

span was significantly greater in the old selection lines ($P < 0.05$) and when the females were made sterile with *ovo^{P1}* ($P < 0.001$). The response of the life span of young-line females to *ovo^{P1}* was significantly ($P < 0.01$) greater than that of the old-line females.

17. T. B. L. Kirkwood, *Nature* **270**, 301 (1977); _____, and R. Holliday, *Proc. R. Soc. London Ser. B.* **205**, 531 (1979).

18. J. W. Vaupel and A. I. Yashin, *Am. Stat.* **39**, 176; J. W. Vaupel *et al.*, *Science* **280**, 855 (1998).

19. P. B. Medawar, *An Unsolved Problem in Biology* (H. K. Lewis, London, 1952); B. Charlesworth, *Evolution in Age-Structured Populations* (Cambridge Univ. Press, Cambridge, ed. 2, 1994); D. Houle *et al.*, *Genetics* **138**, 773 (1994).

20. S. D. Pletcher, D. Houle, J. W. Curtsinger, *Genetics* **148**, 287 (1998).

21. We thank T. Chapman for assistance with irradiation; T. Chapman, K. Fowler, and D. Gems for helpful discussion; and J. Carey, L. Harshman, P. Harvey, S. Pletcher, and J. Vaupel for comments on the manuscript. Supported by the Natural Environment Research Council (UK).

17 September 1999; accepted 24 November 1999

Linkage Disequilibrium and Recombination in Hominid Mitochondrial DNA

Philip Awadalla,¹ Adam Eyre-Walker,^{2*} John Maynard Smith²

The assumption that human mitochondrial DNA is inherited from one parent only and therefore does not recombine is questionable. Linkage disequilibrium in human and chimpanzee mitochondrial DNA declines as a function of the distance between sites. This pattern can be attributed to one mechanism only: recombination.

For many years, it has been accepted that mitochondria are inherited exclusively from the mother in mammals and that their inheritance is therefore clonal. This assumption has been used extensively to date events in human prehistory, including the age of our last common female ancestor, "Eve," and the spread of *Homo sapiens* in Asia and Europe (1). However, mitochondria contain the enzymes necessary for homologous recombination (2), and there are at least two routes by which strict maternal inheritance of mitochondrial DNA (mtDNA) could be circumnavigated: (i) Paternal mitochondria enter the egg at fertilization (3, 4), and (ii) there are copies of mtDNA sequences in the nuclear genome (5) that could be transferred back to the mtDNA. Two reports have presented population genetic evidence of recombination in mtDNA (6, 7). Although doubts were raised about the quality of the data and the conclusions reached in one of these reports (8), subsequent analyses of newer data have reconfirmed the original result (9).

Recombination can also be detected by considering the relation between linkage disequilibrium (LD) and distance. As the distance between sites increases, the effect of recombination on LD should increase, whether recombination is a reciprocal or a nonreciprocal (gene conversion) process. Recombination should therefore manifest itself as a significant decline in LD with distance (10–13).

To measure disequilibrium, we used the Hill and Robertson measure (14)

$$r^2 = \frac{(P_{AB}P_{ab} - P_{aB}P_{Ab})^2}{P_A(1 - P_A)P_B(1 - P_B)} \quad (1)$$

where *A* and *a* represent alleles at one locus, *B* and *b* represent alleles at the other locus, and p_{XY} is the frequency of the *XY* genotype. r^2 can range from 0 to 1, from complete linkage equilibrium to complete disequilibrium, respectively. We only calculated LD values for pairs of sites at which both alleles were segregating at a frequency greater than 10% (in no case were more than two different nucleotides present at a site). There were two reasons for this restriction: (i) Alleles at higher frequencies tend to be older and therefore more likely to show evidence of past recombination events, and (ii) values of $r^2 = 1$ often occur by chance for alleles with low frequencies (15).

We analyzed both sequence and restriction fragment length polymorphism (RFLP) data for complete human mitochondrial genomes. To test whether LD declines with distance, we calculated Pearson's correlation coefficient, ρ , between LD and the distance between sites. To assess significance, we randomized the positions of sites and then recalculated the value of ρ (11, 13, 16). This was repeated 5000 times; the significance is given as the proportion of random ρ values that are more negative than the observed value. For the complete human mitochondrial sequences, we restricted analysis to synonymous variation, because selection can generate LD and there is no evidence of selection on synonymous codon usage in human mitochondrial sequences (7). In the RFLP data, it is generally not possible to determine whether the loss of a site is due to a synonymous or

nonsynonymous change, so all variants in protein, tRNA, and ribosomal RNA coding sequences were analyzed. The human mitochondrial genome is a circular molecule 16,569 base pairs (bp) in length, so the longest distance between pairs is 8284 bp.

In humans, we have five independently collected data sets: 45 complete mtDNA sequences of diverse geographical origin (17) and four RFLP data sets spanning the whole genome, from 147 individuals from around the world (18), 86 Finnish and Swedish individuals (19), 153 Native Siberians (20), and 167 Native Americans (21). The relation between LD and the distance between sites, for four of these data sets, is shown in Fig. 1. In each case, there is an evident decline in LD with increasing distance, and in each case, the decline is significant. For example, Fig. 1A shows the 91 pairwise LD values for the 14 synonymous sites in the DNA sequence data set, plotted as a function of the distance between sites. The relation with distance is highly significant ($P = 0.012$). In the worldwide RFLP data set, there were only four sites segregating mutations at greater than 10%; there is a negative correlation between LD and distance ($\rho = -0.365$), but it is nonsignificant ($P = 0.190$). These patterns are consistent with those observed for nuclear sequences undergoing recombination (10–13) and with population genetic predictions for the decay of LD due to recombination (15, 22, 23).

For chimpanzees (*Pan troglodytes*), we used synonymous variants from the reduced form of nicotinamide adenine dinucleotide dehydrogenase subunit 2 (ND2) locus and data from the control region (CR) from 16 individuals of the subspecies *verus* from western Africa (24). There are about 3800 bp between the two sequences. There were 45 polymorphic sites and 990 pairwise comparisons. Again, there was a significant decline in linkage disequilibrium as distance between pairs of sites increased ($\rho = -0.147$; $P = 0.0076$). The proportion of pairs of sites with LD values significantly different from zero, at the 5% level, was much higher within the ND2 and CR sequences (166 out of 593 pairwise comparisons) than between them (46 out of 396) (Fisher's exact test, $P < 10^{-6}$).

We have demonstrated that LD declines as a function of the distance between sites in humans and chimpanzees. This is expected if there is genetic recombination but hard to

¹Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 1JT, UK.

²Centre for the Study of Evolution and School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK.

*To whom correspondence should be addressed. E-mail: a.c.eyre-walker@sussex.ac.uk

REPORTS

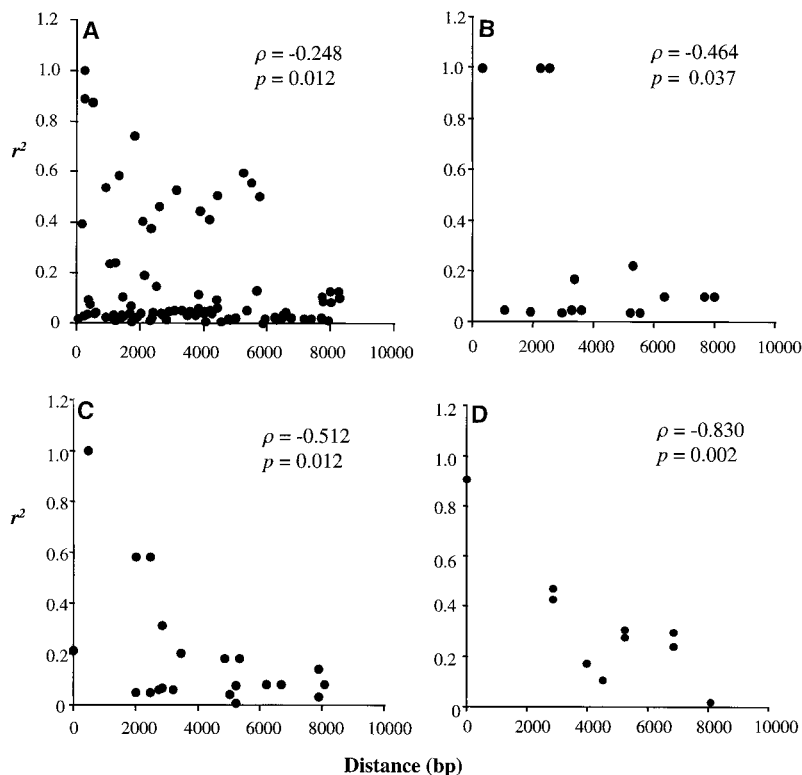


Fig. 1. The relation between LD and the distance between sites in the human mitochondrial genome for four data sets. **(A)** Synonymous variants from 45 complete mtDNA sequences from individuals of diverse origins: 18 Caucasian, 13 Japanese, 2 African, and 12 individuals of unknown origin (17). Sites analyzed are 4985, 6455, 7028, 9540, 10,873, 11,251, 11,299, 11,467, 12,372, 12,705, 13,617, 14,783, 15,043, and 15,301. In this data set, there were no sites with multiple alleles (sites with more than two nucleotides) segregating. RFLP data sets: **(B)** Swedish and Finnish individuals (19). Sites analyzed are 7025, 10,394, 12,308, 13,366, 15,606, and 15,925. **(C)** Native Siberian individuals (20). Sites analyzed are 1715, 5176, 7933, 8391, 10,394, 10,397, and 13,262. **(D)** Native American individuals (27). Sites analyzed are 663, 5176, 10,394, 10,397, and 13,262. In the Native American and Siberian data sets, RFLP sites 10,394 and 10,397 could be caused by a single mutation. However, they segregate at different frequencies, so it seems likely that they are due to different mutations; however, the results remain qualitatively unchanged if either site is excluded (Siberians: excluding 10,394, $\rho = -0.705$, $P = 0.001$; excluding 10,397, $\rho = -0.485$, $P = 0.026$. Native Americans: excluding 10,394, $\rho = -0.583$, $P = 0.089$; excluding 10,397, $\rho = -0.653$, $P = 0.063$).

explain in any other way; sequencing errors would tend to obscure the effect, not generate it. The fact that we obtained significant evidence for recombination in five out of the six hominid data sets we analyzed makes it no longer reasonable to assert that mitochondria do not undergo genetic recombination, particularly because the test we used is not powerful even in a panmictic population (25) and patterns of LD with distance can easily be hidden by LD generated by population subdivision (26).

There are three possible routes by which recombination could occur: (i) paternal leakage, (ii) recombination with copies of mtDNA sequences in the nuclear genome, and (iii) recombination in heteroplasmic individuals. The last possibility is unlikely; not only would two mutations have to occur within one individual and be maintained in heteroplasmy until recombination had occurred, but all four haplotypes (that is, AB, Ab, aB, and ab) would have to

have descendants in our sample for us to detect the recombination event. It is difficult to assess the relative likelihood of the other pathways; paternal leakage is known to occur (3, 4), although in mice, there are efficient mechanisms for destroying paternal mtDNA (4), and there are mtDNA pseudogenes in the nucleus (5).

Many inferences about the pattern and tempo of human evolution (1) and mtDNA evolution (27) have been based on the assumption of clonal inheritance. These inferences will now have to be reconsidered.

References and Notes

1. L. B. Jorde, M. Bamshad, A. R. Rogers, *Bioessays* **20**, 126 (1998).
2. B. Thyagaran, R. A. Padua, C. Campbell, *J. Biol. Chem.* **271**, 27536 (1996).
3. F. Ankel-Simons and J. M. Cummins, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13859 (1996).
4. H. Shitara *et al.*, *Genetics* **148**, 851 (1998).
5. D. C. Wallace, C. Stugard, D. Murdock, T. Schurr, M. D. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 14900 (1997).

6. E. Hagelberg *et al.*, *Proc. R. Soc. London Ser. B Biol. Sci.* **266**, 485 (1999).
7. A. Eyre-Walker, N. H. Smith, J. Maynard Smith, *Proc. R. Soc. London Ser. B. Biol. Sci.* **266**, 477 (1999).
8. V. Macaulay, M. Richards, B. Sykes, *Proc. R. Soc. London Ser. B Biol. Sci.* **266**, 2037 (1999).
9. A. Eyre-Walker, N. H. Smith, J. Maynard Smith, *Proc. R. Soc. London Ser. B Biol. Sci.* **266**, 2041 (1999).
10. N. M. Miyashita and C. H. Langley, *Genetics* **120**, 199 (1988).
11. B. W. Schaeffer and E. L. Miller, *Genetics* **135**, 541 (1993).
12. D. J. Conway *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4506 (1999).
13. P. Awadalla and D. Charlesworth, *Genetics* **152**, 413 (1999).
14. W. G. Hill and A. Robertson, *Theor. Appl. Genet.* **31**, 881 (1968).
15. B. Weir and W. G. Hill, *Am. J. Hum. Genet.* **38**, 776 (1986).
16. We only considered sites at which both alleles segregate at greater than 10%. Let X_i be the position in the mtDNA sequence of the i th site and Y_i be a vector containing the sequence data for that site. During randomization, we randomly allocated X_i values to Y_i vectors without replacement: for example, $[(X_4, Y_1), (X_1, Y_2)]$, and so forth.
17. Forty-five complete, or nearly complete, human mtDNA sequences were taken from Genbank (accession numbers X93334 and D38112) or the literature [D. C. Wallace *et al.*, *Science* **242**, 1427 (1988); M. Yoneda *et al.*, *Biochem. Int.* **21**, 789 (1990); Y. Kobayashi *et al.*, *J. Hum. Genet.* **49**, 590 (1991); T. Ozawa *et al.*, *Biochem. Biophys. Res. Commun.* **177**, 518 (1991); M. D. Brown, A. S. Voljavec, M. L. Lott, I. MacDonald, D. C. Wallace, *FASEB J.* **6**, 2791 (1992); M. D. Brown *et al.*, *Genetics* **130**, 163 (1992); F. M. Reid, G. A. Vernham, H. T. Jacobs, *Hum. Mol. Genet.* **3**, 1435 (1994); T. Ozawa *et al.*, *Biochem. Biophys. Res. Commun.* **207**, 613 (1995); I. Nishino *et al.*, *Biochem. Biophys. Res. Commun.* **225**, 180 (1996); M. J. Rieder, S. L. Taylor, V. O. Tobe, D. A. Nickerson, *Nucleic Acids Res.* **26**, 967 (1998); K. Polyak *et al.*, *Nature Genet.* **20**, 291 (1998); R. M. Andrews *et al.*, *Nature Genet.*, in press]. The data set is different than the one we used previously (7); the data given by S. Marzuki *et al.* [*Hum. Genet.* **88**, 139 (1991)] were removed because they are thought to be incorrect [N. Howell *et al.*, *Am. J. Hum. Genet.* **50**, 1333 (1992)].
18. R. L. Cann, M. Stoneking, A. C. Wilson, *Nature* **325**, 31 (1987). Sequence 110 was removed because this is the "Cambridge" sequence, which is chimeric. Sites whose positions were not firmly resolved were removed if the possible positions were more than 10 bp apart.
19. A. Torroni *et al.*, *Genetics* **144**, 1835 (1996).
20. A. Torroni *et al.*, *Am. J. Hum. Genet.* **53**, 591 (1993).
21. A. Torroni *et al.*, *Genetics* **130**, 153 (1992).
22. T. Ohta and M. Kimura, *Genet. Res.* **13**, 47 (1969).
23. In (14, 15, 22), it is shown that under neutrality, $E(r^2) = 1/(1+4NrD) + 1/n_a$ for $r \ll 1$, where D is the distance between two pairs of sites, N is the effective population size, r is the recombination rate per base pair, and n_a is the sample size.
24. C. A. Wise, M. Sraml, S. Easteal, *Genetics* **148**, 409 (1998).
25. J. Maynard Smith, *Genetics* **153**, 1021 (1999).
26. T. Ohta, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1940 (1982).
27. J. Wakeley, *J. Mol. Evol.* **37**, 613 (1993); M. Hasegawa, A. Di Rienzo, T. D. Kocher, A. C. Wilson, *J. Mol. Evol.* **37**, 347 (1993); L. Excoffier and Z. Yang, *Mol. Biol. Evol.* **16**, 1357 (1999); S. Meyer, G. Weiss, A. von Haeseler, *Genetics* **152**, 1103 (1999).
28. We thank M. Rieder, K. Polyak, and N. Howell for providing their data; J. Hey, W. Hill, G. McVean, J. Slate, and T. Solomon for advice and encouragement; and P. Andolfatto, N. Barton, B. Charlesworth, D. Charlesworth, T. Johnson, P. Keightley, V. Macaulay, M. Turelli, M. Slatkin, N. Smith, and two anonymous referees for comments on the manuscript. P.A. is funded by a University of Edinburgh Faculty of Science and Engineering Research Scholarship. A.E.-W. is funded by the Royal Society.

6 August 1999; accepted 10 November 1999