Questioning Evidence for Recombination in Human Mitochondrial DNA

The possibility of recombination in human mitochondrial DNA (mtDNA), raised recently by Awadalla *et al.* (*1*), holds crucial implications for many evolutionary studies (*2*). Here, we reexamine the data analyzed in (*1*), show that some of those data are likely unreliable, and suggest that the short-distance correlations found by Awadalla *et al.* (*1*) can be more plausibly interpreted phylogenetically.

Awadalla *et al.* (*1*) examined 14 variable positions in 45 mtDNA genomes for distance correlation. The first of those positions, nucleotide 4985, is known to be a sequence error (*3*). In another position, nucleotide 6455, a T has been recorded in eight out of ten sequences from haplogroups M and U (*4*), while in other related haplogroup M genomes, it is a "conventional" C (*5*). We analyzed 48 European and Asian haplogroup M and U mtDNA samples and did not detect any variation at this position; moreover, a C is present at this position in a sequence from sub-Saharan African haplogroup L1 sequence (*6*), as well as in the mitochondrial genomes of *Pan paniscus* and *Pan troglodytes* (*7*). Hence, a sequencing or typing error in (*4*) is a likely explanation. This does not exhaust the list of suspicious polymorphisms used in (*1*); sequence 12 from (*8*), for example, is likely a mosaic of haplogroup T– and H–type mtDNA genomes.

Our next criticism is based on a phylogenetic argument. We have resequenced or typed by restriction enzymes seven sites used in the analysis of Awadalla *et al.* (*1*)—7028, 9540, 10873, 11251, 11467, 12705, and 15043—in 88 mtDNAs of African, Asian, and European origin. All but one of the sites were found to occur only once in the phylogeny of human mtDNA haplogroups. The only exception was at position 15043, where A was found in all haplogroup M and haplogroup I mtDNAs, but only once each in haplogroups L3 and T. We sequenced 14 additional haplogroup T mtDNAs and found that all of them contain G in position 15043, which confirms that this position is slightly polymorphic within haplogroup T. Should this change from G to A be ascribed to recombination? We consider that prospect unlikely, because two other polymorphic sites typical for haplogroup T in the vicinity of 15043, 14905A, and 15607G, are fixed in all haplogroup T mtDNAs examined.

We also stress that substitutions at sites 13366, 15606, and 15925 in haplogroup T [figure 1B in (*1*)], 10394 and 10397 in haplogroup M [figure 1D in (*1*)], and 7933 and 8391 in haplogroup Y [figure 1C in (*1*)], which account for the short distance correlations, segregate in linkage with a number of other haplogroup-specific substitutions that are spread over the entire mtDNA genome (*4, 5, 9, 10*). The latter have, however, remained hidden in the analysis of Awadalla *et al.* (*1*) because of a bias created by particular restriction fragment length polymorphism (RFLP) sites used.

In sum, likely errors in the sequence data used by Awadalla *et al.* (*1*) and the possibility that straightforward phylogenetic explanations can explain the observed correlations make the conclusions drawn in (*1*) weaker than such an exceptionally important problem deserves.

Toomas Kivisild
Richard Villems
Estonian Biocentre and
Department of Evolutionary Biology
Tartu University
23 Riia Street
51010 Tartu, Estonia
E-mail: tkivisil@ebc.ee; rvillems@ebc.ee

References and Notes
1. P. Awadalla, A. Eyre-Walker, J. Maynard Smith, *Science* **286**, 2524 (1999).
2. See, e.g., E. Strauss, *Science* **286**, 2436 (1999).
3. R. M. Andrews *et al.*, *Nature Genet.* **23**, 147 (1999).
4. T. Ozawa, *Biochim. Biophys. Acta* **1271**, 177 (1995).
5. ———— *et al.*, *Biochem. Biophys. Res. Commun.* **177**, 518 (1991).
6. S. Horai, K. Hayasaka, R. Kondo, K. Tsugane, N. Takahata, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 532 (1995).
7. *Pan paniscus*: GenBank accession D38116; *Pan troglodytes*: GenBank accession D38113.
8. M. J. Rieder, S. L. Taylor, V. O. Tobe, D. A. Nickerson, *Nucleic Acids Res.* **26**, 967 (1998).
9. S. Hofmann *et al.*, *Hum. Mol. Genet.* **6**, 1835 (1997).
10. T. G. Schurr *et al.*, *Am. J. Phys. Anthrop.* **108**, 1 (1999).
11. We thank H.-J. Bandelt and M. Bamshad for discussions and comments.

Awadalla *et al.* (*1*) presented an ingenious approach for testing for recombination in the mitochondrial genome. Their analysis of four mtDNA data sets showed a consistent decline of linkage disequilibrium (LD) with physical distance, a phenomenon typically observed in the recombining nuclear genome (*2*). The number of sites they analyzed is quite small in three of the data sets, however, and they used a measure, r^2, that can produce misleading results because of its sensitivity to allele frequency variation (*3*).

We have reanalyzed their data using D'

(*4*), a measure that provides better accuracy and power for LD detection and is less susceptible to the effects of allele frequency variation (*3, 5*). We have also analyzed four additional mtDNA data sets (191 Armenians, 109 Croatians, 388 Turks, and 67 Germans) kindly provided by R. Villems and T. Kivisild. For two polymorphic sites, A and B,

$$D' = (P_{11} - p_1 q_1)/D_{max} \qquad (1)$$

where P_{11} is the population frequency of the haplotype containing alleles A_1 and B_1, p_1 is the frequency of A_1, q_1 is the frequency of B_1, and D_{max} is the maximum value of the numerator in Eq. 1 that is allowed by the allele frequencies. D' is equivalent to r^2/r^2_{max}, where r^2 is the statistic used by Awadalla *et al.* (*1*). Like r^2, D' is expected theoretically to decline as the recombination fraction increases [(*3*), p. 316]. We tested the correlation between physical distance and LD using the Mantel matrix comparison method, a permutation-based approach (*6*).

Although we obtain the same correlations for r^2 with physical distance as Awadalla *et al.* (*1*), we find that, in contrast to their report, only two of their data sets yield significance levels (*P*) of <0.05, and none of four additional data sets that we have analyzed show a significant relationship between r^2 and physical distance (Table 1). Furthermore, neither the data sets analyzed by Awadalla *et al.* nor the four additional sets analyzed here reach statistical significance when D' is used (one set does yield a D' value for which $P < 0.05$, but the correlation is positive rather than negative). Most of the D' values are equal to 1.0, and D' does not show a decline with distance in any of the eight data sets (Fig. 1). Most pairs of sites are at the maximum level of disequilibrium allowed by the allele frequencies. Thus, most of the r^2 values reported by Awadalla *et al.* (*1*) are as high as they can be given the allele frequencies. The apparent decline of r^2 with distance would seem to be primarily an artifact of allele frequency dependency.
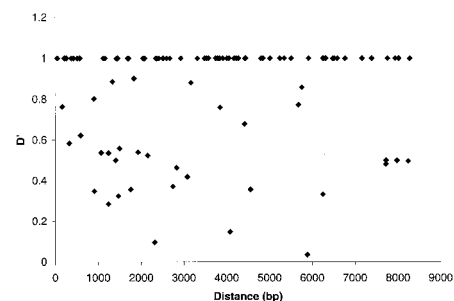
Awadalla *et al.* (*1*) concluded that their

Fig 1. The relationship between linkage disequilibrium, measured by D', and the physical distance between pairs of polymorphisms, measured in number of base pairs (bp), using the data from Awadalla *et al.* [figure 1A of (*1*)].

Table 1. Correlations between LD and physical distance between polymorphic sites. To assess statistical significance, 10,000 random permutations were performed for each data set. As in Awadalla *et al.* (*1*), we used only sites located outside the control region and in which the minor allele had a frequency exceeding 0.10.

Data set	Correlation value for r^2	Correlation value for D'
Awadalla *et al.* [figure 1A of (*1*)]	$-0.245; P > 0.06$	$0.062; P > 0.31$
Awadalla *et al.* [figure 1B of (*1*)]	$-0.464; P > 0.06$	$0.246; P > 0.14$
Awadalla *et al.* [figure 1C of (*1*)]	$-0.512; P < 0.03$	$-0.095; P > 0.33$
Awadalla *et al.* [figure 1D of (*1*)]	$-0.830; P < 0.02$	$-0.036; P > 0.48$
Armenians (6 sites)	$-0.257; P > 0.24$	$0.464; P < 0.04$
Germans (14 sites)	$0.044; P > 0.45$	$0.161; P > 0.12$
Croatians (5 sites)	$-0.231; P > 0.28$	$0.267; P > 0.22$
Turks (10 sites)	$0.139; P > 0.15$	$0.151; P > 0.18$

results imply that mtDNA-based conclusions about human evolution should be "reconsidered." Yet comparisons of worldwide mtDNA variation and autosomal variation have generally yielded consistent results (*7–10*), with the observed differences readily explained by factors such as sex-specific differences in gene flow or the lower effective size of mtDNA. This level of consistency would not be expected if mtDNA variation were affected seriously by recombination.

The possibility of recombination in mtDNA is intriguing and deserves further evaluation. Six of the eight mtDNA data sets examined here fail to show a significant decline of LD with physical distance using the r^2 statistic, however, and none show a decline using the more appropriate D' statistic. Thus, LD patterns provide little support for the hypothesis of mtDNA recombination.

L. B. Jorde
Eccles Institute of Human Genetics
University of Utah School of Medicine
10 North 2030 East
Salt Lake City, UT 84112, USA
E-mail: lbj@genetics.utah.edu

M. Bamshad
Department of Pediatrics
University of Utah School of Medicine
E-mail: mike@genetics.utah.edu

References
1. P. Awadalla, A. Eyre-Walker, J. Maynard Smith, *Science* **286**, 2524 (1999).
2. L. B. Jorde, *Am. J. Hum. Genet.* **56**, 11 (1995).
3. B. Devlin and N. Risch, *Genomics* **29**, 311 (1995).
4. R. C. Lewontin, *Genetics* **49**, 49 (1964).
5. S.-W. Guo, *Hum. Hered.* **47**, 301 (1997).
6. P. E. Smouse, J. C. Long, R. R. Sokal, *Syst. Zool.* **35**, 627 (1986).
7. L. B. Jorde, M. Bamshad, A. R. Rogers, *BioEssays* **20**, 126 (1998).
8. L. B. Jorde *et al.*, *Am. J. Hum. Genet.* **66**, 979 (2000).
9. M. Nei, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6720 (1995).
10. L. L. Cavalli-Sforza, *Trends Genet.* **14**, 60 (1998).

Awadalla *et al.* (*1*) concluded that a negative association between LD and distance between site pairs constitutes evidence for recombination in mtDNA. Our reanalysis of their data reveals major problems with their

analysis and indicates that these data are consistent with mutation and linkage rather than with recombination.

First, the r^2 measure of LD between two loci, used by Awadalla *et al.*, depends not only on recombination but also on allele frequencies at the two loci (*2, 3*). Values of r^2 do not range from 0 to 1 unless allele frequencies are equal at the two loci (*4*), and the most frequent alleles at the two loci are positively associated. For example, in sites 4985 and 11251, in the most extensive data set of 45 complete mtDNA sequences they used, there are 34 AA, 6 AG, 5 GA, and 0 GG sequences (haplotypes), and the value of r^2 is only 0.02. However, these sites actually have the maximum LD possible for the frequencies of the constituent nucleotides.

It is therefore preferable to use the absolute value of the standardized LD measure D', which has a range of 0 to 1 for any set of allele frequencies (*5*). For sites 4985 and 11251, $|D'| = 1$, which occurs whenever at least one of the four possible haplotypes is not present. In the mtDNA sequence data set, 58 of the 91 pairs of the 14 polymorphic sites (63.7%) show the maximum possible LD ($|D'| = 1$). In contrast to the result for r^2 (*1*), no relationship exists between $|D'|$
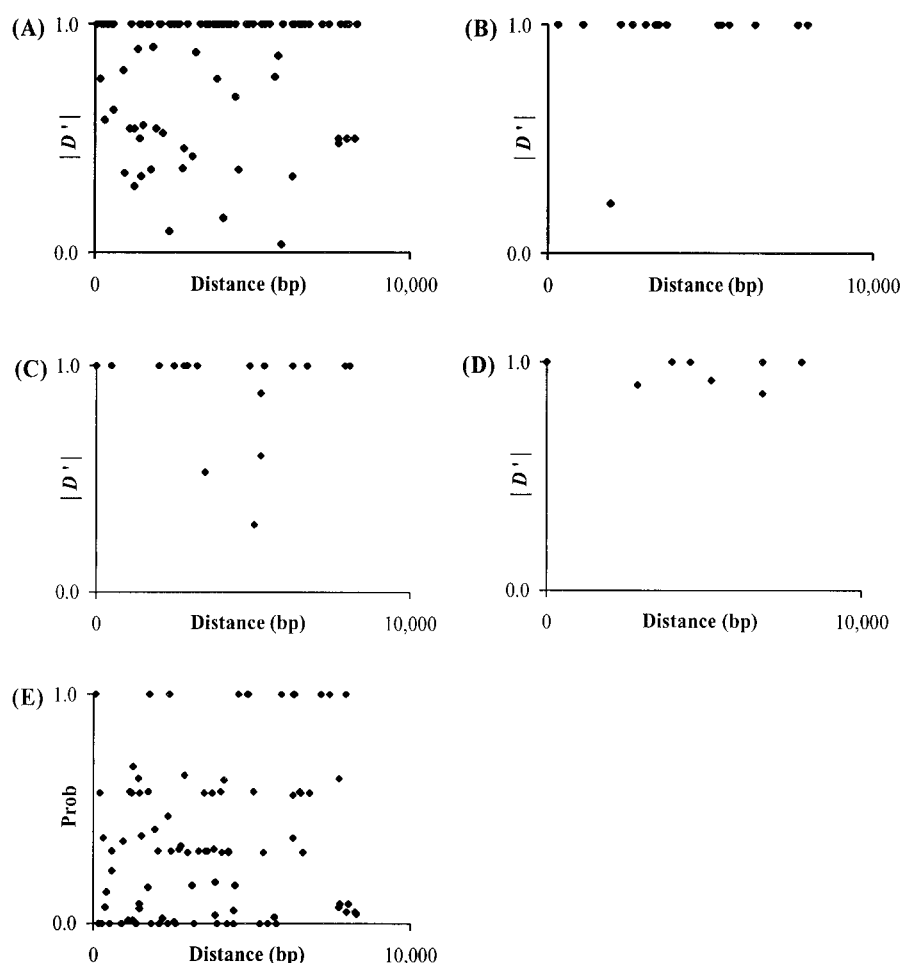

Fig. 1. Relation between distance between site pairs and LD as measured by $|D'|$ for (**A**) synonymous variants from 45 complete mtDNA sequences ($\rho = 0.06$), and for three RFLP data sets obtained from (*1*): (**B**) Swedish and Finnish ($\rho = 0.25$), (**C**) Native Siberians ($\rho = -0.09$), and (**D**) Native Americans ($\rho = 0.04$). (**E**) Fisher's exact-test probability for the data set in (A) ($\rho = 0.20$). Removal of site pairs showing $|D'| = 1$ results in $\rho = -0.13$ for (A) and $\rho = 0.03$ for (E); for other panels, number of observations was insufficient to compute ρ. None of the correlation coefficients are significant at the 5% level.

and physical distance, in base pairs (bp), between sites (Fig. 1A). Indeed, the average $|D'|$ for the 42 site pairs less than 3000 bp apart is 0.79 ± 0.04, smaller than 0.86 ± 0.04 for the 49 site pairs greater than 3000 bp apart. These results and those from the RFLP data (Fig. 1B through 1D) that were also analyzed by Awadalla et al. (1) provide no support for the hypothesis of recombination in human mtDNA. Furthermore, there is no distance-dependent recombination in their mtDNA sequence data set, because site pairs showing $|D'| < 1$ are distributed randomly with distance (Pearson's correlation coefficient, $\rho = -0.13$; $P = 0.27$).

Another approach to measuring LD is to calculate the probability of observing an equal or more extreme two-locus association by chance alone [Fisher's exact test (6)]. Exact probability values can be used to determine the association of LD with physical distance, because the probability value should be higher for more distant sites than for closer sites if recombination is occurring, and thus a positive slope between physical distance and the exact-test probability is expected. No such significant association is observed for the sequence data (Fig. 1E) or the three RFLP data sets (results not shown).

The relationship of r^2 and mutation can be examined by constructing a phylogenetic tree (Fig. 2) using the 14 variable sites analyzed from the 45 sequences. (Only 22 unique sequences actually exist, because six of the haplotypes occur multiple times.) In this tree, internal branches are almost as long as the external branches, an appearance unlike those of other human mtDNA trees (7, 8). This probably reflects the removal of sites at which the nucleotide frequencies were ≤ 0.10, which leaves out any variants represented in less than five sequences. This tree clearly indicates that the nucleotide differences among haplotypes are correlated. Mapping nucleotide substitutions on the phylogenetic tree of 22 unique haplotypes reveals unique transitional changes at four sites (11251, 12372, 14783, and 15043), parallel transitional changes at eight sites (4985, 6455, 7028, 9540, 10873, 12705, 13617, and 15301), and backward changes at the other two sites (11467 and 11299). Pairwise comparisons of sites with unique mutations should not provide any information about recombination; however, r^2 is spuriously close to zero (0.02 to 0.05) for five of six such pairs, even though each pair shows maximum $|D'|$. A similar problem exists for r^2 estimates computed for 28 pairs of parallel

mutation sites (15 pairs show $r^2 < 0.05$). Therefore, unique and parallel mutations by themselves in independent lineages may produce unusually low r^2 values. Furthermore, pairs of sites with either only three of the four possible haplotypes observed, or the fourth haplotype observed only once, occur with a frequency of 84.6% (Fig. 1A), 93.3% (Fig. 1B), 90.5% (Fig. 1C), and 70.0% (Fig. 1D) in the four data sets. These observations support the lack of recombination in human mtDNA, rather than providing evidence for it.

Our reanalysis thus contradicts the contention by Awadalla et al. (1) that recombination is occurring in human mtDNA. Extensive family studies have likewise failed to find any exceptions to strict maternal clonal inheritance of human mtDNA (9–12). There is no need to reconsider inferences about human or mtDNA evolution that have assumed that recombination does not occur in human mtDNA.
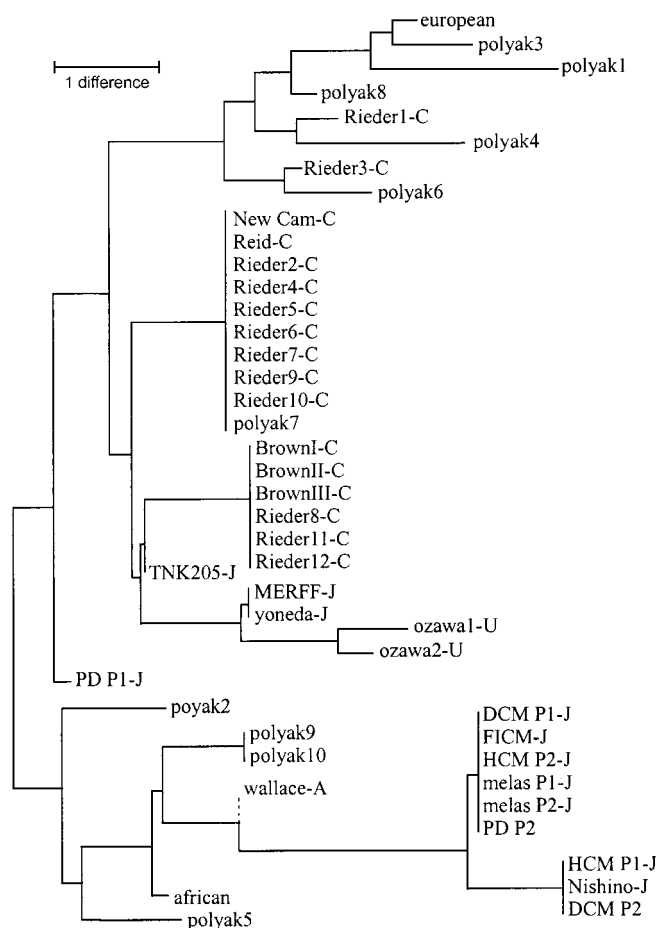
Sudhir Kumar
Philip Hedrick
Thomas Dowling
Department of Biology
Arizona State University
Tempe, AZ 85287, USA

Mark Stoneking
Max Planck Institute for Evolutionary
Anthropology
Inselstrasse 22
D-04103 Leipzig, Germany
E-mail: stoneking@eva.mpg.de

References and Notes
1. P. Awadalla, A. Eyre-Walker, J. Maynard Smith, *Science* **286**, 2524 (1999).
2. M. Nei, *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987), p. 173.
3. P. Hedrick, *Genetics of Populations* (Jones & Bartlett, Boston, 2000), p. 404.
4. ———, *Heredity* **60**, 435 (1988).
5. R. C. Lewontin, *Genetics* **49**, 49 (1964).
6. ———, *Genetics* **140**, 377 (1995).
7. S. B. Hedges, S. Kumar, K. Tamura, M. Stoneking, *Science* **255**, 737 (1992).
8. A. J. Redd and M. Stoneking, *Am. J. Hum. Genet.* **65**, 808 (1999).
9. T. J. Parsons et al., *Nature Genet.* **15**, 363 (1997).
10. H. Soodyall et al., *Am. J. Phys. Anthropol.* **104**, 157 (1997).
11. E. Jazin et al., *Nature Genet.* **18**, 109 (1998).
12. C. Danan et al., *Am. J. Hum. Genet.* **65**, 463 (1999).
13. Because there are only 14 parsimony-informative sites for 22 unique haplotypes, the maximum-parsimony method was not used. Instead, a single tree was obtained by using the neighbor-joining (NJ) method on pairwise distances of 22 haplotypes. The NJ tree is one of the most parsimonious trees.

Recent studies arguing for significant recombination between maternal and paternal mtDNA in humans (1, 2) have generated considerable debate (3–7). Awadalla et al. (8) have added to that debate a new study presenting a significant decline in LD between mtDNA sites with the distance be-

Fig. 2. Phylogenetic tree of human mtDNA haplotypes based on the number of differences observed in the 14 sites analyzed in (1). The neighbor-joining tree is shown with branch lengths denoting the actual number of differences per sequence (13). Wallace sequence is connected to the tree with a dashed line because the sequence has missing data in 3 of 14 sites.

tween sites, for both humans and chimpanzees. Arguing that this effect is difficult to explain in any way other than recombination, Awadalla *et al.* (*8*) concluded that inferences about human and mtDNA evolution based on presumed clonal inheritance "will now have to be reconsidered."

In reference to that conclusion, Awadalla *et al.* cited several important papers that analyzed sequence variation in the mtDNA control region (CR), and that found extreme mutation rate heterogeneity among sites (*9–12*). Other proponents of mtDNA recombination have also suggested that such apparent mutation rate heterogeneity in the CR may be due instead to patterns of recombination (*2*). However, the data analyzed by Awadalla *et al.* come from either the entire mtDNA genome (RFLP data), the entire mtDNA protein-coding region (sequence data, humans), or two widely separated regions (ND2 and the CR, sequence data, chimpanzees). Moreover, their study provided no indication of how frequently recombination would have to occur to produce the negative correlation between linkage disequilibrium and distance they reported. If recombination is causing this effect, it raises an important question: Could that recombination be sufficiently frequent to shape the observed patterns of CR variation, and thereby invalidate the vast body of human evolutionary work that has been based on CR sequences?

To address that question, we repeated the analysis of Awadalla *et al.* (*8*), using a database of hypervariable region 1 and hypervariable region 2 sequences from the CR of 1278 individuals representing a range of ethnic groups (103 African-American, 110 Afro-Caribbean, 98 English Caucasian, 536 U.S. Caucasian, 97 Hispanic, 115 African, 57 U.S. Asian, and 162 Japanese). We analyzed sites that were twofold degenerate within this database, and whose minority variant occurred in at least 5% of the individuals in the database. The positions analyzed were 16069, 16126, 16172, 16187, 16189, 16223, 16224, 16278, 16294, 16304, 16311, 16319, 16362, 73, 146, 150, 152, 153, 182, 195, 198, 204, and 295 [relative to the Cambridge Reference Sequence (*13*)]. For all pairs of sites, we calculated the Hill and Robertson measure of LD (*14*), and analyzed this against the distance between sites. The calculated Pearson's correlation coefficient was a nonsignificantly positive value of 0.062 [significance determined as in Awadalla *et al.* (*8*), with 4129 of 5000 random replicates giving a correlation of 0.062 or lower].

Our analysis shows that in the human mtDNA CR, there is no indication of a negative correlation between LD and distance that would signal the action of recombination. The closer proximity of even the most distant sites in our analysis, 795 bp,

may account for why Awadalla *et al.* detected recombination and we did not. Nonetheless, our results suggest that recombination, if it occurs, is not of a level to leave a trace in a CR data set for which mutation rate heterogeneity among sites is glaringly evident (*9–12*).

In light of that finding, it seems unlikely that our understanding of the pattern and relative rates of sequence evolution within the mtDNA CR will require substantial revision based on the Awadalla *et al.* report. Our analysis also suggests that mtDNA forensic testing will be negligibly impacted by recombination; forensic applications already deal successfully with intergenerational mutation (*15, 16*), clearly a far more significant effect.

Thomas J. Parsons
Jodi A. Irwin
Armed Forces DNA Identification
Laboratory
Armed Forces Institute of Pathology
1413 Research Blvd.
Rockville, MD 20886, USA

References
1. A. Eyre-Walker, N. H. Smith, J. Maynard Smith, *Proc. R. Soc. London Ser. B* **266**, 477 (1999).
2. E. Hagelberg *et al.*, *Proc. R. Soc. London Ser. B* **266**, 485 (1999).
3. P. Arctander, *Science* **284**, 2090 (1999).
4. D. A. Merriweather and F. A. Kaestle, *Science* **285**, 837 (1999).
5. V. Macaulay, M. Richards, B. Sykes, *Proc. R. Soc. London Ser. B* **266**, 2037 (1999).
6. A. Eyre-Walker, N. H. Smith, J. Maynard Smith, *Proc. R. Soc. London Ser. B* **266**, 2041 (1999).
7. E. Strauss, *Science* **286**, 2436 (1999).
8. P. Awadalla, A. Eyre-Walker, J. Maynard Smith, *Science* **286**, 2524 (1999).
9. J. Wakely, *J. Mol. Evol.* **37**, 613 (1993).
10. M. Hasegawa, A. Di Rienzo, T. D. Kocher, A. C. Wilson, *J. Mol. Evol.* **37**, 347 (1993).
11. L. Excoffier and Z. Yang, *Mol. Biol. Evol.* **16**, 1357 (1999).
12. S. Meyer, G. Weiss, A. von Haessler, *Genetics* **152**, 1703 (1999).
13. S. Anderson *et al.*, *Nature* **290**, 457 (1981).
14. W. G. Hill and A. Robertson, *Theor. Appl. Genet.* **38**, 776 (1986).
15. T. J. Parsons *et al.*, *Nature Genet.* 15, **363** (1997).
16. M. M. Holland and T. J. Parsons, *For. Sci. Rev.* **11**, 21 (1999).

2 February 2000; accepted 27 April 2000

Response: We recently showed that LD declines with increasing distance between sites in human and chimpanzee mtDNA (*1*), an observation consistent with genetic recombination in hominid mitochondria. Four groups have questioned our findings for a number of different reasons. Before addressing these arguments in detail, however, we take this opportunity to provide corrected probability values (Table 1), which included slight errors in (*1*). The new figures do not qualitatively affect our conclusions.

Kivisild and Villems argue that our results may stem from errors in the data or a bias in the restriction sites surveyed. This seems unlikely, for several reasons. First, we observe a negative correlation between LD and distance in 8 of the 10 data sets considered by us and by Jorde and Bamshad in their comment, with the probability being <10% in five cases [table 1 of (*1*) and table 1 of Jorde and Bamshad]. Second, random sequencing errors would not be included in our analysis because they would generally appear as singletons, and systematic sequencing errors, though they might increase or decrease LD, would do so without respect to distance. Although there was an error at 4985 in the original Cambridge sequence (*2*), which was not included in our sample, that does not mean that our sequence data are incorrect, and we see no reason to believe that 6455 is in error. If we remove either of these sites from our data, the correlation between LD and distance remains negative, though the significance level is reduced (excluding 4985, $\rho = -0.176$ and $P = 0.151$; excluding 6455, $\rho = -0.234$ and $P = 0.089$; excluding both, $\rho = -0.119$ and $P = 0.275$). The assertion that sequence 12 from (*3*) is incorrect and is "likely a mosaic of haplogroup T– and H–type mtDNA genomes" simply illustrates a dogmatic belief in the clonality of mtDNA. Might it not be a recombinant?

Jorde and Bamshad and Kumar *et al.* suggest that our results may be an artifact of the measure of LD we used, r^2, because r^2 is more dependent upon allele frequencies than other LD statistics such as D'. We believe that this criticism of r^2 is misleading. First, neither group provides a model of allele frequency variation that would generate our results, and such a model is difficult to envisage for a circular chromosome. The decline in r^2 is as expected under a population genetic model with recombination (*4, 5*). Second, both r^2 and D' are affected by the frequency of the alleles (*6, 7*), because the denominators of both are simple functions of allele frequency. In five of the six data sets we analyzed, the reciprocal of the denominator of r^2, $1/p_A(1 - p_A)p_B(1 - p_B)$, is positively correlated with distance (Table 1), which suggests that the negative correlation between r^2 and distance is not caused by an unusual pattern in allele frequencies.

We do not find it surprising that the correlation between D' and distance is not significant, because r^2 should have more power to detect recombination than D'. D' has an extremely skewed distribution at low recombination rates and allele frequencies (*7, 8*), and r^2 provides a more informative estimate of LD. Consider two populations in which the frequencies of three haplotypes, *AB*, *aB* and *Ab*, are (0.33, 0.33, and 0.33) in the first population and (0.49, 0.50, and 0.01) in the second. The evidence for LD is stronger in the first than in the second because the ab-

sence of the fourth haplotype is surprising only in the first sample. This is reflected in the r^2 values, which are 0.33 and 0.01, respectively, but not in D', which is one in each case.

The statistic r^2 also reflects two other aspects of recombination that D' does not. The value of r^2 is greatest when the two sites have similar allele frequencies, and the two rare (common) alleles are in coupling. Sites are more likely to have the same allele frequency, and be in coupling, if they have not recombined, since they then share the same genealogy (9, 10); thus, sites close together are expected to have higher r^2 values if recombination is occurring. D' will also tend to have higher values if the alleles are in coupling if all four haplotypes are present. Hence, r^2 can potentially detect recombination, even when the fourth haplotype is not present and $D' = 1$.

Simulations and data have shown that a high proportion of D' values are expected to equal one in recombining sequences, especially when sites exhibit low allele frequencies (7, 8, 11–13). The striking observation in both the data we used and the data used by Kivisild and Villems in their "phylogenetic argument" is not that many of the pairwise comparisons involve only three haplotypes and $D' = 1$, but that a substantial proportion involve four haplotypes. The fourth haplotype can be produced only by recombination or multiple mutation.

Kumar et al. show that the Fisher's exact test (FET) value is not visibly correlated to distance for our sequence data set, although the correlation is marginally significant ($P = 0.085$). However, the logarithm of the FET value is positively correlated with distance, a relationship that is significant in three of our data sets and marginally significant in two others (Table 1). This is not surprising: r^2 is proportional to the value obtained from a chi-square test for heterogeneity, so the logarithm of the FET value and r^2 are very similar statistics.

Kumar et al. present a phylogenetic analysis that they suggest is more consistent with multiple mutation than recombination. If there is recombination, however, a phylogenetic tree represents nothing physical; it is simply an estimate of the "average" genealogy of the sites being considered. In essence, we find a logical error here: Kumar et al. implicitly assume that recombination does not occur in constructing their tree, and then use the tree to argue that there is no recombination. The most obvious property of the tree is the number of homoplasies it contains, which, again, can be produced only by recombination or multiple mutation.

Recombination has not been observed in pedigree analyses of mtDNA, but the sample sizes are relatively small. Partial control region sequences have been obtained from about 1500 generational events (14–16). Given that there are only 100 mitochondria in the sperm, compared with 100,000 in the egg, this sample size represents the lower limit that would be needed to detect recombination if there were no selection against paternal mitochondria, as there appears to be. Paternal inheritance is rarely tested for, or considered, in these analyses; single nucleotide changes that occur in the pedigree are assumed to reflect new mutations and multiple nucleotide changes are assumed to be due to laboratory error or a problem with the genealogy. Both could be due to paternal inheritance, but that is rarely considered as the cause.

Finally, Parsons and Irwin argue that although there may be evidence of recombination in the mitochondrial genome as a whole, there is no evidence of recombination in the control region, because there is no correlation between LD and distance. We think that this is a dangerous argument to make. First, "an absence of evidence is not evidence of an absence"; many sequences that have undergone recombination do not show a significant decline in LD with distance, and this is likely to be particularly true for short sequences. This does not mean that they are not affected by recombination. Second, they analyzed sequences from a variety of ethnic groups, which may introduce LD due to population subdivision and may thus obscure any patterns in LD due to recombination. Third, as we have argued elsewhere (17, 18), there may be independent evidence of recombination in the control region; phylogenetic trees constructed with control region sequences typically show high levels of homoplasy. Although it is commonly assumed that the homoplasies are caused by hypervariability (19), they may be caused by recombination (17, 18).

In short, we believe that the high level of homoplasy in many mitochondrial data sets in both humans and chimpanzees, and the decline in LD with distance in some, provide good evidence that recombination does occur in mtDNA. The next challenge will be to estimate the rate at which recombination occurs, and test whether hominids are unique in allowing it to happen.

Philip Awadalla
*Institute of Cell, Animal and
Population Biology
University of Edinburgh
Edinburgh EH9 1JT, UK
E-mail: p.awadalla@ed.ac.uk*

**Adam Eyre-Walker
John Maynard Smith**
*Centre for the Study of Evolution and
School of Biological Sciences
University of Sussex
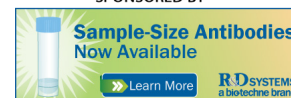Brighton BN1 9QG, UK
E-mail: a.c.eyre-walker@sussex.ac.uk*

References and Notes

1. P. Awadalla, A. Eyre-Walker, J. Maynard Smith, *Science* **286**, 2524 (1999).
2. R. M. Andrews *et al.*, *Nature Genet.* **23**, 147 (1999).
3. M. J. Rieder, S. L. Taylor, V. O. Tobe, D. A. Nickerson, *Nucleic Acids Res.* **26**, 967 (1998).
4. W. G. Hill and A. Robertson, *Theor. Appl. Genet.* **38**, 226 (1968).
5. R. R. Hudson, *Genetics* **109**, 611 (1985).
6. R. C. Lewontin, *Genetics* **120**, 849 (1988)
7. ———, *Genetics* **140**, 377 (1995).
8. P. W. Hedrick, *Genetics* **117**, 331 (1987).
9. R. R. Hudson, *Theor. Popul. Biol.* **19**, 169 (1983).
10. ——— and N. L. Kaplan, *Genetics* **111**, 147 (1985).
11. M. A. Riley, M. E. Hallas, R. C. Lewontin, *Genetics* **123**, 359 (1989).
12. S. W. Schaeffer and E. L. Miller, *Genetics* **135**, 541 (1993).
13. M. Slatkin, *Genetics* **137**, 331 (1994).
14. T. J. Parsons *et al.*, *Nature Genet.* **15**, 363 (1997).
15. E. Jazin *et al.*, *Nature Genet.* **18**, 109 (1998).
16. S. Sigurdardottir, A. Helgason, J. R. Gulcher, K. Stefansson, P. Donnelly, *Am. J. Hum. Genet.* **66**, 1599 (2000).
17. A. Eyre-Walker, N. H. Smith, J. Maynard Smith, *Proc. R. Soc. London Ser. B* **266**, 477 (1999).
18. ———, *Proc. R. Soc. London Ser. B* **266**, 2041 (1999).
19. Thanks to N. Barton, B. Hill, D. Hudson, and G. McVean for discussions.

18 February 2000; accepted 29 May 2000

Table 1. The correlation (ρ) between the distance between sites for the six data sets analyzed by Awadalla *et al.* (*1*) and r^2, the reciprocal of the denominator of r^2, and the logarithm of the probability value from FET. The probability value, *P*, is the proportion of 10,000 randomized data sets for which the correlation is equal to or more negative than the correlation observed in the data. Because of corrections to the chimpanzee data, the level of LD is significant at ≤0.05 for 187 out of 694 pairwise comparisons within the ND2 and control region sequences, but only 64 out of 445 between-region comparisons ($P < 0.0001$).

Data set	r^2		$1/p_A(1-p_A)p_B(1-p_B)$		Log FET	
	ρ	P	ρ	P	ρ	P
Global sequence	−0.245	0.056	0.162	0.823	0.221	0.091
Finns and Swedes	−0.464	0.066	−0.507	0.080	0.410	0.098
Native Siberians	−0.512	0.025	0.134	0.622	0.494	0.040
Native Americans	−0.830	0.008	0.384	0.844	0.794	0.005
Global RFLP	−0.365	0.217	0.730	0.916	0.355	0.211
Chimpanzee	−0.158	0.002	0.167	0.968	0.156	0.009

Science
AAAS

Questioning Evidence for Recombination in Human Mitochondrial DNA

Toomas Kivisild and Richard Villems