Am. J. Hum. Genet. 68:145–153, 2001

Analysis of European mtDNAs for Recombination

J. L. Elson,[1,3] R. M. Andrews,[1,2] P. F. Chinnery,[1] R. N. Lightowlers,[1] D. M. Turnbull,[1,3] and Neil Howell[3,4]

[1]Departments of Neurology and [2]Ophthalmology, The Medical School, and [3]MRC Development Centre for Clinical Brain Ageing, University of Newcastle upon Tyne, Newcastle upon Tyne, and Departments of [3]Radiation Oncology and [4]Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston

The standard paradigm postulates that the human mitochondrial genome (mtDNA) is strictly maternally inherited and that, consequently, mtDNA lineages are clonal. As a result of mtDNA clonality, phylogenetic and population genetic analyses should therefore be free of the complexities imposed by biparental recombination. The use of mtDNA in analyses of human molecular evolution is contingent, in fact, on clonality, which is also a condition that is critical both for forensic studies and for understanding the transmission of pathogenic mtDNA mutations within families. This paradigm, however, has been challenged recently by Eyre-Walker and colleagues. Using two different tests, they have concluded that recombination has contributed to the distribution of mtDNA polymorphisms within the human population. We have assembled a database that comprises the complete sequences of 64 European and 2 African mtDNAs. When this set of sequences was analyzed using any of three measures of linkage disequilibrium, one of the tests of Eyre-Walker and colleagues, there was no evidence for mtDNA recombination. When their test for excess homoplasies was applied to our set of sequences, only a slight excess of homoplasies was observed. We discuss possible reasons that our results differ from those of Eyre-Walker and colleagues. When we take the various results together, our conclusion is that mtDNA recombination has not been sufficiently frequent during human evolution to overturn the standard paradigm.

Introduction

The human mitochondrial genome (mtDNA) has been an important tool in the study of evolution and population genetics. One reason for the widespread use of human mtDNA for analyses of molecular evolution is that its overall rate of divergence is high (particularly within the noncoding control region or D-loop) relative to nuclear genes, thereby providing sufficient signal for phylogenetic analysis, even within short spans of sequence. A second reason is that human mtDNA is inherited maternally, so that mitochondrial lineages are clonal or uniparental, and, thus, its evolutionary history can be reconstructed without the complexity added by biparental recombination (reviewed by Howell 1997, 1999).

This "standard paradigm," however, is a continuing focus of challenge and revision. For example, there is considerable debate over the rate of human mtDNA divergence and whether one can assume a simple universal clock that can be used to accurately time the major events of human evolution and geographical dispersal (e.g., Howell et al. 1996; Parsons et al. 1997; Sigurðardottir et al. 2000). Another persistent question (Howell 1997 and references therein) is whether human mtDNAs have undergone biparental, intermolecular recombination. Eyre-Walker and colleagues have proposed recently that human mtDNA lineages are not clonal and that there has been sufficient recombination to cast doubt on this standard model of human mtDNA evolution (Awadalla et al. 1999, 2000; Eyre-Walker et al. 1999*a*, 1999*b*). If there were excessive complacency with the standard paradigm, this is no longer the case (Strauss 1999; see also the use of the word "panic" in the titles of the reports by Macauley et al. [1999*b*] and by Eyre-Walker et al. [1999*b*]).

A number of investigators have now challenged the conclusions of Eyre-Walker and colleagues (Merriwether and Kaestle 1999; Jorde and Bamshad 2000 ; Kivisild and Villems 2000; Kumar et al. 2000; Parsons and Irwin 2000; Stoneking 2000), and these challenges have been rebutted (Awadalla et al. 2000). Independent evidence for mtDNA recombination in a Melanesian population has been reported, but the data were flawed and the authors have corrected their original conclusion (Hagelberg et al. 1999; 2000). This issue is thus unresolved at the present time, and it is in need of further analysis and critical evaluation.

Received September 20, 2000; accepted for publication November 9, 2000; electronically published December 11, 2000.

Address for correspondence and reprints: Dr. Neil Howell, Biology Division 0656, Department of Radiation Oncology, The University of Texas Medical Branch, Galveston, TX 77555-0656. E-mail: nhowell@utmb.edu

0002-9297/2001/6801-0014$02.00

Two different tests were applied by Eyre-Walker and colleagues to support their contention of human mtDNA recombination. The first test involved phylogenetic analyses of excess homoplasies (parallel mutations at the same nucleotide site) at phenotypically silent sites among a set of human mtDNA coding region sequences (Eyre-Walker et al. 1999*a,* 1999*b*). They concluded that the frequency of homoplasies was much higher than expected on the basis of a single rate of synonymous mutations, and that these "excess" homoplasies were generated by recombination. The second approach of Eyre-Walker and colleagues, which is based on population genetic theory, involves analysis of linkage disequilibrium (LD) among pairs of mtDNA polymorphisms as a function of the distance between sites. They observed a statistically significant inverse correlation between distance and LD, which they concluded was due to mtDNA recombination (Awadalla et al. 1999).

We have determined the complete mtDNA sequences of 64 European and 2 African subjects, and we are using these sequences—in conjunction with a large set of partial sequences—for a comprehensive analysis of mtDNA mutation, segregation, and fixation within the population. In addition, these sequences constitute an independent resource for testing the hypotheses of Eyre-Walker and colleagues. We have applied both of their recombination tests to this set of mtDNA sequences, but we obtain results that differ from those investigators. Possible reasons for the discrepancies between our results and those of Eyre-Walker and colleagues are discussed.

Experimental Procedures

mtDNA Sequences

We have determined the complete nucleotide sequences of 64 European mtDNAs (predominantly from individuals of British descent) and of 2 African mtDNAs. The latter include an L1 haplogroup mtDNA and HeLa cell mtDNA (R.M.A., P.F.C., R.N.L., D.M.T., and N.H., unpublished data). One of the mtDNAs analyzed was a sample of the original placental DNA preparation used for determination of the Cambridge Reference Sequence (Anderson et al. 1981), and we have been able to identify all of the errors in the CRS (Andrews et al. 1999). This revised Cambridge Reference Sequence (rCRS) is used for the studies reported here. The other Europeans were chosen randomly with respect to their mtDNA haplogroup, and they included predominantly normal controls, glaucoma patients, diabetes patients, and patients with Leber hereditary optic neuropathy (LHON).

European mtDNA sequences separate into well-supported phylogenetic clusters, or haplogroups (e.g., Torroni et al. 1996; Macauley et al. 1999*a*). Our 64 European mtDNA sequences could all be classified using the previously identified haplogroup-specific combinations of coding region polymorphisms (Torroni et al. 1996; Macauley et al. 1999*a*), and they included 32 haplogroup H/V sequences, 10 haplogroup J sequences, 8 haplogroup T sequences, 10 haplogroup U/K sequences, 2 haplogroup I sequences, and 1 sequence each of haplogroups W and X. It is now recognized that the previously identified haplogroups H and V, as well as U and K, respectively, are more accurately depicted as subgroups of single haplogroups (see fig. 2 of Macauley et al. [1999*a*]). Among our haplogroup H/V mtDNAs, 28 were haplogroup H and 4 were haplogroup V (the latter mtDNAs carry haplogroup-specific polymorphisms at sites 4580 and 15904; see Macauley et al. [1999*a*]). The frequencies of the different European mtDNA haplogroups in our collection of sequences agree with previous population surveys of Europe (e.g., Richards et al. 1996).

For the results presented here, we limited our analyses to this set of 66 mtDNA sequences, and we have not incorporated any previously published mtDNA sequences, except for the rCRS (Andrews et al. 1999). This decision was made for three main reasons. First, there have been long-standing problems with the accuracy of some of the published sequences (see remarks by Howell et al. [1992] and Macauley et al. [1999*b*]). These problems include, but also extend beyond, the fact that the original CRS both contained a small number of errors and was a "mosaic" of European, African, and bovine mtDNA sequences (Andrews et al. 1999). Every effort has been made to ensure the accuracy of our mtDNA sequences, including extensive confirmatory DNA sequencing in the two laboratories (Newcastle and Galveston) and the use of both automated and manual sequencing methods for the same DNA sample (Andrews et al. 1999). Second, it is important to use an independent set of mtDNA sequences for these analyses. If intermolecular recombination has made a major contribution to human mtDNA evolution and population genetics, then its "footprint" should be present consistently in all population-based sets of sequences. Finally, Awadalla et al. (1999, 2000) contend that LD analysis of mtDNAs from the same ethnic or population group, such as Europeans, provides a sufficiently large number of sites for meaningful analysis without the complications of population subdivision.

Measures of LD

Three measures of LD have been used in this study, and these are defined below. For all of these measures, pairwise combinations of sites are used (i.e., these are two-locus measures of LD), and analysis is restricted to sites with no more than two alleles. In the nomenclature

used here, we have arbitrarily defined the rCRS allele as "1" and the non-rCRS allele as "2", thereby allowing four possible haplotypes for each site pair: p_{11}, p_{12}, p_{21}, and p_{22} ($p_{1+} = p_{11} + p_{12}$, etc.). In this context, therefore, "haplotype" refers to the genotype of the site pairs (see also the same usage by Devlin and Risch [1995]). As an example, 60 of 66 sequences—including the rCRS—have an A residue at nucleotide position 5147 of the L-strand ($p_{1+} = 0.91$), and 6 have a G ($p_{2+} = 0.09$). At nucleotide site 6776, 60 of 66 sequences—including the rCRS—carry a T residue ($p_{+1} = 0.91$), and 6 carry a C ($p_{+2} = 0.09$). The frequencies of the pairwise combinations were: $p_{11} = 0.82$ ($n = 54$); $p_{12} = 0.09$ ($n = 6$); $p_{21} = 0.09$ ($n = 6$); and $p_{22} = 0.00$ ($n = 0$). That is, we did not observe any mtDNAs that carried the non-rCRS allele (polymorphisms) at both nucleotides 5147 and 6776. In most pairwise combinations, only three of the four possible haplotypes were detected in our set of mtDNA sequences.

Awadalla et al. (1999) used the r^2 measure of LD, which is defined in our nomenclature as follows: $r^2 = (p_{11}p_{22} - p_{12}p_{21})^2/p_{1+}p_{2+}p_{+1}p_{+2}$ (see also Devlin and Risch 1995). The Lewontin D' measure is defined as: $D' = (p_{11}p_{22} - p_{12}p_{21})/\min(p_{1+}p_{+2}, p_{+1}p_{2+})$, when $D > 0$, and $D' = (p_{11}p_{22} - p_{12}p_{21})/\min(p_{1+}p_{+1}, p_{+2}p_{2+})$ when $D < 0$ (Devlin and Risch 1995). The important point about this measure is that the absolute value of D' is 1.0 whenever one or more of the four possible haplotypes is not present (Kumar et al. 2000).

Finally, we use the δ measure of LD expressed in terms of haplotype frequencies (Devlin and Risch 1995): $\delta = (p_{11}/p_{21} - p_{12}/p_{22})/(p_{11}/p_{21} + 1)$. The operational constraint on δ is that the value must be positive. Depending upon the values for the four possible haplotypes, calculation may require redefinition of the haplotypes (e.g., p_{22} becomes the frequency of mtDNAs with the CRS alleles at both sites, etc.). Furthermore, there are occasions when two positive values, which can differ, are obtained depending upon how the allele pairs are defined. Usually, one of the positive values is obtained without reassignment of the allele pairs, and this is the one used in the calculations here, but we have not observed that the overall results are changed when one value rather than the other is used in these situations (data not shown).

After calculation of the LD values for the various site pairs, the correlation coefficient of LD versus distance between sites was determined. The statistical significance of this relationship was then ascertained by calculation of the correlation coefficient after randomization of site positions and replication of this routine 5,000 times. The significance is the proportion of random values that are more negative than the experimental value (the same procedure was used by Awadalla et al. [1999]).

Determination of Homoplasies

Eyre-Walker et al. (1999*a*) calculated the number of synonymous homoplasies in their mtDNA sequences by phylogenetic analysis. Among their 29 mtDNA sequences, there were 32 phylogenetically informative sites. The set of >1,500 most-parsimonious trees contained 54 steps, from which a value of 22 homoplasies was obtained. We have used a different approach. The European mtDNA sequences clearly stratify into previously described haplogroups. The resulting phylogenetic tree of these haplogroups is the same as that obtained by others (e.g., Macauley et al. [1999*a*]), although the use of complete coding-region sequences allows the identification of additional haplogroup-specific polymorphisms. Assuming a maximum-parsimony model of evolution, we have detected homoplasies at the following sites:

1. 4736: We detected a polymorphism at this site in 1 of 32 H/V mtDNAs and in 1 of 10 U/K mtDNAs. There is no statistically significant pathway of evolution by which such a mutation could have occurred once at this site.
2. 5426: We detected this polymorphism in both members of a subbranch of haplogroup U mtDNAs and in one of eight haplogroup T mtDNAs.
3. 5471: We detected this polymorphism in one haplogroup H/V mtDNA and 1 of 10 haplogroup J mtDNAs.
4. 6221: This polymorphism was found in the one haplogroup X mtDNA and in HeLa mtDNA; there is no indication that the former arose during evolution from the latter. Analysis of much larger sets of sequences may indicate a single origin of this polymorphism, but we make the conservative assignment of parallel mutational events.
5. 6827: We detected this polymorphism in one African mtDNA and one haplogroup H/V mtDNA.
6. 8251: We detected this polymorphism in both haplogroup I mtDNAs and in the haplogroup W mtDNA. Again, this is a conservative assignment, pending analysis of larger sequence sets.
7. 9947: We detected this polymorphism in one haplogroup T mtDNA and one haplogroup I mtDNA.
8. 10685: We detected this polymorphism in one haplogroup J mtDNA and one haplogroup H/V mtDNA.
9. 13020: We detected this polymorphism in both members of a sub-branch of haplogroup U and in one haplogroup H/V mtDNA.
10. 14470: We detected this polymorphism in one haplogroup U mtDNA and the one haplogroup X mtDNA.
11. 15775: We detected this polymorphism in one haplogroup H/V mtDNA and in the one haplogroup W mtDNA.

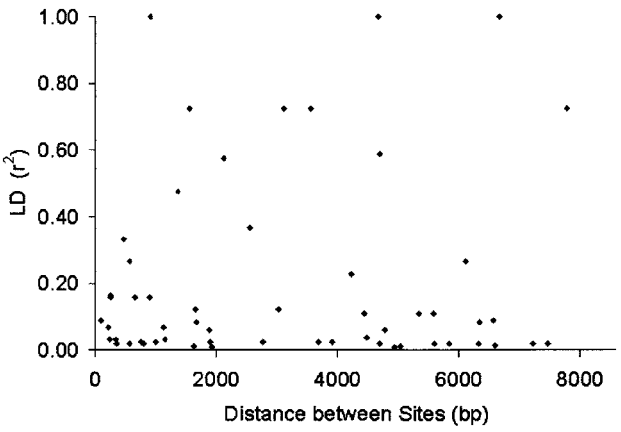

Figure 1 LD as a function of distance for synonymous polymorphisms in the mtDNA coding region. A total of 11 sites were used for analysis, and these included polymorphisms at nucleotides 5147, 6776, 7028, 8697, 11251, 11467, 11719, 11812, 12372, 12612, and 13368. The r^2 estimators of LD were calculated as in Awadalla et al. (1999).

12 and 13. 7028 and 11719: The situation here is more complicated. The non-CRS allele is found at both sites in all non-haplogroup H mtDNAs. The non-CRS allele at site 7028 is also found in all four of our haplogroup V mtDNAs, the "sister" haplogroup to H (Macauley et al. 1999*a*). However, the CRS allele occurs at position 11719 in our haplogroup V mtDNAs. Of our 28 "true" haplogroup H mtDNAs, 1 carries the non-CRS allele at position 11719, and the most likely explanation is homoplasy at this site. Of our 28 haplogroup H mtDNAs, 3 carry the non-CRS allele at position 7028. It is possible that there is an "early" branching within haplogroup H and that this polymorphism has arisen only once within the human population, but we have made the conservative assignment of homoplasy at this site. No other sequence changes in our mtDNAs are candidates for homoplasy.

Results

Analysis of mtDNA LD as a Function of Distance

In our set of 66 complete mtDNAs, there were 11 polymorphic sites that involved synonymous substitutions within the coding region and for which the minority allele or variant was present in ⩾10% of the sequences used for analysis with the r^2 measure of LD (fig. 1). There was no decline in LD as a function of distance between the pairs of polymorphic sites, and we obtained a Pearson's correlation coefficient (ρ) of +0.07, which is not statistically different from zero ($P \simeq .68$). We have also analyzed these sequences with two other measures of LD. When we used the Lewontin D' measure, 54 of 55 site pairs gave absolute values of 1.00, because only three of the four possible haplotypes were detected for the pairwise combination of sites (Kumar et al. 2000). Clearly, therefore, there is no correlation between D' and distance (data not shown). We have also used the δ measure of LD, which is the measure most dependent on the recombination fraction and the least sensitive to differences in allele frequencies (Devlin and Risch 1995; Guo 1997). Of our 55 site pairs, 2 cannot be used, because they contain only two of the four possible haplotypes, and positive values of δ cannot be obtained under such conditions. Using the other 53 site pairs, we found that the correlation coefficient of δ versus distance was −.08, which is not statistically different from a value of zero.

Awadalla et al. (1999) analyzed synonymous changes in the mtDNA coding region sequences to avoid any effects of selection, which can also cause LD in genetic systems with recombination. In their mtDNA RFLP data sets, nonsynonymous changes were included, as well as changes in the rRNA genes and in the noncoding control region. Therefore, we have extended the LD analyses of our sequences to include these additional sites within the mitochondrial genome. As the next step in the analysis, we determined r^2 as a function of distance for all 25 coding region polymorphic sites in our sequence set; this total includes the 11 synonymous polymorphisms and 14 nonsynonymous sequence changes (fig. 2). When r^2 was plotted as a function of distance, a ρ value of +0.04 was obtained, which is indistinguishable from a value

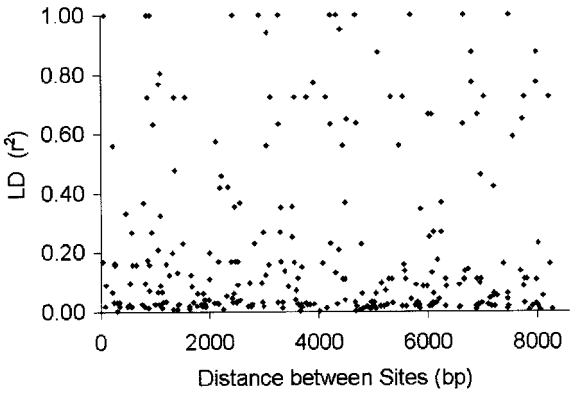

Figure 2 LD as a function of distance for all site changes within the mtDNA coding region. In addition to the 11 synonymous polymorphisms analyzed (see fig. 1), this analysis included the 14 sequence changes at nucleotides 709, 930, 1811, 1888, 2706, 3010, 4216, 10398, 10463, 12308, 13708, 14233, 14766, and 14798. These additional sites included nonsynonymous polymorphisms as well as sequence changes in the rRNA and tRNA genes. The data shown here were obtained using the r^2 measure of LD.

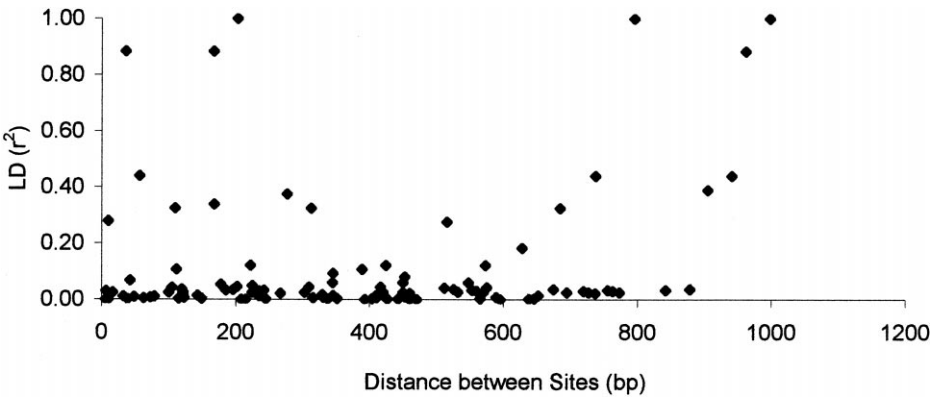

Figure 3 LD as a function of distance for sequence changes in the control region. A total of 15 sites were analyzed, and these included the changes at nucleotides 73, 146, 152, 185, 195, 295, 462, 489, 16069, 16126, 16189, 16294, 16304, 16311, and 16519. The data shown here were obtained using the r^2 measure of LD.

of zero. When only the 14 nonsynonymous changes were analyzed, there was also no decline in LD as a function of distance (data not shown).

Synonymous polymorphisms are more frequent in some mtDNA haplogroups than in others, which reflects their different "ages" or times since the most recent common ancestor (see, e.g., Torroni et al. 1996; Macauley et al. 1999*a*). Furthermore, substantial proportions of such polymorphisms in those analyzed here (8 of 11) were haplogroup specific. As a consequence of these properties, only 1 of the 11 polymorphic sites was specific for haplogroup H, whereas 4 were specific to haplogroup T—although approximately half the sequences (48%) belonged to haplogroup H, whereas haplogroup T accounted for 12%. Although the relative proportions of these haplogroups in our set of sequences are entirely in accord with previous large-scale studies (Richards et al. 1996), we considered the possible effects of sampling on the results of LD analysis. For example, there would be a high probability of obtaining different proportions of the European haplogroups purely as a result of sampling effects. That is, sampling variability in relatively small sample sets could result in the omission of haplogroup-specific polymorphisms if their relative proportions in the sequence set were <10% (the lower limit for inclusion according to the criteria of Awadalla et al. 1999). To investigate this possibility further, we analyzed sequence subsets. It was observed that when haplogroup T mtDNAs were omitted (i.e., mimicking the situation where the proportion of haplogroup T mtDNAs was <10% in a sequence set), the number of silent polymorphisms dropped to 7, and the ρ (r^2 vs. distance) decreased to -0.31, a value that was of borderline statistical significance ($P \simeq .08$). In contrast, omission of haplogroup J mtDNAs—which accounted for only one haplogroup-specific polymorphism in the original group of 11—did not change the results, and the correlation coefficient was $\rho = 0.02$ (data not shown).

We also examined the possibility that recombination occurs in the noncoding control region. Among our set of European and African sequences, there were 15 polymorphic sites that were appropriate for analysis, and the results are shown in figure 3. We observed that, when the r^2 values were plotted against distance, the Pearson's correlation coefficient ρ was $+0.19$. There were an additional six polymorphic sites in which the frequency of the minority allele was >5% but <10%. When these sites were combined with the results for the original 15, the correlation coefficient was $+0.11$. However, when we plotted the δ values against distance, the correlation coefficient was zero (data not shown). Because the r^2 measure of LD is sensitive to both the recombination fraction and the effects of allele frequencies (Devlin and Risch 1995; Guo 1997), it is likely that the positive correlation obtained with this measure was an artifact caused by unequal allele frequencies.

Excess Homoplasy (EH) Test for Recombination

Eyre-Walker et al. (1999*a*, 1999*b*) measured an excess of homoplasic events at silent sites within the mtDNA coding region, which they concluded was caused by recombination. Among our 66 mtDNA sequences, there are 193 sites at which synonymous changes occur in the coding region. If one assumes a single rate of substitution and a total of 3,628 silent sites (Eyre-Walker et al. 1999*a*), one would expect ~5.1 homoplasies. As described previously in "Determination of Homoplasies," we found a total of 13 homoplasies, or an observed/expected ratio of 2.6.

Eyre-Walker et al. (1999*a*) noted that synonymous codon usage in mtDNA is biased, and they compensated for this effect on the number of homoplasies by estimating the effective site number (S_e). The situation is further complicated by the preferential occurrence of transitions relative to transversions, and they assumed, conservatively, that only transitions contribute to homoplasic events, thus obtaining an S_e of 2,269 sites. Using this value, and a total of 193 sites at which synonymous changes occurred in our set of mtDNAs, we estimate that there should be ~8.2 homoplasies and, thus, an observed/expected homoplasy ratio of only 1.6 for our set of mtDNA sequences.

We showed in the previous section that omission of haplogroup T mtDNAs yields a negative correlation between LD and distance. However, omission of these sequences does *not* change the ratio of observed/expected homoplasies. The total number of polymorphic sites in this "minus T haplogroup" sequence subset drops to 171, which yields an expected number of homoplasies of ~4. The observed number of homoplasies was 11 for an observed/expected ratio of 2.7. Use of the S_e correction yields an observed/expected ratio of 1.7.

Discussion

We have used the same tests of mtDNA recombination that have been used by Eyre-Walker and colleagues, but we have applied them to a different, and larger, set of mtDNA sequences. In contrast to their results, analysis of silent polymorphisms in the coding regions of 64 European and 2 African mtDNAs showed no decrease of LD with distance. Furthermore, no evidence of mtDNA recombination was obtained when the analysis was extended both to all coding region sequence changes and to control region polymorphisms. When we used the EH test on our set of sequences, we did find an excess, but it was much lower than that found by Eyre-Walker et al. (1999*a,* 1999*b*). Therefore, there are marked discrepancies between our results and theirs that need to be critically evaluated.

Are the LD and EH Tests Robust and Specific for mtDNA Recombination?

There is debate over whether r^2 is the most appropriate measure of LD (Devlin and Risch 1995; Guo 1997). Jorde and Bamshad (2000) and Kumar et al. (2000) favor Lewontin's D' measure, because it is less sensitive to the confounding effects of unequal allele frequencies, and they find no evidence for mtDNA recombination when this measure is used (see also Merriwether et al. 1991). Awadalla et al. (2000) have made counterarguments to support the use of r^2. The important point is that we obtained no evidence for mtDNA recombination in our set of mtDNA sequences when we used any of three different measures of LD (r^2, D', and δ).

We did observe that the LD test was sensitive to the mtDNA haplogroups that were analyzed, and we obtained a ρ value (-0.31) that was similar to those of Awadalla et al. (1999) when we omitted sequences from one of the haplogroups. However, this sequence subset did not show a change in the ratio of observed/expected number of homoplasies. Their two tests, therefore, are not always congruent. It is possible that the results of Awadalla et al. (1999) may have been biased, because they did not analyze a large enough number of sequences or a large enough set of sites to obtain a sufficiently random sample of mtDNA polymorphic sites (see Conway et al. 1999 and references therein). Furthermore, the inclusion of haplogroup-specific polymorphisms, in the analyses of both our mtDNAs and those of Awadalla et al. (1999), means that a proportion of the LD values are not independent. Weir and Hill (1986) have asserted that the accuracy of the relationship between LD and distance is suspect under such conditions.

There may be additional problems with the LD test. For one thing, the mtDNA sequences should be free of errors (Howell et al. 1992; see also Andrews et al. 1999 and Macauley et al. 1999*b*). One of the polymorphic sites included in the LD analysis of Awadalla et al. (1999) was at nt 4985. However, there is an error at this site in the original CRS (Andrews et al. 1999), and there cannot be substantial polymorphism at this site (see also the same point in Kivisild and Villems 2000). Awadalla et al. (2000) respond that exclusion of this site still yields a negative ρ value (r^2 vs. distance), but it should be noted that the difference between their value and a value of zero is not statistically significant at the 5% level. Kivisild and Villems (2000) also question the occurrence of a polymorphism at nucleotide 6455 in the sequence set of Awadalla et al. (1999). We also do not find polymorphism at this site in our sequences (data not shown). Exclusion of this site again results in a negative, but nonsignificant, ρ value (Awadalla et al. 2000).

Many processes besides recombination affect LD for nuclear genetic markers in humans, including selection and epistasis, random genetic drift, admixture, finite population size, migration, and geographical subpopulations, mutation, coancestry, genetic hitchhiking, and population expansions (Xiong and Guo 1997 and references therein). Slatkin (1994) has demonstrated the marked effect of population growth on LD in the mtDNA control region. Gurven (2000) has shown that LD analysis of mtDNA sequences may allow discrimination between phylogenetically "old" sites and mutational hotspots (see also Stoneking [2000]). Both approaches, however, use LD under the assumption of no mtDNA recombination. Schierup and Hein (2000) have

recently simulated the effects of recombination on phylogenetic analysis and then applied their findings to the control regions from a collection of African mtDNA sequences. Their results are compatible with some level of mtDNA recombination, but—as the authors note—such results may also have been produced by the mtDNA mutation process or by some demographic scenario other than simple exponential growth. It may also be relevant that there is accumulating evidence both that a purely neutral model of human mtDNA evolution is violated (for example, Wise et al. 1998) and that the assumption of site independence is incorrect (for example, Yang 1995; Huelsenbeck and Nielsen 1999; see also the comments in Hey 2000). In summary, therefore, there are several plausible reasons why the decline in LD as a function of distance obtained by Awadalla et al. (1999) is not due to mtDNA recombination.

There are also reasons for being cautious about the EH test for mtDNA recombination. In their initial study, Eyre-Walker et al. (1999*a*) observed 22 homoplasies whereas only 2.2 were expected, an excess of 10-fold. However, their study was flawed by the use of suspect mtDNA sequences (including, but not limited to, failure to correct for the errors in the original Cambridge Reference Sequence). This limitation was pointed out by Macauley et al. (1999*b*), who analyzed a subset of the same mtDNA sequences and found 4 homoplasies where 1.5 were expected (a ratio of 2.7). This excess of homoplasies was not statistically significant. In their second study, Eyre-Walker et al. (1999*b*) analyzed a larger set of mtDNA sequences and obtained an EH ratio of 2.9 (27 observed/9.28 expected), which is statistically significant. For unstated reasons, Eyre-Walker et al. (1999*b*) did not use the effective number of sites (S_e) in their second EH analysis.

The EH test is clearly limited by the accuracy of the sequences and by the accuracy of the phylogenetic analysis that is used to estimate the number of homoplasies. These are not trivial problems, and the sequence set analyzed by Awadalla et al. (1999) is suspect (see the discussion above). Furthermore, in their original set of 29 mtDNA sequences, Eyre-Walker et al. (1999*a;* see their table 3) detected homoplasies at nucleotides 5147, 11251, 11467, 12372, and 12612. However, we did not observe homoplasies at these sites, either in our collection of 64 European and 2 African mtDNAs or in a large set of 300 partial mtDNA sequences (data not shown). These polymorphisms are haplogroup specific, and homoplasy at five such sites in such a small set of mtDNA sequences would not be expected (see also Kumar et al. 2000).

Eyre-Walker et al. (1999*a*) pointed out that one should use the effective number of sites (S_e), rather than the total number of nucleotide sites, to calculate the expected number of homoplasies (see also Maynard Smith 1999). S_e takes into account site variation in mutation rate, and codon bias is assumed to be the major reason for differing probabilities of site change (Eyre-Walker et al. 1999*a;* Maynard Smith 1999; but see also Berg 1999). Eyre-Walker et al. (1999*a*) provide evidence that the excess of homoplasies is not due to hypermutable sites in the human coding region. However, their analyses make assumptions that merit further scrutiny, including the assumption that the same mutational process occurs in all primate groups (see Excoffier and Yang [1999] on site variability in the control region). The slight excess of homoplasies in our mtDNA sequences may reflect some site variation in coding-region divergence rates. At this point, there are substantial uncertainties and complexities that influence the estimation of both the expected and the observed number of homoplasies.

Further Evidence against mtDNA Recombination

Awadalla et al. (1999) suggest that there are two sources of mtDNA sequences that may serve as a substrate for recombination with the maternally transmitted mtDNA: paternal mtDNA and nuclear-inserted mtDNA sequences. The latter possibility is very unlikely, because of the numerous unsuccessful attempts at human mitochondrial transformation with exogenous sequences and because of the lack of evidence for transfer of nuclear DNA to mitochondria during evolution (Thorsness and Fox 1990; Berg and Kurland 2000). Paternal mtDNA is also an unlikely source, and it is worthwhile to note the steps that would be required for paternal mtDNA to undergo recombination. First, paternal mtDNA molecules would have to "evade" the active exclusion/inactivation system that appears to be present in mammals, including humans (reviewed in Howell 1999). Evidence for such a system continues to mount, including the recent finding that there is no paternal mtDNA transmission after intracytoplasmic sperm injection (Danan et al. 1999; see also the recent report of Sutovsky et al. 1999). Second, the paternal and maternal mitochondria must fuse and the mtDNA molecules must then come into sufficient proximity to recombine. This "recombinationally permissive" condition may be very infrequent, because of the sequestration of mtDNA molecules into membrane-bound clusters or nucleoids (Howell 1997). There is recent evidence against the rapid and free exchange of mtDNA molecules within a cell (Enriquez et al. 2000; Jacobs et al. 2000). Third, any recombined mtDNA molecules, present initially as only a small fraction of the total mtDNA pool within the zygote, would then have to undergo clonal expansion to allow transmission to subsequent generations.

In summary, our analyses—which used two tests for mtDNA recombination, those of Eyre-Walker et al.

(1999*a*, 1999*b*) and Awadalla et al. (1999)—yield results that diverge from those of their studies. Our results provide no evidence for evolutionarily significant levels of intermolecular mtDNA recombination in humans. Our results agree with and extend the other studies that have failed to obtain any evidence for human mtDNA recombination (Merriwether et al. 1991; Merriwether and Kaestle 1999; Kivisild and Villems 2000; Kumar et al. 2000; Parsons and Irwin 2000; Stoneking 2000). Therefore, we conclude that there is no compelling reason to overturn the standard paradigm of clonal mtDNA transmission in humans. None of these results or their interpretation, however, should foster complacency about mitochondrial genetics and mtDNA evolution; this is a complex system that is not yet fully understood.

Acknowledgments

This research was supported by a Collaboration Grant from the Wellcome Trust and by grant BCS-9910871 from the National Science Foundation (to N.H.), a Wellcome Trust Programme grant (to R.N.L. and D.M.T.), and a Wellcome Trust Advanced Clinical Training Fellowship (to P.F.C.).

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin JJ, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

Awadalla P, Eyre-Walker A, Maynard Smith J (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286:2524–2525

——— (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931

Jorde LB, Bamshad M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288: 1931

Berg OG (1999) Synonymous nucleotide divergence and saturation: effects of site-specific variations in codon bias and mutation rates. J Mol Evol 48:398–407

Berg OG, Kurland CG (2000) Why mitochondrial genes are most often found in nuclei. Mol Biol Evol 17:951–961

Conway DJ, Roper C, Oduola AMJ, Arnot DE, Kremsner PG, Grobusch MP, Curtis CF, Greenwood BM (1999) High recombination rate in natural populations of *Plasmodium falciparum*. Proc Natl Acad Sci USA 96:4506–4511

Danan C, Sternberg D, Van Steirteghem A, Cazeneuve C, Duquesnoy P, Besmond C, Goossens M, Lissens W, Amselem S (1999) Evaluation of parental mitochondrial inheritance in neonates born after intracytoplasmic sperm injection. Am J Hum Genet 65:463–473

Devlin B, Risch N (1995) A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 29: 311–322

Enriquez JA, Cabezas-Herrera J, Bayona-Bafaluy M, Attardi G (2000) Very rare complementation between mitochondria carrying different mitochondrial DNA mutations points to intrinsic genetic autonomy of the organelles in cultured human cells. J Biol Chem 275:11207–11215

Excoffier L, Yang Z (1999) Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol Biol Evol 16:1357–1368

Eyre-Walker A, Smith NH, Maynard Smith J (1999*a*) How clonal are human mitochondria? Proc R Soc London B 266: 477–483

——— (1999*b*) mtDNA recombination—reasons to panic (reply to Macauley et al.). Proc R Soc London B 266: 2041–2043

Guo S-W (1997) Linkage disequilibrium measures for fine-scale mapping: a comparison. Hum Hered 47:301–314

Gurven M (2000) How can we distinquish between mutational "hot spots" and "old sites" in human mtDNA samples? Hum Biol 72:455–471

Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB, Bowden DK (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Proc R Soc London B 266:485–492

——— (2000) Evidence for mitochondrial DNA recombination in a human population of island Melanesia: correction. Proc R Soc London B 267:1595–1596

Hey J (2000) Human mitochondrial DNA recombination: can it be true? Trends Ecol Evol 15:181–182

Howell N (1997) mtDNA recombination: what do in vitro data mean? Am J Hum Genet 61:19–22

——— (1999) Human mitochondrial diseases: answering questions and questioning answers. Int Rev Cytol 186: 49–116

Howell N, Kubacka I, Mackey DA (1996) How rapidly does the human mitochondrial genome evolve? Am J Hum Genet 59:501–509

Howell N, McCullough DA, Kubacka I, Halvorson S, Mackey D (1992) The sequence of human mtDNA: the question of errors versus polymorphisms. Am J Hum Genet 49: 1333–1337

Huelsenbeck JP, Nielsen R (1999) Effect of nonindependent substitution on phylogenetic accuracy. Syst Biol 48:317–328

Jacobs HT, Lehtinen SK, Spelbrink JN (2000) No sex please, we're mitochondria: a hypothesis on the somatic unit of inheritance of mammalian mtDNA. BioEssays 22:564–572

Kivisild T, Villems R (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931

Kumar S, Hedrick P, Dowling T, Stoneking M (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931

Macauley V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A (1999) The emerging tree of West Eurasian mtDNA: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64: 232–249

Macauley V, Richards M, Sykes B (1999) Mitochondrial DNA

recombination—no need to panic. Proc R Soc London B 266:2037–2039

Maynard Smith J (1999) The detection and measurement of recombination from sequence data. Genetics 153: 1021–1027

Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, Wallace DC (1991) The structure of human mitochondrial DNA variation. J Mol Evol 33:543–555

Merriwether DA, Kaestle FA (1999) Mitochondrial recombination? (Continued). Science 285:837

Parsons TJ, Irwin JA (2000) Questioning evidence for recombination in human mitochondrial DNA. Science 288:1931

Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat Genet 15:363–368

Richards M, Corte-Real H, Forster P, Macauley V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt H-J, Sykes B (1996) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59: 185–203

Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. Genetics 156:879–891

Sigurðardottir S, Helgason A, Gulcher JR, Stefansson K, Donnelly P (2000) The mutation rate in the human mtDNA control region. Am J Hum Genet 66:1599–1609

Slatkin M (1994) Linkage disequilibrium in growing and stable populations. Genetics 137:331–336

Stoneking M (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. Am J Hum Genet 67: 1029–1032

Strauss E (1999) mtDNA shows signs of paternal influence. Science 286:2436

Sutovsky P, Moreno RD, Ramalho-Santos J, Domiko T, Simerly C, Schatten G (1999) Ubiquitin tag for sperm mitochondria. Nature 402:371

Thorsness PE, Fox TD (1990) Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. Nature 346:376–379

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontas M-L, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–1850

Weir BS, Hill WG (1986) Nonuniform recombination with the human β-globin gene cluster. Am J Hum Genet 38:776–778

Wise CA, Sraml M, Eastseal S (1998) Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. Genetics 148: 409–421

Xiong M, Guo S-W (1997) Fine-scale genetic mapping based on linkage disequilibrium: theory and applications. Am J Hum Genet 60:1513–1531

Yang Z (1995) A space-time process model for the evolution of DNA sequences. Genetics 139:993–1005