Widespread Recombination in Published Animal mtDNA Sequences

A. D. Tsaousis, * D. P. Martin, † E. D. Ladoukakis, *, ‡ D. Posada, § and E. Zouros *

*Department of Biology, University of Crete, Iraklio, Crete, Greece; †Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Observatory 7925, Cape Town, South Africa; ‡Centre for the Study of Evolution & School of Life Sciences, University of Sussex, Brighton, United Kingdom; and §Facultad de Biología, Universidad de Vigo, Vigo, Spain

Mitochondrial DNA (mtDNA) recombination has been observed in several animal species, but there are doubts as to whether it is common or only occurs under special circumstances. Animal mtDNA sequences retrieved from public databases were unambiguously aligned and rigorously tested for evidence of recombination. At least 30 recombination events were detected among 186 alignments examined. Recombinant sequences were found in invertebrates and vertebrates, including primates. It appears that mtDNA recombination may occur regularly in the animal cell but rarely produces new haplotypes because of homoplasy. Common animal mtDNA recombination would necessitate a reexamination of phylogenetic and bihistorical inference based on the assumption of clonal mtDNA transmission. Recombination may also have an important role in producing and purging mtDNA mutations and thus in mtDNA-based diseases and senescence.

Introduction

Recombination of mitochondrial DNA (mtDNA) is common in protist, fungi, and plant species (for review, see Gray 1989) but was thought to be absent or rare in animals (Avise 1994). This perception is based on failure to observe recombinant haplotypes in surveys of mtDNA variation in natural populations or animal cell cultures (Zuckerman et al. 1984). Indirect support for this view comes from the questration of mtDNA molecules (Nass 1969; Satoh and Kuroiwa 1991) and the absence of detectable excision repair activity and crossover products in mammalian mitochondria (Clayton, Doda, and Friedberg 1974). However, the demonstration that human mitochondria contain enzymes for the catalysis of homologous recombination (Thyagarajan, Padua, and Campbell 1996) and the observation of mitochondrial fusion in Drosophila (Yaffe 1999) and in rat liver cells (Cortese 1999) suggest the plausibility of recombination. Homologs of the genes involved in Drosophila mitochondrial fusion (fuzzy onions) have been identified in yeast (Fzo 1) and humans (Mfn1 and Mfn2) (Santel and Fuller 2001), but the frequency with which human mitochondria actually fuse is a matter of disagreement (Enriquez et al. 2000; Legros et al. 2002).

Direct evidence of recombination now exists in four animal species: the nematode *Meloidogyne javanica* (Lunt and Hyman 1997), the mussel sister-species *Mytilus galloprovincialis* (Ladoukakis and Zouros 2001a) and *Mytilus trossulus* (Burzynski et al. 2003), the flatfish *Platichthys flesus* (Hoarau et al. 2002), and, most recently, man (Kraytsberg et al. 2004). It is important to emphasize the distinction between intramolecular and intermolecular crossing-over and between nonhomologous and homologous recombination (homologous recombination is by definition intramolecular, but the reverse is not true). It is homologous recombination that has strong implications for the molecular biology and evolution of mitochondrial genomes and whose occurrence in animals has been in doubt. Recombination in *Meloidogyne javanica* (Lunt and Hyman 1997) is an example of intramolecular crossing-over facilitated by the presence of tandem repeats. Size heteroplasmy due to variable number of tandem repeats (VNTRs) is a common feature of the mtDNA control region of many animal species (Densmore, Wright, and Brown 1985; Mortiz and Brown 1987; Bentzen, Legget, and Brown 1988; Gjetvai, Cook, and Zouros 1992) and is thought to be the result of either intramolecular or intermolecular crossing-over (Rand and Harrison 1986; Buroker et al. 1990; La Roche et al. 1990). The study of Hoarau et al. (2002) provides direct evidence for intramolecular crossing-over between arrays of tandem repeats. The reports of recombination in *Mytilus galloprovincialis* and man, on the other hand, are clear cases of homologous recombination.

However, there are several reasons why these diverse observations cannot be considered conclusive evidence for widespread animal mtDNA recombination. Unequal crossovers involving VNTRs are localized to small noncoding parts of mitochondrial genomes, and although VNTRs are common, they are not a general feature of animal mtDNA. It would also be difficult to draw general conclusions about animal mtDNA recombination from studies involving species of the mussel family Mytilidae because they have an uncommon system of mtDNA inheritance known as doubly uniparental inheritance (DUI). DUI is characterized by the presence of two mitochondria genomes, one transmitted through the eggs and the other through the sperm (Skibinski, Gallagher, and Beynon 1994; Zouros et al. 1994). Heteroplasmy (presence of more than one type of mtDNA molecules in an individual) for the two genomes is obligatory in males. The high-frequency recombination observed in these males may, therefore, be considered a peculiarity of DUI.

There is, however, another way to interpret the available evidence for and against widespread animal mtDNA recombination. This is that mtDNA recombination occurs regularly in the animal cell, but its products are difficult to detect under normal circumstances. Strict maternal inheritance in animals ensures homoplasy (presence of a single type of mtDNA molecule in an individual). Any two mtDNA molecules in an organism will, with the exception of postfertilization mutations, have identical sequences and
so too will any molecules resulting from their homologous recombination. All four studies where recombination was observed were designed to take advantage of the fact that recombinants, if they occurred, would be very different from the paternal haplotypes.

It follows therefore that demonstration of widespread animal mtDNA recombination requires the following evidence: (1) it is truly homologous recombination, (2) it is not confined to specific regions of the genome, particularly those that are prone to either intramolecular or intermolecular pairing and crossing-over, (3) it occurs in a variety of distantly related species, and (4) its products can spread and survive in the population. In this paper, we provide such evidence. We have extended the approach of Ladoukakis and Zouros (2001b) by both applying it to a much larger data set and, most importantly, making use of the most powerful statistical tests that are presently available for the detection of recombination in sets of aligned DNA sequences.

Materials and Methods

Sequence Data

Sequences retrieved from public databases were used to generate sets of sequences hereafter referred to as “alignments.” Each alignment consisted of sequences from the same region of an mtDNA protein gene that had been derived from individuals of the same or closely related species. In total, 308 alignments were generated from which we retained 186 for analysis. These consisted of sequences that contained no termination codons and could be unambiguously aligned with no indels from the first to the last codon. These 186 alignments included 7 invertebrate, 22 fish, 10 amphibian, 22 reptile, 14 avian, and 11 mammalian genera, representing a total of 1,091 species. Thirty of the mammalian genera were primates. The number of sequences varied among alignments from 4 to 148, the sequence length from 225 to 1,472 base pairs, the number of informative sites from 5 to 605, and the nucleotide diversity from 0.003 to 0.224. One hundred and twenty-eight alignments coded for cod1, 12 for cod2, 12 for nad2, 8 for nad1, 7 for nad5, 6 for cox1, 6 for nad4, 4 for nad3, 2 for cox3, and 1 for nad6 (see Supplementary Material). A second set of 186 alignments was generated by retaining only third codon positions. The third codon position alignments were used to avoid detection of linkage disequilibria generated by selection rather than recombination.

Detection of Recombination

Alignments were searched for evidence of recombination using nine different methods, which could be classified as global or local methods. Global methods provide a \( P \) value for the occurrence of recombination but without identifying or characterizing particular recombination events, while local methods provide \( P \) values for particular regions of the sequences delimited by putative recombination breakpoints.

Global methods included MaxChi Global (Posada and Crandall 2001), Reticulate (Jakobsen and Eastal 1996), and Geneconv Global (Sawyer 1989). Results of tests performed using these methods were considered to be significant if they produced a \( P \) value smaller than 0.05.

The MaxChi Global method (Posada and Crandall 2001) is an extension of that of Maynard Smith (Maynard Smith 1992). It was implemented in a personal program written in C. MaxChi Global uses a sliding window that is moved along every pair of sequences considering only variable sites. Within each window, a chi square is calculated with the number of matches and mismatches between the two sequences in the first and second halves of the window. The test statistic is the maximum chi square observed in the original alignment. Its null distribution (i.e., under no recombination) is obtained by permuting the columns of the alignment and recalculating its value several times. When applying the test, the step size for moving the sliding window along the sequences was set to 2. The minimum half-size of the sliding window was set to five variable sites, while the maximum half-size of the sliding window was set to the number of variable sites divided by 3. \( P \) values were estimated by randomizing the alignment 1,000 times.

The Reticulate partition compatibility method (Jakobsen and Eastal 1996) was implemented with the program Reticulate (Jakobsen, Wilson, and Eastal 1997) available at \url{http://jcsmr.anu.edu.au/dmn/humgen/ingrid/reticulate.htm}. In this method, a matrix is constructed in which a cell contains information about the compatibility of a pair of informative sites. Two sites are incompatible when the same tree cannot be constructed from these two sites without assuming multiple changes, otherwise they are compatible. The neighbor similarity score is calculated as the number of adjacent cells that are either compatible or not compatible. The statistical significance of this test is calculated by permutation. Here the number of permutations was set to 1,000.

The Geneconv Global method is an extension of that of Sawyer (Sawyer 1989; \url{http://www.math.wustl.edu/~sawyer/geneconv/}). In this method, global \( P \) values are based on a version of the Blast score (Altschul et al. 1990). The global permutation \( P \) value is the proportion of permuted alignments for which some fragment for some pair of sequences has a higher score than the observed fragment. These \( P \) values have a built-in multiple-comparison correction for all sequence pairs in the alignment. The \( g\)-scale parameter was set to 0 in order to be most conservative. The number of permutations was set to 10,000.

\( P \) values for local methods were Bonferroni corrected for multiple tests. For each unique recombination event detected in an alignment, the \( P \) values reported are the smallest corrected probabilities found by each method. A method was considered to indicate significant evidence for recombination when the corrected \( P \) value was smaller than 0.05. Local methods included recombination detection program (RDP; Martin and Rybicki 2000), Geneconv Local, MaxChi Local (Martin, Williamson, and Posada 2004), Chimaera Local (Martin, Williamson, and Posada 2004), SiScan (Gibbs, Armstrong, and Gibbs 2000), and RecScan (Martin, et al. 2005). All local methods were implemented in the program RDP2 (Martin, Williamson, and Posada 2004), available from \url{http://darwin.uvigo.es/rdp/rdp.html} for the Windows OS. The manual of this program contains detailed information on the implementation...
of the different algorithms. Common settings in RDP2 for all methods were that sequences were considered as linear, the P-value cutoff was set to 0.05, the standard Bonferroni correction was used, consensus daughters were found, breakpoints were polished, and only unique events detected by two or more methods were listed (see RDP2’s manual).

The original RDP method (Martin and Rybicki 2000) uses a sliding window to examine variable nucleotide positions of every possible sequence triplet for a change in their percentage identities along an alignment. The significance of this change is approximated using the binomial distribution. This P value is then corrected for multiple significance by multiplying it by the number of independent windows and triplets examined. Here the length of the window was set to 10 variable sites, and the step size was set to 1 nucleotide (nt).

The Geneconv Local method (Martin, Williamson, and Posada 2004) is identical to the Geneconv Global method except that instead of analyzing the entire alignment, it works with three sequences at a time, and Bonferroni corrected Blast-like P values are used. Although in the Geneconv Global test we used a g-scale of 0 to be most conservative, detecting specific breakpoints is much more difficult (there is less data than in the global case) and in this case the g-scale parameter was set to 1.

The MaxChi Local method (Martin, Williamson, and Posada 2004) examines every possible pair of sequences in every triplet in the alignment. This implementation is a modification of the MaxChi Global method but analyzes three sequences at a time. Here the length of the window was set to 30 variable sites, and the step size was set to 1 nt.

The Chimaera Local method (Martin, Williamson, and Posada 2004) is similar to the MaxChi Local method, but in this case, a match occurs when the putative recombinant sequence matches parental A but not parental B, while a mismatch occurs when the putative recombinant sequence matches parental B but not parental A. Here the length of the window was set to 30 variable sites, and the step size was set to 1 nt.

The SiScan method (Gibbs, Armstrong, and Gibbs 2000) implemented in RDP2 examines all combinations of three sequences plus an out-group sequence. This method interprets an increase in the frequency of particular patterns of sites as evidence of recombination. The statistical significance of the test statistic is obtained through permutation of the alignment. Here the size of the window was set to 100, and it was moved 20 nt at a time. Only variable sites were considered. The number of permutations was set to 1,000.

The RecScan method (Martin et al. 2005) is a modification of the Bootscan method (Salminen et al. 1995). It interprets significant alterations in pairwise distance patterns as evidence of recombination. The alterations of the patterns are quantified in terms of their bootstrap support, and the P value is approximated using a Bonferroni corrected version of the binomial distribution. Here the window size was set to 100 sites, while the window step size was set to 20. Candidate events were identified from 100 bootstrap replicates, and the bootstrap cutoff was set to 70%.

Strength of Recombination Signal

Alignments were classified in four arbitrary categories regarding the evidence for recombination:

1. Strong evidence: This category included alignments for which the null hypothesis of no recombination was rejected by at least one of the global methods and by two local methods.
2. Good evidence: This category included alignments for which the hypothesis of no recombination was rejected by more than one test (regardless of being a global or a local method) but did not meet the set of criteria for the first class.
3. Weak evidence: This category included alignments for which only one test (global or local) rejected the hypothesis of no recombination.
4. No evidence: This category included alignments for which no test rejected the hypothesis of no recombination.

Correlation of Recombination Detection Power with Alignment Features

To obtain a relative index of the overall evidence of recombination in an alignment we used the score

\[ Y = -2 \sum_{i=1}^{k} \ln P_i, \]

where \( P_i \) is the probability from the ith recombination detection method and \( k \) is the number of methods. If probabilities from different methods are uncorrelated, \( Y \) should be distributed as a chi square with \( 2k \) degrees of freedom. This score does not provide the overall probability of recombination in an alignment because in fact some methods may be more likely to produce similar probabilities than others. Here it was rather used to evaluate the effect of several varying alignment features (such as number, length, and diversity of sequences) on the probability of detecting recombination when multiple detection methods are used.

Results and Discussion

One strategy for the detection of mtDNA recombination in data sets from population surveys is to search for linkage disequilibria at variable sites in haplotypes from the same species (Awadalla, Eyre-Walker, and Maynard Smith 1999; Elson et al. 2001). One difficulty with this strategy is the confounding effect of homoplasy, which can be overcome by looking at the correlation of linkage disequilibrium with the proximity of variable sites in the genome (Awadalla, Eyre-Walker, and Maynard Smith 1999; Elson et al. 2001). Another difficulty is that the frequencies of the different haplotypes in the sample must be representative of the population frequencies, which would not normally be the case in data sets drawn from public databases or from population surveys that were not designed for this purpose. In fact, results from this strategy have proven either equivocal (Kivisild et al. 2000) or negative (Elson et al. 2001). Another strategy (Ladoukakis and Zouros 2001b) is to look for linkage disequilibria among genomes that were already highly differentiated.
at the time of maternal inheritance may break down in crosses
between different species (Shitara et al. 1998). Hybrids
from these crosses are therefore likely to be heteroplasmic
for highly divergent mitochondrial genomes. If there is no
impediment to mtDNA recombination, fertile offspring
from such crosses will be vectors generating and passing
on recombinant haplotypes to the population that will be
easily distinguishable from the parental types. In due time,
parental and recombinant haplotypes will occur either as
intrasppecific or as interspecific variants. A limitation of this
strategy is that such variants might be rare in natural pop-
ulations because the reduced fitness of hybrids and their
immediate progeny diminishes the long-term survival pros-
pect of recombinant haplotypes. Another limitation is that
postrecombination mutation may obscure the distinction
between recombinant and parental types.

The second strategy was first applied to a small survey
of sequences retrieved from public databases (Ladoukakis
and Zouros 2001b). Only protein-coding mtDNA sequen-
ces were examined because they offered the assurance of
unambiguous alignment. Putative triads of sequences
involved in a recombination event and putative crossover
points were first identified by visual inspection of amino
acid sequences. Statistical testing was based on generating
sequences of the same length as the putatively exchanged
segment by random sampling of nucleotide sites from the
sequence triad and asking whether any such random
sequence could produce a better degree of matching
between donor and recipient sequences than the matching
produced by the putatively exchanged sequence itself.
Three cases of potential recombination were detected:
in the amphipod Gammarus, in the frog Rana, and in the
rodent Apodemus. Maynard Smith and Smith (2002) reex-
amined the data and concluded that the evidence for recomb-
ination was strong for Rana, weak for Apodemus, and
absent for Gammarus.

In this paper, we report the results from a much larger
survey of mtDNA sequences. As in Ladoukakis and Zouros
(2001b), only protein-coding sequences were examined.
The data were analyzed with an array of statistical methods
that have been developed in recent years for the detection
of recombination in large sets of aligned DNA sequences. All
of these methods other than the SiScan method have been
recently evaluated. In fact, they were chosen because they
show reasonable power and low false-positive rates, the lat-
ter being independent of divergence levels and recurrent
mutation (Posada and Crandall 2001; Posada, Crandall,
and Holmes 2002; Martin, Williamson, and Posada 2004).

Based on the arbitrary criteria categorizing the strength
of evidence for recombination (see Materials and Meth-
ods), the 186 alignments were classified as follows: 26 with
strong evidence, 36 with good evidence, 36 with weak evi-
dence, and 88 with no evidence. When third codon position
alignments were examined, these numbers were 21, 14, 26,
and 125 for strong, good, weak, and no evidence, respec-
tively. Thirty alignments were sorted into the first class
when full or third codon position alignments were exam-
ined (table 1). These include the Rana and Apodemus align-
ments, for which evidence for recombination has been
previously detected (Ladoukakis and Zouros 2001b).

One risk in ‘‘real time’’ recombination studies is that
when polymerase chain reactions (PCRs) are presented
with more than one molecule, the polymerase may switch
templates thus creating recombinant products (Paabo,
Irwin, and Wilson 1990). The possibility that this might
have happened in the sequences we have examined must
be very low. Each sequence was extracted from a different
individual, and in many alignments, different sequences
were determined in different laboratories. In an experiment
designed to test the frequency with which ‘‘PCR recombi-
nants’’ occur during the amplification of mixed populations
of undamaged target DNA (as opposed to highly degraded
DNA common in ancient DNA studies), no such recombi-
nants were found (Ladoukakis and Zouros 2001a). Also, if
recombination occurred through PCR jumping, exchanged
regions found in two different sequences should be perfect
copies of one other. Visual inspection of the recombinant
sequences that we identified revealed no perfect matches
between the putatively exchanged parts of these sequences
and those of their most probable parent sequences (fig. 1).

Another possibility is that the nonrandom associations
we have observed are real but that they have resulted from
selection rather than recombination. Because the sequences
in our alignments were drawn from different species, it is
conceivable that selection acting on mtDNA alone or
through nuclear-mtDNA interactions has favored different
combinations of amino acids in different sequences. To
minimize the possibility that the associations are caused
by selection, we performed the same statistical analysis
on alignments that contained only third codon position
nucleotides. Examination of only third codon position
nucleotide sequences is an unduly conservative constraint
on the detection of recombination because the number of
compared sites is reduced to approximately one third,
and the retained sites are those that are most likely to accu-
ulate postrecombination mutations and thus reduce the
power of the tests. Even so, the statistical evidence for
recombination from the full sequence alignments was
strongly correlated with the evidence when only third
codon base sequences were used (the regression of \( Y \)
calculated from full sequences against the \( Y \) calculated
from third codon base sequences gave \( r = 0.806, P = 1 \times
10^{-35} \) on \( n = 142 \)). Also, 17 of the 30 alignments in table
1 were sorted into the first class whether full sequences or
third codon base sequences were used. This implies that
the mtDNA recombination events that we have detected are
unlikely to be the products of selection-induced linkage
disequilibria.

The preponderance of cytochrome b sequences among
the alignments of the first class is simply a reflection of this
gene’s high frequency in the 186 alignments tested (\( \chi^2 =
2.084, df = 1, P = 0.149 \)). Thus, there is no reason to sus-
pect that, at least for the protein-coding part of the genome,
recombination is more likely in one region than in another.
There are 2 invertebrate, 7 fish, 1 amphibian, 2 reptile, 2
avian, 11 nonprimate mammalian, and 4 primate species
among the 29 listed in table 1 (one genus, Myotis, is rep-
resented twice). Again, this distribution is a good represen-
tation of the frequency of the taxa examined (\( \chi^2 = 6.199,
df = 6, \) and \( P = 0.401 \)) and does not indicate that recombi-
nation is likely to be more common in some major divisions.
of animals than in others. In figure 1 we present for visual inspection the first recombination example in table 1 for an invertebrate, fish, amphibian, reptile, bird, and primate. For Collocalia it is apparent that both recombination points lie within the sequence examined, whereas only one breakpoint is included for the other examples. It is difficult to determine whether one or both breakpoints are included in the Bactrocer a sequence. In 22 of the 30 cases listed in table 1, all three sequences involved in the recombination event could be identified. In 11 of these, each of the three sequences came from three different species; in 9 two sequences belonged to a different species, in 9 two sequences belonged to the same species and one to another, and in 9 three sequences came from three different species. In 22 of the 30 cases listed in table 1, all three sequences involved in the recombination event could be identified. In 11 of these, each of the three sequences came from three different species; in 9 two sequences belonged to a different species, in 9 two sequences belonged to the same species and one to another, and in 9 three sequences came from three different species.

### Table 1

Thirty Events of Recombination with the Strongest Statistical Support in the 186 Animal mtDNA Alignments Examined

<table>
<thead>
<tr>
<th>Genus (Common Name)</th>
<th>Data Set Number</th>
<th>Number of Sequences</th>
<th>Gene Sequence Length</th>
<th>Number of Diagnostic Sites</th>
<th>Nucleotide Divergence</th>
<th>Supporting Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cephalophus3</td>
<td>037</td>
<td>23</td>
<td>cob</td>
<td>1,140</td>
<td>307 (253)</td>
<td>0.082 (0.208)</td>
</tr>
<tr>
<td>Mammal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Collocalia3</td>
<td>044</td>
<td>15</td>
<td>cob</td>
<td>1,143</td>
<td>156 (104)</td>
<td>0.049 (0.109)</td>
</tr>
<tr>
<td>Bird</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Python3</td>
<td>149</td>
<td>28</td>
<td>cob</td>
<td>313</td>
<td>93 (—)</td>
<td>0.094 (—)</td>
</tr>
<tr>
<td>Reptile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, Mi, C, S</td>
</tr>
<tr>
<td>Rana6</td>
<td>150</td>
<td>14</td>
<td>cob</td>
<td>572</td>
<td>213 (176)</td>
<td>0.157 (0.411)</td>
</tr>
<tr>
<td>Amphibian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Aerodromas3</td>
<td>002</td>
<td>12</td>
<td>cob</td>
<td>1,143</td>
<td>116 (—)</td>
<td>0.043 (—)</td>
</tr>
<tr>
<td>Bird</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Anguilla3</td>
<td>010</td>
<td>39</td>
<td>cob</td>
<td>1,140</td>
<td>300 (251)</td>
<td>0.074 (0.188)</td>
</tr>
<tr>
<td>Fish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Serrislou1</td>
<td>163</td>
<td>7</td>
<td>cob</td>
<td>1,141</td>
<td>318 (156)</td>
<td>0.121 (0.182)</td>
</tr>
<tr>
<td>Fish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Tragelaphus3</td>
<td>179</td>
<td>13</td>
<td>cob</td>
<td>1,140</td>
<td>319 (239)</td>
<td>0.099 (0.234)</td>
</tr>
<tr>
<td>Mammal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Arteius1</td>
<td>011</td>
<td>22</td>
<td>cob</td>
<td>278 (141)</td>
<td>164 (119)</td>
<td>0.07 (0.168)</td>
</tr>
<tr>
<td>Mammal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Artrusopides4</td>
<td>020</td>
<td>21</td>
<td>nad4</td>
<td>689</td>
<td>303 (—)</td>
<td>0.053 (0.127)</td>
</tr>
<tr>
<td>Reptile</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Rhabdiana1</td>
<td>153</td>
<td>121</td>
<td>cob</td>
<td>1,138</td>
<td>369 (280)</td>
<td>0.053 (0.127)</td>
</tr>
<tr>
<td>Fish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Kous5</td>
<td>088</td>
<td>9</td>
<td>cob</td>
<td>1,140</td>
<td>191 (151)</td>
<td>0.072 (0.177)</td>
</tr>
<tr>
<td>Mammal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Semnopithecus4</td>
<td>162</td>
<td>8</td>
<td>cob</td>
<td>1,140</td>
<td>206 (141)</td>
<td>0.073 (0.151)</td>
</tr>
<tr>
<td>Mammal—Primate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Bactrocer1</td>
<td>021</td>
<td>47</td>
<td>cox1</td>
<td>636</td>
<td>243 (191)</td>
<td>0.112 (0.282)</td>
</tr>
<tr>
<td>Insect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Macaca3</td>
<td>097</td>
<td>36</td>
<td>cob</td>
<td>859</td>
<td>303 (—)</td>
<td>0.079 (—)</td>
</tr>
<tr>
<td>Mammal—Primate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Ovis3 Mammal</td>
<td>122</td>
<td>4</td>
<td>cob</td>
<td>1,143</td>
<td>97 (87)</td>
<td>0.045 (0.12)</td>
</tr>
<tr>
<td>Ctenomys3</td>
<td>049</td>
<td>65</td>
<td>cob</td>
<td>1,140</td>
<td>470 (317)</td>
<td>0.085 (0.194)</td>
</tr>
<tr>
<td>Mammal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Phalanger4</td>
<td>132</td>
<td>18</td>
<td>nad2</td>
<td>1,036</td>
<td>605 (171)</td>
<td>0.224 (0.194)</td>
</tr>
<tr>
<td>Mammal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Chironomus5 Insect</td>
<td>042</td>
<td>46</td>
<td>cob</td>
<td>672</td>
<td>319 (—)</td>
<td>0.147 (—)</td>
</tr>
<tr>
<td>Mammal—Primate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Apodocus4 Mammal</td>
<td>011</td>
<td>10</td>
<td>cob</td>
<td>1,140</td>
<td>385 (300)</td>
<td>0.147 (0.361)</td>
</tr>
<tr>
<td>Aciapensere9 Fish</td>
<td>001</td>
<td>26</td>
<td>cob</td>
<td>1,141</td>
<td>308 (—)</td>
<td>0.056 (—)</td>
</tr>
<tr>
<td>Tilapia3 Fish</td>
<td>176</td>
<td>14</td>
<td>nad2</td>
<td>1,047</td>
<td>370 (—)</td>
<td>0.126 (—)</td>
</tr>
<tr>
<td>Trachiptereres5</td>
<td>178</td>
<td>12</td>
<td>cob</td>
<td>825</td>
<td>231 (—)</td>
<td>0.121 (—)</td>
</tr>
<tr>
<td>Mammal—Primate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Sperminus3 Mammal</td>
<td>168</td>
<td>148</td>
<td>cob</td>
<td>1,140</td>
<td>528 (—)</td>
<td>0.112 (—)</td>
</tr>
<tr>
<td>Mammal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Myotis3 Mammal</td>
<td>113</td>
<td>66</td>
<td>cob</td>
<td>1,140</td>
<td>525 (363)</td>
<td>0.146 (0.343)</td>
</tr>
<tr>
<td>Mammal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Nandopsis5 Fish</td>
<td>116</td>
<td>5</td>
<td>cob</td>
<td>1,062</td>
<td>779 (—)</td>
<td>0.126 (—)</td>
</tr>
<tr>
<td>Mammal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Trachinota5 Fish</td>
<td>177</td>
<td>7</td>
<td>cob</td>
<td>1,140</td>
<td>— (216)</td>
<td>0.247 (—)</td>
</tr>
<tr>
<td>Mammal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M, R, D, Gl, RS, Mi, C, S</td>
</tr>
<tr>
<td>Pecos1 Mammal</td>
<td>137</td>
<td>21</td>
<td>nad1</td>
<td>766</td>
<td>— (178)</td>
<td>0.258 (—)</td>
</tr>
</tbody>
</table>

Note. — Information in parentheses refers to third base sequence alignments. Open dashes indicate that recombination was not detected in full sequence alignments, and dashes in parentheses indicate that the recombination was not detected in third base sequence alignments. Superscripts denote origin of sequences involved in recombination: 1: all three sequences came from three different species; 2: the recombinant sequence came from one species and both parental sequences came from another species; 3: the recombinant and one parental sequence came from the same species, and in one of the other two sequences further breakpoint was found; 4: both parental sequences came from the same species, and in one of the other two sequences further breakpoint was found; 5: only one sequence could be identified. Data set numbers refer to Appendix provided as online Supplementary Material to this paper. Abbreviation for genes are as follows: cob: cytochrome b, cox1: cytochrome oxidase 1, nad1: NADH dehydrogenase subunit 1, nad2: NADH dehydrogenase subunit 2, nad4: NADH dehydrogenase subunit 4. Abbreviations for statistical tests are as follows: M: MaxChi, R: Reticulate, G: Geneconv, D: RDP, Gl: Geneconv Local, RS: RecScan, Ml: MaxChi Local, C: Chimaera, S: SiScan. Sequence accession numbers can be obtained from A.D.T.
All three sequences belonged to the same species. This illustrates the importance of recombination in generating mtDNA variation that eventually becomes an interspecific difference.

The number of sequences, the sequence length, the number of informative sites, and the degree of nucleotide divergence varied widely among alignments. To evaluate the effect of these variables on the likelihood of detecting recombination, we implemented a simple linear regression with each of these factors as the independent variable and $Y$ (see Materials and Methods) as the dependent variable. The probability of detecting recombination was not correlated with the number of sequences in the alignment ($r = 0.065, P = 0.393$), and it was weakly correlated with sequence length ($r = 0.168, P = 0.027$). The probability of detecting recombination was, however, strongly correlated with both the nucleotide diversity and number of informative sites in the alignment ($r = 0.251, P = 8 \times 10^{-4}$ and $r = 0.343, P = 3.7 \times 10^{-6}$, respectively).

This result is in agreement with simulation studies evaluating many of the recombination detection methods used here (Posada and Crandall 2001), which showed that recombination detection power increases with the number of variable/phylogenetically informative sites in an alignment.

Our results satisfy the requirements necessary to suggest that recombination is a common event in animal mtDNA. It is clear from figure 1 that the recombinant sequences we have identified are products of true homologous recombination. Five different mitochondrial genes and all major divisions of the animal kingdom are represented in table 1. Finally, the fact that recombinant and parental types were found in different individuals of different species (whose mtDNA variation was surveyed for reasons other than the detection of recombination) suggests that recombinant haplotypes can spread within populations. A very recent survey of animal mtDNA recombination by Piganeau, Gardner, and Eyre-Walker (2004) produced similar results. These authors also used the strategy of searching for evidence of recombination among sequences of animal mtDNA protein genes of the same or different species (whose mtDNA variation was surveyed for reasons other than the detection of recombination) suggests that recombinant haplotypes can spread within populations. A very recent survey of animal mtDNA recombination by Piganeau, Gardner, and Eyre-Walker (2004) produced similar results. These authors also used the strategy of searching for evidence of recombination among sequences of animal mtDNA protein genes of the same or different species.
tests and were not used here for the reasons explained above. Piganeau, Gardner, and Eyre-Walker (2004) examined 279 alignments and found that in 14.2% of them there existed evidence for recombination supported by one or more of the four tests used. Interestingly, they found no such evidence among the 12 asexual species included in their survey. The two studies are not strictly comparable because of the different statistical tests used and also because of the different sampling effort devoted to different animal taxa. The studies do, however, supplement one another in demonstrating that animal mtDNA recombination might be much more common than previously anticipated and that it is becoming increasingly more difficult to maintain the view that animal mtDNA is intrinsically different from that of other eukaryotes with regard to mechanisms and frequencies of recombination. The paucity of examples of recombination in the animal kingdom seems to be more a reflection of the difficulty with which its products can be detected than of the frequency of its occurrence. In two known cases of heteroplasmy, one in mussels resulting from normal biparental inheritance and another in humans resulting from paternal mtDNA leakage (Schwarz and Vissing 2002) recombination has been detectable (Ladoukakis and Zouros 2001a; Kratsberg et al. 2004).

Our study was not designed to provide estimates of mtDNA recombination in the animal kingdom but only to detect its presence. The number of sampled alignments is biased in favor of vertebrates and in favor of mammals within vertebrates. Moreover, there is a large degree of heterogeneity among the alignments with respect to the number of sequences they contain, the number of species from which the sequences were obtained, and whether the sequences that belong to the same species were derived from one or more populations. We believe that most alignments gathered from sequences in GeneBank cannot provide reliable estimates of haplotype frequencies in natural populations, which are essential for the estimation of population parameters (like 4N_r, where N is the effective population size and r is the recombination rate per gene per generation) or linkage disequilibrium. Therefore, we did not try to estimate the rate of recombination (4N_r) Kuhner, Yamato, and Felsenstein 2000; McVean, Awadalla, and Fearnhead 2002; Li and Stephens 2003) or the decay of linkage disequilibrium with increasing distance between makers (Awadalla, Eyre-Walker, and Maynard Smith 1999).

The implications of widespread animal mtDNA recombination are diverse. Recombination constitutes a means of evading uniparental inheritance and thus clonal transmission. With regard to the use of mtDNA for evolutionary studies, it would mean that we should not, as it has been common practice until now, draw conclusions about the evolutionary history of the entire mitochondrial genome by looking at parts of it. Such extrapolations are, for example, implicit in studies of hybridization and introgression in natural populations (Ferris et al. 1983; Powell 1983; Harrison 1989) and of mtDNA selection in natural or laboratory populations (Clark and Lykkegaard 1988; MacRae and Anderson 1988; Ballard and Kreitman 1995). Heteroplasmy due to mutation cannot be considered as the sole or even the most important factor that confounds rates of mtDNA divergence and phylogenetic inference. Estimates of effective population size for mtDNA may have to be revised upward, and the theory of hierarchical analysis of mtDNA variation, from the mitochondrion to the species through the intermediate levels of the cell, the organism, and the population (Birky, Maruyama, and Fuerst 1983; Birky, Fuerst, and Maruyama 1989: Zouros and Rand 1999), will need to be reconsidered. Equally important is that the mutation load of single genomes and the purging of mutations from the mtDNA pool will not be governed by the rules of strict clonal transmission. Nonclonal transmission of the animal mitochondrial genome may, therefore, have important consequences in the etiology of mtDNA-based conditions for the organism, such as senescence and diseases caused by mtDNA deletions and mutations.

**Literature Cited**


Richard Thomas, Associate Editor

Accepted December 21, 2004