

Letter to the Editor

The Correlation Between Linkage Disequilibrium and Distance: Implications for Recombination in Hominid Mitochondria

Julien Meunier* and Adam Eyre-Walker†

*Ecole Normale Supérieure de Lyon, Lyon, France; and †Centre for the Study of Evolution and School of Biological Sciences, University of Sussex, Brighton, England

It is generally believed that the mitochondrial genome is inherited from a single parent in animals and higher plants and that the inheritance of mitochondrial DNA (mtDNA) is therefore clonal (Birky 1995). However, the clonality of human mtDNA has recently been questioned (Awadalla, Eyre-Walker, and Maynard Smith 1999; Eyre-Walker, Smith, and Maynard Smith 1999a; Hagelberg et al. 1999) and keenly debated (Eyre-Walker, Smith, and Maynard Smith 1999b; Macaulay, Richards, and Sykes 1999; Awadalla, Eyre-Walker, and Maynard Smith 2000; Eyre-Walker 2000; Jorde and Bamshad 2000; Kivisild and Villems 2000; Kumar et al. 2000; Parsons and Irwin 2000).

The evidence of recombination in human mtDNA comes from two sources. First, many phylogenetic trees constructed using mtDNA contain a large amount of homoplasy (e.g., Vigilant et al. 1991; Ingman et al. 2000), which has generally been attributed to the presence of hypervariable sites in the mtDNA (Hasegawa et al. 1993; Wakeley 1993). However, Eyre-Walker, Smith, and Maynard Smith (1999a, 1999b) have suggested that the homoplasies could be due to recombination. Second, Awadalla, Eyre-Walker, and Maynard Smith (1999) found that linkage disequilibrium (LD) declined as a function of the distance between sites in several human data sets and one chimpanzee mtDNA data set, a pattern which is highly consistent with recombination. However, their analysis was criticized on several grounds. In particular, Jorde and Bamshad (2000) and Kumar et al. (2000) argued that the measure of LD used by Awadalla, Eyre-Walker, and Maynard Smith (1999), r^2 , the squared correlation of allele frequencies (Hill and Robertson 1968), was inappropriate since it is dependent on allele frequencies; they argued that the absolute value of D' , the absolute value of D over its maximum value (Leventon 1964), was a more appropriate statistic, and they noted that $|D'|$ was not correlated to the distance between sites for any of the data sets considered by Awadalla, Eyre-Walker, and Maynard Smith (1999). Awadalla, Eyre-Walker, and Maynard Smith (2000) subsequently gave a number of reasons why they believed that r^2 might be a better measure of LD for detecting recombination, but ultimately this is an argument which can only be resolved by investigation. Here, we present the results of some population genetic simulations we

used to investigate our ability to detect recombination from the correlation between r^2 or $|D'|$ and distance.

We simulated a population of N haploid circular genomes of length L under a model of mutation, random genetic drift, and recombination. In each generation, U mutations and R recombination events were distributed across the population, where U and R were random numbers drawn from Poisson distributions with means of NLu and Nr , respectively, where u is the nucleotide mutation rate and r is the recombination rate per chromosome. For each recombination event, two genomes were chosen at random, along with a starting point. From the randomly chosen starting point, z nucleotides were transferred between the genomes in a gene conversion-type process. Most simulations were performed using reciprocal recombination, but we also investigated nonreciprocal recombination by transferring the information from one chromosome to the other but not reciprocating the process (this yielded very similar results; not shown). To form the next generation, we randomly sampled N genomes with replacement.

Initial simulations showed that the model parameterized in Nu and Nr as expected. We therefore performed all subsequent simulations with a population size of 500 genomes, assuming a chromosome length of 16,500 bases, the approximate length of the human mtDNA. The simulation was allowed to run for $10N$ generations to equilibrate, after which 100 samples of 49 genomes were sampled at intervals of $2N$. These samples could be considered largely independent, since a haploid population is expected to coalesce, on average, every $2N$ generations; this was checked by calculating the autocorrelation between successive samples. For each sample, we calculated the correlation between r^2 (and $|D'|$) and the distance between sites. Significance was assessed using a one-tailed Mantel test with 500 randomized data sets (see Awadalla, Eyre-Walker, and Maynard Smith [1999] for details).

To begin our investigations, we followed Awadalla, Eyre-Walker, and Maynard Smith (1999) and restricted our analysis to sites at which both alleles were present at a frequency of $>10\%$, and we set the mutation rate so that ~ 50 sites were polymorphic at that level. We arbitrarily chose to exchange 5,000 nt in each recombination event and varied the recombination rate across a broad range of values. The results are presented in table 1. As expected, the probability of detecting recombination increased as the rate of recombination increased, and there was no evidence of excessive type I error when there was no recombination. Over all rates of recombination, r^2 and $|D'|$ performed at levels that were very similar to each other. We then investigated a broad range of other parameter combinations by varying

Address for correspondence and reprints: Adam Eyre-Walker, Centre for the Study of Evolution and School of Biological Sciences, University of Sussex, Brighton BN1 9QG, United Kingdom. E-mail: a.c.eyre-walker@sussex.ac.uk.

Mol. Biol. Evol. 18(11):2132–2135. 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Table 1
Proportion of Samples (and standard error) for Which the Correlation Between r^2 or $|D'|$ and Distance Is Significant at the 5% Level for Different Levels of Recombination

R	$ D' $	r^2	r^2 Without 4-Genotype Pairwise Comparisons
5	0.96 (0.03)	0.93 (0.04)	0.88 (0.05)
1.5	0.65 (0.07)	0.70 (0.07)	0.59 (0.07)
0.5	0.30 (0.07)	0.28 (0.06)	0.24 (0.06)
0.15	0.09 (0.04)	0.12 (0.05)	0.10 (0.04)
0.05	0.05 (0.03)	0.06 (0.03)	0.05 (0.03)
0	0.00 (0.00)	0.02 (0.02)	0.02 (0.02)

NOTE.—The mutation rate was set such that ~50 polymorphisms were at a frequency of $\geq 10\%$, and 5,000 bp were transferred in each recombination event.

(1) the recombination tract length from 500 to 8,000 bp, (2) the frequency above which alleles were included in the analysis from 0 to 0.2, and (3) the mutation rate such that either 50 or 10 polymorphisms were, on average, included in the analysis. While changing many of these parameters changed the probability of detecting recombination, it did not alter the relative performances of r^2 and $|D'|$, which were very similar in all cases.

Our simulations suggest that $|D'|$ and r^2 have similar abilities to detect recombination when they are correlated against the distance between sites, at least under the conditions we simulated. Why, then, is there a negative correlation between r^2 and distance in many hominid data sets, while there is no correlation between $|D'|$ and distance? The most likely explanations are that r^2 is detecting recombination, but from a different source of information than $|D'|$, or that r^2 is being misled by some other signal in the data. $|D'|$ detects recombination

from the pairs of sites which have all four genotypes; if there are no such pairs of sites, then it cannot detect recombination, since $|D'| = 1$ for all pairwise comparisons (PWCs). However, r^2 could potentially obtain some, if not all, of its information from the two and three genotype PWCs. This appears to be the case; the proportion of samples in which r^2 detects recombination is little affected by the removal of the four-genotype PWCs from the analysis (table 1).

A similar pattern is observed in the human and chimpanzee data sets analyzed by Awadalla, Eyre-Walker, and Maynard Smith (1999) and others; removing the PWCs with four genotypes from the analysis makes little difference in the value of the correlation between r^2 and distance or the significance of the result (table 2). Thus, the r^2 test appears to make little use of the four-genotype cases, and this may be the reason $|D'|$ and r^2 give different results in the hominid mtDNA data sets—they use different information. But this leaves us with two questions: (1) if recombination is occurring, then why does $|D'|$ not detect it in the hominid data sets, and (2) how does r^2 detect recombination when it ignores the most obvious evidence of recombination, the presence of all four genotypes?

There are several processes that were not modeled in our simulation: (1) hypervariable sites, (2) population size expansion, and (3) population subdivision. Any one of these could potentially be responsible for the different results obtained with r^2 and $|D'|$ in the hominid data sets, but we suspect that hypervariable sites are the most likely candidate. Both population subdivision and population expansion will reduce the proportion of PWCs that have four genotypes and therefore reduce the efficiency of $|D'|$, but all the hominid data sets show a reasonable proportion of four-genotype PWCs, and population sub-

Table 2
The Correlation of $|D'|$ or r^2 and the Distance Between Sites for a Number of Hominid mtDNA Data Sets

Data	Reference ^a	$ D' $	r^2	r^2 Without 4-Genotype Pairwise Comparisons
Global	1	0.062	-0.245*	-0.238*
Global minus 4985 ^b	1	0.050	-0.176	-0.216
Global ^c	2	0.109	0.115	0.108
Finns and Swedes	3	0.246	-0.464*	-0.527**
Native Siberians	4	-0.095	-0.512**	-0.492**
Native Americans	5	-0.036	-0.830***	-0.836**
Caucasians	6	0.113	-0.355	-0.749
Germans	7	0.170	0.044	-0.060
Adygei and Druze	8	0.212	0.155	0.242
Southeast Asians	9	-0.180	-0.462	-0.999
Africans	10	0.123	0.045	0.022
Tibetans	11	-0.124	-0.280**	-0.441***
Chimpanzees	12	-0.092	-0.158***	-0.159***

NOTE.—Only sites at which the minor allele was segregating at $>10\%$ were included.

^a 1—Awadalla, Eyre-Walker, and Maynard Smith (1999); 2—Ingman et al. (2000); 3—Torrioni et al. (1996); 4—Torrioni et al. (1993); 5—Torrioni et al. (1992b); 6—Torrioni et al. (1992a); 7—Hofmann et al. (1997); 8—Macaulay et al. (1999); 9—Ballinger et al. (1992); 10—Chen et al. (1995); 11—Torrioni et al. (1994); 12—Wise, Sraml, and Eastaale (1998).

^b It has been suggested that the polymorphism at site 4985 in the data set of Awadalla, Eyre-Walker, and Maynard Smith (1999) is the result of a sequencing error (Kivisilid and Vilems 2000).

^c Only the synonymous variants from the data set of Ingman et al. (2000) were used.

* $P < 0.10$.

** $P < 0.05$.

*** $P < 0.01$.

division will also affect the efficiency of r^2 by introducing LD which is unassociated with distance. In contrast, hypervariable sites may introduce a source of noise into the distribution of the four-genotype PWCs, which does not greatly affect r^2 . We ran our simulations at a mutation rate at which almost all the mutations occurred at sites which were monomorphic (i.e., the simulations conformed to the infinite-sites assumption); therefore, almost all of the PWCs with four genotypes were produced by recombination. However, if some of them were generated by repeated mutation, they would be distributed at random with respect to the distance between sites, and this would act as noise. This may affect the ability of $|D'|$ to detect recombination, and it may explain the differences between r^2 and $|D'|$ in hominid data sets, particularly since it seems likely that some of the four-genotype PWCs are produced by multiple mutation, and not recombination, in the hominid data. The rate of change in mtDNA is such that some four-genotype PWCs are produced by multiple mutation, even if we assume that all sites are equally likely to change (Eyre-Walker, Smith, and Maynard Smith 1999a, 1999b); since Stoneking (2000) has shown that there is variation in the rate of change across sites, in the control region at least, a substantial fraction of the four-genotype pairwise comparisons could be due to multiple mutations.

Resolving the different results obtained with r^2 and $|D'|$ in hominid mitochondria will require further work, but it seems likely that they are associated with the different information that the two statistics employ.

Acknowledgments

We thank John Maynard Smith, Philip Awadalla, Eddie Holmes, and three anonymous referees for helpful discussion and comments. A.E.-W. was funded by the Royal Society.

LITERATURE CITED

- AWADALLA, P., A. EYRE-WALKER, and J. MAYNARD SMITH. 1999. Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science* **286**:2524–2525.
- . 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**:1931a.
- BALLINGER, S. W., T. G. SCHURR, A. TORRONI, Y. Y. GAN, J. A. HODGE, K. HASSAN, K.-H. CHEN, and D. C. WALLACE. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient Mongoloid migrations. *Genetics* **130**:139–152.
- BIRKY, C. W. J. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. USA* **92**:11331–11338.
- CHEN, Y. S., A. TORRONI, L. EXCOFFIER, A. S. SANTACHIARA-BENERECETTI, and D. C. WALLACE. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* **57**:133–149.
- EYRE-WALKER, A. 2000. Do mitochondria recombine in humans? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**:1573–1580.
- EYRE-WALKER, A., N. H. SMITH, and J. MAYNARD SMITH. 1999a. How clonal are human mitochondria? *Proc. R. Soc. Lond. B Biol. Sci.* **266**:477–483.
- . 1999b. Reply to Macaulay et al.: mtDNA recombination—reasons to panic. *Proc. R. Soc. Lond. B Biol. Sci.* **266**:2041–2042.
- HAGELBERG, E., N. GOLDMAN, P. LIO, S. WHELAN, W. SCHIEFENHOVEL, J. B. CLEGG, and D. K. BOWDEN. 1999. Evidence for mitochondrial recombination in a human population of island Melanesia. *Proc. R. Soc. Lond. B Biol. Sci.* **266**:485–492.
- HASEGAWA, M., A. DI RIENZO, T. D. KOCHER, and A. C. WILSON. 1993. Toward a more accurate time scale for the human mitochondrial DNA tree. *J. Mol. Evol.* **37**:347–354.
- HILL, W. G., and A. ROBERTSON. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* **38**:226–231.
- HOFMANN, S., M. JAKSCH, R. BEZOLD, S. MERTENS, S. AHOLT, A. PAPROTTA, and K. D. GERBITZ. 1997. Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlations with D loop variants and association with disease. *Hum. Mol. Genet.* **6**:1835–1846.
- INGMAN, M., H. KAESSMAN, S. PAABO, and U. GYLLENSTEN. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**:708–713.
- JORDE, L. B., and M. BAMSHAD. 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**:1931a.
- KIVISILID, T., and R. VILLEMS. 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**:1931a.
- KUMAR, S., P. HEDRICK, T. DOWLING, and M. STONEKING. 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**:1931a.
- LEWONTIN, R. C. 1964. The interaction of selection and linkage. I. Genetic considerations; heterotic models. *Genetics* **49**:49–67.
- MACAULAY, V., M. RICHARDS, E. HICKEY, E. VEGA, F. CRUCIANI, V. GUIDA, R. SCOZZARI, B. BONNE-TAMIR, B. SYKES, and A. TORRONI. 1999. The emerging tree of west Eurasian mtDNAs: a synthesis of control region and RFLPs. *Am. J. Hum. Genet.* **64**:232–249.
- MACAULAY, V., M. RICHARDS, and B. SYKES. 1999. mtDNA recombination—no need to panic. *Proc. R. Soc. Lond. B Biol. Sci.* **266**:2037–2040.
- PARSONS, T. J., and J. A. IRWIN. 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**:1931a.
- STONEKING, M. 2000. Hypervariable sites in the mtDNA control region are mutational hotspots. *Am. J. Hum. Genet.* **67**:1029–1032.
- TORRONI, A., K. HUOPONEN, P. FRANCALACCI, M. PETROZZI, L. MORELLI, R. SCOZZARI, D. OBINU, M. L. SAVONTAUS, and D. C. WALLACE. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**:1835–1850.
- TORRONI, A., M. T. LOTT, M. F. CABELL, Y. CHEN, L. LAVERGNE, and D. C. WALLACE. 1992a. mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-Loop region. *Am. J. Hum. Genet.* **55**:760–776.
- TORRONI, A., J. A. MILLER, L. G. MOORE, S. ZAMUDIO, J. ZHUANG, T. DROMA, and D. C. WALLACE. 1994. Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am. J. Phys. Anthropol.* **93**:189–199.
- TORRONI, A., T. G. SCHURR, C.-C. YANG et al. (11 co-authors). 1992b. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were

- founded by two independent migrations. *Genetics* **130**:153–162.
- TORRONI, A., R. I. SUKERNIK, T. G. SCHURR, Y. B. STARIKORSKAYA, M. F. CABELL, M. H. CRAWFORD, A. G. COMUZZIE, and D. C. WALLACE. 1993. mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* **53**:591–608.
- 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.
- WAKELEY, J. 1993. Substitution rate heterogeneity variation among sites in hypervariable region 1 of human mitochondrial DNA. *J. Mol. Evol.* **37**:613–623.
- WISE, C. A., M. SRAML, and S. EASTEAL. 1998. Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. *Genetics* **148**:409–421.

EDWARD HOLMES, reviewing editor

Accepted July 16, 2001