

# A Broad Survey of Recombination in Animal Mitochondria

Gwenaël Piganeau,<sup>1</sup> Michael Gardner,<sup>2</sup> and Adam Eyre-Walker

Centre for the Study of Evolution, School of Life Sciences, University of Sussex, Brighton, UK

Recombination in mitochondrial DNA (mtDNA) remains a controversial topic. Here we present a survey of 279 animal mtDNA data sets, of which 12 were from asexual species. Using four separate tests, we show that there is widespread evidence of recombination; for one test as many as 14.2% of the data sets reject a model of clonal inheritance and in several data sets, including primates, the recombinants can be identified visually. We show that none of the tests give significant results for obligate clonal species (apomictic parthogens) and that the sexual species show significantly greater evidence of recombination than asexual species. For some data sets, such as *Macaca nemestrina*, additional data sets suggest that the recombinants are not artifacts. For others, it cannot be determined whether the recombinants are real or produced by laboratory error. Either way, the results have important implications for how mtDNA is sequenced and used.

## Introduction

In animals, it is generally thought that mitochondrial DNA (mtDNA) is inherited from one parent, usually the mother (Birky 1995; Avise 2000), and that its inheritance is therefore clonal. However, paternal inheritance of mtDNA, that is, transmission of the paternal mtDNA into the egg and its survival in the adult organism, has been demonstrated in a number of species (Kondo et al. 1990; Gyllensten et al. 1991; Magoulas and Zouros 1993; Skibinski, Gallagher, and Beynon 1994; Zouros et al. 1994; Kvist et al. 2003), including our own (Schwartz and Vissing 2002). Furthermore, there is evidence of recombination in variety of taxa (Ladoukakis and Zouros 2001a, 2001b; Hoarau et al. 2002; Maynard Smith and Smith 2002; Andolfatto, Scriber, and Charlesworth 2003). The strength of this evidence varies considerably. The most convincing evidence of recombination comes from a human individual who has been shown to have both paternal and maternal mtDNA in their muscle tissue, along with a variety of recombinants (Kraytsberg et al. 2004). There is also direct evidence of recombination in mussels, where it has been possible to sequence parental and recombinant mtDNA molecules (Ladoukakis and Zouros 2001a). However, mussels have an unusual mode of inheritance of mitochondrial DNA, which makes them particularly prone to recombination.

To investigate the question of whether recombination occurs further, and if it does, how prevalent it is, we compiled 267 data sets of protein-coding sequences from mtDNA, in which multiple individuals from a species had been sequenced. We subjected these data sets to four tests of recombination: (1) the *maxchi* test of Maynard Smith (1992), (2) the *geneconv* test of Sawyer (1989) (<http://www.math.wustl.edu/~sawyer/geneconv/>), (3) the correlation between  $r^2$  (Hill and Robertson 1968), a measure of linkage disequilibrium, and distance (henceforth, the *LD $r^2$*  test), and (4) the correlation between  $|D'|$  (Lewontin 1964), another measure of linkage disequilibrium, and distance (the

*LD $|D'|$*  test). We chose to use four tests, instead of one, to investigate whether any evidence of recombination was method dependent. The tests we chose represent different approaches to the detection of recombination and are among the most efficient methods available for low recombination rates (Posada and Crandall 2001).

## Material and Methods

### Data

Data were compiled by searching GenBank with ACNUC (Gouy et al. 1984) for animal species in which multiple alleles of a mitochondrial gene had been sequenced. We restricted ourselves to one data set per species. The sequences were trimmed of any non-protein-coding sequence because the LD-distance tests,  $r^2$  and  $|D'|$ , can be sensitive to variation in the mutation rate (Innan and Nordborg 2002). Data sets with fewer than 10 polymorphisms were removed. We then scanned the data for evidence of nuclear-encoded copies of the mtDNA (*numts*). First, all sequences were translated, and any data set or sequence with either of the universal stop codons (TAA and TAG) was removed. Second, we removed any data set in which the number of nonsynonymous polymorphisms exceeded the number of synonymous polymorphisms, without correcting for the number of sites. We also noticed, in an initial survey of the data, cases in which putative recombinants were generated by two frame-shift mutations, usually a 1-bp insertion, followed several nucleotides later, by a deletion—this generated a series of polymorphisms that tended to give significant results under *maxchi* and the  $r^2$  tests. These do not appear to be recombinants, so they were removed. After these procedures, we were left with 267 data sets across a broad range of animal species: 156 vertebrates, 57 arthropods, 29 mollusks, 12 nematodes, 11 echinoderms, and two annelids. We did not include human sequences in the data set because these have been subjected to exhaustive analysis by us and others (Ingman et al. 2000; Jorde and Bamshad 2000; Elson et al. 2001; McVean, Awadalla, and Fearnhead 2002; Piganeau and Eyre-Walker 2004).

### Tests of Recombination

The data were subjected to four tests of recombination. (1) In the *maxchi* test of Maynard Smith (1992), we follow Maynard Smith's original implementation of his test, rather than the recent version suggested by Posada and Crandall (2001). In their method, they only survey the middle third of

<sup>1</sup> Present address: UMR 7628, Laboratoire Arago, 66651 Banyuls sur mer, France

<sup>2</sup> Present address: School of Biological Sciences, University of Bristol, Bristol, UK

Key words: clonal inheritance, *Macaca nemestrina*, mtDNA, recombination.

E-mail: gwenael.piganeau@obs-banyuls.fr.

*Mol. Biol. Evol.* 21(12):2319–2325. 2004

doi:10.1093/molbev/msh244

Advance Access publication September 1, 2004

**Table 1**  
**Results from Four Tests of Recombination on 267 Data Sets of Animal mtDNA**

	<i>maxchi</i>	<i>geneconv</i>	$LDr^2$	$LD D'$
All sexual species				
Number of data sets	267	267	267	227
% Significant at 5%	11.9%***	10.4%***	14.2%***	8.4%*
Combined probability	< 0.0001	< 0.0001	< 0.0001	0.003
Synonymous sexual species				
Number of data sets	249	249	249	212
% Significant at 5%	10.0%***	7.2%	9.2%***	7.5%
Combined probability	<0.0001	0.003	<0.0001	0.048
All Asexual species				
Number of data sets	12	12	12	5
% Significant at 5%	0.0%	0.0%	0.0%	0.0%
Combined probability	0.79	0.90	0.71	0.62
Probability of sexual vs. asexual	0.023	0.023	0.045	0.318

NOTE.—Numbers of data sets are lower for  $LD|D'$  because we excluded data sets in which  $|D'| = 1$  for all pairwise comparisons (see *Materials and Methods*). \*  $P < 0.05$ , \*\*\*  $P < 0.001$ .

the alignment for evidence of recombination. We survey the entire region but prevent anomalous behavior of the statistic by requiring that all expected values in the chi-square are at least 2. This requirement prevents large chi-square values being produced when the expected values are very small. This test appears to be considerably more powerful than the implementation of Posada and Crandall (2001). Simulations, however, suggest that the type I error is not affected. Data sets in which all chi-square tests, in the original data, fail because of small expected values were excluded from further analysis. (2) In Sawyer's *geneconv* test (Sawyer 1989), we considered global  $P$  values, which automatically correct for multiple comparisons, for internal fragments. We used the default parameters because there has been no systematic analysis of the behavior of the *geneconv* test. (3) We considered the relationship between the measure of linkage disequilibrium,  $r^2$  (Hill and Robertson 1968), and distance between sites, with the significance assessed by a Mantel test (the  $LDr^2$  test). (4) We considered the relationship between the measure of linkage disequilibrium,  $|D'|$  (Lewontin 1964), and distance between sites, with the significance assessed by a Mantel test (the  $LD|D'|$  test). A data set was excluded from further analysis with the  $LD|D'|$  if  $|D'| = 1$  for all pairwise comparisons, including these data sets would make the test conservative because all randomized data sets have the same correlation as the original data set (i.e., a correlation of 0). One thousand randomizations were performed for each test to assess significance; however, if none of the randomized data sets exceeded the observed value, we repeated the analysis with 10,000 and then 100,000 randomizations. If the test still yielded no randomized data sets that exceeded the observed value, we set the probability value to 0.00001 in all further analyses. These methods are available at [www.lifesci.sussex.ac.uk/CSE/test](http://www.lifesci.sussex.ac.uk/CSE/test).

#### Statistical Analysis

To combine results across data sets we used Fisher's method of combining probabilities by summing  $-2\ln(p)$

across data sets. Because  $-2\ln(p)$  is chi-square distributed with 2 degrees of freedom, the sum is chi-square distributed with  $2n$  degrees of freedom, where  $n$  is the number of data sets.

To test whether there was more evidence of recombination in sexual species than in asexual species we calculated the combined probability value for each data set for each test (i.e., by summing  $-2\ln(p)$ , where  $p$  is the probability value). We divided the value by the number of data sets to obtain the average and considered the difference between the sexual species and the asexual species as our test statistic  $Z = -2\{(\sum_{i=1}^{n_{sex}} \ln(p_i)/n_{sex}) - (\sum_{i=1}^{n_{asex}} \ln(p_i)/n_{asex})\}$  where  $n_{sex}$  and  $n_{asex}$  are the numbers of sexual and asexual data sets. To find the distribution of this statistic under the null hypothesis, that there is no difference between sexual and asexual species, we pooled the probability values from the sexual and asexual data sets and then randomly selected, without replacement,  $n_{sex}$  and  $n_{asex}$  data sets and recalculated  $Z$ . We repeated this procedure 1,000 times.

ANOVA analyses were performed on the probability values from each test that were transformed as  $\text{Sqrt}(\ln(p))$ , which was found to be approximately normally distributed. Kruskal-Wallis tests were also performed.

#### Results

We compiled 267 animal mtDNA data sets and subjected them to four tests of recombination. All four tests of recombination are highly significant if we combine probabilities across data sets (table 1); furthermore, the proportion of tests that are significant at the 5% level is significantly greater than expected for all individual tests. Note that the majority of data sets come from different genera, so there is little chance of taxa sharing polymorphisms and, therefore, being nonindependent. The proportion of tests that are significant at the 5% level varies from just over 8.4% for  $LD|D'|$  up to 14.2% for  $LDr^2$ . The results remain qualitatively unchanged if we restrict the analysis to synonymous polymorphisms (table 1): all four recombination detection methods reject the null hypothesis when combining probabilities.  $LDr^2$  and *maxchi* still show a significant excess of tests that are significant at the 5% level, whereas  $LD|D'|$  and *geneconv* are marginally significant ( $P = 0.06$ ); this is not surprising because removing nonsynonymous polymorphisms reduces the number of polymorphisms analyzed by about one third.

Although all four tests of recombination are expected to be fairly robust, in that they make few apparent assumptions (see *Discussion*), there may be factors in mtDNA data sets that tend to generate false positives, such as epistatic selection (Wallis 2000), clustering of hypermutable sites (Innan and Nordborg 2002), or correlated mutations (Hey 2000). The fact that the results remain highly significant for most tests when we restrict the analysis to synonymous variants suggests that epistatic selection is not responsible for the evidence of recombination. However, to further test these potential pitfalls, we compiled data from as many animal apomictic parthenogens as possible. Apomictic parthenogens reproduce asexually by the

**Table 2**  
**Twenty Species Showing the Strongest Evidence of Recombination**

Phylogeny	Species	<i>maxchi</i>	<i>geneconv</i>	$LDr^2$	$LD D' $	Rank	
Arthropods							
Hexapoda	Collembolla	<i>Gomphiocephalus Hodgsoni</i>	0.001	0.005	0.155	0.568	11
Insecta	Odonata	<i>Libellula Quadrimaculata</i>	0.05	0.425	0.002	0.004	7
Malacostraca	Decapoda	<i>Alpheus lottini</i>	$10^{-4}$	0.076	0.429	0.339	12
Vertebrates							
Aves	Gruiformes	<i>Grus antigone</i>	0.190	0.134	0.030	0.008	20
	Passeriformes	<i>Passerella iliaca</i>	0.120	0.045	0.023	0.028	17
		<i>Campylorhynchus brunneicap</i>	0.002	0.107	0.021	0.903	15
		<i>Dendroica petechia</i>	0.058	0.023	0.013	—	8
Mammalia	Primates	<i>Papio papio</i>	0.260	0.010	0.008	0.782	14
		<i>Macaca nemestrina</i>	$<10^{-5}$	0.097	$2 \times 10^{-5}$	0.24	3
		<i>Mandrillus sphinx</i>	$10^{-4}$	0.320	0.139	0.034	6
	Rodentia	<i>Microtus longicaudus</i>	0.0001	$<10^{-5}$	0.200	0.372	4
		<i>Apodemus sylvaticus</i>	0.001	0.061	$7 \times 10^{-4}$	0.900	9
Sauria	Squamata	<i>Bradypodion occidentale</i>	0.011	0.389	0.022	—	16
Teleostei	Gadiformes	<i>Merlangius merlangus</i>	0.066	0.042	0.007	0.320	18
	Perciformes	<i>Micropterus salmoides</i>	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$	0.676	2
		<i>Macrodon ancylodon</i>	0.14	0.388	0.023	0.001	13
Mollusks							
Bivalvia	Mytiloidea	<i>Mytilus galloprovincialis</i>	0.031	0.312	0.005	0.274	10
	Veneroidea	<i>Vesicomya pacifica</i>	0.021	0.003	0.008	0.528	5
Cephalopoda		<i>Gonatus onyx</i>	0.52	0.074	0.003	—	19
Nematoda	Tylenchinda	<i>Bursaphelenchus conicaudatus</i>	$<10^{-5}$	$<10^{-5}$	$<10^{-5}$	0.002	1

NOTE.—Data sets in which  $|D'| = 1$  for all pairwise comparisons were not analyzed for  $LD|D'|$  (see *Materials and Methods*). GeneBank accession numbers for each data set are available in table 2 in Supplementary Material online.

suppression of meiosis; the inheritance of mtDNA must, therefore, be strictly clonal unless there is recombination within heteroplasmic individuals. Data were available for one annelid: *Octolasion tyrtaeum*; eight arthropods: *Aramigus tessellatus santafecinus*, *Aramigus tessellatus viridipallens*, *Rhopalosiphum padi*, *Lysiphlebus testaceipes*, *Timema genevieve*, *Trichoniscus pusillus*, *Darwinula stevensoni*, and *Eucypris virens*; and three mollusks: *Campeloma parthenum*, *Campeloma geniculum*, *Campeloma limum*. Although our sample size is small, none of the four tests reject a model of clonal inheritance in any of these species, and the combined probability value from sexual species is significantly greater than that in the asexual species for all tests except  $LD|D'|$  (table 1); that is, sexual species show significantly more evidence of recombination than do asexual species. This is as expected if the signal we observed is caused by recombination and not by alternative mtDNA-specific processes. There are no significant differences between sexual species and asexual species in the length of sequences and the number polymorphisms, but asexual data sets have significantly fewer sequences (Mann-Whitney test,  $P < 0.05$ ) and the of the asexual data sets come from one genus of gastropods.

Only four of the data sets are individually significant because we have performed four tests on 267 species; that is, applying the Bonferroni correction means that a test has to be individually significant at a probability level of  $0.05/(4 \times 267) = 4 \times 10^{-5}$  to provide significant evidence of recombination. These data sets are from a plant parasitic nematode, *Bursaphelenchus conicaudatus*; a primate, *Macaca nemestrina*; a fish, *Micropterus salmoides*; and

a rodent, *Microtus longicaudus*. These data sets all contain recombinant sequences between two distinct haplotypes, which suggests that they may be recombinants between subspecies or species (see below). However, there is strong evidence of recombination even if these data sets are removed, because the combined probability values are highly significant for each test (*maxchi*,  $P = 0.001$ ; *geneconv*,  $P = 0.014$ ;  $LDr^2$ ,  $P < 10^{-4}$ ; and  $LD|D'|$ ,  $P = 0.013$ ). The problem is simply that we cannot say precisely which species has undergone recombination. Let us imagine that we performed a statistical test on 100 independent data sets and found one to be significant at a probability value of 0.0001 and 20 to be significant at a probability value of 0.05. Clearly the first data set is individually significant because the probability of observing one or more data sets at that level of significance is very small, even when we have 100 data sets ( $P = 0.01$ ). In contrast, we expect to observe at least five data sets that are significant at 5%. However, the probability of observing 20 or more is highly unlikely ( $P < 0.0001$ ). So we know that there is evidence of recombination in the remaining 99 data sets, and it is likely to be in one of the data sets that have a probability value of 0.05. However, we cannot tell which data set has actually undergone recombination.

In table 2, we list the species that contribute most to the evidence of recombination. A complete list of the outcome of the four tests for the 267 data sets is available in table 1 in Supplementary Material online. To give a measure of the overall evidence of recombination, we summed  $-2\text{Ln}(p)$  across tests, where  $p$  is the probability value. This value cannot be converted into an overall

a) *Macaca nemestrina*

Group	243	253	254	274	280	306	315	317	333	348	351	357	366	375	390	393	420	436	446	462	463	471	489	507	510	546	615	618	621	628	633	636	643	664	681	684	685		
1	1	C	A	T	T	G	G	C	T	T	A	C	C	G	A	G	G	A	T	C	C	G	T	G	C	C	A	T	G	C	A	A	C	G	G	T	G	G	
2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	Rec1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	T	T	A	C	-	T	T	C	-	-	T	G	G	T	A	-	C	A	-	-	
6	Rec2	-	-	-	-	-	-	-	-	-	-	-	G	-	-	-	-	C	T	-	-	-	-	T	-	C	-	A	T	G	-	-	-	-	-	A	-	-	
7	2	T	G	C	C	A	A	T	A	C	T	T	A	A	G	A	A	C	C	T	T	A	C	A	T	T	C	C	A	T	G	G	T	A	A	C	A	A	
8	2	T	G	C	C	A	A	T	A	C	T	T	A	A	G	A	A	C	C	T	T	A	C	A	T	T	C	C	A	T	G	G	T	A	A	C	A	A	
9	2	T	G	C	C	A	A	T	A	C	T	T	A	A	G	A	A	C	C	T	T	A	C	A	T	T	C	C	A	T	G	G	T	A	A	C	A	A	
10	2	T	G	C	C	A	A	T	A	C	T	T	A	A	G	A	A	C	C	T	T	A	C	A	T	T	C	C	A	T	G	G	T	A	A	C	A	A	
type		s	r	r	s	r	s	s	r	s	s	s	s	s	s	s	s	s	r	s	r	s	s	s	s	s	s	s	s	r	s	s	r	r	s	s	r		

b) *Papio papio*

	49	89	109	126	138	204	205	207	252	270	330	347	421	450	454
1	T	C	C	A	A	A	C	C	G	G	C	A	C	C	G
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-	-	-	-	T	-	-	
4	-	-	-	G	G	-	-	-	-	-	T	-	-	-	
5	-	-	-	-	-	-	-	-	-	-	T	T	T	-	
6	-	-	-	-	-	-	-	-	-	A	-	-	-	C	
7	-	-	-	-	-	-	-	-	-	A	-	-	-	C	
Rec	C	G	T	G	G	-	-	T	T	A	-	-	-	C	
<i>P. anubis</i>	C	-	T	G	-	-	-	T	T	A	A	T	T	C	
type	s	r	s	s	s	s	r	s	s	s	s	r	r	s	

FIG. 1.—Two species showing evidence of recombination. (a) Polymorphic sites in *Macaca nemestrina*. The putative recombinants are labeled as Rec1 and Rec2. The polymorphisms shown are those that differentiate the two major haplogroups of sequences, listed as group 1 and group 2. We have also omitted sites before nucleotide 225. Before this point, there are a further 10 sites in the putative recombinants that match the group 1 and group 2 sites [1, AF350388; 2, AF350389; 3, AF350390; 4, AF350391; 5, AF350396; 6, AF350397; 7, AF350394; 8, AF350395; 9, AF350398; 10, AF350399]. (b) Polymorphic sites in the *Papio papio* data set aligned with a *Papio anubis* sequence. The putative recombinant is shaded. Accession numbers are as follows: 1, AY212049; 2, AY212053; 3, AY212050; 4, AY212051; 5, AY212052; 6, AY212054; 7, AY212055; 8, AY212056; *P. anubis*, AY212040.

probability value because the tests are not statistically independent; however, it provides a metric that is related to the overall support for recombination.

The list of most likely recombinant species has a number of interesting features (table 2). First, there is significant evidence of recombination throughout the animal kingdom. Second, paternal inheritance of mtDNA has been observed in four of the orders in which there is evidence of recombination (Primates [Schwartz and Vissing 2002], Rodentia [Gyllensten et al. 1991], Passeriformes [Kvist et al. 2003], and Mytiloidea [Skibinski, Gallagher, and Beynon 1994; Zouros et al. 1994]). Although, paternal leakage is not the only route by which recombination can occur, it is the most likely (Eyre-Walker 2000). Third, *Mytilus galloprovincialis* only ranks 10th on our list, and yet recombination is well documented in this species (Ladoukakis and Zouros 2001), which undergoes double uniparental inheritance and in which leakage of the maternal mtDNA into the paternal line has been observed in several data sets (Hoeh et al. 1997).

There does not appear to be any evidence that the frequency of recombination differs across our data set; an ANOVA of the transformed probability values shows no evidence of differences between phyla, classes, or orders. These results remained unchanged if the analyses were

restricted to groups (i.e., phyla, classes, or orders) that had more than 20 species.

A visual inspection of the data sets that show the strongest evidence of recombination, reveals a small number of cases in which the recombinant molecules can be identified visually. Two of these are in primates. In *Macaca nemestrina*, one of the data sets that is individually significant, there are two distinct groups of haplotypes, with a third group of two sequences that appear to be recombinants between the first two (figure 1). Both putative recombinants are significant according to the *maxchi* test ( $P < 10^{-5}$  and  $P = 0.01$  for *Rec1* and *Rec2*, respectively). The putative points of recombination are between nucleotides 420 and 436 for *Rec1* and between 489 and 507 for *Rec2*. If we just consider those positions that are fixed in either group 1 or group 2 (i.e., that define the two major haplotypic groups), then before the putative breakpoint, there are 27 matches and two mismatches to the group 1 sequences, and after the putative breakpoint, there are five matches and 15 mismatches. For the second recombinant sequence, there are 30 matches and six mismatches before the breakpoint and eight matches and six mismatches after the breakpoint. The fact that there are nucleotides, both before and after the breakpoint, that match the “wrong” parental group of sequences might be

**Table 3**  
**Analysis of Macaque mtDNA Data Sets**

Data Set	Reference	<i>maxchi</i>	<i>geneconv</i>	$LD_r^2$	$LD D'$
<i>Macaca nemestrina</i> (Mn2)	Roos et al. (2003)	0.97	0.53	0.55	0.40
<i>Macaca nemestrina</i> (Mn3)	Evans et al. (1999)	$10^{-5}$	0.01	0.001	0.001
Sulawesi macaques	Evans et al. (1999)	$10^{-5}$	$10^{-4}$	0.001	0.22

the result of the formation of heteroduplex during recombination (Szostak et al. 1983), parallel mutations, or further recombination. Interestingly, recombination in this species seems to have been between two quite different haplotypes that may be the main subspecies of *M. nemestrina*: *M. nemestrina nemestrina* and *M. nemestrina leonina*. All the group 1 sequences are listed as the former, whereas two of the four group 2 sequences are given as *M. nemestrina leonina*.

There are two other data sets of *M. nemestrina* sequences in GenBank. These were sequenced by two different groups from different DNA samples (i.e., all data sets were independently obtained). One of the data sets (Mn3) shows evidence of recombination in all of the recombination tests, whereas the other data set shows no evidence of recombination (Mn2) (table 3). The data set that does not show evidence of recombination is from a region of the *cytb* gene that overlaps the putative breakpoint identified above in the Mn1 data set. However, *M. nemestrina* is a very broadly distributed species found throughout Southeast Asia and the islands of Sumatra and Borneo. Data set Mn2 was sampled from Sumatra alone, whereas data set Mn3 was sampled from much of the species range. We do not know the precise sampling of data set Mn1, but it too appears to have been sampled broadly because there are divergent haplotypes in the data set, and some individuals are listed as *M. n. leonina*.

Each of the recombinants in *M. nemestrina* appears to be between distinct subspecies. This raises the question of whether recombination between subspecies is common. To investigate this further, we tested for recombination among the seven species of macaques that inhabit the island of Sulawesi, which is east of Borneo (*Macaca nigra*, *M. nigrescens*, *M. hecki*, *M. tonkeana*, *M. maura*, *M. ochreata*, and *M. brunescens*). These seven species are thought to have evolved in isolation but to have come into secondary contact (Evans et al. 2003). Three of the four tests of recombination are significant (table 3), and a visual inspection of the data reveals at least one recombination event, where the recombinant molecule has subsequently spread through several of the species.

There is also evidence of recombination in another primate, *Papio papio*, the Guinea baboon. In this species, one individual is distinctly different from the others up to nucleotide 270, after which it becomes identical to some of the other sequences. Interestingly the sequence up to nucleotide 270 is more similar to that of several other baboon species, including *P. anubis*, the olive baboon,

whose range is adjacent to that of *P. papio* (Newman, Jolly, and Rogers 2003). However, the sequence of *P. papio* is not identical to that of *P. anubis*, which suggests that either recombination occurred sometime in the past or there has been subsequent recombination within *P. papio*.

However, cases in which recombination events can be identified are exceptional; in the vast majority of cases we cannot identify recombinants visually. This is perhaps not surprising, because recombination is only obvious to the eye when there have been very few recent recombination events between sequences that are quite different.

## Discussion

We have shown that the evidence for recombination is pervasive in animal mtDNA data sets. However, it is unclear whether the recombinants are real or whether they have been introduced during the sequencing process. Either way, the results are important for how mtDNA is used and sequenced.

There are a variety of ways in which artifactual evidence of recombination could be produced. First, there may be *numts* in the data; these are nuclear-encoded copies of the mtDNA. We have gone to some length to exclude *numts* by eliminating all data sets in which there are stop codons or the ratio of nonsynonymous to synonymous polymorphism is unreasonable (data sets in which this ratio was greater than approximately one third were excluded). These strategies will exclude most data sets in which all sequences are *numts*, but they may not be an effective guard against data sets in which a small number of sequences are recent *numts*. Second, there could be contamination between two DNA samples. This could lead to recombination in two ways. First, the polymerase may jump from one template to another during PCR. Second, different primers may bind to the two templates with different efficiencies; for example, the first set of primers may bind to the start of the gene from individual A very well, but the second half of the gene may bind very poorly. Third, recombination could be introduced during sequence assembly. This would likely produce perfect recombinants with a sequence that is composed of a perfect match, with one sequence followed by a perfect match to another sequence. Each of these should be easy to catch if the DNA has been sequenced in both directions, but there can be no guarantee that this has happened or that both strands have been sequenced to high standard. Ultimately, confirmation of the recombinants we have discovered will require careful resequencing. Testing for recombination may be one way to assess data quality.

It is also possible that the evidence of recombination is not caused by experimental error but by a problem with the tests of recombination. Innan and Nordborg (2002) have pointed out that the linkage disequilibrium methods can generate evidence of recombination if there is a concentration of hypermutable sites, at which the infinite sites assumption is violated. It also seems likely that *maxchi* and *geneconv* will be susceptible to this bias. However, there are a number of reasons for believing that clusters of hypermutable sites are not responsible for the evidence of recombination. First, hypermutable sites cannot generate the sort of patterns of recombination that we have detected

in *M. nemestrina*, *P. papio*, and a number of other species. Second, sexual taxa show significantly more evidence of recombination than do asexual taxa, and, yet, there is no reason why asexual taxa should not have clusters of hypermutable sites. Third, it seems likely that the clustering would have to be strong to generate the strong evidence of recombination we have detected (it should be emphasized that all data sets are from protein-coding genes so that none of them contain control region sequence).

If recombination is occurring in nature, then it has some implications, but possibly in areas in which it is often not thought about. Clonality is often explicitly assumed in analyses of demography; for example, mtDNA has been used extensively to trace the spread of humans across the world. It is likely that much of this work will be unaffected because the patterns are established by migration and limited gene flow, so the mitochondrial molecules involved in the pattern never have the chance to recombine. However, recombination may affect inferences about changes in population size because recombination can mirror some of the patterns induced by population size expansion. MtDNA has also been used to date various events. Under some circumstances, these dates will be affected by recombination. For example, mtDNA has been used to date our most recent matrilineal ancestor. Recombination will affect this date—it will generally mean that the date has been underestimated (Eyre-Walker 2000)—but by how much is unclear. However, the area in which mtDNA is used most often, molecular systematics, is the area in which recombination is rarely considered, and, yet, it might have important implications. In the *M. nemestrina* and *P. papio* sequences shown in figure 1, there are clearly recombinants between two subspecies, or genetically distinct groups of animals. Without an appreciation of recombination, the phylogenetic status of some individuals would be incorrect. Furthermore, the obvious evidence of introgression would be missed.

It is likely that some of the evidence for recombination is a consequence of laboratory error, but unless the quality of DNA sequencing is very poor, recombination in mtDNA is moderately frequent and occurs both within and between species and subspecies.

## Acknowledgments

We would like to thank John Welch, Meg Woolfit, David Waxman, Sebastien Gourbiere, Noël Smith, Peter Keightley, and John Maynard-Smith for stimulating discussions and comments. We are grateful to FEBS (G.P.) BBSRC (G.P., M.G., and A.E.W.) and the Royal Society (A.E.W.) for funding.

## Literature Cited

Andolfatto, P., J. Striber, and B. Charlesworth. 2003. No association between mitochondrial DNA haplotypes and a female-limited mimicry phenotype in *Papilio glaucus*. *Evol. Int. J. Org. Evol.* **57**:305–316.

Avise, J. C. 2000. *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, Mass.

Birky, C. W. J. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proc. Natl. Acad. Sci. USA* **92**:11331–11338.

Elson, J. R., R. M. Andrews, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull, and N. Howell. 2001. Analysis of European mtDNAs for recombination. *Am. J. Hum. Genet.* **68**:145–153.

Evans, B. J., J. Carlos Morales, J. Supriatna, and D. J. Melnick. 1999. Origin of the Sulawesi macaques (Cercopithecidae: *Macaca*) as suggested by mitochondrial DNA phylogeny. *Biol. J. Linn. Soc.* **66**:539–560.

Evans, B. J., J. Supriatna, N. Andayani, and D. J. Melnick. 2003. Diversification of Sulawesi macaque monkeys: decoupled evolution of mitochondrial and autosomal DNA. *Evolution* **57**:1931–1946.

Eyre-Walker, A. 2000. Do mitochondria recombine in humans? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**:1573–1580.

Gouy, M., F. Milleret, C. Mugnier, M. Jacobzone, and C. Gautier. 1984. ACNUC: a nucleic acid sequence data base and analysis system. *Nucleic Acids Res.* **12**:121–127.

Gyllensten, U., D. Wharton, A. Josefsson, and A. C. Wilson. 1991. Paternal inheritance of mitochondrial DNA in mice. *Nature* **352**:255–257.

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

Hill, W. G., and A. Robertson. 1968. Linkage disequilibrium in finite populations. *Theoret. Appl. Genet.* **38**:226–231.

Hoarau, G., S. Holla, R. Lescasse, W. T. Stam, and J. L. Olsen. 2002. Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Mol. Biol. Evol.* **19**:2261–2264.

Hoeh, W. R., D. T. Stewart, C. Saavedra, B. W. Sutherland, and E. Zouros. 1997. Phylogenetic evidence for role-reversals of gender-associated mitochondrial DNA in *Mytilus* (Bivalvia: Mytilidae). *Mol. Biol. Evol.* **14**:959–967.

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

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

Jorde, L. B., and M. Bamshad. 2000. Questioning evidence for recombination in human mitochondrial DNA. *Science* **288**:1931a.

Kondo, R., Y. Satta, E. T. Matsuura, H. Ishiwa, N. Takahata, and S. I. Chigusa. 1990. Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. *Genetics* **126**:657–663.

Kraytsberg, Y., M. Schwartz, T. A. Brown, K. Ebraldise, W. S. Kunz, D. A. Clayton, J. Vissing, and K. Khrapko. 2004. Recombination of human mitochondrial DNA. *Science* **304**:981.

Kvist, L., J. Martens, A. A. Nazarenko, and M. Orell. 2003. Paternal leakage of mitochondrial DNA in the great tit (*Parus major*). *Mol. Biol. Evol.* **20**:243–247.

Ladoukakis, E. D., and E. Zouros. 2001a. Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Mol. Biol. Evol.* **18**:1168–1175.

———. 2001b. Recombination in animal mitochondrial DNA: evidence from published sequences. *Mol. Biol. Evol.* **18**:2127–2131.

Lewontin, R. C. 1964. The interaction of selection and linkage. I. Genetic considerations; heterotic models. *Genetics* **49**:49–67.

Magoulas, A., and E. Zouros. 1993. Restriction-site heteroplasmy in anchovy (*Engraulis encrasicolus*) indicates incidental biparental inheritance of mitochondrial DNA. *Mol. Biol. Evol.* **10**:319–325.

- Maynard Smith, J. M. 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **34**:126–129.
- Maynard Smith, J., and N. H. Smith. 2002. Recombination in animal mitochondrial DNA. *Mol. Biol. Evol.* **19**:2330–2332.
- McVean, G., P. Awadalla, and P. Fearnhead. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**:1231–1241.
- Newman, T. K., C. J. Jolly, and J. Rogers. 2003. Mitochondrial phylogeny and systematics of baboons (*Papio*). *Am. J. Phys. Anthropol.* **122**.
- Piganeau, G., and A. Eyre-Walker. 2004. A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. *Heredity* **92**:282–288.
- Posada, D., and K. A. Crandall. 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. USA* **98**:13757–13762.
- Roos, C., T. Ziegler, J. K. Hodges, H. Zischler, and C. Abegg. 2003. Molecular phylogeny of Mentawai macaques: taxonomic and biogeographic implications. *Mol. Phylogenet. Evol.* **29**:139–150.
- Sawyer, S. 1989. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**:526–538.
- Schwartz, M., and J. Vissing. 2002. Paternal inheritance of mitochondrial DNA. *N. Engl. J. Med.* **347**:576–580.
- Skibinski, D. O. F., C. Gallagher, and C. M. Beynon. 1994. Mitochondrial DNA inheritance. *Nature* **368**:817–818.
- Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl. 1983. The double-strand-break model for recombination. *Cell* **33**:25–35.
- Wallis, G. P. 2000. Mitochondrial recombination or coevolution of sites? *Trends Ecol. Evol.* **15**:470–471.
- Zouros, E., A. O. Ball, C. Saavedra, and K. R. Freeman. 1994. Mitochondrial DNA inheritance. *Nature* **368**:818.

Michael Nachman, Associate Editor

Accepted August 23, 2004