count under resolutions Ia and Ib for humans and chimpanzees, respectively, and the number in parentheses is the count under resolutions IIa and IIb. The probability levels are presented in the same manner, with the first probability referring to resolution I and the probability in parentheses referring to resolution II.

gories and collapses the tip and interior categories into a single polymorphic category, yielding a 2×2 table. FETs were used to test the null hypothesis that the ratio of nonsynonymous to synonymous nucleotide differences is the same within and between species. The tests reveal a significantly higher nonsynonymous to synonymous ratio within humans than is seen between species (FET probability = 0.0005), but not within chimpanzees (FET probability = 0.4033).

The contingency test for the full model can be subdivided to investigate the evolutionary dynamics of the different mutational categories (Templeton 1987, 1996). First, the impact of the structural region of the molecule upon the evolutionary dynamics of nonsynonymous and synonymous mutations can be examined by a contingency test of the first and second rows of Table 3 (which contrasts the evolutionary dynamics of nonsynonymous mutations in the transmembrane *vs.* surface regions) and a separate contingency analysis of the third and fourth rows (synonymous mutations across the two structural regions). The permutational probability values for the contingency analysis of nonsynonymous mutations in the transmembrane *vs.* surface regions are 1.000 and 0.369 for the human and chimpanzee data, respectively. For synonymous mutations across structural regions, the permutational probabilities are 0.379 and 0.456 for human network resolutions Ia and IIa, respectively, and 0.580 and 0.599 for chimpanzee network resolutions Ib and IIb, respectively. None of these results is significant at the 5% level, and this may be caused in part by the small numbers of observations in some categories in the contingency table. To enhance statistical power, the data were further pooled into polymorphic (tip and interior) *vs.* fixed, and young (tip) *vs.* old (interior and fixed) categories. None of the tests was significant.

A second nested series of additional contingency tests examines the evolutionary dynamics of nonsynonymous *vs.* synonymous mutations within the transmembrane regions (rows one and three of Table 3) and within the surface regions (rows two and four of Table 3; Templeton 1987, 1996). The transmembrane results are highly significant for the human data (permutational probability = 0), but neither network resolution yields significant results for the chimpanzee data (permutational probabilities = 0.372 and 0.411 for resolutions Ib and IIb, respectively). None of the surface results is significant (permutational probability values are 1.000 for both human network resolutions and 0.429 for the chimpanzee data). The pooling categories of polymorphic/fixed and young/old were also only significant for the human transmembrane data (FET probabilities = 0.0001 and 0.0037, respectively).

The Tajima and the Fu and Li tests of neutrality: The Tajima (1989) test examines whether the average number of pairwise nucleotide differences between sequences (π) is larger or smaller than expected from the observed number of polymorphic sites (θ). Under the assumption of a random mating population at equilibrium, the difference between π and θ (D_T) is expected to be zero. A positive value of D_T indicates possible balancing selection or population subdivision. A negative value suggests recent directional selection, a population bottleneck, or background selection of slightly deleterious alleles (Tajima 1989). The Fu and Li (1993) test takes a genealogical approach and is based on the principle of comparing the number of mutations on internal branches with those on external branches. Compared with a neutral model of evolution, directional selection would result in an excess of external mutations, while balancing selection would result in an excess of internal mutations. Ideally, an outgroup is used so that the number of mutations in the external branches can be determined unambiguously. Since it is not clear which tests are most powerful, Tajima's (1989) D_T test, as well as the *D* and *F* tests proposed by

TABLE 4

D_T, *D*, and *F* calculated for the human and chimpanzee *ND2* gene

	D_T	*D*	*F*
Humans ($n = 21$)			
All sites	−2.071 (s)	−2.858 (s)	−3.185 (s)
Nonsynonymous	−1.855 (s)	−2.321 (s)	−2.605 (s)
Synonymous	−1.950 (s)	−2.537 (s)	−2.884 (s)
Chimpanzees ($n = 19$[a])			
All sites	−0.190	−1.043	−1.378
Nonsynonymous	−0.590	−1.380	−1.516
Synonymous	−0.045	−0.816	−1.187

D_T, Tajima's (1989) test statistic; *D* and *F*, Fu and Li's (1993) test statistics.

s, significant at the 5% level.

[a] Individuals Pt175 and Pt176 were excluded from the analyses because they are not from the *P. t. verus* subspecies.

Fu and Li (1993), were used to investigate departures from neutrality at the *ND2* gene in humans and chimpanzees (Table 4).

A significantly negative D_T, *D*, and *F* is observed for the human *ND2* data. This is consistent with a pattern of there being too many rare nucleotide polymorphisms with respect to predictions of the neutral theory (see *e.g.*, Braverman *et al.* 1995). In the case of chimpanzees, none of the tests is significant, and thus, by this criterion, the data are consistent with a neutral model of mtDNA evolution.

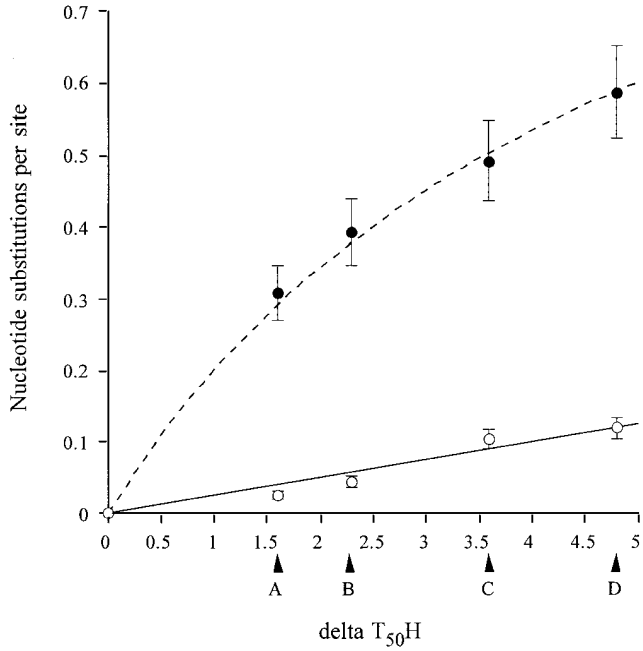


Figure 3.—Accumulation of synonymous (broken line) and nonsynonymous (solid line) nucleotide substitutions in the *ND2* gene with $\Delta T_{50}H$ values derived from DNA–DNA hybridization (Sibley and Ahlquist 1987). (A) Human × chimpanzee comparison. (B) Human + chimpanzee × gorilla comparisons. (C) Human + chimpanzee + gorilla × orangutan comparisons. (D) Human + chimpanzee + gorilla + orangutan × siamang comparisons.

DISCUSSION

The contingency approach to testing neutrality depends on accurate and unbiased counts of the numbers of mutations in various categories (Templeton 1996). The network approach of Bandelt *et al.* (1995) enables identification of any ambiguity in the tree topology that will affect this analysis. For the *ND2* data set, the human and chimpanzee networks each contained two alternatives that affected the numbers of mutations in some of the categories (Figure 2 and Table 3). The contingency analyses were therefore repeated over all network alternatives. The conclusions about the evolution of the *ND2* gene in humans and chimpanzees are not affected by this topological ambiguity.

The contingency tests do not apply for highly diverged sequences: when fully saturated, the ratio of nonsynonymous to synonymous nucleotide differences is expected to be as much as two times greater in between-species than in within-species comparisons. This effect is not important when the species to be compared are closely related and the sequences are not close to saturation (Maynard-Smith 1994). Thus, an analysis of mutational saturation of nonsynonymous and synonymous substitutions should be done before interpreting a rejection of the null hypothesis as evidence for selection. Such an analysis of substitutions in the *ND2* gene is presented in Figure 3 and Table 5. Nonsynonymous substitutions are far from saturated between divergent species. Although synonymous substitutions do not appear to be saturated in the human × chimpanzee comparison, they are likely to be undercounted, resulting in conservative tests.

The data presented here for the full contingency analysis (Table 3) provide a clear rejection of the null hypothesis that the human *ND2* gene is evolving, according to a strictly neutral model of molecular evolution. This strong departure from neutrality is also seen in the McDonald and Kreitman (1991) test, and it is consistent with an excess of nonsynonymous polymorphisms (or deficiency of synonymous polymorphisms) within humans compared with interspecific differences. This pattern has previously been documented for the human *ND2* gene (based on a limited sample size), and it appears to be widespread in the human mitochondrial genome (Nachman *et al.* 1996).

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 (Templeton 1987, 1996). The first nested series examined the distribution of nonsynonymous and synonymous mutations in the transmem-

TABLE 5

Estimates of the number of nonsynonymous and synonymous nucleotide substitutions per site between five species of primates using the method of Li (1993)

	Human	Chimpanzee	Gorilla	Orangutan	Siamang
Human	—	0.308 ± 0.039	0.423 ± 0.049	0.527 ± 0.061	0.617 ± 0.066
Chimpanzee	0.025 ± 0.006	—	0.361 ± 0.043	0.473 ± 0.054	0.606 ± 0.068
Gorilla	0.044 ± 0.008	0.044 ± 0.008	—	0.474 ± 0.053	0.640 ± 0.068
Orangutan	0.103 ± 0.013	0.100 ± 0.013	0.109 ± 0.014	—	0.486 ± 0.053
Siamang	0.108 ± 0.013	0.115 ± 0.014	0.113 ± 0.014	0.141 ± 0.016	—

Values below and above the diagonal represent nonsynonsymous and synonymous nucleotide substitutions, respectively.

brane *vs.* surface regions. None of the contingency tests involving synonymous mutations resulted in a rejection of the null hypothesis, nor did the tests for nonsynonymous mutations. There are very few nonsynonymous mutations overall (especially in the surface regions), however, so this lack of significance could be caused by the much lower statistical power in this case as compared to the synonymous mutation case (Templeton 1996). One way of regaining power in a contingency framework is to pool categories. The standard pooling of "polymorphic *vs.* fixed" (Templeton 1987) did not produce a significant result, nor did pooling into "young *vs.* old" (Templeton 1996).

The second nested series examined the distribution of nonsynonymous *vs.* synonymous mutations within the transmembrane and surface regions separately. The null hypothesis of neutrality is not rejected in the surface regions, but it is strongly rejected in the transmembrane regions of the human *ND2* gene. An examination of the data reveals a greater number of nonsynonymous polymorphisms within human transmembrane regions than expected based on interspecific comparisons. Within human transmembrane regions, 66.7% of the polymorphisms are nonsynonymous mutations. In contrast, between humans and chimpanzees, only 15.2% of the fixed differences are nonsynonymous mutations. Furthermore, there is an excess of nonsynonymous mutations in the young category (63.6%) compared with the old category (18.6%).

The results presented here are generally consistent with other studies that have used DNA sequence data to test the hypothesis that mtDNA variation is neutral. In Drosophila, non-neutral patterns have been documented for *ND5* (Rand *et al.* 1994) and *cyt b* (Ballard and Kreitman 1994). Similar patterns have also been observed for *ND3* in mice (Nachman *et al.* 1994), as well as *ND3* (Nachman *et al.* 1996) and *COII* (Templeton 1996) in humans and chimpanzees. In all of these studies, the ratio of nonsynonymous to synonymous nucleotide differences is greater within species than between species (Table 6). What might account for the observed pattern?

TABLE 6

Summary of variation at nonsynonymous (NS) and synonymous (S) sites within and between different species

		Fixed (between)			Polymorphic (within)		
Species	Locus	NS	S	$^{NS}/_{S}$	NS	S	$^{NS}/_{S}$
Drosophila	*ND5*[a]	15	54	0.28	7	11	0.64
	cyt b[b]	10	97	0.10	6	12	0.50
Mouse	*ND3*[c]	2	23	0.09	11	13	0.85
Chimpanzee	*ND2*[d]	11	84	0.13	7	32	0.22
	ND3[e]	4	31	0.13	4	3	1.33
Human	*ND2*[d]	11	84	0.13	10	11	0.91
	ND3[e]	4	31	0.13	4	7	0.57

In all cases, the ratio of nonsynonymous to synonymous polymophisms within species is greater than the ratio for fixed differences between species. The tests are not significant for *ND5* in Drosophila, *ND2* in chimpanzees, and *ND3* in humans.

[a] Rand *et al.* (1994).

[b] Ballard and Kreitman (1994).

[c] Nachman *et al.* (1994).

[d] This study. The chimpanzee data is from the west African subspecies *P. t. verus.*

[e] Nachman *et al.* (1996).

One possible explanation for these observations is that some form of balancing selection is maintaining amino acid variability. This hypothesis is considered unlikely for the human *ND2* data because Tajima's (1989) D_T test, as well as the *D* and *F* tests proposed by Fu and Li (1993), are significantly negative, indicating an excess of low frequency polymorphisms (Table 4). Under a model of balancing selection, some polymorphisms would be maintained in the population at intermediate frequencies, thus leading to positive test values.

A second possible explanation for the results is that there has been a recent relaxation of selective constraint in the human lineage. This would result in some previously deleterious mutations becoming neutral and being incorporated into the population as polymorphisms (Takahata 1993a). Takahata (1993a) has argued that deleterious mutations in the human population may have became harmless under the changed (improved) environment throughout the Pleistocene. This hypothesis, however, does not adequately explain the results of Tajima's and Fu and Li's tests (Table 4). A relaxation of selective constraint is expected to increase the rate at which mutations at nonsynonymous sites are introduced into the population, but it is expected to have very little such effect on mutations at synonymous sites. Under this scenario, we would expect to observe negative test values for nonsynonymous sites but not for synonymous sites. The results presented here, however, show that D_T, *D*, and *F* are significantly negative for both nonsynonymous and synonymous sites (Table 4). Negative test values have also been observed for the noncoding control region (Jorde *et al.* 1995; C. A. Wise, unpublished results).

The pattern of mtDNA variation is consistent with a model of a population bottleneck followed by an expansion in population size (*e.g.*, Di Rienzo and Wilson 1991; Rogers and Harpending 1992; Harpending *et al.* 1993; Sherry *et al.* 1994; Rogers and Jorde 1995). This model can be used to explain the negative values of Tajima's and Fu and Li's tests; however, it does not explain the significant contingency test results. Furthermore, if human mitochondrial genome diversity reflects historical patterns of population size change, then similar patterns are expected of nuclear genome diversity. This appears not to be the case, and differences in the patterns of mitochondrial and nuclear genome diversity have recently been interpreted as evidence against the population expansion scenario (Hey 1997). A population bottleneck in the human lineage also appears to be incompatible with the unusual polymorphism at the major histocompatibility complex (*Mhc*) loci (Takahata 1990, 1993b; Klein *et al.* 1993; Ayala *et al.* 1994; Ayala 1995; Ayala and Escalante 1996), despite some criticism of the details of some of these analyses (Erlich *et al.* 1996). It is also inconsistent with the pattern of nucleotide polymorphism at the β-globin locus (Harding *et al.* 1997), and of *Alu* repeat and microsatellite variation (H. Harpending, personal communication).

Another possibility is that amino acid mutations at *ND2* are slightly deleterious (*e.g.*, Ohta 1992). Slightly deleterious mutants may persist within populations for brief periods, but they are unlikely to rise in frequency or become fixed. Slightly deleterious models of molecular evolution have previously been invoked as potential explanations for patterns of mitochondrial genome evolution in Drosophila (DeSalle and Templeton 1988; Ballard and Kreitman 1994), mice (Nachman *et al.* 1994), and humans (Nachman *et al.* 1996; Templeton 1996). The test results presented here reveal a significant excess of young nonsynonymous polymorphisms within human *ND2* transmembrane regions, suggesting that they may be deleterious. This model is also consistent with the negative values of Tajima's and Fu and Li's tests.

The relative contribution of slightly deleterious mutations to heterozygosity increases as effective population size, N_e, decreases (Kimura 1983). Thus, if nonsynonymous mutations in the mitochondrial genome are slightly deleterious, we would expect a relative increase, compared to neutral synonymous mutations, as N_e decreases. The results presented here for the *ND2* gene indicate that the nonsynonymous to synonymous ratio is significantly greater within humans than within chimpanzees (0.91:0.22, Table 6; FET probability = 0.0196). The results for the much smaller *ND3* gene are not significant (0.57:1.33, Table 6; FET probability = 0.6305) based on a comparison of 61 humans and five chimpanzees (Nachman *et al.* 1996).

Under neutrality, diversity should be low in small populations and high in large ones. Thus, the lower level of mitochondrial diversity in humans compared with chimpanzees (Table 2; Ferris *et al.* 1981; Morin *et al.* 1994; Ruvolo *et al.* 1994; Nachman *et al.* 1996; Wise *et al.* 1997) may reflect a smaller N_e in humans. This is consistent with the results described above. However, the lower level of nuclear genome diversity in chimpanzees (Wise *et al.* 1997, and references therein) implies that the N_e of chimpanzees is smaller than humans, thus we would expect to observe a higher nonsynonymous to synonymous ratio within chimpanzees. This is inconsistent with the slightly deleterious model presented above. The contingency test results could, however, reflect the occurrence of slightly deleterious mutations if effective population size had been reduced for the mitochondrial genome by a selective sweep that did not affect most or all of the nuclear genome.

The low levels of human mtDNA diversity have been used as support for the out-of-Africa replacement hypothesis (Cann *et al.* 1987; Vigilant *et al.* 1991); however, directional selection could also explain the reduced mtDNA diversity in humans compared with chimpanzees (π = 0.24% for humans and 0.73% for the chimpanzee subspecies *P. t. verus*, Table 2). Because

there is no apparent genetic recombination in mtDNA, this depletion of variation could be the result of an advantageous mutation anywhere within the mitochondrial genome sweeping through the human population. The results of Tajima's and Fu and Li's tests are consistent with the occurrence of directional selection in the human mitochondrial genome. If an advantageous mutation has recently become fixed in the population, then the majority of mutations in the population are expected to be "young," thus leading to negative test values. The disparate pattern of variation in the nuclear genome (Harding *et al.* 1997; Hey 1997) is also consistent with this explanation since even though the entire mitochondrial genome would be affected the nuclear genome would be unaffected except possibly at specific genes interacting epistatically with mitochondrial genes.

We are grateful to D. C. Gajdusek and C. J. Gibbs, Jr., for supplying the chimpanzee blood samples, to L. Croft for technical assistance, and to G. Chelvanayagam for discussions and assistance with the secondary structure modeling. This work was funded by Australian Research Council grant A59332440 to S.E.

LITERATURE CITED

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.

Aquadro, C. F., and B. D. Greenberg, 1983 Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics **103:** 287–312.

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb *et al.*, 1987 Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Sys. **18:** 489–522.

Ayala, F. J., 1995 The myth of eve: molecular biology and human origins. Science **270:** 1930–1936.

Ayala, F. J., and A. A. Escalante, 1996 The evolution of human populations: a molecular perspective. Mol. Pylogenet. Evol. **5:** 188–201.

Ayala, F. J., A. Escalante, C. O'hUigin and J. Klein, 1994 Molecular genetics of speciation and human origins. Proc. Natl. Acad. Sci. USA **91:** 6787–6794.

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

Bandelt, H., P. Forster, B. C. Sykes and M. B. Richards, 1995 Mitochondrial portraits of human populations using median networks. Genetics **141:** 743–753.

Board, P. G., C. J. Gibbs and D. C. Gajdusek, 1981 Polymorphism of erythrocyte glyoxalase II in anthropoid primates. Folia Primatologica **36:** 138–143.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley and W. Stephan, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140:** 783–796.

Brown, W. M., 1980 Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc. Natl. Acad. Sci. USA **77:** 3605–3609.

Cann, R. L., M. Stoneking and A. C. Wilson, 1987 Mitochondrial DNA and human evolution. Nature **325:** 31–36.

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

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

Di Rienzo, A., and A. C. Wilson, 1991 Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc. Natl. Acad. Sci. USA **88:** 1597–1601.

Erlich, H. A., T. F. Bergstrom, M. Stoneking and U. Gyllensten, 1996 HLA sequence polymorphism and the origin of humans. Science **274:** 1552–1554.

Excoffier, L., 1990 Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. J. Mol. Evol. **30:** 125–139.

Ferris, S. D., W. M. Brown, W. S. Davidson and A. C. Wilson, 1981 Extensive polymorphism in the mitochondrial DNA of apes. Proc. Natl. Acad. Sci. USA **78:** 6319–6323.

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

Gonder, M. K., J. F. Oates, T. R. Disotell, M. R. J. Forstner, J. C. Morales *et al.*, 1997 A new west African chimpanzee subspecies? Nature **388:** 337.

Greenberg, B. D., J. E. Newbold and A. Sugino, 1983 Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene **21:** 33–49.

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.

Harpending, H., S. T. Sherry, A. R. Rogers and M. Stoneking, 1993 The genetic structure of ancient human populations. Curr. Anthropol. **34:** 483–496.

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

Horai, S., and K. Hayasaka, 1990 Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. Am. J. Hum. Genet. **46:** 828–842.

Horai, S., Y. Satta, K. Hayasaka, R. Kondo, T. Inoue *et al.*, 1992 Man's place in Hominoidea revealed by mitochondrial DNA genealogy. J. Mol. Evol. **35:** 32–43.

Horai, S., R. Kondo, Y. Nakagawa-Hattori, S. Hayashi, S. Sonoda *et al.*, 1993 Peopling of the Americas, founded by four major lineages of mitochondrial DNA. Mol. Biol. Evol. **10:** 23–47.

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

Johnson, M. J., D. C. Wallace, S. D. Ferris, M. C. Rattazzi and L. L. Cavalli-Sforza, 1983 Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. J. Mol. Evol. **19:** 255–271.

Jorde, L. B., M. J. Bamshad, W. S. Watkins, R. Zenger, A. E. Fraley *et al.*, 1995 Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am. J. Hum. Genet. **57:** 523–538.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking effect" revisited. Genetics **123:** 887–899.

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

Klein, J., N. Takahata and F. J. Ayala, 1993 MHC polymorphism and human origins. Sci. Am. **269:** 46–51.

Li, W.-H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:** 96–99.

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 hitch-hiking effect of a favourable gene. Genet. Res. **23:** 23–35.

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

Merriwether, D. A., A. G. Clark, S. W. Ballinger, T. G. Schurr, H. Soodyall *et al.*, 1991 The structure of human mitochondrial DNA variation. J. Mol. Evol. **33:** 543–555.

Morin, P. A., J. J. Moore, R. Chakraborty, L. Jin, J. Goodall *et al.*, 1994 Kin selection, social structure, gene flow, and the evolution of chimpanzees. Science **265:** 1193–1201.

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.

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

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

Persson, B., and P. Argos, 1994 Prediction of transmembrane segments in proteins utilising multiple sequence alignments. J. Mol. Biol. **237:** 182–192.

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

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

Rogers, A. R., and L. B. Jorde, 1995 Genetic evidence on modern human origins. Hum. Biol. **67:** 1–36.

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.

Sherry, S. T., A. R. Rogers, H. Harpending, H. Soodyall, T. Jenkins *et al.*, 1994 Mismatch distributions of mtDNA reveal recent human population expansions. Hum. Biol. **66:** 761–775.

Sibley, C. G., and J. E. Ahlquist, 1987 DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. J. Mol. Evol. **26:** 99–121.

Slatkin, M., 1985 Gene flow in natural populations. Annu. Rev. Ecol. Syst. **16:** 393–430.

Smith, S. W., R. Overbeek, C. R. Woese, W. Gilbert and P. M. Gillevet, 1994 The genetic data environment and expandable GUI for multiple sequence analysis. Comp. Appl. Biosci. **10:** 671–675.

Stephan, W., T. H. E. Wiehe and M. W. Lenz, 1992 The effect of strongly selected substitutions on neutral polymorphism-analytical results based on diffusion theory. Theor. Pop. Biol. **41:** 237–254.

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

Tajima, F., 1993 Measurement of DNA polymorphism, pp. 37–59 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Japan Scientific Societies Press, Tokyo.

Takahata, N., 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA **87:** 2419–2423.

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

Takahata, N., 1993b Allelic genealogy and human evolution. Mol. Biol. Evol. **10:** 2–22.

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., 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.

Thyagarajan, B., R. A. Padua and C. Campbell, 1996 Mammalian mitochondria possess homologous DNA recombination activity. J. Biol. Chem. **271:** 27536–27543.

Vigilant, L., R. Pennington, H. Harpending, T. D. Kocher and A. C. Wilson, 1989 Mitochondrial DNA sequences in single hairs from a southern African population. Proc. Natl. Acad. Sci. USA **86:** 9350–9354.

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.

Watson, E., K. Bauer, R. Aman, G. Weiss, A. von Haeseler *et al.*, 1996 mtDNA sequence diversity in Africa. Am. J. Hum. Genet. **59:** 437–444.

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

Watterson, G. A., 1978 The homozygosity test of neutrality. Genetics **88:** 405–417.

Whittam, T. S., A. G. Clark, M. Stoneking, R. L. Cann and A. C. Wilson, 1986 Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. Proc. Natl. Acad. Sci. USA **83:** 9611–9615.

Wilson, A. C., R. L. Cann, S. M. Carr, M. George, U. B. Gyllensten *et al.*, 1985 Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. **26:** 375–400.

Wise, C. A., M. Sraml, D. C. Rubinsztein and S. Easteal, 1997 Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. Mol. Biol. Evol. **14:** 707–716.

Zaykin, D. V., and A. I. Pudovkin, 1993 Two programs to estimate significance of χ^2 values using pseudo-probability tests. J. Hered. **84:** 152.

Communicating editor: R. R. Hudson