Heredity (2004) 92, 282–288
© 2004 Nature Publishing Group All rights reserved 0018-067X/04 $25.00
www.nature.com/hdy

A reanalysis of the indirect evidence for recombination in human mitochondrial DNA

G Piganeau and A Eyre-Walker
Center for the Study of Evolution, School of Biological Sciences, Biols, Sussex University, BN1 9QG Brighton, UK

In an attempt to resolve the controversy about whether recombination occurs in human mtDNA, we have analysed three recently published data sets of complete mtDNA sequences along with 10 RFLP data sets. We have analysed the relationship between linkage disequilibrium (LD) and distance between sites under a variety of conditions using two measures of LD, r^2 and $|D'|$. We find that there is a negative correlation between r^2 and distance in the majority of data sets, but no overall trend for $|D'|$. Five out of six mtDNA sequence data sets show an excess of homoplasy, but this could be due to either recombination or hypervariable sites. Two additional recombination detection methods used, *Geneconv* and *Maximum Chi-Square*, showed nonsignificant results. The overall significance of these findings is hard to quantify because of nonindependence, but our results suggest a lack of evidence for recombination in human mtDNA.

Heredity (2004) **92,** 282–288, advance online publication, 28 January 2004; doi:10.1038/sj.hdy.6800413

Keywords: recombination in mtDNA; linkage disequilibrium

Introduction

There has been considerable debate about whether recombination occurs in mitochondrial DNA. Recombination has never been directly observed in human mtDNA but paternal inheritance has been recently observed (Schwartz and Vissing, 2002). Two lines of indirect evidence suggest that recombination might have occurred. The first line of evidence comes from the excess of homoplasies that are observed in phylogenetic trees, that is, identical mutation events occurring independently in different parts of a phylogeny (Eyre-Walker *et al*, 1999). However, this excess of homoplasies could also be due to hypervariable sites and it thus remains unclear whether recombination or heterogeneity in the mutation rate is involved (McVean, 2001; Wiuf, 2001). The second line of evidence for recombination comes from the observation of a negative relationship between linkage disequilibrium (LD) and distance in some human mtDNA data sets (Awadalla *et al*, 1999). However, the analysis of other data sets has not corroborated this observation (Ingman *et al*, 2000; Jorde and Bamshad, 2000; Elson *et al*, 2001; Herrnstadt *et al*, 2002a), and if the relationship is observed, it is only observed when LD is measured with r^2 as opposed to $|D'|$ (Jorde and Bamshad, 2000). This has led to an intense discussion about whether recombination occurs in human mitochondria (Jorde and Bamshad, 2000; Kivisild and Villems, 2000; Kumar *et al*, 2000; Parsons and Irwin, 2000; McVean, 2001; Wiuf, 2001; Innan and Nordborg, 2002). McVean (2001) suggested a method to get consistent result between the two statistics. This is to perform the analysis only on pairs of sites that are informative about recombination. However, the choice of the informative sites requires a prior knowledge of the recombination rate in the data set analysed, which makes the interpretation of the outcome of this test difficult. Further work on methods of detecting recombination from polymorphism data showed that the power for detecting recombination, that is, the probability of detecting recombination when there is recombination, by estimating the number of homoplasies (Posada and Crandall, 2001) or the relationship between LD and distance (Meunier and Eyre-Walker, 2001; Wiuf, 2001) is very low for small rates of recombination. However, simulation (Posada and Crandall, 2001) and experimental data analysis (Posada, 2002) showed that the most powerful methods to detect recombination from polymorphism data are the homoplasy test (Maynard-Smith and Smith, 1998), *Geneconv* (Sawyer, 1999) and *Maximum Chi-Square* (Maynard-Smith, 1992). Unfortunately, the power of the LD-distance test was not assessed along with these three methods.

We decided to extensively reanalyse the evidence of recombination in human mtDNA because different groups have used different data sets and different methods to analyse their data set. For example, Awadalla *et al* (1999) restricted their analysis to synonymous variants segregating at greater than 10%, Herrnstadt *et al* (2002a, b) to variants segregating at greater than 5% whereas Ingman *et al* (2000) included all polymorphisms. Awadalla *et al* (1999) reasoned that restricting the analysis to polymorphisms segregating at higher frequency would increase the probability of detecting recombination by focusing the analysis on older mutations. However, this has never been tested, and theoretical work actually suggests that the ability to detect recombination is independent of the frequency of the alleles included in the analysis, at least in populations

Correspondence: G Piganeau, Center for the Study of Evolution, School of Biological Sciences, Biols, Sussex University, BN1 9QG Brighton, UK.
E-mail: g.v.piganeau@sussex.ac.uk

Received 1 November 2002; accepted 19 September 2003

that are stationary in size (Meunier and Eyre-Walker, unpublished results). Furthermore, several large data sets of complete human mtDNA sequences have recently been published (Ingman *et al*, 2000; Finnila *et al*, 2001; Herrnstadt *et al*, 2002a, b), but these have not been analysed in depth. In this paper we analyse the relationship between LD and distance, using both r^2 and $|D'|$ to measure LD, excluding mutations segregating below several different thresholds. We also run the *homoplasy test* (Maynard-Smith and Smith, 1998), *Geneconv* (Sawyer, 1999) and *Maximum Chi-Square* (Maynard-Smith, 1992).

Methods

Data

We extracted the protein coding sequences from four recently published data sets of complete mtDNA sequences (45 sequences from Awadalla *et al*, 1999; 53 sequences from Ingman *et al*, 2000; 192 sequences from Finnila *et al*, 2001 and 560 sequences from Herrnstadt *et al*, 2002a). Elson *et al* (2001) recently analysed a data set of 64 European sequences and two Africans; this data set was not publicly available for analysis. The sequences compiled by Awadalla *et al* and those published by Ingman *et al* and Herrnstadt *et al* are globally distributed, whereas the sequences published by Finnila *et al* are from Finnish individuals. Since there are many polymorphisms in the Herrnstadt data we split the data set into three population groups (African 56 sequences, Asian 69 sequences and European 435 sequences), which we analysed separately. From the protein coding sequence we extracted biallelic synonymous polymorphisms segregating in codons which did not contain any nonsynonymous polymorphisms. We discarded nonsynonymous segregating sites because selection on these sites may introduce noise into the relationship between LD and distance. If there are epistatic interactions between nonsynonymous mutations, then two mutations may be favoured together, increase substantially in frequency and generate LD. McVean (2001) also suggested that adaptive substitution could lead to a correlation between LD and distance. The details of the data are given in Table 1. The data are available upon request.

We also compiled 10 data sets of human mtDNA restriction fragment length polymorphism (RFLP) (Table 2) from various geographic regions. It is not generally possible to determine whether an RFLP is due to a synonymous and nonsynonymous mutation without additional sequencing—this information was only provided for data set 3. To analyse the relationship between LD and distance, we removed the polymorphic sites contained in the D-loop (500 bp of each site of the replication origin) as a disproportionate number of the polymorphisms are located in the Dloop (Ingman *et al*, 2000); furthermore, the rate of mutation in the D-loop is higher than at synonymous sites (comparison of the proportion of segregating sites in quartet and in the D-loop by a Fisher exact test in the Finnila *et al*, data set, $P<0.001$) and this alone may generate trends in LD with distance (Awadalla *et al*, 1999; Innan and Nordborg, 2002).

Relationship between LD and distance

We estimated LD using two measures, $|D'|$ and r^2 for all pairs of polymorphic sites. For one pair of biallelic loci $A_1|A_2$ and $B_1|B_2$: LD, D and $|D'|$ and r^2 are defined as follows (Lewontin, 1964; Hill and Robertson, 1968):

$$D = f_{A1B1}f_{A2B2} - f_{A1B2}f_{A2B1}$$

$$|D'| = |D|/\mathrm{Min}(f_{A1}f_{B1}, f_{A2}f_{B2})$$

$$r^2 = D^2/(f_{A1}f_{A2}f_{B1}f_{B2})$$

where f_{A1}, f_{A2}, f_{B1} and f_{B2} are the frequencies of the A1, A2, B1 and B2 alleles.

We calculated the correlation coefficient between LD and the distance between polymorphic sites excluding

Table 1 Nucleotide polymorphism data set used in the analysis

Data set	*Population sampled*	*Synonymous polymorphism*
Awadalla *et al* (1999)	45 Eurasians and Africans	200
Ingman *et al* (2000)	53 Eurasians, Asians (17) and Africans (21)	320
Finnila *et al* (2001)	192 Finns	179
Herrnstadt *et al* (2002a)	69 Asians	235
	56 Africans	151
	435 Europeans	506

Table 2 RFLP data sets used in the analysis

Data set	*Population*	*Individuals*	*Haplotypes*	PS
1. Ballinger *et al* (1992)	Southeast Asian	153	115	169
2. Chen *et al* (1995)	Africans	140	79	118
3. Hofmann *et al* (1997)	German	67	67	41
4. Macaulay *et al* (1999)	West Eurasians	95	95	82
5. Torroni *et al* (1994a)	Caucasian from the US	174	117	111
6. Torroni *et al* (1994b)	Tibetans	54	42	61
7a. Torroni *et al* (1992)	Native Americans	167	50	56
7b. Torroni *et al* (1993a)	Native Americans	379	92	109
8. Torroni *et al* (1993b)	Siberians	153	34	47
9. Torroni *et al* (1996)	Finns and Swedes	86	52	84

PS: total polymorphic sites of each data set.
Data set 7a is a subset of data set 7b.

sites at which the rare allele was segregating below a certain cutoff frequency: no cutoff, no singletons, 0.05, 0.10 and 0.20. Previous studies used Pearson's correlation coefficient to assess the significance of the relationship between LD and distance, we also computed Spearman's rank correlation coefficient because the relationship between LD and distance is not necessarily linear (Wiuf, 2001). We then assessed the significance of the relationship by a Mantel test (Sokal and Rohlf, 1995; Awadalla *et al*, 1999). The Mantel test was performed by maintaining the data matrix and randomly permuting the position of sites; for each randomly permuted data set we calculated the correlation between LD and distance to obtain the null distribution of correlation coefficients.

Homoplasy test (Maynard-Smith and Smith, 1998)

We used MEGA (Kumar *et al*, 2001) to estimate the number of homoplasies from the most parsimonious tree for the six nucleotide polymorphism data sets using only the synonymous informative variants. We then estimated the number of expected homoplasies under clonality and tested for an excess of homoplasies using the homoplasy test described by Maynard-Smith and Smith (1998). This test implies the calculation of the effective number of sites. The number of effective sites equals the number of sites (nA, nC, nT, nG) by their probability of mutations (pA, pC, pT, pG). There are 3411 synonymous sites in human mitochondria, 16% are T, 44% C, 36% A and 5% G. Assuming that the rate of transversion is negligible, and that base composition is at equilibrium, the effective number of sites equals 2172 in human mitochondria (4*3411*[0.6*0.26*0.74 + 0.4*0.125*0.875]) (Maynard-Smith and Smith, 1998).

Geneconv (Sawyer, 1999)

The program GENECONV 1.81. (http://www.math.wustl.edu/~sawyer/geneconv/index.html) was employed. The global permutation *P*-values are based on BLAST like global scores (10 000 replicates).

Maximum Chi-Square (Maynard-Smith, 1992)

The computer program MaxChi2 was kindly provided by David Posada, implementing a modification of the Maximum Chi-Square method suggested by Wiuf *et al* (2001). The statistic employed was the Maximum Chi-Square in the original alignment. For each pair of sequences, this statistic was calculated on a sliding window that moved one nucleotide at a time and included only variable sites. The width of the window was arbitrarily set to the total number of variable sites divided by 1.5 (following Posada and Crandall, 2001). The significance of the putative recombination events identified by Maximum Chi-Square was assessed by randomly permuting the positions of sites 1000 times. We used Bonferroni correction to correct for multiple tests.

Results

The relationship between LD and distance

We have investigated the relationship between LD and distance for 16 human mtDNA data sets, six of these are data sets of complete mtDNA sequences while the other 10 are RFLP data sets. We have performed 20 analyses on each nucleotide data set and 16 analyses on each RFLP data set—we have investigated the correlation between LD, as measured by either r^2 or $|D'|$, and the distance between sites using both Pearson's and Spearman's correlation coefficients, including all sites, and excluding sites at which the rare allele is a singleton, or below a frequency of 5, 10 (RFLP data set) and 20% (nucleotide data set).

The results from the analysis of the complete mtDNA data sets are given in Table 3. We restricted our analysis to synonymous polymorphisms to minimise the effects of LD generated by selection. For the data set of

Table 3 Relationships between LD and distance for pairs of synonymous sites for different cutoff frequencies

Cutoff	*No.*	*Awadalla et al (1999)*				*No.*	*Ingman et al (2000)*				*No.*	*Finnila et al (2001)*			
		r^2		$\|D'\|$			r^2		$\|D'\|$			r^2		$\|D'\|$	
		P	S	P	S		P	S	P	S		P	S	P	S
⩾0.2	6	−0.63*	−0.58*	−0.01	−0.03	16	−0.14	−0.13	−0.08	−0.14	5	0.09	0.09	0.63*	0.65*
⩾0.1	14	−0.24	−0.18	0.06	0.1	30	0.07	0.08	0.08	0.05	16	−0.14	−0.22*	−0.12	−0.07
⩾0.05	28	−0.03	−0.02	0.07	0.07*	77	−0.006	−0.01	−0.01	−0.02	30	−0.06	−0.1	0.035	0.04
$>1/n$	49	−0.02	−0.03	0.03	0.03	131	−0.02	−0.01	−0.02	−0.02	111	−0.02	−0.04	0.004	0.01
0	185	−0.01	0.005	0.003	−0.005	307	0.003	0.03?	−0.01	−0.03?	175	−0.03*	−0.03	0.002	0.002
Cutoff	*No.*	*Africans (Herrnstadt et al, 2002a)*				*No.*	Asians (Herrnstadt *et al*, 2002a)				*No.*	*Europeans (Herrnstadt et al, 2002a)*			
		r^2		$\|D'\|$			r^2		$\|D'\|$			r^2		$\|D'\|$	
		P	S	P	S		P	S	P	S		P	S	P	S
⩾0.2	22	−0.04	−0.05	0.09	0.008	15	−0.05	−0.1	0.1	0.12+	4	−0.1	−0.1	−0.41	−0.23
⩾0.1	46	−0.01	−0.02	−0.004	−0.03	22	−0.002	−0.02	0.06	0.07	14	0.04	−0.03	0.02	0.03
⩾0.05	72	−0.004	−0.04	−0.01*	−0.05*	24	−0.006	−0.02	0.02	0.05	21	−0.01	−0.07	−0.06	−0.03
$>1/n$	111	−0.01	−0.01	−0.02	−0.03*	42	−0.02	0.02	0.01	0.01	191	−0.01	−0.01	−0.01	−0.01
0	207	−0.02*	−0.013	−0.002	−0.006	135	−0.003	-10^{-4}	0.001	−0.01	444	−0.005?	−0.01?	−0.004?	−0.004?

*Significant ($P<0.05$) correlation coefficient. No.: number of polymorphic sites; *n*: number of sequences per data set; *P*: Pearson correlation's coefficient; *S*: Spearman's correlation coefficient?; significance could not be assessed through permutations because of nonpractical simulation times.

Awadalla *et al* (1999), the correlation between r^2 and distance is almost always negative and it is significant for polymorphisms segregating at greater than 20%. In contrast, the correlation between $|D'|$ and distance is almost always small and nonsignificant, except for polymorphisms segregating at $>5\%$ when Spearman's correlation coefficient is used, when it is significantly positive. In the data set of Ingman *et al* (2000) there is an overall tendency towards negative correlations but none of the correlations are significant. In the data set of Finnila *et al* (2001), r^2 is generally negatively correlated to distance and often significantly so; however, the $|D'|$ is significantly positively correlated to distance for polymorphisms segregating at $>20\%$. In the three data sets from Herrnstadt *et al* (2002a), the correlations are generally very small, with r^2 being almost always negatively correlated with distance; $|D'|$ is generally negatively correlated for the Africans and Europeans, but positively correlated for the Asians. Few of the correlations are significant.

The 10 RFLP data sets show no consistent trend between LD and distance (Table 4). Data set 5 shows no trend, data sets 2 and 4 show a positive significant correlation between r^2 and distance. Data sets 3, 6, 7a and

Table 4 Relationships between R^2 and $|D'|$ and distance between polymorphic sites (PS) for the 10 data sets (Table 1)

Data set	*Cutoff*	*No.*	R^2 P	R^2 S	$\|D'\|$ P	$\|D'\|$ S
1	0	150	−0.06**	−0.07**	0.1**	0.11**
	1/*n*	71	−0.18**	−0.18**	0.18**	0.21**
	0.05	14	−0.17	−0.15	−0.14	−0.05
	0.1	5	−0.14	−0.19	−0.33	−0.4
2	0	118	0.01	0.02	0.08	0.08
	1/*n*	69	0.01	0.04	−0.13	−0.02
	0.05	26	−0.04	0.1*	−0.005	−0.01
	0.1	11	−0.02	−0.06	0.16	0.09
3	0	21	−0.09	−0.1	0.07	0.05
	1/*n*	19	−0.11	−0.16*	0.09	0.06
	0.05	13	−0.03	0.07	0.06	−0.02
	0.1	7	−0.15	−0.03	0.19	0.14
4	0	82	−0.003	0.04*	0.0002	−0.01
	1/*n*	39	0.005	0.009	0.014	−0.004
	0.05	22	0.03	0.03	0.06	0.03
	0.1	9	0.16	0.12	0.21	0.21
5	0	111	0.006	−0.002	0.006	0.008
	1/*n*	49	0.016	0.02	0.007	0.01
	0.05	12	0.1	0.04	−0.06	−0.02
	0.1	4	−0.1	0.11	0.19	0.56
6	0	50	−0.111	0.01	−0.08**	0.04
	1/*n*	22	−0.29*	−0.21**	−0.3**	−0.26**
	0.05	13	−0.43**	−0.29**	−0.34**	−0.15*
	0.1	8	−0.6**	−0.66**	−0.13	−0.04
7a	0	46	0.01	0.01	−0.01	−0.01
	1/*n*	29	0.09	−0.002	−0.01	−0.007
	0.05	5	−0.8**	−0.62*	−0.03	0.05
	0.1	5	−0.8**	−0.62*	−0.03	0.05
7b	0	85	0.03*	−0.003	−0.03*	−0.03*
	1/*n*	51	0.03	−0.02	−0.05*	−0.05*
	0.05	5	−0.71*	−0.5	0.02	0.05
	0.1	5	−0.71*	−0.5	0.02	0.05
8	0	41	−0.007	−0.05	−0.01	−0.02
	1/*n*	26	−0.03	−0.02	−0.04	−0.003
	0.05	12	−0.18	−0.16	−0.06	−0.02
	0.1	7	−0.5*	−0.33	−0.09	−0.17
9	0	69	−0.02	−0.01	0.01	0.014
	1/*n*	22	−0.09*	−0.02	0.02	0.04
	0.05	8	−0.18	−0.12	0.03	0.07
	0.1	6	−0.4	−0.17	0.24	0.3

*Significant ($P<0.05$).

**Highly significant correlation coefficients.

We restricted the analysis to synonymous sites (data 3) and polymorphic sites not located in the D loop (position from 0 to 500 bp and greater than 16 000 bp were excluded). No.: number of polymorphic sites *S*: Spearman's rank correlation coefficient, *P*: Pearson's correlation coefficient.

P-value were assessed by 1000 permutations of data sets (as in Awadalla *et al*, 1999).

8 largely show negative correlations with several of the correlations being significant or highly significant. In data set 9 the correlation between r^2 and distance is negative for all frequency classes, whereas it is positive for $|D'|$. Surprisingly, there is a highly significant negative correlation between r^2 and distance in data set 1, but a highly significant positive correlation between $|D'|$ and distance, when all mutations are included in the analysis, or only singletons are excluded.

The results of the analyses are summarised in Table 5. Overall, there is an excess of negative correlations for r^2 (51 out of 60) on the synonymous polymorphism where all significant results are for a negative relationship. For the RFLP polymorphism analysed, the proportion of negative correlation between r^2 and distance is less (55 out of 80 test performed) and the significant relationship are positive or negative. The correlation between $|D'|$ and distance is as likely to be negative as positive, and significant correlations are found equally in both directions both for nucleotide and RFLP polymorphism. It is important to appreciate that none of the data sets or analyses are independent so it is difficult to assess the overall significance of these findings: none of the correlations are significant if we correct for multiple comparisons.

The correlation between LD and distance is very similar for both Pearson's and Spearman's correlation coefficient, although the correlations using the latter are usually a little larger, and slightly more significant.

Homoplasy test

Recombination is not only expected to lead to a decline in LD with distance but also to generate homoplasies in phylogenetic trees; that is, instances in which the same mutation, or the reverse mutation, occurs at the same site in different parts of the tree. Homoplasies can be generated by both recombination and multiple mutations. Maynard-Smith and Smith (1998) have devised a method to predict the number of homoplasies due to multiple mutations when there is no recombination, taking into account codon usage bias. They have also devised a test of whether the observed number is greater than the number expected under clonality. We have applied their methods to synonymous variants segregating in each of the mtDNA sequence data sets; we restrict our analysis to DNA sequence data sets since it is only possible to predict the number of homoplasies for synonymous polymorphisms. The results are presented in Table 6. In each case the observed number of homoplasies is higher than the number expected under clonality, and with the exception of the data set of Finnila *et al*, the difference is significant.

Table 5 Comparison of the synonymous SNP and RFLP analyses

	SNP	*RFLP*
Number of test performed	60	80
Negative correlations r^2-distance	51 (85%)	55 (69%)
Significant negative correlations	5	19
Significant positive correlations	0	2
Negative correlations $|D'|$-distance	30 (50%)	38 (47%)
Significant negative correlations	3	9
Significant positive correlations	3	4

Geneconv and Maximum Chi-Square

No recombination was detected by *Geneconv* with any of the six sequence data sets. No recombination event was detected by *Maximum Chi-square* after correction for multiple tests.

Discussion

We have analysed 16 human mtDNA data sets for indirect evidence of recombination using four approaches; we have investigated the correlation between pairwise LD and the distance between sites, the level of homoplasy in phylogenetic trees, and the clustering of substitutions by *Geneconv* and *Maximum Chi-Square*. Overall, there is a tendency towards a negative correlation between LD, when it is measured by r^2, and distance. Of the 140 analyses we have performed, 104 showed a negative correlation. Furthermore, the significant correlations were all negative for the nucleotide polymorphism. Unfortunately, because many of the tests are nonindependent, it is not possible to assess the overall significance of these results. In contrast, approximately half the analyses showed a positive correlation between LD, when it was measured using $|D'|$, and distance, and the significant correlations were both positive and negative. The analysis of simulated data sets under simple evolutionary models (constant population size, no selection) suggests that there should not be any conflict between the two statistics (Meunier and Eyre-Walker, 2001). Interestingly, McVean *et al* (2002) showed that the r^2 statitistics is more powerful than the $|D'|$ statitistics under the finite sites model. As the finite site model describes well, the high rate of polymorphism on synonymous sites in human mtDNA, the r^2 statistics should be favoured.

In contrast to the rather confused picture offered by the analysis of LD, there is a clear excess of homoplasy in five out of six data sets for which this analysis was performed. This excess of homoplasy could be due to

Table 6 Homoplasy test for the synonymous nucleotide polymorphism data sets

	Polymorphic sites	*Informative sites*	*Observed homoplasies*	*Expected homoplasies*	*P-value*
Awadalla *et al* (1999)	200	49	27	9.3	0.0001
Finnila *et al* (2001)	179	115	10	7.8	0.18
Ingman *et al* (2001)	320	131	38	24	0.0052
Africans Herrnstadt *et al* (2002a)	235	124	20	13.6	0.04
Asians Herrnstadt *et al* (2002a)	151	45	11	5.5	0.01
Europeans Herrnstadt *et al* (2002a)	506	213	110	70	0.0001

Table 7 Evidence for recombination from the synonymous polymorphism data (no frequency cutoff) from the five recombination detection methods used

Data set	R^2	$\vert D'\vert$	*Homoplasy*	*Geneconv*	*Max Chi-square*
Awadalla *et al* (1999)	NS	NS	**	NS	NS
Ingman *et al* (2000)	NS	NS	**	NS	NS
Finnila *et al* (2001)	*	NS	NS	NS	NS
Herrnstadt *et al* (2002a) Africans	*	NS	*	NS	NS
Herrnstadt *et al* (2002a) Asians	NS	NS	*	NS	NS
Herrnstadt *et al* (2002a) Europeans	ND	ND	**	NS	NS

*For $P<0.05$.
**For $P<0.01$.
ND: no data.
NS: nonsignificant.

recombination or it could be due to hypervariable sites. Stoneking (2000) has recently demonstrated that hypervariable sites exist in the mtDNA control region (see also additional analysis by Eyre-Walker and Awadalla, 2001). However, the control region has unique mutational dynamics, including a four-fold higher mutation rate than on synonymous positions so it is not clear whether hypervariable sites exist in the coding region of the mtDNA. Eyre-Walker *et al* (1999) performed a number of analyses that were aimed at testing whether hypervariable sites exist in the coding region; they found no evidence, but none of their analyses were particularly powerful.

The outcome of the five recombination detection methods used is summarised in Table 7. If we exclude the homoplasy test because it generates a high rate of false positives if there is mutation rate heterogeneity (Posada and Crandall, 2001), no complete sequence data set shows evidence for the recombination from two tests. Empirical data analysis suggests that one should not rely on a single test to infer the presence of recombination (Posada, 2002). Therefore, there is a lack of evidence for recombination in human mtDNA from our analysis, although two out of six sequence data sets show evidence for recombination in one method (excluding the Homoplasy test). This indicates either no recombination or very rare recombination events. As paternal inheritance of mtDNA in humans has recently been observed (Schwartz and Vissing, 2002) recombination in human mtDNA is a real possibility. However, our study shows that it is hardly detectable from sequence data with available recombination detection methods.

Acknowledgements

We are very thankful to Yohan Gourbiere for his contribution to the analysis. We are grateful to Kari Majamaa for providing nucleotide polymorphism data sets and to David Posada for his C code of the homoplasy test and MaxChi2. GP is funded by a FEBS long-term fellowship and AE-W by the Royal Society and the BBSRC.

References

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

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K *et al* (1992). Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* **130**: 139–152.

Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (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.

Elson JL, Andrews RM, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (2001). Analysis of European mtDNAs for recombination. *Am J Hum Genet* **68**: 145–153.

Eyre-Walker A, Awadalla P (2001). Does human mtDNA recombine? *J Mol Evol* **53**: 430–435.

Eyre-Walker A, Smith NH, Smith JM (1999). How clonal are human mitochondria? *Proc Roy Soc London B* **266**: 477–483.

Finnila S, Lehtonen MS, Majamaa K (2001). Phylogenetic network for European mtDNA. *Am J Hum Genet* **68**: 1475–1484.

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C *et al* (2002a). Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* **70**: 1152–1171.

Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C *et al* (2002b). Erratum. *Am J Hum Genet* **71**: 448.

Hill WG, Robertson A (1968). The effects of inbreeding at loci with heterozygote advantage. *Genetics* **60**: 615–628.

Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, Paprotta A *et al* (1997). Population genetics and disease susceptibility characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. *Hum Mol Genet* **6**: 1835–1846.

Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* **408**: 708–713.

Innan H, Nordborg M (2002). Recombination or mutational hot spots in human mtDNA? *Mol Biol Evol* **19**: 1122–1127.

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

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.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**: 1244–1245.

Lewontin RC (1964). The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**: 49–67.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V *et al* (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**: 232–249.

Maynard-Smith J (1992). Analyzing the mosaic structure of genes. *J Mol Evol* **34**: 126–129.

Maynard-Smith J, Smith N (1998). Detecting recombination from gene trees. *Mol Biol Evol* **15**: 590–599.

McVean G, Awadalla P, Fearnhead P (2002). A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**: 1231–1241.

McVean GA (2001). What do patterns of genetic variability reveal about mitochondrial recombination? *Heredity* **87**: 613–620.

Meunier J, Eyre-Walker A (2001). The correlation between linkage disequilibrium and distance: implications for recombination in hominid mitochondria. *Mol Biol Evol* **18**: 2132–2135.

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

Posada D (2002). Evaluation of methods for dctecting recombination from DNA Sequences: empirical data. *Mol Biol Evol* **19**: 708–717.

Posada D, Crandall KA (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA* **98**: 13757–13762.

Sawyer SA (1999). GENECONV: a computer package for statistical detection of gene conversion. Distributed by the author available at, http://www.math.wustl.edu/ ~ sawyer.

Schwartz M, Vissing J (2002). Paternal inheritance of mitochondrial DNA. *N Engl J Med* **347**: 576–580.

Sokal J, Rohlf R (1995). Biometry. 3rd edn. WH Freeman and Company: New York. pp 887.

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

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R *et al* (1996). Classification of European mtDNAs from an analysis of three European populations. *Genetics* **144**: 1835–1850.

Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC (1994a). 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, Miller JA, Moore LG, Zamudio S, Zhuang J, Droma T *et al* (1994b). 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, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M *et al* (1993a). Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* **53**: 563–590.

Torroni A, Schurr TG, Yang CC, Szathmary EJ, Williams RC, Schanfield MS *et al* (1992). 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, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH *et al* (1993b). mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am J Hum Genet* **53**: 591–608.

Wiuf C (2001). Recombination in human mitochondrial DNA? *Genetics* **159**: 749–756.

Wiuf C, Christensen T, Hein J (2001). A simulation study of the reliability of recombination detection methods. *Mol Biol Evol* **18**: 1929–1939.