Mitochondrial and Nuclear Genes Present Conflicting Portraits of Human Origins

Jody Hey

Department of Ecology, Evolution, and Natural Resources, Rutgers University

Human mitochondrial DNA (mtDNA) sequences reveal an abundance of polymorphic sites in which the frequencies of the segregating bases are very different. A typical polymorphism involves one base at low frequency and the other base at high frequency. In contrast, nuclear gene data sets tend to show an excess of polymorphisms in which both segregating bases are at intermediate frequencies. A new statistical test of this difference finds significant differences between mtDNA and nuclear gene data sets reported in the literature. However, differences in the polymorphism patterns could be caused by different sample origins for the different data sets. To examine the mtDNA–nuclear difference more closely, DNA sequences were generated from a portion of the X-linked pyruvate dehydrogenase $E1\alpha$ subunit (PDHA1) locus and from a portion of mitochondrial control region I (CRI) from each of eight individuals, four from sub-Saharan Africa. The two genes revealed a significant difference in the site frequency distribution of polymorphic sites. PDHA1 revealed an excess of intermediate-frequency polymorphisms, while CRI showed an excess of sites with the low-high frequency pattern. The discrepancy suggests that mitochondrial variation has been shaped by natural selection, and may not be ideal for some questions on human origins.

Introduction

A distribution of DNA polymorphism frequencies can be used to ask questions about recent natural selection and changing population sizes (Tajima 1989a, 1989b; Rogers and Harpending 1992; Fu and Li 1993). When a site is polymorphic, usually just two bases are observed in a sample of sequences from a population, and it is possible that one base may have a low frequency (and thus the other base a high frequency) or both bases may have intermediate frequencies. Repeatedly, studies of human mitochondrial sequences have revealed a relative abundance of polymorphisms with the low-high pattern, and a paucity of intermediate-frequency polymorphisms (Excoffier 1990; Merriwether et al. 1991; Vigilant et al. 1991; Ruvolo, Zehr, and von Dornum 1993). This departure from neutral expectations is consistent with a model of a recently growing population (Tajima 1989a; Slatkin and Hudson 1991; Rogers and Harpending 1992) and has been regarded as evidence of a recent and rapid human population expansion, perhaps within the past 200,000 years (Stringer and Andrews 1988; Stoneking 1994). Representative large and small mitochondrial studies are shown in figure 1aand b.

If the mitochondrial polymorphism pattern is primarily due to historical patterns of population size change, then similar patterns are expected of nuclear polymorphisms. However, if natural selection on mitochondria is the primary cause of the mtDNA polymorphism pattern, then other genes may not reveal the same pattern. To compare mitochondrial and nuclear genes, it was necessary to avoid those nuclear gene studies that

Abbreviations: PDHA1, pyruvate dehydrogenase $E1\alpha$ subunit; CRI, mitochondrial control region I.

Key words: human origins, *Homo sapiens*, mitochondrial DNA, natural selection, statistical tests, PDHA1.

Address for correspondence and reprints: Jody Hey, Department of Ecology, Evolution, and Natural Resources, Rutgers University, Nelson Labs, P.O. Box 1059, Piscataway, New Jersey 08855-1059. E-mail: hey@mbcl.rutgers.edu.

Mol. Biol. Evol. 14(2):166-172. 1997

© 1997 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

focused on polymorphisms that were first discovered within just a small sample (Long et al. 1990; Bowcock et al. 1991). These polymorphisms necessarily tend to appear at intermediate frequencies and are biased for the purpose of site frequency analysis. A literature search returned two large nuclear gene studies in which variation was measured uniformly for all sampled individuals (fig. 1c and d). The nuclear gene studies have an excess of intermediate frequency polymorphisms while the mitochondrial studies have an excess of polymorphisms with the low-high pattern. These differences could be due to the different populations sampled for the different studies. There exists considerable population structure in global patterns of human genetic variation (Cavalli-Sforza, Menozzi, and Piazza 1993). In particular, populations from Africa are the most variable and seem to possess some of the oldest variation (Merriwether et al. 1991; Vigilant et al. 1991). About two thirds of the sampled sequences for figure 1a and b had African origins. In contrast, each of the nuclear gene studies in figure 1c and d included just one non-African population.

To address whether the differences between mitochondria and nuclear genes are due to different sample origins, a small sample of African and non-African individuals were studied at a portion of the nuclear genome and a portion of the mitochondrial genome.

Materials and Methods

DNA Sequencing

All DNA templates were from males to avoid sequencing diploid genomic templates. Sequence names and sources of genomic DNA: H-NA1, Coriell Institute for Medical Research, cell line repository # NA05963C, Amish Pedigree 884; H-NA2, Coriell Institute repository # NA07016, Utah pedigree K-1331; H-V1, Coriell Institute repository # NA11035, Venezuelan pedigree 104; H-V2, Coriell Institute repository # NA04477, Venezuelan pedigree 102; H-B1, Bantu genomic DNA male 1; H-B2, Bantu genomic DNA male 2; H-K1, Khoisan genomic DNA male 1; H-K2, Khoisan genomic



FIG. 1.—The frequency distribution of polymorphic sites. For each data set, n is the number of sequence; L is the length of the DNA sequence; and s is the number of polymorphic base pair positions, not including insertions and deletions. For comparison across data sets, π and θ have been divided by L to give estimates of diversity per base pair. Nearly all polymorphisms involve just two of the four nucleotide bases. To avoid redundancy, only the frequency of the less common base is shown. Thus, a polymorphism with one sequence that differs from the remaining n - 1 sequences is counted in frequency class 1. For the few three- and four-base polymorphisms, the frequencies of the two or three least common bases, respectively, are also included in s. For the large data sets, frequency classes are in groups of five. White bars are the observed frequencies, and total bar height sums to s. Black bars are expected levels under constant population size (Tajima 1989b). Hatched bars are expected levels found by simulations for a recently growing population (see *Materials and Methods* and *Results*). a, Mitochondrial cytochrome oxidase subunit II gene (Ruvolo, Zehr, and von Dornum 1993). b, Mitochondrial control region I and II data (Vigilant et al. 1991). c, β -globin data collected from the Vanuatu islands (Fullerton et al. 1994). d, Elastin data from a British Caucasian population (Raybould, Bierley, and Hulten 1995). L is not given for Elastin, which is restriction site data, and π and θ have not been standardized in this case.

DNA male 2. For the X-linked pyruvate dehydrogenase E1 α subunit (PDHA1) locus, sequences were obtained by polymerase chain reaction (PCR) amplification using primers corresponding to base numbers 14084–14103 and 16387–16406 (Koike et al. 1990). The region sequenced corresponds to the span from position 14210 to position 15957 of the reference sequence (Koike et al. 1990). For the mitochondrial control region I (CRI), initial amplification primers corresponded to positions 15975–15996 and 16401–16422 (Anderson et al. 1981). PCR products were converted to single-stranded form

(Higuchi and Ochman 1989) and sequenced directly. This was done for both strands. The GenBank accession numbers are U75933–U75940 for the PDHA1 sequences, and U75941–U75948 for the CRI sequences.

Simulations

Conventional coalescent simulations (Hudson 1990) were used to determine the distribution of Tajima's D (Tajima 1989b), and the distribution of the difference, Δ , between D values (see *Results*). Each value of Δ was determined from two independent, constant population size coalescent simulations, one for each of two loci. For each locus, the mean of the observed values of π and $\hat{\theta}$ (see fig. 1 and *Results*) was used as an estimate of the population mutation parameter, 2Gu (see *Results*), that was input to the simulations.

Coalescent simulations were also conducted under a population growth scenario. As in the case of constant population size simulations, a distribution of Δ was determined from multiple pairs of independent coalescent simulations. Each population growth simulation proceeded with a specified initial population size N_0 , final population size N_1 , and time of population growth, in generations, t. Rogers and Harpending (1992), in their table 1, provide estimates of these quantities based on a large worldwide sample of mitochondrial DNAs (Cann, Stoneking, and Wilson 1987). One of the benefits of coalescent simulation is that, apart from sample size, the population mutation parameter is the only required parameter, and it is not necessary to represent the actual population size in the simulation (Hudson 1990). This simplification, in the case of a coalescent simulation of sudden population growth, means that the actual estimates of N_0 , N_1 , and t are scaled by the population size at the beginning of the simulations (i.e., N_1 , since coalescent simulations proceed backward in time). With scaling, the estimates of Rogers and Harpending (1992) (rows 1 and 2 of their table 1) become $N_1 = 1$, $N_0 =$ 0.00594, and t = 0.01753. The scaled population size terms, N_1 and N_0 , can be applied to the mitochondrial simulations as well as to those for nuclear genes (assuming the population expansion did not affect the sexes differently). However, the scaled time parameter that is based on mitochondria cannot be applied directly to nuclear genes. Under the model, the time of population expansion was necessarily the same for all portions of the genome, and the scaled value of t must be adjusted for the differences in effective population size that are expected for different portions of the genome. An X chromosome gene is expected to have about three times-and a nuclear gene about four times-the effective population size of mitochondria. In effect, the relative time of the expansion, if it were scaled by the effective population size for the nuclear gene regions, is more recent than for the case of mitochondria. For the simulations representing the X-linked locus (PDHA1), t was set to 0.005843 (i.e., 0.01753/3). For the nuclear genes (β -globin and Elastin), t was set to 0.004383 (i.e., 0.01753/4). The coalescent simulations also require a population mutation parameter. However, π and θ (see fig. 1 and Results) are only appropriate estimates if the population size is constant, and they are not appropriate for the case of recent population expansion. To generate estimates under this model, preliminary population growth simulations were conducted (using N_1 , N_0 , and t as described above) to find values of the population mutation parameter that generated, on average, the same number of polymorphic sites that were observed in the data sets. These values are 22 for PDHA1, 102 for the β -globin data set, 24.5 for the Elastin data set, 42 for CRI, 25 for the mitochondrial COII data set, and 170

PDHA1

а

PDHAI		CRI	
		Position in	111111111
Position in	111	DNA sequence	666666666
DNA sequence	3046	-	000111111
	4687		569122466
	8197		192169878
H-NA1	CGTT	H-NA1	GCCCTCCCC
H-NA2		H-NA2	ATT-CG
H-V1		H-V1	A-TT-A
H-V2		H-V2	A - T AT - T
H-B1	TCCC	H-B1	A-TA
н-в2	T-CC	н-в2	A-TG
H-K1		H-K1	A-TG-T-
н-к2	TCCC	H-K2	A-TA

h

FIG. 2.—a, Polymorphic base positions among eight PDHA1 sequences. A "-" means the sequence matches H-NA1 at that position. During sequencing it was found that all eight sequences differed at multiple locations from the reference sequence (Koike et al. 1990). For this reason, the base positions simply reflect a count from the 5' end of the aligned sequences. H-NA1 had an insertion of eight bases beginning at position 1364, with respect to the other seven sequences. b, Polymorphic base positions refer to the reference sequence (Anderson et al. 1981).

for the large data set on mitochondrial control regions I and II. The results from these simulations were also used to generate expected site frequency distributions under the model of rapid population expansion (fig. 1).

Results

Figure 2 shows the DNA sequence polymorphisms found at PDHA1 and CRI. Polymorphism summaries and site frequency distributions are shown in figure 3. The amount of base pair heterozygosity at PDHA1 (π = 0.00113) is very close to the average of 0.0011 reported for four-fold degenerate sites within amino-acidcoding regions (Li and Sadler 1991). However this value is higher than for most X chromosomal genes, which on average appear to have about one third as much variation as autosomal genes (Hofker et al. 1986). Four polymorphisms were found at PDHA1, all within introns, and all among the four African sequences (H-B1, H-B2, H-K1, and H-K2). This pattern is roughly consistent with an out-of-Africa model of human origins. Interestingly, all four polymorphisms were also at intermediate frequency. In contrast, most of the CRI polymorphisms have the low-high frequency pattern. In short, the PDHA1 and CRI data reveal the same kind of difference (fig. 3a and b) that had been found in the other data sets (fig. 1).

A conventional chi-square test for differences in the frequency distributions cannot be made because linkage causes nonindependence among the polymorphisms within a locus. Instead, a simple two-step approach was developed: for each gene, the departure from null model expectations was assessed using Tajima's D (Tajima 1989b), and then the statistical significance of the difference between D values for pairs of genes was determined. D is proportional to the difference between two commonly used measures of nucleotide variation, π and θ . Under neutrality and constant population size, both measures have an expected



FIG. 3.—The frequency distribution of polymorphic sites at (a) PDHA1 and (b) CRI. See figure 1 legend for explanation.

value that is equal to the population mutation parameter 2Gu (often referred to as θ), where G is the effective number of gene copies in the population at the locus under investigation, and u is the mutation rate at the locus per generation. π is the average number of differences between all pairs of sequences (Nei and Taiima 1981). A polymorphism of intermediate frequency appears as a difference between many pairs of sequences, and so contributes more to π than does a polymorphism with the low-high frequency pattern. In contrast, θ is a function only of the number of DNA sequences (n) and the number of polymorphic sites (s), and it does not depend on the frequency of polymorphisms (Hudson 1990). For each of the mtDNA data sets (figs. 1a, 1b, and 3b), the excess of low-frequency polymorphisms results in a value of π that is less than θ , and D is negative. In the case of the nuclear genes (figs. 1c, 1d, and 3a) there is an excess of intermediate-frequency polymorphisms compared to the expected values, and D values are positive.

Let $\Delta_{n1,n2} = D_{n1} - D_{n2}$, where n1 and n2 are the sample sizes in the two data sets. For convenience, the D values are ordered so that $\Delta_{n1,n2}$ is positive. If the loci are unlinked, D_{n1} and D_{n2} are independent and the probability density of $\Delta_{n1,n2}$ can be found using a convolution of the probability densities for the two D's (expression 47 in Tajima 1989b). The general form of the convolution, and the probability that $\Delta_{n1,n2}$ takes a particular value δ , is

$$P(\delta) = \int \phi_{n1}(x)\phi_{n2}(x-\delta) \, dx, \qquad (1)$$

where $\phi_{n1}(x)$ is the probability that D_{n1} is equal to x (eq. 47 in Tajima 1989b). The probability p of observing a

value that is more extreme than is observed can be found by evaluating

$$p = \int_{\Delta_{n2,n2}}^{\infty} P(\delta) \ d\delta + \int_{-\infty}^{-\Delta_{n2,n2}} P(\delta) \ d\delta. \qquad (2)$$

By considering values greater than or equal to the observed value of $\Delta_{n1,n2}$ and less than or equal to $-\Delta_{n1,n2}$, the test becomes two-tailed. Both expressions (1) and (2) were evaluated numerically to determine the statistical significance of observed values of $\Delta_{n1,n2}$. Because the theoretical distribution of D is actually an approximation (Tajima 1989*b*), a simulated distribution of $\Delta_{n1,n2}$ was also used. The results of the statistical tests are shown in table 1. With the exception of the contrasts involving the Elastin data set, all of the tests that include a mitochondrial gene study and a nuclear gene study are statistically significant.

These tests are based on the assumption of a constant population size (Tajima 1989b). However, given the good fit between the mitochondrial data and models of population growth (Rogers and Harpending 1992; Takahata 1993), null models of population expansion were also considered. If a population is growing very rapidly, then the gene tree histories will approach a star phylogeny (Tajima 1989*a*; Slatkin and Hudson 1991). Under a star phylogeny, π and θ have a simple relationship to one another. Consider a star phylogeny connecting *n* sequences, and let there be x_i mutations on the branch leading to sequence *i*. Then the number of polymorphic sites, *s*, is equal to

$$\sum_{i=1}^{n} x_i, \tag{3}$$

and the average number of pairwise differences, π , is equal to

Table 1 Δ Values and Tests of Significance

	Nuclear Gene Data Sets			
– Mitochondrial Data Sets	PDHA1	β-globin (Fullerton et al. 1994)	Elastin (Raybould, Bierley, and Hulten 1995)	
CRI				
Δ^a p^b p-simulated ^e p-growing population ^d	2.781 0.0447* 0.0278* 0.0042**	2.866 0.0391* 0.0228* 0.0016**	1.762 0.2212 0.1904 0.0200*	
COII (Ruvolo, Zehr, and von Dornum 1993)				
Δ p p-simulated p-growing population	2.803 0.0417* 0.0220* 0.0022**	2.888 0.0364* 0.0208* 0.0084**	1.784 0.2175 0.1886 0.0448*	
Control regions I and II (Vigilant et al. 1991)				
Δ p p-simulated p-growing population	3.161 0.0214* 0.0076** 0.0210*	3.246 0.0204* 0.0124* <0.0002**	2.142 0.1312 0.1010 0.0578	

* $P \le 0.05$.

** $P \leq 0.01$.

^a Δ is the absolute value of the difference between D for a nuclear gene data set and D for a mitochondrial gene data set. D values are given in figures 1 and 3.

^b p is the probability of observing a value greater than the observed Δ , plus the probability of observing a value less than negative Δ (a two-tailed test), calculated with expression (2).

^c *p*-simulated is based on 5,000 independent simulated values of Δ assuming a constant population size (see *Materials and Methods*).

^d p-growing population is based on 5,000 independent values of Δ assuming a recent population expansion (see Materials and Methods).

$$\frac{\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}x_{i}+x_{j}\right)}{\binom{n}{2}}.$$
(4)

Substitution of (3) into (4) shows that $\pi = 2s/n$. Furthermore, since

$$\hat{\theta} = s / \sum_{i=1}^{n-1} \frac{1}{i}$$
(5)

(Hudson 1990), it follows that

$$\pi = \frac{2\hat{\theta} \sum_{i=1}^{n-1} 1/i}{n}.$$
 (6)

Thus, under a star phylogeny, π and θ are related by a scalar and will have a correlation of one. In contrast, under a constant population size, π and θ can vary with respect to each other (Tajima 1989b). This means that the quantity $\pi - \theta$, to which D is proportional, has a source of variation under a constant population size that is not present when a population is growing rapidly. In short, the two extremes of very rapid population growth and constant population size are associated with relatively low and high variances for D, respectively. It follows that the observed values of Δ would be judged even more unlikely under the rapid-growth model than under the constant-population-size model.

As an additional check, simulations were conducted under a specific rapid-population-growth model that had previously been fitted to mitochondrial data. The simulations incorporated values of a recent large population size, an initial small population size, and a time of population size change that had previously been estimated (Rogers and Harpending 1992) using a large worldwide mtDNA data set (Cann, Stoneking, and Wilson 1987). The probabilities associated with the observed values of Δ are considerably less under this population-growth model (table 1), as expected if the variance of *D* is less than under a constant-population-size model.

Discussion

These analyses suggest that variation at nuclear genes and mtDNA are not both consistent with a common demographic history. In particular, the PDHA1– CRI contrast effectively eliminates simple explanations based on the genetic structure among human populations, although it remains possible that more complex demographic models could be found to fit the data.

The statistically significant findings arise even in contrasts among genes for which sample sizes and estimates of 2Gu were quite low. There appear to be two reasons for this. First is the simple reason that the observed values of Δ are large. Second is that the variance of D is essentially not a function of either sample size or 2Gu. Indeed, for samples with more than five sequences and 2Gu greater than one (for the entire sequenced region, not per base pair) the mean and variance of D are nearly constant and close to zero and one, respectively (Tajima 1989b). Since D is only slightly a function of sample size and 2Gu, the same goes for Δ . In short, statistical tests based on D and Δ do not require large sample sizes.

In principle, natural selection at either nuclear or mitochondrial genes (or both) could be responsible for the observed differences. For example, balancing selection at nuclear genes will increase the levels of intermediate-frequency polymorphisms and elevate D for these loci. Indeed, PDHA1 does reveal more variation than most X-chromosomal genes (Hofker et al. 1986) and it is possible that this is due to balancing selection. However, the mtDNA is the most likely candidate for a selective explanation for the observed values of Δ . In contrast to the three unlinked nuclear genes, the D values from the different mitochondrial studies are not independent because the mtDNA is a single linkage group. The mtDNA is a also an excellent candidate for a byproduct of natural selection that occurs when multiple loci are tightly linked. Under tight linkage, natural selection cannot act effectively on multiple polymorphic sites, and this leads to a reduction of the effective number of gene copies experienced by the region of tight linkage (Felsenstein 1974). Recent analyses of mitochondrial variation in several organisms, including humans, suggest that they are segregating an excess of slightly deleterious variation in protein-coding genes (Nachman, Boyer, and Aquadro 1994; Rand, Dorfsman, and Kann 1994; Nachman et al. 1996). These mutations are expected to create a background selection effect (Charlesworth, Morgan, and Charlesworth 1993) for the entire mitochondrial genome, simultaneously knocking down the effective population size for mitochondria and shifting the site frequency distribution toward an excess of the low-high pattern, as is observed. Background selection against strongly deleterious mutations probably could not explain the observation (Charlesworth, Morgan, and Charlesworth 1993; Charlesworth, Charlesworth, and Morgan 1995). This type of selection has only a small effect on D, and the mitochondrial genome has relatively few genes that present targets for strongly deleterious mutations. However, deleterious mutations of very slight effect, or that are nearly or completely recessive in their effect, can lead to a shift in the site frequency distribution and a negative D (Charlesworth, Charlesworth, and Morgan 1995).

Recent data from the human Y chromosome also do not have the same pattern of polymorphism found in mitochondria. A diverse sample of 16 individuals revealed π greater that θ (Hammer 1995). As with the other nuclear genes, *D* is positive (0.2306) but it is not significantly different from the mitochondrial genes. The Y chromosome region that was sequenced was from the nonrecombining region, so it may also be subject to an interaction of linkage and natural selection.

The patterns of mtDNA variation have previously been interpreted as evidence of rapid population growth that followed a period of small population size about 200,000 years ago (Cann, Stoneking, and Wilson 1987; Vigilant et al. 1991), and the analysis of Rogers and Harpending (1992) suggested a date between 60,000 and 120,000 years. This historical portrait is consistent with the estimated dates of the earliest modern human fossils and has provided support for the hypothesis of a recent African origin for modern humans (Stringer and Andrews 1988). The PDHA1 polymorphisms are consistent with the African-origin portion of this hypothesis as they were found among the sequences of African origin, suggesting that African populations are more variable at this locus. However, the site frequency distributions of the nuclear genes are in conflict with the population expansion scenario. The positive D values from nuclear genes suggest that human populations were relatively large and not subject to population bottlenecks during the time that human ancestors evolved into modern form.

Acknowledgments

I thank M. Bulmer, D. Rand, A. Rogers, P. Smouse, M. Stoneking and N. Takahata for helpful comments, and N. Leahy and A. Zaidi for laboratory assistance. The Bantu and Khoisan DNAs were kindly supplied by M. Stoneking and H. Soodyall. This work was supported by a Johnson & Johnson Discovery Award and by National Science Foundation grant DEB-9306625.

LITERATURE CITED

- ANDERSON, S., A. T. BANKIER, B. G. BARRELL et al. (14 coauthors). 1981. Sequence and organization of the human mitochondrial genom. Nature 290:457–465.
- BOWCOCK, A. M., J. R. KIDD, J. L. MOUNTAIN, J. M. HEBERT, L. CAROTENUTO, K. K. KIDD, and L. L. CAVALLI-SFORZA. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc. Natl. Acad. Sci. USA 88:839-843.
- CANN, R. L., M. STONEKING, and A. C. WILSON. 1987. Mitochondrial DNA and human evolution. Nature **325**:31–36.
- CAVALLI-SFORZA, L. L., P. MENOZZI, and A. PIAZZA. 1993. Demic expansions and human evolution. Science **259**:639–646.
- CHARLESWORTH, B., M. T. MORGAN, and D. CHARLESWORTH. 1993. The effect of deleterious mutations on neutral molecular evolution. Genetics 134:1289–1303.
- CHARLESWORTH, D., B. CHARLESWORTH, and M. T. MORGAN. 1995. The pattern of neutral molecular variaton under the background selection model. Genetics **141**:1619–1632.
- 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.
- FELSENSTEIN, J. 1974. The evolutionary advantage of recombination. Genetics 78:737-756.
- FU, Y. X., and W. H. LI. 1993. Statistical tests of neutrality of mutations. Genetics 133:693–709.
- FULLERTON, S. M., R. M. HARDING, A. J. BOYCE, and J. B. CLEGG. 1994. Molecular and population genetic analysis of allelic sequence diversity at the human β -globin locus. Proc. Natl. Acad. Sci. USA **91**:1805–1809.
- HAMMER, M. F. 1995. A recent common ancestor for human Y chromosomes. Nature **378**:376–378.
- HIGUCHI, R. G., and H. OCHMAN. 1989. Production of singlestranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. 17: 5865.
- HOFKER, M. H., M. I. SKRAASTAD, A. A. B. BERGEN, M. C. WAPENAAR, E. BAKKER, A. MILLINGTON-WARD, G. J. B. VAN OMMEN, and P. L. PEARSON. 1986. The X chromosome shows less genetic variation at restriction sites than the autosomes. Am. J. Hum. Genet. 39:438-451.
- HUDSON, R. R. 1990 Gene genealogies and the coalescent process. Pp 1–44 in P. H. HARVEY and L. PARTRIDGE, eds. Oxford surveys in evolutionary biology. Vol. 7. Oxford University Press, New York.
- KOIKE, K., Y. URATA, S. MATSUO, and M. KOIKE. 1990. Characterization and nucleotide sequence of the gene encoding the human pyruvate dehydrogenase α -subunit. Gene **91**: 307–311.
- LI, W. H., and L. SADLER. 1991. Low nucleotide diversity in man. Genetics **129**:513–523.
- LONG, J. C., A. CHAKRAVARTI, C. D. BOEHM, S. E. ANTON-ARAKIS, and H. H. KAZAZIAN. 1990. Phylogeny of human B-globin haplotypes and its implications for recent human evolution. Am. J. Phys. Anthropol. **81**:113–130.
- MERRIWETHER, D. A., A. G. CLARK, S. W. BALLINGER, T. G. SCHURR, H. SOODYALL, T. JENKINS, S. T. SHERRY, and D. C. WALLACE. 1991. The structure of human mitochondrial DNA variation. J. Mol. Evol. 33:543-555.
- NACHMAN, M. W., S. N. BOYER, and C. F. AQUADRO. 1994. Nonneutral evolution at the mitochondrial NADH dehydro-

genase 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., and F. TAJIMA. 1981. DNA polymorphism detectable by restriction endonucleases. Genetics **97**:145–163.
- RAND, D. M., M. DORFSMAN, and L. M. KANN. 1994. Neutral and non-neutral evolution of Drosophila mitochondrial DNA. Genetics 138:741–756.
- RAYBOULD, M. C., A. J. BIERLEY, and M. HULTEN. 1995. Molecular variation of the human elastin (ELN) gene in a normal human population. Ann. Hum. Genet. 59:149–161.
- ROGERS, A. R., and H. HARPENDING. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9:552–568.
- RUVOLO, M., S. ZEHR, and M. VON DORNUM. 1993. Mitochondrial COII sequences and modern human origins. Mol. Biol. Evol. **10**:1115–1135.
- SLATKIN, M., and R. R. HUDSON. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. Genetics 129:555–562.

- STONEKING, M. 1994. In defense of "eve"—a response to Templeton's critique. Am. Anthropol. 96:131-141.
- STRINGER, C. B., and P. ANDREWS. 1988. Genetic and fossil evidence for the origins of modern humans. Science 239: 1263–1268.
- TAJIMA, F. 1989*a*. The effect of change in population size on DNA polymorphism. Genetics **123**:597–601.
- . 1989b. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.
- TAKAHATA, N. 1993. Allelic genealogy and human evolution. Mol. Biol. Evol. 10:2–22.
- 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.

CHARLES F. AQUADRO, reviewing editor

Accepted October 31, 1996