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.

18 January 2000; accepted 27 April 2000

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

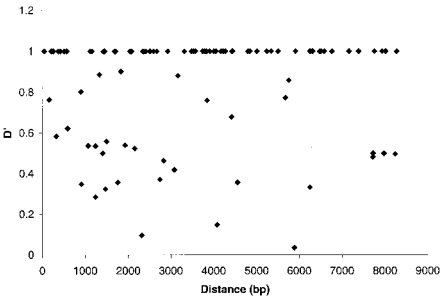

Fig 1. The relationship between linkage disequilibrium, measured by D', and the physical distance between pairs of polymorphisms, measured in number of base pairs (bp), using the data from Awadalla *et al*. [figure 1A of (*1*)].