

Mitochondrial DNA polymorphism: Evidence that variants detected by restriction enzymes differ in nucleotide sequence rather than in methylation

(mtDNA evolution/mtDNA cloning/sequence analysis/rats)

FRANK J. CASTORA, NORMAN ARNHEIM, AND MELVIN V. SIMPSON

Department of Biochemistry, State University of New York, Stony Brook, New York 11794

Communicated by Philtp Stekevitz, July 22, 1980

ABSTRACT Restriction enzyme analysis of mtDNAs for the purpose of determining sequence divergence rests on the assumption that variant recognition sites differ with respect to sequence and not methylation. This assumption was tested on two mtDNAs, A and B, which are distributed throughout the laboratory rat population and which can be distinguished by a number of restriction enzymes. The mtDNAs were cloned and the nucleotide sequences of corresponding small *Hind*III fragments, in which a variant *Eco*RI site occurs, were determined. Evidence that the fragments differ in sequence and not methylation is as follows: (i) The cloned mtDNA yielded the same fragment pattern as did native mtDNA when treated with *Eco*RI, *Hha* I, *Hinf*I, and *Hae* III; (ii) three nucleotide replacements were found in the 169-base pair fragment, A·T ↔ G·C, A·T ↔ G·C, T·A ↔ G·C; (iii) one of these replacements, A·T ↔ G·C at position 80, accounts for the presence of the *Eco*RI site in the type A and its absence in the type B mtDNA. Examination of the sequence leads to the suggestion that these three nucleotide replacements are silent; i.e., they would not lead to amino acid substitutions in a possible encoded protein.

The phenomenon of intraspecific mtDNA polymorphism in mammals as detected by restriction endonucleases is well established (1-11) and is proving valuable as an approach to a number of important problems. For example, the discovery of two types of mtDNA in the laboratory rat population (3) was used to conclusively prove the operation of maternal inheritance in mammals (5, 6, 12, 13). Also, the high degree of this intraspecific mtDNA variation in all the species that have been studied, and particularly in *Rattus rattus* (ref. 11; unpublished data), suggests a rapid rate of evolution for mtDNA, and this has been proposed previously (14, 15) in an interspecific comparison of mtDNAs. Thus, the restriction enzyme analysis of this rapidly evolving DNA offers an approach at high resolution to the study of the evolution of closely related species (15) and even of the population genetics within a species (refs. 10 and 11; unpublished data).

The valid application of restriction endonuclease studies of mtDNA to these and associated problems rests upon a number of assumptions, two of which are relevant here. One is that the distinction between variant mtDNAs by restriction enzymes results from single base changes rather than from deletions, insertions, or rearrangements. The other is that this distinction is based on differences in nucleotide sequence rather than on differential methylation. The latter process could well play a role in the discrimination among otherwise identical DNA molecules by many bacterial restriction endonucleases, which could act in a manner analogous to their behavior in restriction-modification systems. DNA methylation is a general cellular phenomenon. In animal nuclear DNA, 5-methylcytosine is found (in CpG doublets) as the modified base (16). The de-

tailed nature of the methylation sites of mtDNA is, however, still unsettled (17-20), although the involvement of cytosine is clear. In bacteria, relevant because of the possible prokaryotic origin of mitochondria, the methylation of guanine, adenine, and cytosine is well documented. Moreover, recent studies on the mechanism of maternal inheritance in chloroplasts (21, 22) have brought to light a methylation-restriction enzyme system and have shown that DNA methylation can affect the susceptibility of *Eco*RI (G-A-A-T-T-C), *Bam*HI (G-G-A-T-C-C), and *Hpa* II (C-C-G-G) recognition sites to cleavage by these respective enzymes.

For these reasons, it was imperative to obtain direct evidence with respect to the nature of the observed variations in nucleotide sequence. We have obtained such evidence, reported here, by cloning mtDNA in *Escherichia coli* and by determining the nucleotide sequences of corresponding small fragments of types A and B mtDNA, the two types found in the laboratory rat (*Rattus norvegicus*) population (3-8).

In addition, certain coding aspects of the sequence are discussed in relation to the evolution of mtDNA. An abstract of these results containing the two sequences has been published (23). Similar conclusions have been published in abstract form by others* without, however, the inclusion of sequence data.

MATERIALS AND METHODS

Isolation of mtDNA. Rats possessing type A or type B mtDNA were obtained from our own Sprague-Dawley type A and B colonies (5). Mitochondria were isolated from liver by differential centrifugation as described (24). mtDNA was purified by sucrose density gradient sedimentation (21,000 rpm, 16 hr, Beckman SW 27 rotor, 4°C) followed by two CsCl/ethidium bromide isopycnic centrifugations (35,000 rpm, 40 hr, Beckman SW 50.1 rotor, 20°C). The lower band material was collected, dialyzed several hours in the dark with three changes of STE (0.15 M NaCl/0.01 M Tris-HCl/0.001 M EDTA, pH 7.5) to remove the Cs salt, and then extracted twice with STE-saturated phenol. This treatment removes bound and unbound dye and permits succeeding restriction enzyme digestion with all enzymes we have used thus far. The DNA was concentrated and freed from salt by precipitation with 2.5 vol of ethanol, and then taken up and stored at 4°C in TE buffer (0.01 M Tris-HCl/0.001 M EDTA, pH 7.5).

Plasmid pBR322 was grown in *E. coli* strain K 803 (*recA*⁻). It was purified by two CsCl/ethidium bromide centrifugations. Lower band material was collected and dye was removed by passage through a Dowex 50W-X8 cation-exchange column.

Abbreviations: kb, kilobase(s); bp, base pair(s); TE buffer, Tris/EDTA buffer.

*Kroon, A. M., Bakker, H., deVos, W. M., Pepe, G. & Saccone, C. (1979) 11th International Congress of Biochemistry, Toronto, ON, July 8-10, p. 27 (abstr.). See Note Added in Proof.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

The DNA was precipitated with ethanol and was then taken up and stored in TE buffer at -20°C .

Ligation and Transformation. The purified type A and type B mtDNAs were partially digested with restriction endonuclease *Bam*HI so that there were approximately equimolar amounts of linear mtDNA [15.9 kilobases (kb)], large *Bam*HI fragment (10.8 kb), and small *Bam*HI fragment (5.1 kb) present. This mixture of mitochondrial fragments was ligated to *Bam*HI-linearized pBR322 with 2 μl of phage T4 DNA ligase (New England BioLabs, 0.5 unit/ μl) by using 600 ng of mtDNA and 20 ng of plasmid in a volume of 130 μl of ligation solution (0.1 M Tris-HCl, pH 7.8/0.05 M NaCl/0.012 M MgCl_2 /0.01 M dithiothreitol/0.154 mM ATP). Transformation of *E. coli* strain K 803 cells was carried out as described (25). No full-length mitochondrial inserts were obtained but an equal proportion of small and large *Bam* fragment inserts of both types A and B were recovered from the 600 colonies screened by growing the ampicillin-resistant colonies on tetracycline plates. Recombinant DNA was handled under P2/EK1 conditions as described in the National Institutes of Health guidelines.

Isolation of Hybrid Molecules. Small-scale preparations of all ampicillin-resistant tetracycline-sensitive colonies were obtained by growing 50-ml samples overnight in the presence of chloramphenicol. The cells were collected, lysed, and cleared by a 1-hr centrifugation at 20,000 rpm in a Beckman SW 27 rotor. This cleared lysate was extracted twice with an equal volume of 1:1 mixture of phenol and chloroform. Dissolved phenol was removed by two extractions with diethyl ether and the DNA was precipitated by addition of 2.5 vol ethanol.

The cloned DNA used for the sequence studies was prepared by growing appropriate clones in 1-liter volumes in the presence of chloramphenicol and isolating DNA as above with the following modifications. After preparation of the cleared lysate, CsCl and ethidium bromide were added to the supernate and plasmid DNA was purified by two CsCl/ethidium bromide centrifugations. The lower band material was recovered, dialyzed, and extracted with phenol as described above. This DNA was then precipitated, collected, and stored in TE buffer at -20°C .

Restriction Digests and Gel Electrophoresis. All restriction endonuclease digestions were performed as described (5). Restriction digestions designed to characterize the cloned DNA with respect to type and orientation were performed upon DNA isolated from 50-ml plasmid preparations. Because these samples contained a large amount of contaminating RNA, we routinely added 1 μl of heat-treated RNase A (10 mg/ml in 10 mM Tris-HCl, pH 7.5) to each restriction digest.

Agarose gel electrophoresis was performed in 0.7% or 1% agarose in E buffer (40 mM Tris acetate/20 mM sodium acetate/2 mM EDTA, pH 8.3) or in ME buffer (40 mM Tris-HCl/5 mM magnesium acetate/0.5 mM EDTA, pH 8.0). Electrophoresis was carried out for 3 hr at 90 V or 6 hr at 60 V. The fragment sequenced in this study was isolated from a *Hind*III digest of cloned mtDNA separated by overnight electrophoresis at 150 V in a 5% polyacrylamide gel in TBE buffer (50 mM Tris borate/5 mM EDTA, pH 8.3). This small [169-base pair (bp)] fragment was visualized by UV shadowing and was excised and extracted by the crush and soak method (26).

DNA Sequence Analysis. The *Hind*III fragments from types A and B mtDNAs were isolated from the appropriate clones (pFC111A and pFC101B), which contained the small *Bam*HI inserts. End labeling, strand separation, and sequence analysis were performed according to Maxam and Gilbert (26).

The sequences of both complementary strands of each *Hind*III fragment were determined from their 5' ends so that confirmation was obtained for each strand. Furthermore, because it was our expectation that there would be very few se-

quence differences between the two *Hind*III fragments, we repeated the sequencing procedure several times to ensure that any differences observed did not result from errors in reading the autoradiograms.

RESULTS

Experimental Design. Two types of mtDNA, termed A and B, are found in the laboratory rat population (3–8); any individual possesses only one of these types. They can be differentiated by a number of restriction enzymes, and their sequence divergences can be calculated (27) to be about 1.6% (ref. 11 and unpublished). We wished to obtain direct evidence that a variant restriction enzyme site differed in nucleotide sequence in the two DNAs. To do this, we made use of a small restriction fragment containing a variant restriction site.

A variant *Eco*RI site (type A mtDNA possesses two more *Eco*RI sites than does type B) lies between two closely spaced invariant *Hind*III sites, as can be seen in the restriction map in Fig. 1 (ref. 11; unpublished data). Thus, after *Hind*III treatment of the two DNAs, the two corresponding fragments can be isolated and their sequences can be determined. To obtain sufficient amounts of this small *Hind*III fragment, the mtDNAs were first cloned.

Cloning of the Types A and B mtDNAs: Initial Evidence Against Methylation. The types A and B *Hind*III fragments whose sequences we wished to compare are found in the smaller of the two moieties produced when mtDNA is digested with *Bam*HI, and these fragments were cloned as described. Each recombinant plasmid was shown to contain an insert of mitochondrial origin by comparison of a *Bam*HI restriction digest of this DNA to an identical digest of native mtDNA. Of the recombinant molecules screened in this way, seven contained inserts of the small (5.1-kb) *Bam*HI fragment (lane 1, Fig. 2A), while eight contained the large (10.8-kb) *Bam*HI fragment (lane 2, Fig. 2A).

In order to determine which clones were derived from type A or type B DNA, the recombinants were each digested with *Bam*HI followed by *Eco*RI. From the detailed restriction maps of these mtDNAs (Fig. 1), it was expected that the following fragments would be liberated by this double digestion: type A large *Bam*HI recombinant, eight fragments distinguished by a band of 1.7 kb; type A small *Bam*HI recombinant, three fragments; type B large *Bam*HI recombinant, seven fragments distinguished by a band at 2.45 kb; type B small *Bam*HI recombinant, two fragments. The gel patterns depicted in Fig. 2B indicate that these expected results were in fact obtained.

From these cloning results, it can be seen that the differential cleavage of types A and B native mtDNA by *Eco*RI is maintained in the cloned DNA. Thus, these findings provide valuable initial evidence about the sequence vs. methylation alternative. The results clearly speak against methylation being a factor in the loss of *Eco*RI sites in type B mtDNA; the *E. coli* strain used to grow the plasmid contains no *Eco*RI methylases. This view receives further support from treatment of the plasmid with

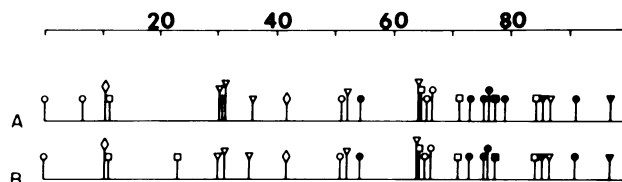


FIG. 1. Restriction maps of *R. norvegicus* types A and B mtDNA. The circular maps have been linearized and the scale (% of total length) begins at the origin of heavy strand replication. Replication proceeds from left to right. Sites are: ●, *Eco*RI; ▼, *Sst* II; ■, *Sst* I; ○, *Hind*II; ▽, *Hind*III; □, *Hha* I; ◇, *Bam*HI.

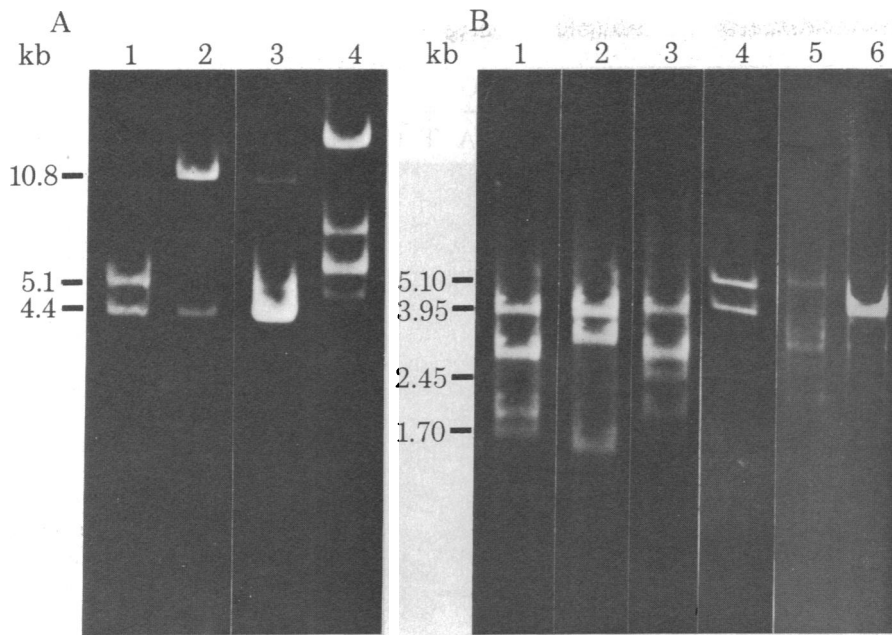


FIG. 2. Characterization of recombinant plasmid DNAs by restriction analysis. (A) Agarose gel patterns of recombinant DNAs digested with *Bam*HI. Lanes: 1, recombinant molecules containing small *Bam*HI mitochondrial insert; 2, recombinant molecules containing large *Bam*HI mitochondrial insert; 3, control *Bam*HI digest of a mixture of pBR322 and native mtDNA; 4, *Eco*RI marker digest of phage λ DNA. (B) Agarose gel patterns of recombinant plasmid DNAs from a *Bam*HI/*Eco*RI double digest. Lanes: 1, type A large *Bam*HI recombinants; 2, type A small *Bam*HI recombinants; 3, type B large *Bam*HI recombinants; 4, type B small *Bam*HI recombinants; 5, double digest control of a (nonequimolar) mixture of types A and B mtDNA; 6, double digest control of pBR322.

Hha I, *Hae* III, and *Hinf*I, enzymes which also distinguish between types A and B mtDNA. Such treatment yields identical results with native and plasmid DNAs (results not shown).

Sequence Analysis: Evidence for Sequence Differences Between Types A and B mtDNA. In order to provide conclusive evidence that nucleotide sequence variations lie at the basis of type A-B mtDNA differences, the sequences of the A and B small *Hind*III fragments obtained from plasmids pFC111A and pFC101B, respectively, were determined. Such variations were in fact found, and Fig. 3 shows one of many sequencing gels, which was run under conditions that would permit the three observed nucleotide replacements to be viewed on the same gel. These replacements are indicated in the figure and in the legend.

Several observations can be made about the complete sequence (Fig. 4). First, the size of the fragment is 169 bp. Second, the sequences of the A and B DNAs are identical except for base pair changes occurring at positions 56 (A-T \leftrightarrow G-C), 80 (A-T \leftrightarrow G-C), and 122 (T-A \leftrightarrow G-C). Finally, the change at position 80 occurs within the *Eco*RI site.

DISCUSSION

Evolutionary and other studies using restriction analysis of mtDNA have rested on the assumption that the variations detected by these enzymes amongst sets of mtDNAs are based on nucleotide sequence differences. This assumption has been tested in our work. The system used consists of cloned variant type A and B mtDNAs, from which corresponding *Hind*III fragments were isolated and their sequences were determined.

To validate the above assumption, the results of the test experiment should fulfill three predictions. First, the cloned mtDNAs should show behavior toward *Eco*RI (and toward other restriction enzymes that differentiate between the two DNAs) identical to that of native mtDNA. Second, on the basis of the 1.6% sequence divergence between the two DNAs estimated by restriction enzyme analysis, and assuming a reasonably random distribution of variant sites, we should observe 3 base substitutions in this 169-bp fragment (whose size we had previously estimated to be about 180 bp). Finally, and most critically, the sequence of the *Eco*RI recognition site contained in the type A DNA fragment should be altered to a nonrecognition sequence in the B fragment (probably by only one base substitution).

The results showed that, first, the two cloned mtDNAs behave toward *Eco*RI, *Hha* I, *Hae* III, and *Hinf*I in a differential manner which is identical to that of native mtDNAs. Second, the size of the *Hind*III fragment was found by sequencing to be 169 bp and 3 nucleotide replacements were detected. Third, the *Eco*RI recognition site was altered from G-A-A-T-T-C in the type A DNA to the nonrecognition sequence G-G-A-T-T-C in the type B molecule. Thus, the three predictions have been borne out, and we conclude that the differences between the two polymorphic forms of mtDNA result primarily from nucleotide replacements rather than methylation. The possibility exists, of course, that a few differential methylation sites could occur in the mtDNA molecule. Although 5-methylcytosine has been found to be quite common in mtDNA of animals (20), studies with the *Msp* I/*Hpa* II restriction enzyme pair have shown that none of the C-C-G-G sequences in mtDNA of rat or other species (31) have a 5-methylcytosine in the internal CpG doublet. On the other hand, the nearest neighbors of 5-methylcytosine are not known for mtDNA and therefore methylation of cytosine in other environments is still possible. Our results on *Hae* III (G-G-C-C), *Hha* I (G-C-G-C), and *Hinf*I (G-A-N-T-C), show that, at least with respect to these enzymes, polymorphism studies will not be affected by base modification.

Calculations of sequence divergence from restriction enzyme data (27, 32) are based on the assumption that the observed differences between DNAs result from single base substitutions rather than from deletions, insertions, or rearrangements. Although deletions or insertions probably do occur rarely (11), the results presented here, albeit still limited, suggest that the differences between mtDNAs arise primarily from single base substitutions.

We infer that the conclusions derived from these studies are also applicable to the many polymorphic forms of mtDNA that we and others have observed in the wild rat (10, 11) and in other animal (1, 2, 9) and human (1) populations. Studies on wild rat populations should substantiate this view.

The rapid evolution of mtDNA is evidenced by the rather high intraspecific sequence divergences that have been found by restriction analysis of various mammalian DNAs, as high as 10.5% in *R. rattus* (ref. 11; unpublished data). This raises the question of the extent, within a species, to which these nucleotide replacements are actually expressed as phenotypic

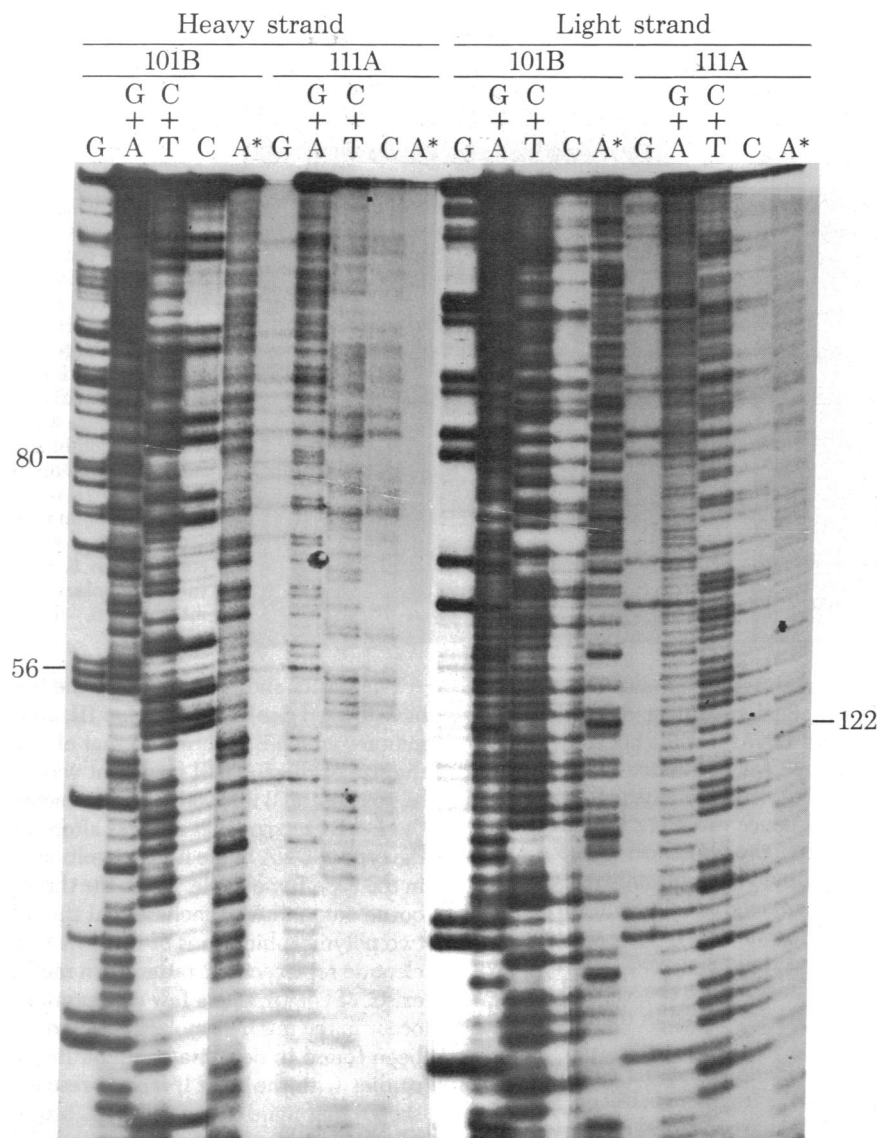


FIG. 3. Sequencing gels of corresponding small *Hind*III fragments of types A and B mtDNA. Nucleotide replacements are indicated at position 56 (see the two G columns, heavy strands), position 80 (see the two G columns, heavy strands), and position 122 (see the two C+T columns, light strands).

changes in gene products and the extent to which they may be silent. The two sequences reported here offer an opportunity of obtaining a preliminary answer to this question.

Saccone *et al.* (33) have found that a region of mtDNA that contains the 169-bp piece hybridizes to poly(A)-containing mtRNA, suggesting that the fragment whose sequence we determined is part of a coding region. We have examined both strands with respect to this possibility. The upper strand (Fig. 4) is high in G+T (60%), suggesting that this strand is the heavy strand, and it is the heavy strand that has been implicated in almost all mtDNA template activity (34–36).

Of the three possible heavy strand reading frames that might code for a hypothetical protein, the one starting at position 168 would yield an mRNA transcript that would contain multiple termination codons starting at positions 147, 144, 129, 105, 66, 60, 42, 27, and 9. The frame starting at 167 has, for mammalian mtDNA, an uncharacteristically high G content in third positions [21% as contrasted with 7% for a human mtDNA gene (28)]. Moreover, this frame would code for a relatively low level (31%) of hydrophobic amino acids (see discussion of third reading frame below). We thus exclude both these frames from further consideration. The third reading frame (starting with position 169) is low in Gs in third positions (3.6%), and possesses

no termination codons. [The two opal codons, UGAs at positions 19 and 16, probably code for tryptophan (28–30).] Moreover, this frame would code for a high level, 62%, of hydrophobic amino acids, which is characteristic of mitochondrially encoded polypeptides [compare cytochrome oxidase of humans, 57% (28), cattle, 51% (37), and yeast, 54% (30), and of yeast ATPase, 78% (38)]. Thus, this is our preferred frame, and if it is used to assess the phenotypic effects of the three variant codons, UCA ↔ UCC, AUU ↔ AUG, and GCU ↔ GCC, we see that all the changes are in the third position and that in no case would there be a change in the amino acid encoded.

Analysis of the complementary (lower) strand of the 169-bp fragment reveals only one reading frame lacking multiple termination codons. This frame, starting at position 3, could code for 30 amino acids. Two of the codons that distinguish the A and B mtDNAs fall into this region. In both of these cases, CGA ↔ CGG and UGA ↔ UGG, the changes are in the third positions of the codons and would not result in amino acid replacements. Thus, although the data are still limited, the results suggest that a high proportion of the evolutionary base changes are silent. To our knowledge, the only case involving the mitochondrial genome in which an actual comparison can be made between evolutionary base changes and resulting amino

