Length and Sequence Variation in Evening Bat D-Loop mtDNA

Gerald S. Wilkinson and Alyson M. Chapman

Department of Zoology, University of Maryland, College Park, Maryland 20742

Manuscript received November 12, 1990
Accepted for publication March 21, 1991

ABSTRACT

Length variation in D-loop mitochondrial DNA was observed after amplification with the polymerase chain reaction (PCR) in 28% of 195 evening bats, *Nycticeius humeralis*, from seven colonies. Nucleotide sequences of PCR products show that this heteroplasmy is characterized by an 81-bp region which is tandemly repeated five to eight times. Southern blots using PCR products as probes on *Hae*III genomic digests confirm the presence of heteroplasmy. Furthermore, densitometry of electrophoresed PCR products from 109 mother-offspring pairs indicate that heteroplasmy is stably transmitted from mother to offspring with one exception: a heteroplasmic offspring had a homoplasmic mother and sib. Nucleotide sequences from this family reveal that a repeat duplication and deletion occurred. The observed mutation rate per generation, μ, for length polymorphism is comparable to an independent estimate, $\mu = 10^{-2}$, based on hierarchical diversity statistics. With the exception of the repeat nearest the proline tRNA gene, sequence similarities between repeats within bats are consistent with a model of concerted evolution due to unidirectional replication slippage. Selection is inferred to act on the first repeat because in comparison to other repeats it has the least sequence divergence among bats, the fewest transversions, and the lowest minimum free energy associated with folding.

MITOCHONDRIAL DNA of most animals ranges in size from 16 to 18 kb and contains 13 protein genes, 22 transfer RNAs, 2 ribosomal RNAs, and a regulatory region known as the control region in invertebrates or the displacement loop (D-loop) in vertebrates. The D-loop lies between the phenylalanine tRNA ($tRNA^{Phe}$) and the proline tRNA ($tRNA^{Pro}$) and is so named because the two parent strands are displaced by a short, variable length replication product known as 7S mtDNA (CLAYTON 1982). The presence of this single-stranded component enables this region of the mtDNA molecule to bind hydrophobic regulatory proteins (ALBRING, GRIFFITH and ATTARDI 1977) and to undergo duplication and deletion events when repeats are present (BUROKER *et al.* 1990). Comparison of rat, mouse, human, cow and *Xenopus laevis* D-loop sequences reveal that most of the D-loop is A + T rich with the exception of a central G + C rich region containing an open reading frame that shows substantial similarity across species at the level of amino acid function (SACCONE, ATTIMONELLI and SBISA 1987).

Substantial length variation has recently been found in or near the control region or D-loop of many animal mtDNAs. For example, three species of bark weevils possess mitochondria ranging in size from 30 to 36 kb (BOYCE, ZWICK and AQUADRO 1989) because an 0.8–2.0-kb sequence is tandemly repeated adjacent to the control region. All individuals sampled had more than one mtDNA form, *i.e.* were heteroplasmic. Heteroplasmy has also been attributed to variable copy number of tandem repeats in the control region of sea scallop (LA ROCHE *et al.* 1990), *Drosophila mauritiana* (SOLIGNAC, MONNEROT and MOUNOLOU 1986), and field crickets (RAND and HARRISON 1989), and in the D-loop of shad (BENTZEN, LEGGETT and BROWN 1988), sturgeon (BUROKER *et al.* 1990), whiptail lizards (MORITZ and BROWN 1987), and rabbit (MIGNOTTE *et al.* 1990). In humans, heteroplasmy is rare but has been detected in association with several disorders as an 8-kb duplication encompassing the D-loop (POULTON, DEADMAN and GARDINER 1989) and as a series of multiple deletions within the D-loop (ZEVIANI *et al.* 1989). Length variation between humans, in contrast, is well-known for D-loop mtDNA (CANN and WILSON 1983; GREENBERG, NEUBOLD and SUGINO 1983). The apparent scarcity of heteroplasmy among mammals has prompted some investigators (*e.g.* RAND and HARRISON 1989) to speculate that homeotherms, due to their higher metabolic rates, may experience stronger selection for smaller and less variable mtDNAs than poikilotherms.

In addition to length variation, substantial nucleotide sequence variability in the D-loop has been recorded both within (AQUADRO and GREENBERG 1983) and between species (FORAN, HIXSON and BROWN 1989). Because the polymerases for both replication and transcription are nuclear-coded, the promoters for transcription of both strands lie between the open reading frame of the D-loop and the $tRNA^{Phe}$, and

replication of the heavy (H) strand is primed by D-loop light (L) strand RNA (CHANG and CLAYTON 1985), several workers (GREENBERG, NEUBOLD and SUGINO 1983; BROWN 1985; CHANG *et al.* 1985; FORAN, HIXSON and BROWN 1989) have claimed that the species-specificity of the D-loop sequence is evidence for nuclear-mitochondrial genome coevolution. This interpretation assumes that mitochondrial D-loop nucleotide sequences have been under selection for their ability to facilitate nuclear enzyme binding. An alternative, although not exclusive, hypothesis is that D-loop sequence variability between species may be the result of concerted evolution, *i.e.* the creation of tandemly repeated sequences through replication slippage or unequal crossing over (OHNO 1970; HILLIS *et al.* 1991). Concerted evolution could generate species-specific D-loop sequences independent of functional differences if separate lineages accumulate substitutions independently.

In this paper we use the polymerase chain reaction (PCR) to characterize length and sequence variation in the D-loop of evening bat (*Nycticeius humeralis*) mitochondrial DNA. By examining PCR product sizes and sequences between known mother-pup pairs we verify that length heteroplasmy is transmitted maternally and is caused by a duplication or deletion of an 81-bp sequence that is tandemly copied five to eight times in the D-loop. By convention we use the term heteroplasmy to refer to multiple mtDNA genomes within an individual rather than within a cell. But, in contrast to prior studies in which length variation is assessed using restriction enzyme digests of DNA extracted from an organ or the entire animal, we can localize heteroplasmy to relatively few cells within an individual using PCR. By comparing nucleotide sequence similarity between 81-bp repeats within bats to sequence similarity of corresponding repeats in different bats, we evaluate concerted evolution and nuclear-mitochondrial coevolution as possible alternative explanations for creating and maintaining the substantial nucleotide sequence variation observed between bats both between and within nursery colonies.

MATERIALS AND METHODS

Study sites: Tissue samples were taken from adult female evening bats, a small (10 g) insectivorous Vespertilionid bat, at seven attic nursery colonies (Table 1), six in northern Missouri and southern Iowa and one in North Carolina (G. S. WILKINSON, unpublished) during the summer between 1987 and 1990. At these sites, females migrate in April from winter hibernacula and faithfully return to their natal nursery colonies in either hollow trees or attics (WATKINS 1970).

To document transmission of length and sequence variants, 58 pregnant females were kept in captivity through parturition until their young were large enough to sample. In 1988 we kept ten bats and in 1989 nine bats from the Hutton colony and in 1989 we also kept nine bats from the Grim colony. In 1990 we kept 30 females captured in an attic in Edenton, North Carolina. All bats were released near the site of capture before the young were old enough to feed independently.

Sample collection and preparation: Bats were captured with hoop or mist nets as they departed from attics at dusk. After banding and measuring, each adult bat's chest fur was clipped, a 3-mm excision was made, and approximately 1–3 mg of pectoral muscle was excised while applying ethylene chloride topically as a local anesthetic. Tissue samples were stored in liquid nitrogen until returned to the laboratory where they were kept at −80° until DNA was extracted.

Muscle biopsies were minced in 300 μl of buffer (0.05 M Tris/HCl pH 8.0, 0.1 M EDTA, 0.1 M NaCl, and 1% SDS), incubated overnight at 55° with proteinase K (0.5 μg/ml), and then kept at 37° for 1 h with RNase (0.1 μg/ml) prior to several phenol/phenol:chloroform:isoamyl/chloroform:isoamyl alcohol extractions (SAMBROOK, FRITSCH and MANIATIS 1989). The resulting supernatant contained 50–1000 ng of genomic DNA and was purified and concentrated using Centricon-30 microconcentrators following manufacturer's instructions (Amicon Division, Danvers, Massachusetts).

PCR primers and reactions: Three pairs of 20-bp sequences were used as primers to amplify and sequence mtDNA (Figure 1). The location, abbreviation, sequence and orientation (L or H) of each strand is cytochrome B (C): L, 5′-TGAATTGGAGGACAACCAGT-3′, $tRNA^{Pro}$ (P): L, 5′-TCCTACCATCAGCACCCAAAGC-3′, initiation of the repeat array (I): L, 5′-TGAAAAAACTACACA-CATGTAC-3′, termination of the repeat (T): H, 5′-T-TGACTGTATGGGGTATG TAC-3′, conserved sequence block F (F): H, 5′-GTTGCTGGTTTCA-CGGAGGTAG-3′, and conserved sequence block E (E): H, 5′-CCTGAAGTAGGAACCAGATG-3′. Conserved sequence blocks (CSB) are highly homologous regions that have been found in dolphin, cow, human and mouse mtDNA (SOUTHERN, SOUTHERN and DIZON 1988). While the I and T primers are specific for evening bats, the other two pairs of primers will amplify D-loop mtDNA from at least five families of bats (G. S. WILKINSON and A. M. CHAPMAN, unpublished).

Double-stranded amplifications using the PCR were performed following standard protocol (Perkin-Elmer-Cetus, Norwalk, Connecticut) and included a control with no template DNA. Forty cycles of 95° for 1 min, 55° for 1.5 min and 72° for 2 min were followed by 7 min at 72°. PCR product size was determined by agarose gel electrophoresis and ethidium bromide staining (see Figures 2 and 3).

To quantify the relative amount of each PCR product, Polaroid photographs of UV illuminated gels were digitized with a video camera, 16-bit frame-grabber board, and Macintosh computer and then measured using the program Image 1.24. The average inverse pixel value within an 8 by 40 pixel rectangle positioned over each band was recorded for each fragment and for an adjacent background sample. To compute relative frequencies, the background value was subtracted before normalizing each fragment value for an individual to sum to one. Only relative frequencies greater than 0.05 are recorded. This technique provided a rapid and highly repeatable measure of fragment frequency.

Southern blot analysis: To corroborate estimates of heteroplasmy based on PCR products we digested purified genomic DNA with *Hae*III. This enzyme cuts the mtDNA outside each end of the tandem repeat array. Digested DNA was electrophoresed through 1.0% agarose, blotted to Zeta-Probe (Bio-Rad, Richmond, California) nylon and probed with a PCR product obtained using the C and F primers.

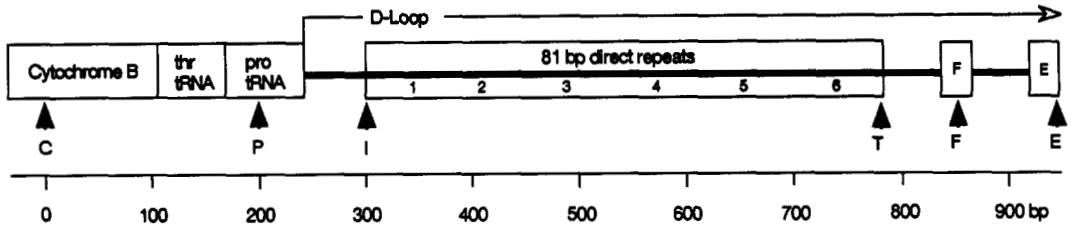

FIGURE 1.—Locations of PCR primers (arrows) and tandem repeats within evening bat mtDNA. See text for primer sequences.

The probe was labeled with ^{32}P using a random primer labeling kit (United States Biochemical, Clevelend, Ohio). Prehybridization and hybridization were carried out using Denhardt's reagent as a blocking agent (SAMBROOK, FRITSCH and MANIATIS 1989). Mitochondrial DNA fragments that hybridized to the PCR probe were visualized after 3 days of exposure to x-ray film.

DNA sequencing: Sequencing of double-stranded (ds) and single-stranded (ss) DNA was performed by the dideoxy chain termination method (SANGER, NICKLEN and COULSON 1977) using a Sequenase kit (United States Biochemical). Double-stranded DNA was obtained from PCR products that had been cut from an agarose gel and purified using glass beads (GENECLEAN, BIO 101 Inc., La Jolla, California). Single-stranded DNA was created either by denaturing dsDNA or by asymmetrical PCR (GYLLENSTEN and EHRLICH 1988). Single-stranded binding protein (SSB, 0.5 μg) was added to the labeling mix to eliminate compression zones in the gel. The SSB was inactivated by incubating each termination reaction at 95° for 15 min with 0.1 μg of proteinase K. The sequencing reaction products were separated in an 8.0% acrylamide/urea gel for which the top buffer was 0.5 × TBE (SAMBROOK, FRITSCH and MANIATIS 1989) while the bottom buffer was 1 × TBE and 3 M sodium acetate in a 2:1 ratio. These procedures allowed us to score approximately 350 bp from each 50-cm lane.

Sequences were obtained starting with either the C or P primer and moving into the D-loop or using the E or T primers and sequencing toward the $tRNA^{Pro}$ gene. These primer pairs allowed us to sequence complimentary strands through the region of overlap in the middle of the tandem repeats and provided an average of 64 bp of single-copy sequence between the C primer and the I primer that included part of the cytochrome *b* gene and the $tRNA^{Thr}$ gene. Substantial sequence similarity between evening bat and other mammalian sequences for these genes confirmed that we were amplifying and sequencing mtDNA.

Statistics: To apportion the variability in fragment number within individuals, among individuals within colonies, and among colonies we used the diversity indices (BIRKY, MARUYAMA and FUERST 1983), $K_i = 1 - S\, x_{ij}^2$, where x_{ij} is the frequency of the *j*th size class in the *i*th level. In this study we consider three levels for *i*: individual, colony and region. To maintain consistency with BIRKY MARUYAMA and FUERST (1983) and RAND and HARRISON (1989) we denote the diversity index within individuals as K_b, among individuals within colonies as K_c, and among colonies within regions as K_d. To quantify relative variation in diversity at each level we used three *C* statistics (RAND and HARRISON 1989). The fraction of diversity found within individuals is given by $C_I = \bar{K}_b/K_d$. The fraction among individuals within colonies is $C_{IC} = (K_c - \bar{K}_b)/K_d$ and the fraction among colonies within regions is given by $C_{CR} = (K_d - \bar{K}_c)/K_d$.

To test for differences in genotype or size class frequencies among colonies we used a Monte Carlo randomization procedure (ROFF and BENTZEN 1989) for computing the significance associated with obtaining a contingency chi-square value as large or larger than that observed. This technique permits inclusion of categories in which cell values are small by randomizing elements while keeping row and column totals constant.

To assess the difference between nucleotide sequences from the same repeat we calculate the expected number of substitutions per site using the Jukes-Cantor distance,

$$d = \tfrac{3}{4} \ln (3/(4q - 1))$$

where q is estimated as the fraction of homologous nucleotides which have the same base. We use this distance metric rather than $1 - q$ because q overestimates true similarity due to multiple substitution events and because d scales linearly with time if substitution rates are equivalent at all sites (JUKES and CANTOR 1969).

We apportion the similarity between sequences to different regions analogous to the diversity statistics described above by computing the average number of nucleotide substitutions either between pairs of bats from the same colony, from different colonies or from different regions. The proportion of DNA divergence attributable to each level is then found by taking the difference in average substitution rates and scaling by the substitution rate for bat pairs from the highest level, *i.e.* in different regions (NEI 1987).

Descriptive statistics are given as mean ± one standard error.

RESULTS

Repeat inheritance and mutation: PCR resulted in amplification products that varied in length both within and between individuals (Figures 2 and 3). Direct sequencing of each product revealed that fragments differ in size due to the addition or deletion of an 81-bp sequence that is tandemly repeated five to eight times in the D-loop (Figure 1). Southern blot analysis of *Hae*III digested genomic DNA obtained from muscle tissue and probed with PCR product confirmed the same pattern of fragments among and within individuals as obtained by PCR (Figure 2).

PCR band patterns are very similar among females and pups from the same family (Figure 3). To quantify this similarity we estimated the repeatability, R, within families of the arcsin-square root transform of the six repeat frequency obtained by densitometry. With an average of 2.9 bats measured in each of 58 families, $R = 0.989 \pm 0.003$ (BECKER 1975). Densitometric estimates of six repeat frequencies in pups closely correspond to those of their mothers (Figure 4) with one exception. The box in Figure 4 indicates an offspring that produced two amplification products

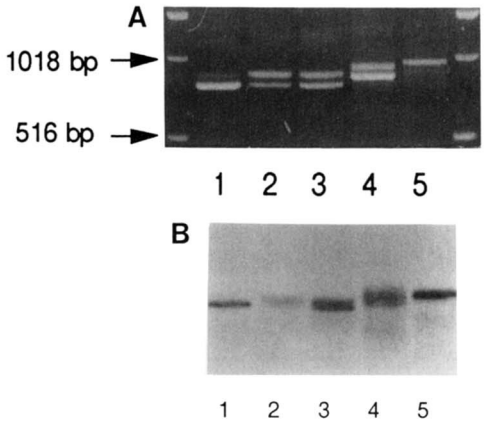

FIGURE 2.—PCR products using the C-E primer pair (A) and Southern blots of *Hae*III-digested genomic DNA probed with a P-E PCR product (B) from five bats. Lane 1 contains a blot from a bat that was homoplasmic for five repeats, lanes 2 and 3 contain siblings' blots that were heteroplasmic for five and six repeats, lane 4 contains a blot of a bat heteroplasmic for six and seven repeats, and lane 5 contains a blot of a bat homoplasmic for seven repeats. Outside lanes in (A) contain length standards.

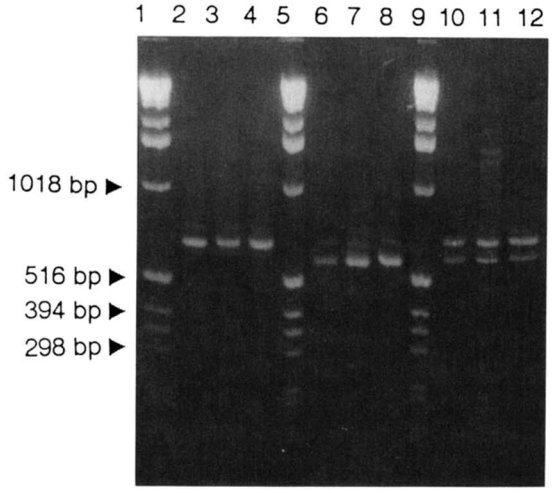

FIGURE 3.—PCR products from three families resulting from amplification using the P-E primer pair. Lanes 1, 5 and 9 are length standards. Lanes 2, 6 and 10 contain samples from adult females while the two lanes following them contain samples from their two pups. Lanes 10–12 display heteroplasmy for fragments with five and six 81-bp repeats.

while its mother and sibling produced just one. Thus, a length mutation must have occurred in the mother. Given that we scored 109 meiotic events, the effective rate at which such mutations can be scored in each zygote by PCR is $1/109 = 0.0092 \pm 0.0091$. Note that this is not the mutation rate per mitochondrion because we have not assessed the number or distribution of mitochondrial types per gamete. Furthermore, this is a lower estimate of mutation rates per zygote because PCR cannot detect simultaneous duplication and deletion events. As shown below, these do occur.

Because each maternal repeat had a characteristic pattern of base substitutions, we were able to deter-

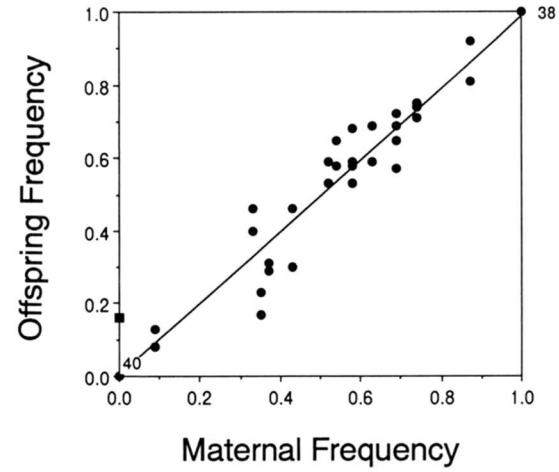

FIGURE 4.—Frequency of six repeat PCR products for 109 mother-offspring pairs. The box indicates an offspring which was heteroplasmic with a homoplasmic mother. Numbers indicate replicate values for homoplasmic pairs. The least squares regression equation is $y = 0.0029 + 0.986x$, $r^2 = 0.98$.

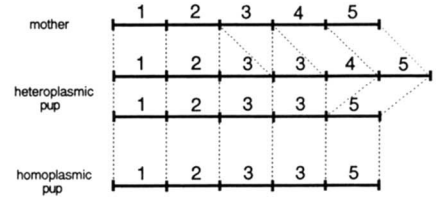

FIGURE 5.—Diagram of the positions of identical repeat sequences in the one family in which a homoplasmic mother had a heteroplasmic offspring. Note that repeat 3 duplicated in all offspring arrays while repeat 4 was deleted in the 5 repeat offspring fragments.

mine how the length change in this family occurred. The six repeat offspring fragment is one repeat longer than the maternal fragment because repeat three has been duplicated (Figure 5). Furthermore, the five repeat fragment for the heteroplasmic pup and its homoplasmic sib share this duplication but lack maternal repeat four. Thus, the five repeat fragments of both pups also show a deletion and concordant duplication of internal maternal repeats while the six repeat fragment of the heteroplasmic pup is due just to a duplication.

Over 500 bp of nucleotide sequence for five other families, *i.e.* all repeats for a mother and her two pups, were also obtained. In all mother-offspring comparisons, including a five repeat family and a seven repeat family, both pup sequences were identical to the maternal sequence.

Occurrence of heteroplasmy: Of the 195 adult bats sampled, 27.7% of the individuals were heteroplasmic (Table 1). Only three heteroplasmic animals had evidence of three different size classes; the remaining heteroplasmic individuals contained two size classes (Table 2). Although the North Carolina colony sample appears to have a lower incidence of heteroplasmy than most of the Missouri colonies, this difference is

TABLE 1

Frequency of mtDNA size classes and heteroplasmy, and diversity indices among sites

Site (n)	$p(5)$	$p(6)$	$p(7)$	$p(8)$	$p(H)$	$\bar{K}_b$	K_c	Arcsin$\sqrt{p}(6)$
Total (195)	0.313	0.643	0.040	0.004	0.277	0.125	0.487	1.005 ± 0.046
Missouri (166)	0.279	0.677	0.039	0.004	0.295	0.133	0.462	1.059 ± 0.048
Busby (17)	0.218	0.714	0.068	0.000	0.294	0.126	0.439	1.103 ± 0.148
Grim (14)	0.268	0.696	0.036	0.000	0.357	0.186	0.443	1.064 ± 0.149
Easton (37)	0.264	0.693	0.044	0.000	0.324	0.139	0.449	1.081 ± 0.098
Smith (23)	0.133	0.705	0.130	0.031	0.304	0.135	0.467	1.114 ± 0.132
Zion (48)	0.205	0.790	0.005	0.000	0.292	0.135	0.333	1.243 ± 0.071
Hutton (27)	0.601	0.397	0.002	0.000	0.222	0.094	0.481	0.627 ± 0.134
North Carolina (29)	0.509	0.446	0.045	0.000	0.172	0.080	0.540	0.692 ± 0.134

Site indicates the location where the sample was obtained, *i.e.* either state or colony. n = sample size, $p(5)$ = frequency of the size class containing 5 tandem repeats, $p(H)$ = frequency of heteroplasmic individuals, $\bar{K}_b$ = mean diversity index over all individuals at that site, K_c = diversity index for that site, and arcsin$\sqrt{p}(6)$ = mean arcsin$\sqrt{p}(6)$ over all individuals.

TABLE 2

Frequencies of mtDNA genotypes at each site

Site (n)	$f(5)$	$f(6)$	$f(7)$	$f(56)$	$f(67)$	$f(567)$	$f(78)$
Total (195)	40	97	3	38	12	3	2
Missouri (166)	27	87	2	34	11	3	2
Busby (17)	3	9	0	1	3	1	0
Grim (14)	2	6	0	4	2	0	0
Hutton (27)	13	8	0	5	0	1	0
Easton (37)	5	19	1	9	3	0	0
Smith (23)	1	14	1	3	2	0	2
Zion (48)	3	31	0	12	1	1	0
North Carolina (29)	13	10	1	4	1	0	0

Site indicates the location where the sample was obtained, *i.e.* either state or colony. n = sample size $f(56)$ = number of individuals which are heteroplasmic for 5 and 6 tandem repeats, $f(5)$ number of individuals which are homoplasmic for 5 repeats.

not sufficient to cause heteroplasmy frequency to vary significantly among colony sites ($X^2 = 1.77$, d.f. = 6, $P = 0.939$).

These results are consistent both with an analysis of variance on K_b and with the hierarchical C statistics. The within individual diversity, K_b, does not differ significantly among colonies ($F_{6,188} = 0.596$, $P = 0.733$). In terms of relative diversity, the variation among colonies, $C_{CR} = 7.2\%$, is much less than either the diversity among individuals within colonies, $C_{IC} = 66.8\%$ or the diversity within individuals, $C_I = 25.6\%$.

Population subdivision: In contrast to the lack of variation in the proportion of heteroplasmic individuals within each colony, significant variation does occur among colonies in genotype frequencies ($X^2 = 70.431$, $P < 0.001$, 1000 randomizations, Table 2) and in the arcsin-square root transform of the frequency of six repeats ($F_{6,188} = 4.523$, $P = 0.0003$, Table 1). Note that the genotype frequencies only score the presence or absence of a size class while the frequency of six repeats is calculated from the individual frequencies obtained from densitometric measurements. Application of Fisher's partial least squares differences (PLSD) posthoc comparison to the arcsin-square root transformed frequencies indicates that this variation is caused by two colonies, Hutton and the North Carolina colony, which have lower frequencies of six and higher frequencies of five repeats than the remaining Missouri colonies (Table 1). Similar results are obtained for genotype frequencies upon posthoc comparison (Table 2). The genotype frequencies at the Hutton and Edenton colonies are independent of colony ($X^2 = 3.266$, $P = 0.91$, 500 randomizations) as are those at the five other Missouri colonies ($X^2 = 29.256$, $P = 0.17$, 500 randomizations). Thus, although differences in both genotype and repeat frequency can be demonstrated between colonies, they do not increase with geographic distance.

Sequence similarity among sites: To determine if there were more nucleotide substitutions between bats from different colonies or different regions than within colonies we computed the Jukes-Cantor distance between all possible pairs of the 52 unrelated bats for which sequence data was obtained for six repeats. This gave 549 pairs of bats from the same colony, 630 pairs from two different colonies in the same region, and 147 pairs from different regions–one from a Missouri colony and the other from the North Carolina colony. We tested for differences between these three sets of pairs by taking 1000 samples of 549 pairs of distances at random without regard to the identity of the bats, ordering the means, and then counting the number of sampled means less than the observed mean obtained from bats, in this case, from the same colony. Identical randomization tests were conducted with 630 pairs and 147 pairs to determine if two bats from different colonies in the same state or two different states, respectively, have more or fewer nucleotide substitutions than expected by chance.

Although very little sequence variation exists among bats between the cytochrome *b* and the $tRNA^{Pro}$ gene sequences, there were some differences between individuals in the number of adenosine nu-

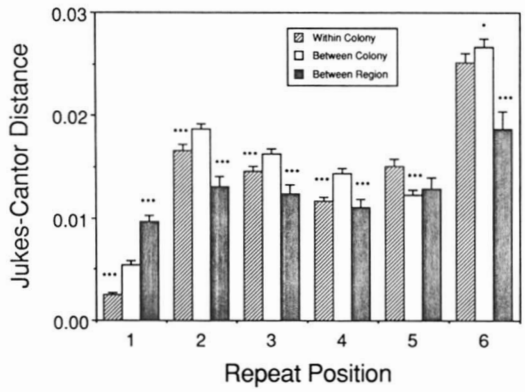

FIGURE 6.—Jukes-Cantor distances between pairs of bats in the same colony, in different colonies in Missouri, or between bats from Missouri and North Carolina colonies. The distances for pairs within a colony include sequences for 33 bats at the Zion colony, four bats from the Hutton colony, and three bats at each of the other five colonies. Sequences were obtained either from bats homoplasmic for six repeats or from the six repeat fragment of a heteroplasmic bat. Asterisks indicate those sets of paired distances that were significantly different from the overall mean distance according to randomization tests as described in the text. Repeat position 1 is closest to the $tRNA^{Pro}$ gene.

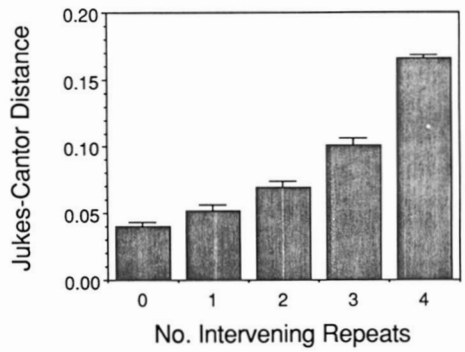

FIGURE 7.—Jukes-Cantor distances between the sequences of two 81-bp repeats within each of 52 bats. The number of repeats separating the two repeats is indicated on the abscissa. Sequences were obtained either from bats homoplasmic for six repeats or from the six repeat fragment of a heteroplasmic bat. Error bar indicates one standard error obtained by bootstrapping.

cleotides at the beginning of the D-loop. Most individuals had seven adenosine residues at this position, but both six and eight were recorded. However, randomization tests indicated that no significant differences exist between any of the three sets of pairs and the overall mean. The overall average Jukes-Cantor distance between bat pairs was 0.0044 ± 0.0003. In contrast, if the sequence data from all six repeats is pooled, a significant difference is obtained for pairs of bats within colonies versus pairs between colonies or between regions ($P < 0.001$). The average Jukes-Cantor distance within the repeat region is 0.0143 ± 0.0002 indicating greater sequence divergence between bats within the repeat region than outside the repeats. To determine if any nucleotide sequence heterogeneity exists among the six repeats, we analyze each repeat separately below.

Although heterogeneity exists among the mean distances for all repeats depending on whether the pairs were from the same or different colonies, the most instructive difference between these groups occurs in repeat one, that repeat closest to the $tRNA^{Pro}$ gene (Figure 6). This repeat is the most conserved repeat in that it shows significantly less divergence between pairs of bats than any of the other repeats. In comparison to the within repeat nucleotide substitution rate between regions, 25.5% of the genetic differentiation observed at repeat one is due to variation within colonies, 30.9% is due to variation between colonies, and 43.6% is due to sequence variation between regions. In contrast, repeat six shows much greater divergence between bats, although bats within the same colony do not have consistently more similar sequences at repeat six than bats from different colonies (Figure 6). Internal repeats show distances between individuals which are intermediate to those for repeats one and six, and also fail to show consistently greater divergence between bats from different colonies.

Sequence similarity among repeats within bats: If either unequal crossing over or replication slippage causes heteroplasmy, then the greatest sequence similarity should occur between those repeats that most frequently undergo duplication events (OHTA 1980). Under this premise, duplication almost certainly is restricted to neighboring repeats because of the 52 bats sequenced, adjacent repeats are most similar and the distances between repeat sequences diverge at an exponentially increasing rate as the number of intervening repeats increases (Figure 7).

Because the adjacent repeat category in Figure 7 pools all pairs of adjacent repeats together, any effect of repeat position on adjacent repeat sequence similarity is obscured. Figure 8 shows that repeat location does affect the similarity between adjacent repeats. Adjacent pairs of repeats in the middle of the tandem repeat region show the most sequence similarity whereas adjacent repeats at each end of the repeat region show significantly greater distances. However, this divergence in sequence similarity is asymmetrical in that the repeat furthest from the $tRNA^{Pro}$, repeat six in most individuals, shows much greater differentiation from its neighboring repeat than does the repeat closest to the $tRNA^{Pro}$. Note, however, that repeat one is much less similar to repeat two than internal adjacent repeats. Thus, repeat one is unlikely to be the result of a recent duplication event.

Substitution bias and location within repeats: Base pair substitutions have not occurred at random among repeats with respect to base pair identity or position. Within the consensus repeat sequence, *i.e.* that sequence obtained by using the base pair observed most often in all bats at each of the 81 positions, there are

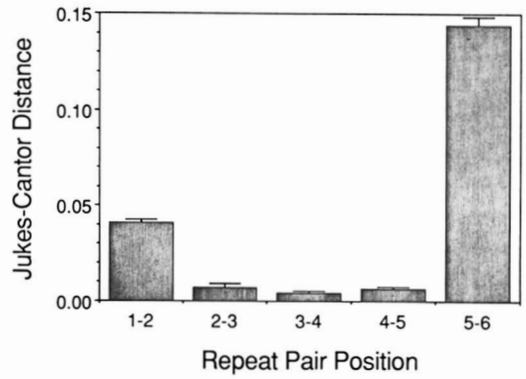

FIGURE 8.—Jukes-Cantor distances between adjacent 81-bp repeats within each of 52 bats. Sequences were obtained either from bats homoplasmic for six repeats or from the six repeat fragment of a heteroplasmic bat. Repeat positions are the same as in Figure 6. Error bar indicates one standard error obtained by bootstrapping.

TABLE 3

Substitution bias within each repeat using all pairs of 52 sequences

Repeat	Transversions	Transitions	Percent Transversions
1	0	497	0
2	100	1718	5.5
3	192	1411	12.0
4	194	1130	14.6
5	378	1050	26.5
6	1253	784	61.5
Total	2117	6590	32.1

5 (6.1%) guanosine, 9 (11.1%) cytosine, 29 (35.8) thymidine and 38 (46.9) adenosine bases. Given these base pair frequencies, transversions should represent 78.4% of all substitutions. Instead, only 32.1% of all substitutions are transversions. Transversions increase progressively from 0% in repeat one to 61.5% in repeat six (Table 3).

Furthermore, the distribution of transversions is not random with respect to nucleotide position within a repeat (Figure 9). The transversions in repeats two through five are all due to an A-T substitution at position 40 in the repeat. Although substitutions have occurred at least once at 21 of the 81 sites in a repeat (Figure 9) only 11 sites have altered in repeats two through five. Nine of those 11 sites have also changed in repeat six, but only two of the 11 sites have ever changed in repeat one. Both repeat one and repeat six have substitutions at an additional five sites each. Some, but not all, of these substitutions occur in bulge areas or are paired with complimentary substitutions in stem regions of secondary structures (Figure 10) that can be computed, as described below, for each repeat.

Repeat secondary structures: The most common sequence of the L strand first repeat, as well as all other repeat sequences, forms a secondary structure (Figure 10) with a stem and terminal loop when the sequence is folded to minimize the free energy of the structure (ZUKER and STIEGLER 1981). Comparison of the minimum free energies associated with all first, second and end repeats, of which there are 9, 13 and 17 haplotypes, respectively, shows a significant difference among the mean binding energies ($F_{2,155}$ = 68.14, P = 0.0001). Fisher PLSD tests indicate that significant differences exist between all three repeats. Note, however, that repeat one shows noticeably lower average free energies than either repeat two or six (Table 4).

The sequence in the terminal loop of all structures formed is of interest because it includes the first 7 bp, 5′-ACATAAA-3′, which DODA, WRIGHT and CLAYTON (1981) identified in humans as being 51–53 bp upstream from the termination of replication of the 7S daughter strand. In the mouse, there are four such termination-associated sequences (TAS), one of which is the same as the human, and the other three differ by 1 bp, *i.e.* 5′-ACATTAA-3′. The 7-bp human TAS also appears along one side of the stem in the secondary structures. The probability of finding these 7 bp together anywhere in the 81-bp repeat, given the frequency of each nucleotide in the most common first repeat, is 0.069. The probability of finding these 7 bp in the terminal loop is 0.00086. Thus, the consistent location of this 7-bp sequence in the terminal loop can be considered a nonrandom event.

As with sturgeon mtDNA repeats (BUROKER *et al.* 1990), lower free energies are obtained when more than one repeat is allowed to fold into a secondary structure. For example, folding repeat one and two for one bat gave a minimum free energy of −25.4, repeats one through three gave −40.9, repeats one through four −51.2, and repeats one through five −64.4. All of these structures were two-branched stem-loops with 5′-ACATAA-3′ in each terminal loop.

DISCUSSION

The results presented in this paper demonstrate that PCR can be used to quantify length variation in mtDNA among populations. The advantages of this technique over traditional restriction fragment length polymorphism analysis are several: PCR is quicker, does not require radioactivity, can be conducted on minute samples obviating the need to sacrifice small animals and can be used for direct sequencing. Furthermore, estimates of sequence differences between individuals based on restriction maps may be misleading if there is duplication and deletion of repeats in a tandem array. Frequent duplication/deletion events will cause substitution rates to appear higher than they really are. On the other hand, the presence of substantial length and nucleotide variation means that

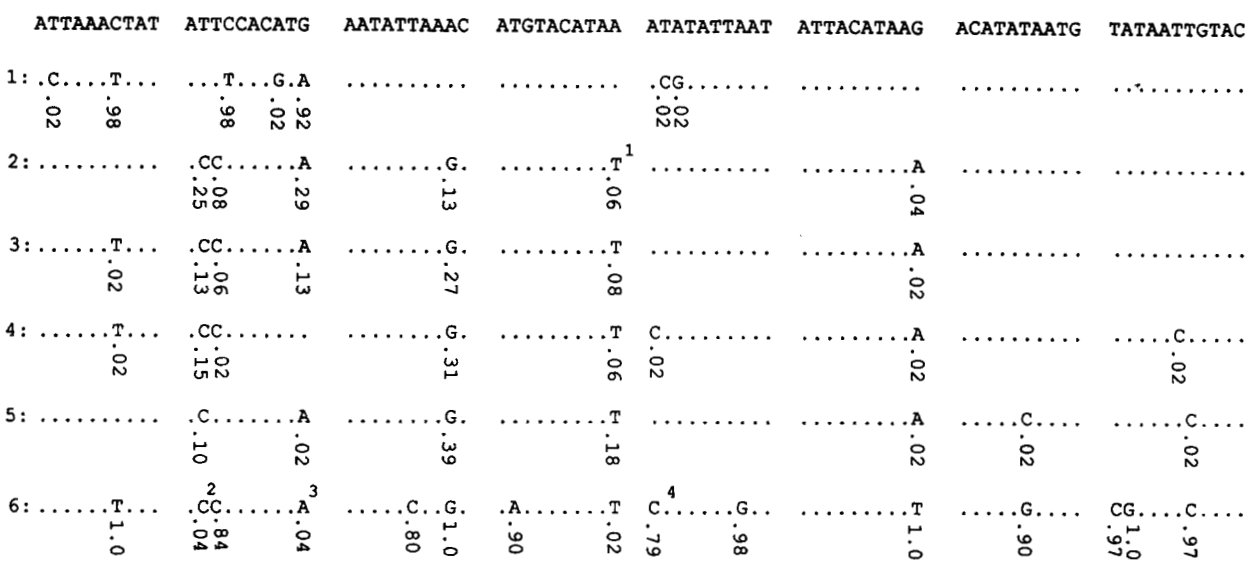

FIGURE 9.—Location and frequency of substitutions within the 3′ to 5′ consensus sequence of the light strand of the 81-bp repeat. The proportion of the 52 bats carrying a particular substitution in each repeat is indicated below the respective base. Each substitution involved only one base pair change as indicated with four exceptions: [1]) one of three changes was A to G, [2]) one of two changes was T to A, [3]) one of two changes was G to C, and [4]) one of 39 changes was A to G.

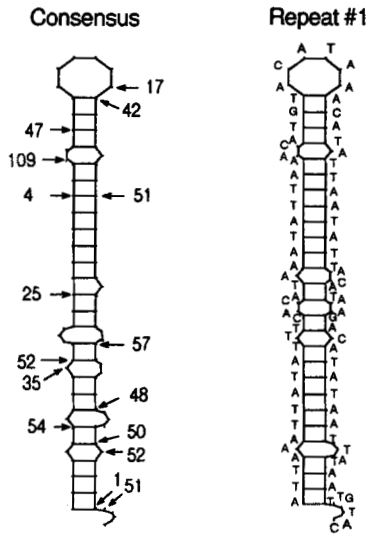

FIGURE 10.—Secondary structures for the most common first repeat and the consensus sequence obtained by minimizing the free energy according to the Zuker-Stiegler method. The location and number of substitutions observed among repeats two through six for 52 bat sequences are indicated on the consensus repeat.

TABLE 4

Minimum free energies associated with stem-loop secondary structures of each repeat

Repeat	Mean (kcal/mole)	SE	No. of different structures
1	−9.78	0.09	9
2	−8.03	0.13	13
6	−7.44	0.22	17

PCR of mtDNA can be used in parentage studies when maternal relationships are unknown.

Utility for matriline assignment: A high mutation rate for length variation reduces genetic differentiation of colonies. In the evening bat D-loop, mutation rates for length polymorphism are too high to allow fixation of different length variants in isolated populations, but not high enough for most of the variation to be present within an individual. Thus, the frequency of mtDNA size classes within an individual can be used as a first approximation to assign individuals to matrilines because the most likely source for a length variant is through a common maternal ancestor rather than from a mutational event. However, a more accurate classification can be made by comparing the sequence patterns within repeats because there are many more unique repeat genotypes than length variants and the per generation mutation rate for a nucleotide substitution in each repeat is much smaller than the mutation rate for length polymorphism. Thus, by compiling a catalog of repeat haplotypes and then comparing the sequences of the repeats, rather than the nucleotides, it is not only possible to identify individuals to matrilines (G. S. WILKINSON unpublished), but also to track repeat evolution.

Estimation of length mutation rates: The existence of up to eight copies of a direct repeat suggests that mammal mtDNA may not experience strong selection for reduced mtDNA size as has been proposed (HARRISON 1989). Because PCR has only recently been used to assess mtDNA length variation, we predict that more examples of heteroplasmy in mammals will be found. For example, in addition to evening bats, we have found tandem repeats and similar frequencies of length heteroplasmy in two other Vespertilionid

bats, *Myotis lucifugus* and *Eptesicus fuscus* (G. S. WILKINSON and A. M. CHAPMAN, in preparation). Rabbits exhibit similar length variation except that the tandem repeats lie between the CSBs and the $tRNA^{Phe}$ (MIGNOTTE *et al.* 1990). Although the frequency of heteroplasmy in evening bats, 28%, is lower than that estimated in *Gryllus firmus*, 60%, or *Gryllus pennsylvanicus*, 45%, (RAND and HARRISON 1989), the diversity indices of all hybrid crickets, $K_c = 0.471$ and $\bar{K}_c = 0.149$, are very similar to the diversity indices for all evening bats, $K_c = 0.487$ and $\bar{K}_b = 0.125$. These indices show that most of the length variation, 66.8% in evening bats and between 61.3 and 67.4% for crickets, occurs within a colony, or in the case of crickets, lineage. To maintain this level of length variation in the absence of selection, the mutation rate for length polymorphism must be relatively high and/or the population size must be large to counteract the effects of random assortment of mitochondria into gametes which will rapidly remove variation unless the pool of mitochondria per gamete is very large (CLARK 1988).

If the effective number of females is known and transmission is strictly maternal, then $K_c = 2N\mu/(2N\mu + 1)$ where N is the number of breeding females in the population and μ is the mutation rate per individual (BIRKY, MARUYAMA and FUERST 1983). If we substitute the observed mutation rate of 0.0092 and the observed K_c of 0.487 into this equation then $N = 52$. Because this estimate of population size is very close to the average number of adult females in the seven colonies we sampled, we conclude that the per generation mutation rate is near 9×10^{-3}. This mutation rate is about 20 times greater than that estimated for field crickets (RAND and HARRISON 1989).

The length mutation event described in this paper required both a duplication and a deletion event. Initially, a duplication must have occurred in repeat three which presumably affected at least some of the maternal mtDNA molecules in the ovaries. In other words, the mother must have had different mtDNA forms in different tissues. This would explain why the maternal sequence amplified from a chest muscle biopsy could differ from both offspring sequences. The duplication event was apparently followed by a deletion of repeat four in some mtDNA molecules. If an ovum received one mtDNA type, a homoplasmic offspring would result; if it received both types, the offspring would be heteroplasmic.

Mechanisms for length variation: Four different mechanisms have been proposed to account for length variation in mitochondrial DNA: intra- and intermolecular recombination (RAND and HARRISON 1989), slipped mispairing (STREISINGER *et al.* 1966; EFSTRATIADIS *et al.* 1980), illegitimate elongation (BUROKER *et al.* 1990), and transposition (RAND and HARRISON 1989). The recent documentation of DNA being transferred into mitochondria by a protein (VESTWEBER and SCHATZ 1989) opens the possibility that transposition could occur among mitochondria. Transposition would not, however, produce both a duplication and deletion nor would it result in adjacent internal repeats being more similar than external repeats. Thus, transposition cannot account for the patterns of sequence variation we observe in *N. humeralis*, but it may account for the origin of the tandem repeat unit. Although both intra- and intermolecular recombination is frequent in plant mtDNA (SEDEROFF 1987), no direct evidence of recombination has yet been found for animal mtDNA. As RAND and HARRISON (1989) point out, intermolecular recombination results in molecules of differing sizes which have not yet been found. However, intramolecular recombination between, for example, 7S and parent mtDNA, need not alter molecule size and would be very difficult to distinguish from replication slippage.

Although the slipped mispairing model was presented to explain deletions which are flanked by short direct repeats (EFSTRATIADIS *et al.* 1980), a similar process could also produce duplications. Instead of the parental strand forming a single-stranded loop which gets excised before replication occurs, formation of a single-stranded loop in the daughter strand will result in a repeat duplication after DNA replication and resolution of the heteroduplex molecule. This is, in fact, the essence of the illegitimate elongation model (BUROKER *et al.* 1990). Both processes are clearly a form of replication slippage. BUROKER *et al.* (1990) claim that illegitimate elongation is unique to mitochondrial replication because of its triplex nature. An alternative and perhaps more revealing distinction is that mtDNA, unlike nuclear DNA, replicates each strand independently. In the D-loop this process is unidirectional for 7S mtDNA because the L strand does not begin replicating until H strand replication reaches the L strand origin of replication, a conserved noncoding region between the $tRNA^{Asn}$ and the $tRNA^{Cys}$ genes (CLAYTON 1982). Unidirectional replication is significant because repeat duplication can only occur in one direction. Thus, the most recently duplicated repeat should be at the end of an array unless the last repeat in the array is protected in some way from undergoing a deletion event.

The increased nucleotide divergence between repeats one and two, as compared to adjacent internal repeats (Figure 8), indicates that repeat one has not duplicated as recently as have the internal repeats. The even greater divergence between the last repeat (number five, six, seven or eight depending on the size of the repeat array) and the adjacent internal repeat suggests that the last repeat also has not dupli-

cated recently. Presumably, the first and last repeats in the array undergo duplication and deletion events at much lower rates, if at all, than the internal repeats.

Evidence for selection: If a duplication event is not always accompanied by a corresponding deletion, then copy number will change in one strand and heteroplasmy will result. Thus, partial independence between duplications and deletions can account for heteroplasmy, but cannot explain why we usually found six but never less than five or more than eight repeats. If our estimate of μ is correct and the occurrence of tandem repeats predates the origin of the genus, as their presence in both *Myotis* and *Eptesicus* indicates, then there has been ample time for more variation in repeat copy number to occur unless selection also acts to maintain an optimal number of copies. One possible reason for an intermediate number of repeats is that a large secondary structure composed of five, six or more repeats may form when a 7S strand binds to a protein. If such protein binding occurs, then an optimal number of repeats probably exists and more or fewer repeats should decrease binding efficiency. Under this scenario we might expect to find repeat number, as well as nucleotide sequence, differing between related species.

The explanation just provided is insufficient by itself, however, to explain the patterns of nucleotide sequence similarity we found between bats at each repeat. If substitution rates are equal along each repeat and selection is absent, then all repeats, even those that duplicate, should diverge between bats at equal rates. Thus, in the absence of selection the nucleotide differences between bats from different states should be the same for each repeat. However, the repeat nearest the $tRNA^{Pro}$ is much more conserved between bats than the other five repeats indicating that substitutions are either less frequent or less tolerated in this repeat. The hypothesis that repeat one is under purifying selection is consistent with the absence of transversions in this repeat but not in other repeats and with repeat one having the lowest average free energy associated with a folded stem-loop structure. Therefore, an alternative interpretation of these findings is that only one repeat is necessary for protein binding. Some other explanation must then be provided to explain the apparent limitation on copy number within a tandem array. If true, this latter scenario implicates concerted evolution as the mechanism causing sequence similarity among internal repeats and within, but not necessarily between, bats while nuclear-mitochondrial coevolution is likely to be responsible for sequence similarity between bats at the first repeat. Comparison of repeat sequences among closely related species should help to resolve where and how selection acts on length and sequence variation in this region.

We thank W. STEPHAN, A. CLARK and two anonymous reviewers for helpful comments on the manuscript, F. MAYER, D. O'REILLY, A. SCHERRER, S. STEELE, P. TRAIL, R. WILLIAMS and L. WOLFENBARGER for assistance in the field, and the Smith, Grim, Hutton, Busby, Easton, and McLain families for access to bat colonies. We are grateful to T. KOCHER for providing the C and E primers and to W. RASBAND for providing the image analysis program. Supported by a Searle Scholar Award/Chicago Community Trust and by a biomedical research support grant from the University of Maryland.

LITERATURE CITED

ALBRING, M., J. GRIFFITH and G. ATTARDI, 1977 Association of a protein structure of probable membrane derivation with HeLa cell mitochondrial DNA near its origin of replication. Proc. Natl. Acad. Sci. USA **74:** 1348–1352.

AQUADRO, C. F., and B. D. GREENBERG, 1983 Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics **103:** 287–312.

BECKER, W. A., 1975 *Manual of Quantitative Genetics*, Ed. 3. Washington State University, Pullman.

BENTZEN, P., W. C. LEGGETT and G. G. BROWN, 1988 Length and restriction site heteroplasmy in the mitochondrial DNA of American shad (*Alosa sapidissima*). Genetics **118:** 509–518.

BIRKY, C. W. J., T. MARUYAMA and P. FUERST, 1983 An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics **103:** 513–527.

BOYCE, T. M., M. E. ZWICK and C. F. AQUADRO, 1989 Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics **123:** 825–836.

BROWN, W. M., 1985 The mitochondrial genome of animals, pp. 95–130 in *Molecular Evolutionary Genetics*, edited by R. J. MACINTYRE. Plenum, New York.

BUROKER, N. E., J.R. BROWN, T. A. GILBERT, P. J. O'HARA, A. T. BECKENBACH, W. K. THOMAS and M. J. SMITH, 1990 Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics **124:** 157–163.

CANN, R. L., and A. C. WILSON, 1983 Length mutations in human mitochondrial DNA. Genetics **104:** 699–711.

CHANG, D. D., and D. A. CLAYTON, 1985 Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc. Natl. Acad. Sci. USA **82:** 351–355.

CHANG, D. D., T. W. WONG, J. E. HIXSON and D. A. CLAYTON, 1985 Regulatory sequences for mammalian mitochondrial transcription and replication, pp. 135–144 in *Achievements and Perspectives of Mitochondrial Research*, edited by E. QUAGLIARIELLO, E. C. SLATER, F. PALMIERI, C. SACCONE and A. M. KROON. Elsevier, New York.

CLARK, A. G., 1988 Deterministic theory of heteroplasmy. Evolution **42:** 621–626.

CLAYTON, D. A., 1982 Replication of animal mitochondrial DNA. Cell **28:** 693–705.

DODA, J. N., C. T. WRIGHT and D. A. CLAYTON, 1981 Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78:** 6116–6170.

EFSTRATIADIS, A., J. W. POSAKONY, T. MANIATIS, R. M. LAWN, C. O'CONNELL, R. A. SPRITZ, J. K. DERIEL, B. G. FORGET, S. M. WEISSMAN, J. L. SLIGHTOM, A. E. BLECHL, O. SMITHIES, F. E. BARALLE, C. C. SHOULDERS and N. J. PROUDFOOT, 1980 The structure and evolution of the human β-globin gene family. Cell **21:** 653–668.

FORAN, L. A., J. E. HIXSON and W. M. BROWN, 1989 Similarities of ape and human sequences that regulate mitochondrial DNA transcription and their DNA synthesis. Nucleic Acids Res. **17:** 5841–5861.

GREENBERG, B. D., J. E. NEUBOLD and A. SUGINO, 1983 Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochonrial DNA. Gene **21:** 33–49.

GYLLENSTEN, U. B., and H. A. EHRLICH, 1988 Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQa locus. Proc. Natl. Acad. Sci. USA **85**:7652–7655.

HARRISON, R. G., 1989 Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. TREE **4:** 6–11.

HILLIS, D. M., C. MORITZ, C. A. PORTER and R. J. BAKER, 1991 Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science **251:** 308–310.

LA ROCHE, J., M. SNYDER, D. I. COOK, K. FULLER and E. ZOUROS, 1990 Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *Placopecten magellanicus*. Mol. Biol. Evol. **7:** 45–64.

JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules., pp. 21–132 in *Mammalian Protein Metabolism*, edited by H. N. MUNRO. Academic Press, New York.

MIGNOTTE, F., M. GUERIDE, A.-M. CHAMPAGNE and J.-C. MOUNOLOU, 1990 Direct repeats in the noncoding region of rabbit mitochondrial DNA: involvement in the generation of intra and inter-individual heterogeneity. Eur. J. Biochem. **194:** 561–571.

MORITZ, C., and W. M. BROWN, 1987 Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA **84:** 7183–7187.

NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University, New York.

OHNO, S., 1970 *Evolution by Gene Duplication*. Springer Verlag, Berlin.

OHTA, T., 1980 *Evolution and Variation in Multigene Families*. Springer Verlag, Berlin.

POULTON, J., M. E. DEADMAN and R. M. GARDINER, 1989 Tandem duplications of mitochondrial DNA in mitochondrial myopathy: analysis of nucleotide sequence and tissue distribution. Nucleic Acids Res. **17:** 10223–10229.

RAND, D. M., and R. G. HARRISON, 1989 Molecular population genetics of mtDNA size variation in crickets. Genetics **121:** 551–569.

ROFF, D. A., and P. BENTZEN, 1989 The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. Mol. Biol. Evol. **6:** 539–545.

SACCONE, C., M. ATTIMONELLI and E. SBISA, 1987 Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. J. Mol. Evol. **26:** 205–211.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2 (3 vols.). Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SEDEROFF, R. R., 1987 Molecular mechanism of mitochondrial genome evolution in higher plants. Am. Nat. **130:** S30-S45.

SOLIGNAC, M., M. MONNEROT and J.-C. MOUNOLOU, 1986 Concerted evolution of sequence repeats in *Drosophila* mitochondrial DNA. J. Mol. Evol. **24:** 53–60.

SOUTHERN, S. O., P. J. SOUTHERN and A. E. DIZON, 1988 Molecular characterization of a cloned dolphin mitochondrial genome. J. Mol. Evol. **28:** 32–42.

STREISINGER, G., Y. OKADA, J. EMRICH, J. NEWTON, A. TSUGITA, E. TERZAGHI and M. INOUYE, 1966 Frameshift mutations and the genetic code. Cold Spring Harbor Symp. Quant. Biol. **31**:77–84.

VESTWEBER, D., and G. SCHATZ, 1989 DNA-protein conjugates can enter mitochondria via the protein import pathway. Nature **338:** 170–172.

WATKINS, L. C., 1970 Observations on the distribution and natural history of the evening bat (*Nycticeius humeralis*) in northwestern Missouri and adjacent Iowa. Trans. Kans. Acad. Sci. **72:** 330–336.

ZEVIANI, M., S. SERVIDEI, C. GELLERA, E. BERTINI, S. DIMAURO and S. DIDONATO, 1989 An autosomal dominant disorder with multiple deletions of mitochondrial DNA starting at the D-loop region. Nature **339:** 309–311.

ZUKER, M., and P. STIEGLER, 1981 Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9:** 133–148.

Communicating editor: A. G. CLARK