Molecular Characterization of Lengthy Mitochondrial DNA Duplications From the Parasitic Nematode *Romanomermis culicivorax*

Joyce L. Beck Azevedo[1] and Bradley C. Hyman

Department of Biology, University of California, Riverside, California 92521

Manuscript received June 28, 1992
Accepted for publication December 23, 1992

ABSTRACT

Complete nucleotide sequences, precise endpoints and coding potential of several 3.0-kilobase mitochondrial DNA (mtDNA) repeating units derived from two isofemale lineages of the mermithid nematode *Romanomermis culicivorax* have been determined. Endpoint analysis has allowed us to infer deletion and inversion events that most likely generated the present day repeat configuration. Each amplified unit contains the genes for NADH dehydrogenase subunits 3 and 6 (ND3 and ND6), an open reading frame (ORF 1) that represents a cytochrome P450-like gene, and three additional unidentified open reading frames. The primary nucleotide sequences of the *R. culicivorax* mt-repeat copies within individual haplotypes are highly conserved; three nearly complete copies of the repeat unit vary by 0.01% at the nucleotide level. These observations suggest that concerted evolution mechanisms may be active, resulting in sequence homogenation of these lengthy duplications.

BECAUSE of limited gene coding capacity and the absence of introns and intergenic noncoding sequences, animal mitochondrial DNA (mtDNA) molecules have been considered highly economized (SEDEROFF 1984; ATTARDI 1985; BROWN 1985). Generally, these compact genomes encode information for two ribosomal RNAs and 22 tRNAs that contribute to an organelle-specific translation system, and 12 or 13 protein-coding genes whose products are assembled into multisubunit complexes involved with electron transport and oxidative phosphorylation. Each mitochondrial genome also contains a control region composed of sequences that regulate mtDNA replication and transcription (BROWN 1985; CLAYTON 1982, 1984). Length variations confined to the control region are primarily responsible for the size range observed among animal mitochondrial (mt-) genomes (MORITZ and BROWN 1987). In contrast to the mtDNAs of higher metazoa which exhibit conservation of genome size, gene content and organization (BROWN *et al.* 1982; MORITZ, DOWLING and BROWN 1987), invertebrate mt-genomes display broader variation in size, coding potential and gene arrangement (WOLSTENHOLME *et al.* 1990).

Previous assertions of extreme economization and conserved organization (WALLACE 1982; SEDEROFF 1984; ATTARDI 1985; BROWN 1985) have been challenged by numerous recent reports describing animal mtDNAs with large-scale size variation attributable to repetitive sequences. Length variation is most often a consequence of variable copy numbers of tandemly repeated, noncoding sequences within the control region. Such is the case with mtDNA duplications in vertebrates such as Cnemidophorus lizards [64 base pair (bp); DENSMORE, WRIGHT and BROWN 1985)], Alosa fish [1.5 kilobase (kb); BENTZEN, LEGGETT and BROWN 1988)], Acipenser fish (82 bp; BUROKER *et al.* 1990), Oryctolagus rabbits (20 bp and 153 bp; MIGNOTTE *et al.* 1990), Macaca monkeys (160 bp; HAYASAKA, ISHIDA and HORAI 1991), and Nycticeius bats (81 bp; WILKINSON and CHAPMAN 1991); and in invertebrates, including Drosophila (470 bp; SOLIGNAC, MONNEROT and MOUNOLOU 1986), Gryllus crickets (220 bp; RAND and HARRISON 1989), Pissodes weevils (0.8–2.0 kb; BOYCE, ZWICK and AQUADRO 1989), Placopecten scallops (1.4 kb; SNYDER *et al.* 1987; LAROCHE *et al.* 1990), and Meloidogyne nematodes (8, 63 and 102 bp; OKIMOTO *et al.* 1991).

In only a few cases has mtDNA length variation been attributable to tandem repeats encompassing *coding* sequences. Coding regions have been observed in tandemly repeated sequences from the mt-genomes of Cnemidophorus lizards (1.5–8.0 kb repeats, encoding the control region, tRNA and rRNA genes, and the gene for cytochrome *b*; MORITZ and BROWN 1986, 1987), Triturus newts (1.1–8.5 kb repeats, encompassing the large and small rRNA genes and subunits 1 and 2 of NADH-dehydrogenase, ND1 and ND2; WALLIS 1987) and Heteronotia geckos (1.2–10.4 kb repeats, including both large and small rRNA genes, the control region, and most of the gene for ND1; MORITZ 1991; ZEVERING *et al.* 1991). The apparent scarcity of higher vertebrate mtDNA sequence amplification and size-variant heteroplasmy has prompted the suggestion that homeotherms, due to their elevated metabolic rates, may be under a stronger selec-

[1] Present address: Department of Biology, Southern College, Collegedale, Tennessee 37315-0370.

tion for smaller and less variable mtDNAs than poikilotherms (RAND and HARRISON 1989).

All known duplications of animal mtDNA are present as intact tandem, direct repeat units except those in the nematode *Romanomermis culicivorax*, where inverted copies are found disjunct from a tandemly repeated set (HYMAN, BECK and WEISS 1988; BECK and HYMAN 1988). To gain a better understanding of the nature and occurrence of this unusual repeat unit organization, it became necessary to characterize, at the primary sequence level, independent copies of the repeating unit and their adjoining unique sequences.

In this study we report the complete nucleotide sequences of 3.0-kb repeat units isolated from two independently propagated *R. culicivorax* lineages (3B4 and NAC6) subcultured from a mixed laboratory population (POWERS, PLATZER and HYMAN 1986a). This information represents the first lengthy duplicated mtDNA segment encompassing *coding* regions to be characterized in detail at the molecular level. We also compare nucleotide sequences of several repeat copies independently cloned from separate genomic locations within the same mitochondrial lineage. Each complete repeating unit, irrespective of its source, has the potential to encode several genes typically encoded within animal mtDNA as well as potentially novel products, including a cytochrome P450-like protein. Further nucleotide sequence analysis reveals that the incidence of base substitutions occurring between repeating units from each lineage is strikingly similar to that observed within an unamplified gene locus. However, when repeating units within a lineage are compared, unexpectedly reduced levels of nucleotide divergence are detected. In conjunction with the precise localization of junctions between unique and amplified mtDNA sequences, our results collectively provide a perspective on the mechanisms and frequencies of events which may have generated this unusual mt-genome organization.

MATERIALS AND METHODS

Cultivation of *R. culicivorax* laboratory cultures: The *R. culicivorax* isofemale lineages 3B4 and NAC6 were derived from a heterogenous field population originally isolated from Lake Charles, Louisiana (PETERSEN, CHAPMAN and WOODWARD 1968), and subsequently maintained as laboratory stocks. Establishment and maintenance of these isofemale lineages were previously described (POWERS, PLATZER and HYMAN 1986a).

Plasmids and bacterial strains: The bacterial plasmids pUC118, pUC119 (VIEIRA and MESSING 1987) and pBluescript SK(−) (SHORT *et al.* 1988) were used as vectors for molecular cloning and DNA sequencing experiments. *Escherichia coli* strains MV1190 (VIEIRA and MESSING 1987), XL1-Blue (BULLOCK, FERNANDEZ and SHORT 1987), and PLK-F′ (KRETZ and SHORT 1989) were rendered competent for DNA-mediated transformation and were used as recipients in molecular cloning experiments (SAMBROOK, FRITSCH and MANIATIS 1989).

Nucleic acid isolation, restriction enzyme digestion and gel electrophoresis: *R. culicivorax* mtDNA was purified by ethidium bromide-CsCl or Hoechst 33258-CsCl isopycnic centrifugation from lysed post-parasite stage nematodes as previously described (POWERS, PLATZER and HYMAN 1986a,b). Plasmid DNAs were isolated using standard procedures (SAMBROOK, FRITSCH and MANIATIS 1989). *R. culicivorax* total cellular RNA was isolated via a guanidinium thiocyanate/phenol-chloroform single step extraction (CHIRGWIN *et al.* 1979; CHOMCZYNSKI and SACCHI 1987).

Restriction enzyme digestions and gel electrophoresis were carried out using standard procedures (SAMBROOK, FRITSCH and MANIATIS 1989). Formamide-treated RNA was fractionated through denaturing agarose gels using 2.2 M formaldehyde in the presence of 50% formamide (LEHRACH *et al.* 1977; SAMBROOK, FRITSCH and MANIATIS 1989). DNA and RNA were visualized by staining gels in 1 μg/ml ethidium bromide followed by illumination using a Fotodyne 3-3000 transilluminator.

Nematode cDNA library construction: A cDNA library was constructed from total nematode polyadenylated poly(A^+) RNA which had been eluted from an oligo(dT) cellulose column. The poly(A^+) RNA population was converted into cDNA molecules and introduced into the λ ZAP cloning vector (SHORT *et al.* 1988; SAMBROOK, FRITSCH and MANIATIS 1989). The λ library was packaged *in vitro* into phage particles utilizing the Gigapack™II Gold packaging extract (KRETZ and SHORT 1989; Stratagene, La Jolla, California) and amplified in the *E. coli* strain PLK-F′. MtDNA-specific cDNA clones were identified by plaque hybridization using highly purified *R. culicivorax* mtDNA molecules which had been ^{32}P-labeled *in vitro* by random priming (SAMBROOK, FRITSCH and MANIATIS 1989).

DNA sequencing and sequence analysis: DNA sequences of sequential exonuclease III-generated deletions (HENIKOFF 1987; SAMBROOK, FRITSCH and MANIATIS 1989) were obtained by the dideoxy-chain termination procedure (SANGER, NICKLEN and COULSON 1977). Nucleotide sequences were analyzed using the Genetics Computer Group (GCG) software package (DEVEREUX, HAEBERLI and SMITHIES 1984) operating on a VAX 8700, and by CLUSTAL (HIGGINS and SHARP 1988), RDF (LIPMAN and PEARSON 1985) and BLASTP (ALTSCHUL *et al.* 1990) programs. Modifications of the universal genetic code used in computer-assisted translation of the mt-nucleotide sequences were: AGA and AGG=serine; TGA=tryptophan; ATA=methionine (OKIMOTO *et al.* 1992). Nematode mt-protein genes were identified by alignment of predicted translations with amino acid sequences of known mt-protein genes and by hydropathic profile comparisons (CHOU and FASMAN 1978; KYTE and DOOLITTLE 1982). Nucleotide sequences are available in GenBank, accession #L08174.

RESULTS

Repeat unit similarity within 3B4 mtDNA: A simplified physical map of the *R. culicivorax* mitochondrial genome is displayed in Figure 1. Individual molecules may vary in size because they contain different copy numbers of a 3.0-kb duplication (BECK and HYMAN, 1988) Spatial separation of repeating unit copies is revealed by their map locations within the *Hin*dIII-A and *Hin*dIII-B restriction fragments (Figure 1; HYMAN, BECK and WEISS 1988). Nucleotide sequences have been determined for repeat units cloned from these separate areas of the 3B4 mitochondrial genome. Alignment of these different repeat sequences reveals extensive nucleotide similarity.

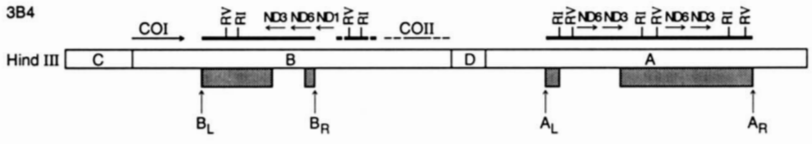

FIGURE 1.—Simplified physical map of the 3B4 *R. culicivorax* mitochondrial genome. *Hin*dIII restriction fragments are lettered based upon descending molecular length. Bold horizontal lines above the restriction map designate locations of the 3.0-kb duplications. *Eco*RI (RI) and *Eco*RV (RV) represent cleavage sites found within the amplified DNA. Thin horizontal arrows denote location and transcriptional orientation of identified mt-genes. Dashed lines indicate ambiguity in positioning the endpoints of various sequences. Stippled boxes indicate sequenced regions of repeated mtDNA. Endpoints of *Hin*dIII-A and *Hin*dIII-B repeat clusters are shown as A_L, A_R, B_L and B_R and are noted by small vertical arrows below the sequence. A more detailed restriction map can be found in BECK and HYMAN (1988).

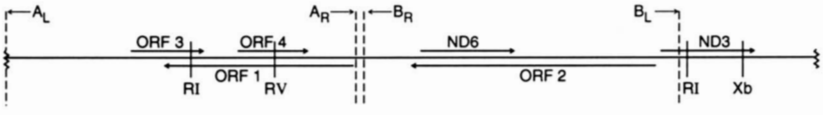

FIGURE 2.—Linear map of a single 3.0-kb repeating unit, with restriction site and endpoint designations as in Figure 1. Map commences (vertical saw-tooth lines) with the left endpoint of the repeat cluster located within the *Hin*dIII-A restriction fragment (A_L; see Figure 1). A_R, B_L, B_R indicate position and orientation (arrow) of sequences identical to the endpoints of each major duplication cluster found within intact repeating units.

Less than 0.01% nucleotide divergence is evident over 8,368 nucleotides sequenced, which collectively represent nearly three complete, independent copies of the amplified locus (Figure 1). These nucleotide changes do not affect open reading frames (ORFs) 2 or 4, coding sequences for ND3, ND6, or the cytochrome P450-like gene (ORF 1; Figure 2), but one substitution generates a frameshift causing ORF 3 to be reduced in size from 92 to 31 amino acids in two of three repeating unit copies. As described below, it is this same ORF that is severely altered by a frameshift mutation in the NAC6 isofemale lineage.

Repeat sequence similarity between nematode lineages: The complete nucleotide sequences have been determined for cloned duplications from strains 3B4 (3032 bp) and NAC6 (3065 bp). These repeating units are 95.3% identical, and have a 78% A/T base composition. We had previously identified (BECK and HYMAN 1988) differences in length between the repeats in these two lineages to be primarily a result of an imperfect 23-bp tandem motif which is present twice in NAC6 but only once in 3B4. Mechanisms whereby this duplication/deletion may have arisen are discussed in BECK and HYMAN (1988). The remainder of the size difference between the 3B4 and NAC6 repeat units is due to several small insertions/deletions which range from single nucleotides to a 12-mer (Table 1). These account for 9% of total mismatches when each is scored as a single event independent of size. The remaining 91% of mismatches are substitutions (142) which consist of 73% transitions and 27% transversions, and represent each of the possible nucleotide changes (Table 1).

TABLE 1

Repeating unit sequence differences between isofemale lineages

Type of mismatch	No. of occurrences
Insertion/deletion	14 (9%)
Single nucleotide	5
Dinucleotide	3
Trinucleotide and larger	6
Substitution	142 (91%)
Transition	
A ↔ G	55
T ↔ C	49
Transversion	
C ↔ G	8
C ↔ A	12
T ↔ A	10
T ↔ G	8

Of the 14 insertion/deletion mismatches between the 3B4 and NAC6 repeating units, 8 appear to be simple duplications. Three of the single nucleotide insertion/deletion events are an A ↔ AA or T ↔ TT, and occur within A or T homopolymer runs. One of the dinucleotide insertion/deletion events is a TC ↔ TCTC, the trinucleotide insertion/deletion mismatch is a TTA ↔ TTATTA, a 6-nucleotide insertion/deletion event is the direct duplication of ATTTTA, and 12- and 23-bp insertions are each imperfect direct repeats.

Coding potential of the repeated units: There are six ORFs in the amplified unit which are large enough to encode some of the polypeptides typically encoded in animal mtDNA; the four smaller ORFs (90–117-amino acid residues in length) are located in one strand while the two largest ones (238 and 293 amino

FIGURE 3.—CLUSTAL sequence alignment of ND3 and ND6 polypeptides from *R. culicivorax* lineage 3B4 and NAC6 repeats with those from *A. suum* and *C. elegans* mtDNAs (OKIMOTO *et al.* 1992). Gaps between residues are inserted to optimize the alignments. Identical (*) and conserved (•) amino acids are indicated (HIGGINS and SHARP 1988).

acids) are encoded in the opposite strand (Figure 2).

ND3 and ND6: Two of the smaller ORFs have been identified as protein-coding genes expected to be found in animal mitochondrial genomes. These include the coding sequences for NADH-dehydrogenase subunits 3 and 6 (ND3, ND6). These ORFs were identified by amino acid alignments (Figure 3) and hydropathy profile comparisons (data not shown) with *Ascaris suum*, *Caenorhabditis elegans* and *Drosophila yakuba* ND3 and ND6 genes (OKIMOTO *et al.* 1992; CLARY and WOLSTENHOLME 1985).

Multiple sequence alignments generated by CLUSTAL of ND3 and ND6 are presented in Figure 3, in which polypeptides derived from *R. culicivorax*, *A. suum* and *C. elegans* are compared. For both ND3 and ND6, similarity values varied depending upon the *R. culicivorax* lineage. When ND3 derived from isofemale lineage 3B4 was compared with *C. elegans* and *A. suum*, similarity values measured 26% and 32%, respectively (Table 2). Higher similarity scores were obtained when the NAC6 ND3 protein was aligned with *C. elegans* (30%) and *A. suum* (38%). In contrast, the ND6 protein from both lineages revealed strikingly lower levels of similarity with ND6 polypeptides from *C. elegans* and *A. suum* (Table 2).

Predicted amino acid sequences for the ND3 and ND6 polypeptides indicate that the two protein subunits are longer in the 3B4 genome than their cognates in NAC6. For ND3, a difference of 39-amino acid residues results from a putative insertion event within the NAC6 gene copy that creates a premature translation stop codon (TAA) in the NAC6 gene. The ND6 coding sequences have two features which contribute to their length difference. In addition to a substitution which creates a TAA translation stop

TABLE 2

Interspecific comparison (%) of ND3 and ND6 amino acid similarities

Assay	A. suum[a]	C. elegans[a]	D. yakuba[b]	X. laevis[c]
ND3				
3B4	32	26	28	25
NAC6	38	30	29	27
ND6				
3B4	16	12	18	16
NAC6	23	16	29	16

[a] OKIMOTO *et al.* (1992).
[b] CLARY and WOLSTENHOLME (1985).
[c] ROE *et al.* (1985).

codon, there is an in-frame 12-bp tandem duplication within this NAC6-derived ORF. Together, these contribute to a difference of 16 amino acids between the ND6 polypeptides. Half of the substitutions between ND3 or ND6 coding sequences are first codon position changes, of which 53% are conservative substitutions. All third position substitutions result in silent mutations. The ND6 coding sequence appears to be the least conserved among repeat unit ORFs, with 88% identity at the nucleotide level and 71% amino acid identity (Table 3).

ORF 1: Translations of the two longest ORFs encoded in the complementary strand are displayed in Figure 4. Predicted amino acid sequences from these ORFs are also longer in the 3B4 repeating unit when compared to their cognates from NAC6. For ORF 1, a single base insertion in 3B4 mtDNA causes this putative 3B4 peptide to potentially be extended by seven additional residues at its amino terminus relative to that found in the NAC6 repeating unit. Of the 21-codon differences between these ORF 1 polypeptides, nine are generated by third position silent substitutions while two of the remaining mutations generate conservative amino acid changes (Figure 4).

GenBank database searches using the BLAST algorithm reveals that ORF 1 (239 amino acids) contains a region with significant similarity to a highly conserved heme-binding domain common among cytochrome P450 proteins. CLUSTAL alignment for this portion of the nematode cytochrome P450-like gene with cytochrome P450 sequences from human, mouse and rat is depicted in Figure 5. Through a Monte Carlo analysis, the RDF statistical program (NBRF-PIR; LIPMAN and Pearson 1985) generated *z*-values higher than 9 for comparisons of this entire cytochrome P450-like sequence (239 amino acids) with typical cytochrome P450s, indicating significant similarity.

The protein product predicted from the ORF 1 sequence differs from common cytochrome P450 polypeptides in two respects. The overall length of the ORF 1 translation product is predicted to be 239 amino acids, which is approximately half the average

TABLE 3

mtDNA ORF comparisons between isofemale lineages

Assay	Nucleotide[a]	Amino acid[a]	Length[b]
ORF 1 (P450-like)	96	94	239/232
ORF 2	93	85	293/222
ORF 4	98	99	90/90
ND3	98	97	118/79
ND6	88	71	117/98
ND1	98	98	93/122

[a] Expressed as percentage identity between ORFs found in 3B4 and NAC6 repeating units.

[b] Expressed as the number of amino acids in ORFs from isofemale lineage 3B4 relative to that in NAC6 mtDNA (3B4/NAC6).

length of authentic cytochrome P450 peptides. In addition, an invariant cysteine considered essential for heme-ligand binding has been replaced by phenylalanine, a consequence of a single base substitution in ORF 1 (Figure 5).

ORF 2: Several features contribute to the observed difference in the ORF 2 length between 3B4 and NAC6 mt-repeats. A 4-base insertion in 3B4 mtDNA results in a peptide with 66 additional residues at its amino terminus relative to that encoded in NAC6 repeat units. An in-frame 12-bp duplication within the NAC6 encoded polypeptide adds four amino acids to the NAC6 peptide while another single base deletion causes the NAC6 translation product to be nine residues shorter. Collectively, these features generate a 3B4 polypeptide sequence 72 amino acids longer then its NAC6 cognate. Most of the third position substitutions (23 out of 24) in ORF 2 are silent while 21% of the remaining codon changes are conservative.

ORFs 3 and 4: The predicted polypeptide representing ORF 3 in the 3B4 duplication does not appear to have a cognate in the sequenced NAC6 mt-repeat due to an insertion within the NAC6 repeat that creates a translational frame-shift. The putative ORF 4 polypeptide is identical in length between the repeat units derived from the two mtDNA types. Greater than 60% of the substitutions in ORF 4 are third position silent mutations. Identity comparisons reveal ORF 4 to be the most conserved at both the nucleotide (98% identity) and amino acid (99%) levels when ORFs derived from both isofemale lineages are compared (Table 3).

Codon usage: The percentage of *R. culicivorax* mt-codons that end in G, A, or T is more similar to that of *C. elegans* than to the values found for *A. suum*, while the percentage of codons ending in C is higher in *R. culicivorax* than in either of the other nematode mtDNAs (OKIMOTO *et al.* 1992). Similar values for the incidence of leucine among mt-proteins were found among the three nematodes, however the *C. elegans* and *A. suum* mtDNAs have a much higher preponderance of leucine codons beginning with T than the leucine codon usage in *R. culicivorax*.

Identification of repeating unit endpoint sequences: Endpoints of amplified regions, represented by junctures of duplicated and unique mtDNA sequences, were identified at four locations within the 3B4 mt-genome. Two of these represent endpoints for the largest tandem repeat set (located entirely within the *Hin*dIII-A restriction fragment, A_L and A_R; Figure 1) while the remaining two form the boundaries for the large inverted repeat cluster found within the *Hin*dIII-B restriction fragment (B_L and B_R; Figures 1 and 2). Three of the four endpoints reside external to the six ORFs contained within each repeat unit, while the fourth (B_L) occurs near the 5′-end of the ND3 coding sequence (Figure 2).

The 72-bp segment immediately adjacent to endpoint A_L of the *Hin*dIII-A repeat cluster shares greater than 80% nucleotide identity to an internal portion of the amplified unit (Figure 6a), suggesting

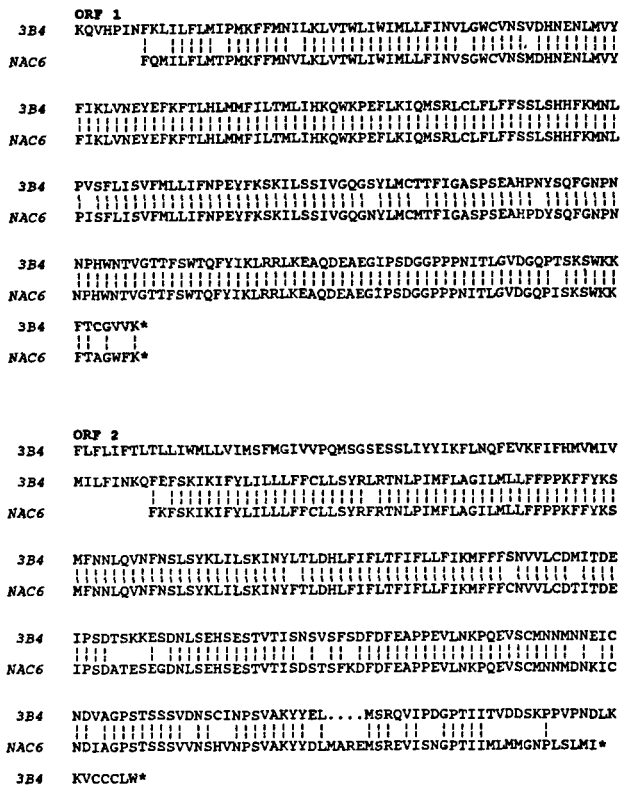

FIGURE 4.—Alignment of 3B4 and NAC6 ORF 1 (cytochrome P450-like gene) and ORF 2 predicted amino acid sequences. (•) indicates the absence of an amino acid. Vertical dashed lines indicate identities between 3B4 and NAC6 ORFs.

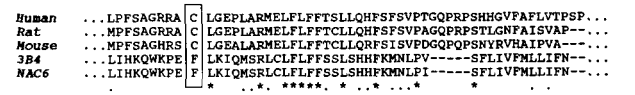

FIGURE 5.—CLUSTAL sequence alignment of 3B4 and NAC6 ORF 1 (49-amino acid internal domain) with the highly conserved heme-binding sequence element from human (GONZALEZ *et al.* 1988), mouse (ICHIKAWA, ITAKURA and NEGISHI 1989) and rat (MATSUNAGA *et al.* 1989) cytochrome P450 sequences. Annotations are as described in Figure 3. Three dots at the beginning and end of each sequence indicate continuation of amino acids in each direction. Box identifies position of highly conserved heme-binding residue in authentic cytochrome P450 proteins.

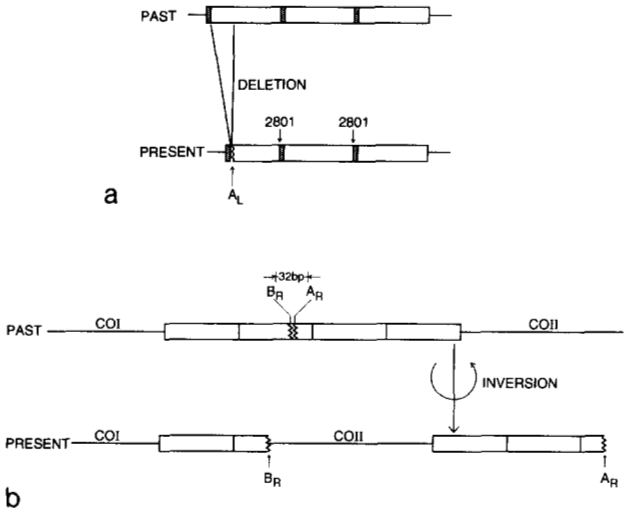

FIGURE 6.—(a) Generation of the A_L endpoint by a deletion. Stippled boxes represent the 72-bp sequence immediately adjacent to the cluster and its cognate internal to the repeat unit. (b) Generation of endpoints A_R and B_R after inversion. The breakpoint is located within the 32-bp sequence separating the A_R and B_R endpoints.

a portion of a duplication has been deleted in the present repeat configuration. In addition, sequences identical to the right endpoints of each cluster (A_R, B_R; Figure 1) are separated by 32-bp within the intact repeating unit (Figure 2). Collectively, this information can be employed to infer molecular mechanisms for the generation of both inverted repeat copies as well as members of tandemly repeated clusters that appear truncated. These models (Figure 6, a and b) are discussed in detail below.

Comparison of unique coding sequence information between 3B4 and NAC6 mt-genomes: Nucleotide sequence information for 368 bp representing a portion of the NADH-dehydrogenase subunit 1 (ND1) protein coding sequence, identified by amino acid alignment with known ND1 polypeptides (J. AZEVEDO and B. HYMAN, manuscript in preparation), has been obtained from both 3B4 and NAC6 mtDNAs (Figure 1). This gene lies external to the repeating units in each mt-genome and might be expected to be under different selective pressures when compared with amplified gene sequences (BROWN 1985).

This portion of the ND1 gene shares 97.6% nucleotide identity, commensurate with a 95.3% level calculated for repeating unit identity between these isofemale lineages. The ND1 coding region is 80% A/T in 3B4 and 79% in NAC6 (Figure 7). Of nine mismatches in the 368-bp region (Figure 7), all are single-base substitutions, with seven (77.8%) transitions and two (22.2%) transversions. Six of the nine mismatches are in the third codon position, resulting in silent mutations. Of the remaining three mismatches, two are in the second codon position and one is in the first. A transition creates a translation stop (TAA) codon in the 3B4 ND1 sequence which causes the

```
    ...L  S  F  E  I  L  I  L  L  L  I  V  T  P  F
3B4 ...CTA AGT TTT GAA ATT TTA ATC TTA TTA TTA ATT GTA ACT CCT TTT   47
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
NAC6 ...CTA AGT TTT GAA ATT TTA ATC TTA TTA TTA ATT GTA ACT CCT TTT
    ...L  S  F  E  I  L  I  L  L  L  I  V  T  P  F

    M  I  T  N  K  A  S  L  T  L  I  N  Q  F  S
    ATA ATT ACT AAT AAA GCC TCA TTA ACT CTT ATT AAT CAA TTT TCT   92
    ||| ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ATA ATT TCT AAT AAA GCC TCA TTA ACT CTT ATT AAT CAA TTT TCT
    M  I  S  N  K  A  S  L  T  L  I  N  Q  F  S

    V  I  N  V  N  L  M  F  I  F  F  F  L  S  I
    GTA ATT AAT GTA AAT TTA ATA TTT ATT TTT TTT TTT TTA AGT ATT   137
    ||| ||| ||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    GTA ATT AAT GTA AGT TTA ATA TTT ATT TTT TTT TTT TTA AGT ATT
    V  I  N  V  S  L  M  F  I  F  F  F  L  S  I

    L  E  C  Q  R  S  P  M  D  L  S  E  G  E  S
    CTT GAA TGT CAA CGC TCA CCT ATA GAT TTG TCT GAA GGA GAG AGG   182
    ||| ||| ||| ||| ||  ||| ||| ||| ||  ||| ||| ||| ||| |||
    CTT GAA TGT CAA CGC TCG CCT ATA GAT TTG TCG GAA GGA GAG AGG
    L  E  C  Q  R  S  P  M  D  L  S  E  G  E  S

    E  L  V  S  G  Y  N  I  E  L  S  S  V  M  F
    GAA CTT GTT AGG GGT TAT AAT ATT GAG TTA AGA AGA GTC ATA TTT   227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||  ||| |||
    GAG CTT GTT AGG GGT TAT AAT ATT GAG TTA AGA AGA GTT ATA TTT
    E  L  V  S  G  Y  N  I  E  L  S  S  V  M  F

    I  F  I  F  L  S  E  Y  N  T  M  I  F  M  M
    ATT TTT ATT TTT TTA AGG GAG TAT AAT ACT ATA ATT TTT ATA ATA   272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    ATT TTT ATT TTT TTA AGG GAG TAT AAT ACT ATA ATT TTT ATA ATA
    I  F  I  F  L  S  E  Y  N  T  M  I  F  M  M

    S  I  M  *
    AGA ATT ATA TAA TTT ATT TAT TTA AAA ATA ATA AGT TTA TTA TTA   317
    ||| ||| ||| |  ||| ||| ||  ||| ||| ||| ||| ||| ||| ||| |||
    AGA ATT ATA TGA TTT ATT TAC TTA AAA ATA ATA AGT TTA TTA TTA
    S  I  M  W  F  I  Y  L  K  M  M  S  L  L  L

    TGT TTA TCT TTA TTT TTA ATA CTA TTA ATT CGT TCA TGT TTT CCT   362
    ||| ||| ||| ||| ||| ||  ||| ||| ||| ||| ||| ||| ||| |||
    TGT TTA TCT TTA TTT TTA ATG CTA TTA ATT CGT TCA TGT TTT CCT
    C  L  S  L  F  L  M  L  L  I  R  S  C  F  P

    CGA ATT...   368
    ||| |||
    CGA ATT...
    R  I  ...
```

FIGURE 7.—NADH-dehydrogenase subunit 1 (ND1) comparison between *R. culicivorax* 3B4 and NAC6 mt-genomes. The nucleotide sequences and putative translations of 368 bp of the ND1 coding sequences from 3B4 and NAC6 mt-genomes are aligned 5′ to 3′ (from left to right). Numbering of these sequences commences with the beginning of a cloned segment of nematode mtDNA, and represents a region near the 3′ end of known ND1 coding sequences. The predicted amino acid sequences of the ND1 genes are shown above the 3B4 sequence and below the NAC6 sequence; boxed-in areas indicate divergent codons. Annotations are as in Figure 4.

peptide to be 58 residues shorter than predicted for either the *A. suum* or *C. elegans* ND1 polypeptides (OKIMOTO *et al.* 1992), while the second codon base transition at position 106 results in the exchange of an asparagine for a serine. The single first codon position mutation results in a conservative substitution (serine for threonine; Figure 7).

DISCUSSION

Molecular characterization of invertebrate and poikilotherm vertebrate mtDNAs has changed the common perception that the animal mt-genome is an extremely economized unit (ATTARDI 1985). Mitochondrial genomes greater than 20 kb have been found in several phylogenetically distant groups (reviewed by GJETVAJ, COOK and ZOUROS 1992), but there seems to be no correlation between mtDNA size and taxonomic relationship. Common among these organelle genomes is the presence of duplicated

mtDNA loci which directly contributes to their large size.

Early generalizations describing the small size of animal mitochondrial genomes (SEDEROFF 1984; ATTARDI 1985; CLARK-WALKER 1985), were substantiated by examples where smaller representatives of size-polymorphic mtDNA populations appear to accumulate (WALLACE 1982; SOLIGNAC *et al.* 1984; HALE and SINGH 1986; RAND and HARRISON 1986, 1989; BOURSOT, YONEKAWA and BONHOMME 1987; WALLIS 1987). However, in most polymorphic mtDNA populations, molecules with two or more repeat copies appear to be the predominant form (DENSMORE, WRIGHT and BROWN 1985; RAND and HARRISON 1986, 1989; SNYDER *et al.* 1987; BENTZEN, LEGGETT and BROWN 1988; BOYCE, ZWICK and AQUADRO 1989; LAROCHE *et al.* 1990; GJETVAJ, COOK and ZOUROS 1992). Such is the case with the *R. culicivorax* mt-genomes described here. Detectable levels of molecules that exclusively contain a single 3.0-kb repeat copy have not been observed in our laboratory populations.

Complete nucleotide sequence information for lengthy mtDNA duplications is available only from a limited number of organisms. The 1.45-kb repeated element of sea scallops has been completely sequenced but does not contain any obvious coding sequences although it does harbor several ORFs, the longest of which is 369 bp long (LAROCHE *et al.* 1990).

Inter-genomic repeating unit comparisons: The large *R. culicivorax* mitochondrial genome size is the consequence of DNA sequence amplification involving approximately 3.0 kb. In our cultured field population, individual nematodes are monomorphic for one of several mtDNA size variants (POWERS, PLATZER and HYMAN 1986b; BECK and HYMAN 1988). We previously reported that the length differences between the mt-genomes of two isofemale lineages, 3B4 and NAC6, are primarily due to the number of repeating units generated during the amplification event(s) (BECK and HYMAN 1988). NAC6 mtDNA contains four complete repeat copies, while the 3B4 mitochondrial genome contains only three intact repeats.

In addition to reiteration frequency, the size of the individual repeating unit in 3B4 compared to that of NAC6 varies. These two repeats differ in length by 33 bp. This difference is primarily the result of an imperfect "mini-repeat" of 23 bp found tandemly duplicated in the NAC6 unit and only as a single copy in the 3B4 repeat. The mechanism most likely to have generated this sequence difference would involve slip-strand mispairing (SSM; HYMAN and SLATER 1990).

Numerous additional examples of likely strand-slippage events are observed upon further alignment of these two repeating units. Eight of 14 insertion/deletions are duplications, likely to be generated by SSM (Table 1). Local denaturation and displacement of DNA strands may be facilitated in this AT-rich mtDNA repeat unit which contains numerous poly-A and poly-T tracts. Short insertions or deletions may result if strand-slippage is immediately followed by replication or repair events (STREISINGER *et al.* 1966; ALBERTINI *et al.* 1982; PALMER 1985; LEVINSON and GUTMAN 1987). A tendency for small length mutations of this sort to occur preferentially in AT-rich spacer regions of chloroplast genomes has also been observed (ZURAWSKI, CLEGG and BROWN 1984; PALMER 1985).

Repeating unit characteristics and coding potential: The *R. culicivorax* mt-repeat represents the first lengthy duplicated animal-mtDNA segment encompassing coding regions to be characterized in detail at the molecular level. We have found it to contain six ORFs, two of which (ND3 and ND6) are expected to be encoded in animal mt-genomes. The predicted ND3 and ND6 polypeptides are both shorter than those encoded in most animal mitochondrial genomes, as is the case for ND3 and ND6 genes characterized from two other nematodes, *A. suum* and *C. elegans* (OKIMOTO *et al.* 1992).

The two ND6 genes representing each isofemale lineage exhibit the lowest amino acid similarity values among all the *R. culicivorax* mt-gene pairs we have characterized. This result is consistent with the rapidly evolving nature of the ND6 protein (BROWN 1985; OKIMOTO *et al.* 1992). ND6 amino acid similarities between *C. elegans* and *A. suum* have been measured at 58.3% (OKIMOTO *et al.* 1992). However, intraspecific variation at the level detected here was unexpected. This observation may indicate that separation of the two isofemale lineages is an ancient event, or that ND6 in particular represents a unique hypervariable locus similar to the large sequence divergence observed among menhaden fish and chuckwalla lizard populations (AVISE, BOWEN and LAMB 1989).

The similarity of ORF 1 to a highly conserved heme-binding region of cytochrome P450 genes was not anticipated, since this gene is typically found in the nuclear genome of metazoans (SPURR *et al.* 1991). The putative polypeptide generated from this coding sequence is unlikely to have conventional cytochrome P450 activity because it is approximately half the length of most cytochrome P450 polypeptides and it no longer contains a critical, highly conserved heme-binding cysteine (Figure 5).

Cytochrome P450-enzyme activity has not been found in several different genera of parasitic helminths, including all nematodes examined to date (PRECIOUS and BARRETT 1989). Due to deviations from the universal genetic code employed by the mitochondrial translation system (reviewed in BROWN 1985; CLARY and WOLSTENHOLME 1985; JACOBS *et al.* 1988; OKIMOTO *et al.* 1992), it is unlikely that a

nuclear-derived cytochrome P450 gene could be effectively translated within the interior of the organelle if gene transfer from the nucleus had occurred. Codon usage similar to the remainder of the *R. culicivorax* mt-genes indicates such a putative transfer must have been an ancient event. Alternatively, this ORF could represent the remnant of a gene that did not escape the ancestral mtDNA molecule. Irrespective of origin, this sequence may have accumulated sufficient mutations to render a protein with completely different function or alternatively, a cytochrome P450 pseudogene. We have now characterized several independently cloned cDNAs for this cytochrome P450-like gene, providing direct evidence for its transcription and persistance of mRNA at detectable steady-state levels.

The coding potential of several other ORFs within the NAC6 mtDNA repeat is affected by mutations. All except ORF 4 are shorter in the NAC6 repeat unit than their cognates in 3B4 (Table 3). It is possible that the single sequenced duplication representing NAC6 mtDNA may be a nonfunctional copy, and that other repeating units from NAC6 may have potentially functional ORFs. However, significant restriction site polymorphism among the NAC6 repeats has not been observed (BECK and HYMAN 1988).

Duplication boundaries: A striking feature in our initial characterization of these nematode mt-genomes was the existence of incomplete tandem repeat copies (HYMAN, BECK and WEISS 1988; BECK and HYMAN 1988). Analysis of all four endpoints delimiting the two major inverted repeat clusters allows us to speculate on the molecular events that resulted in the generation of incomplete tandem repeats less than 3.0 kb in length.

It appears the region bordering the *Hin*dIII-A left endpoint (A_L) may have been an initial endpoint for repeat unit amplification, because a subsequent deletion event (Figure 6a) between this region and a 72-bp sequence presently sharing 80% nucleotide identity within the repeat could have generated the current A_L endpoint (Figures 1 and 6a). MtDNA deletions most likely sponsored by SSM events occurring between two separated, short direct repeats have been well documented in *R. culicivorax* (HYMAN and SLATER 1990), as well as in human mt-genomes (WALLACE 1992).

Alignment of A_R and B_R with the 3.0-kb complete repeat unit reveals that copies of these same sequences are separated by only 32 bp in an inverted orientation within an intact repeat (Figure 2). The remarkably close proximity and orientation of these sequences suggests that the two large repeat arrays, now spatially separated within *Hin*dIII-A and *Hin*dIII-B restriction fragments, may have originally comprised a single large tandemly repeated cluster. When this large putative ancestral tandem array is split by inversion in the vicinity of the 32-bp region, spatially inverted repeat clusters may be generated (Figure 6b), resulting in the present-day configuration (Figure 1).

Divergence among intra-genomic repeats: After duplication of mitochondrial genes creates redundant information, one or more copies are expected to be under relaxed selection pressures. Function may be lost or altered without consequence assuming a single remaining copy maintains its original coding potential. Evolutionary independence of duplicated sequence copies is observed for Cnemidophorus lizards and Heteronotia gecko mtDNA repeats (MORITZ and BROWN 1987; MORITZ 1991). This differs from the situation for the phylogenetically conserved inverted repeat of land-plant chloroplast DNA where copy-correction is almost immediate (reviewed by PALMER 1985).

Gecko mtDNA duplicated coding sequences provide evidence for both large- and small-scale changes, including deletions ranging from 0.19 to 2.5 kb, and numerous restriction site gains and losses (MORITZ 1991; ZEVERING *et al.* 1991). Some of the deletions involve coding sequences and are likely to disrupt gene function but are not deleterious because functional gene copies are encoded in other repeat units and presumably remain unaffected.

The large duplicated nematode mtDNA sequences reveal a paucity of intra-molecular divergence (0.01%). This observation indicates that the mtDNA amplifications reported here either represent very recent events or that an active homogenizing mechanism may be eliminating sequence divergence as it occurs. Several types of evidence suggest that the duplications and rearrangements within *R. culicivorax* mtDNA are *not* recent events. Amplification and repeat unit inversion are present in two different lineages indicating that these events were probably generated within a common ancestor. These lineages have traversed sufficient generations to allow fixation of different mtDNA haplotypes into homoplasmic nematodes (HYMAN and SLATER 1990). Moreover, observed divergence among repeat-encoded ORFs (including a significant difference of 29% for the amino acid residues in ND6) and the single copy ND1 gene (2%) suggests generation and separation of these two lineages have not occurred recently.

Many short noncoding repeats in animal mtDNA appear to show concerted evolution among copies residing in the same genome, maintained by recurrent cycles of deletion and reduplication (SOLIGNAC, MONNEROT and MOUNOLOU 1986; BOYCE, ZWICK and AQUADRO 1989; RAND and HARRISON 1989; BUROKER *et al.* 1990; WILKINSON and CHAPMAN 1991). Heteroplasmic individuals are frequently generated which contain mt-genomes harboring varying copy numbers of the repeat sequence and often represent a significant proportion of a population. This process may not

be expected to operate in the larger coding sequence duplications because they appear to be generated less frequently, are usually present in fewer copies per molecule, and high copy numbers resulting in extreme mt-genome sizes may not easily be tolerated (MORITZ and BROWN 1987; MORITZ 1991). This seems to be the situation for the *R. culicivorax* mtDNA repeats, which are exceptionally large (3.0 kb). We have not been able to observe heteroplasmic individuals containing mt-molecules with varying numbers of repeat units (HYMAN and SLATER 1990).

mtDNA restriction site polymorphism also has not been observed among nematode subpopulations separated in laboratory culture for more than 200 generations. However, in this same time period segregation of homoplasmic lineages that contain a length variant not found in the 3B4 and NAC6 populations have arisen (HYMAN and SLATER 1990). This is consistent with earlier studies that indicate length variants occur at a frequency higher than point mutations (HARRISON 1989; LAROCHE *et al.* 1990). These observations, along with the nucleotide sequence data reported here, suggest that mechanisms for homogenization of lengthy mtDNA repeats are available in this nematode.

We wish to thank E. PLATZER for the use of his facilities to propagate nematodes, D. BIRD and M. SPRINGER for advice on computer analyses, L. WALLING and M. CLEGG for critical reading of this manuscript, and M. MCCLURE (University of California-Irvine) for assistance in performing the BLASTP sequence analysis at the NCBI. This work was supported by U.S. Department of Agriculture Competitive grant 85-CRCR-1-1629, U.S. Agency for International Development grant DPE-5542-6-55-7059-00, U.S. Public Health Service Biomedical Research funds, the Agricultural Experiment Station and the Research Committee of the University of California, Riverside, and a University of California Conservation Grant (to B.C.H.); and by State of California Mosquito Research, Sigma Xi, and University of California, Riverside Newall Research and Dissertation grants (to J.L.B.A.)

LITERATURE CITED

ALBERTINI, A. M., M. HOFER, M. P. CALOS and J. H. MILLER, 1982 On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell **29:** 319–328.

ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

ATTARDI, G., 1985 Animal mitochondrial DNA: an extreme example of genetic economy. Int. Rev. Cytol. **93:** 93–145.

AVISE, J. C., B. W. BOWEN and T. LAMB, 1989 DNA fingerprints from hypervariable mitochondrial genotypes. Mol. Biol. Evol. **6:** 259–269.

BECK, J. L., and B. C. HYMAN, 1988 Role of sequence amplification in the generation of nematode mitochondrial DNA polymorphism. Curr. Genet. **14:** 627–636.

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.

BOURSOT, P., H. YONEKAWA and F. BONHOMME, 1987 Heteroplasmy in mice with deletion of a large coding region of mtDNA. Mol. Biol. Evol. **4:** 46–55.

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 Press, New York.

BROWN, W. M., E. M. PRAGER, A. WANG and A. C. WILSON, 1982 Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18:** 225–239.

BULLOCK, W. O., J. M. FERNANDEZ and J. M. SHORT, 1987 XL1-Blue: a high efficiency plasmid transforming recA *Escherichia coli* strain with β-galactosidase selection. Bio/Technology **5:** 376.

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.

CHIRGWIN, J. M., A. E. PRZYBYLA, R. J. MACDONALD and W. J. RUTTER, 1979 Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18:** 5294–5299.

CHOMCZYNSKI, P., and N. SACCHI, 1987 Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:** 156–159.

CHOU, P. Y., and G. D. FASMAN, 1978 Prediction of the secondary structure of proteins from their amino acid sequence. Adv. Enzymol. **47:** 45–147.

CLARK-WALKER, G. D., 1985 Basis of diversity in mitochondrial DNAs, pp. 277–297 in *The Evolution of Genome Size*, edited by T. C. SMITH. John Wiley & Sons, New York.

CLARY, D. O., and D. R. WOLSTENHOLME, 1985 The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. **22:** 252–271.

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

CLAYTON, D. A., 1984 Transcription of the mammalian mitochondrial genome. Annu. Rev. Biochem. **53:** 573–594.

DENSMORE, L. D., J. W. WRIGHT and W. M. BROWN, 1985 Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenetic and bisexual lizards (genus *Cnemidophorus*). Genetics **110:** 689–707.

DEVEREUX J., P. HAEBERLI and O. SMITHIES, 1984 A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:** 387–395.

GJETVAJ, B., D. I. COOK and E. ZOUROS, 1992 Repeated sequences and large-scale size variation of mitochondrial DNA: a common feature among scallops (*Bivalvia:Pectinidae*). Mol. Biol. Evol. **9:** 106–124.

GONZALEZ, F. J., F. VILBOIS, J. P. HARDWICK, O. W. MCBRIDE, D. W. NEBERT, H. V. GELBOIN and U. A. MEYER, 1988 Human debrisoquine 4-hydroxylase (P450IID1): cDNA and deduced amino acid sequence and assignment of the CYP2D locus to chromosome 22. Genomics **2:** 174–179.

HALE, L. R., and R. S. SINGH, 1986 Extensive variation and heteroplasmy in size of mitochondrial DNA among geographic populations of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **83:** 8813–8817.

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

HAYASAKA, K., T. ISHIDA and S. HORAI, 1991 Heteroplasmy and polymorphism in the major noncoding region of mitochondrial DNA in Japanese monkeys: association with tandemly repeated sequences. Mol. Biol. Evol. **8:** 399–415.

HENIKOFF, S., 1987 Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol. **155:** 156–165.

HIGGINS, D. G., and P. M. SHARP, 1988 CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene **73:** 237–244.

HYMAN, B. C., J. L. BECK and K. C. WEISS, 1988 Sequence amplification and gene rearrangement in parasitic nematode mitochondrial DNA. Genetics **120:** 707–712.

HYMAN, B. C., and T. M. SLATER, 1990 Recent appearance and molecular characterization of mitochondrial DNA deletions within a defined nematode pedigree. Genetics **124:** 845–853.

ICHIKAWA, T., T. ITAKWIA and M. NEGISHI, 1989 Functional characterization of two cytochrome P-450s within the mouse, male-specific steroid 16α-hydroxylase gene family: expression

in mammalian cells and chimeric proteins. Biochemistry **28:** 4779–4784.

JACOBS, H. T., D. J. ELLIOTT, V. B. MATH and A. FARQUHARSON, 1988 Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. Mol. Biol. **202:** 185–217.

KRETZ, P. L., and J. M. SHORT, 1989 Gigapack™II: restriction free (*hsd*$^-$, *mcrA*$^-$, *mcrB*$^-$, *mrr*$^-$) lambda packaging extracts. Stratagies **2:** 25–26.

KYTE, J., and R. F. DOOLITTLE, 1982 A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:** 333–348.

LAROCHE, 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.

LEHRACH, H., D. DIAMOND, J. M. WOZNEY and H. BOEDTKER, 1977 RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochemistry **16:** 4743–4751.

LEVINSON, G., and G. A. GUTMAN, 1987 Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4:** 203–221.

LIPMAN, D. J., and W. R. PEARSON, 1985 Rapid and sensitive protein similarity searches. Science **227:** 1435–1441.

MATSUNAGA, E., U. M. ZANGER, J. P. HARDWICK, H. V. GELBOIN, U. A. MEYER and F. J. GONZALEZ, 1989 The CYP2D gene subfamily: analysis of the molecular basis of the debrisoquine 4-hydroxylase deficiency in DA rats. Biochemistry **28:** 7349–7355.

MIGNOTTE, F., C. BIJU-DUVAL, A.-M. CHAMPAGNE, M. GUERIDE, M. MONNEROT and J. C. MOUNOLOU, 1990 Repeats and variations in the control region of rabbit mtDNA, pp. 97–101 in *Structure, Function and Biogenesis of Energy Transfer Systems*, edited by E. QUAGLIARIELLO, S. PAPPA, F. PALMIERI and C. SACCONE. Elsevier Science Publishers B.V., The Netherlands.

MORITZ, C., 1991 Evolutionary dynamics of mitochondrial DNA duplications in parthenogenetic geckos, *Heteronotia binoei*. Genetics **129:** 221–230.

MORITZ, C., and W. M. BROWN, 1986 Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science **233:** 1425–1427.

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.

MORITZ, C., T. DOWLING and W. M. BROWN, 1987 Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu. Rev. Evol. Syst. **18:** 269–292.

OKIMOTO, R., H. M. CHAMBERLIN, J. L. MACFARLANE and D. R. WOLSTENHOLME, 1991 Repeated sequence sets in mitochondrial molecules of root knot nematodes (*Meloidogyne*): nucleotide sequences, genome location and potential for host-race identification. Nucleic Acids Res. **19:** 1619–1626.

OKIMOTO, R., J. L. MACFARLANE, D. O. CLARY and D. R. WOLSTENHOLME, 1992 The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics **130:** 471–498.

PALMER, J. D., 1985 Evolution of chloroplast and mitochondrial DNA in plants and algae, pp. 131–240 in *Monographs in Evolutionary Biology: Molecular Evolutionary Genetics*, edited by R. J. MACINTYRE. Plenum Press, New York.

PETERSEN, J. J., H. C. CHAPMAN and D. B. WOODWARD, 1968 The bionomics of a mermithid nematode of larval mosquitoes in southwestern Louisiana. Mosq. News **28:** 346–352.

POWERS, T. O., E. G. PLATZER and B. C. HYMAN, 1986a Large mitochondrial genome and mitochondrial DNA size polymorphism in the mosquito parasite, *Romanomermis culicivorax*. Curr. Genet. **11:** 71–77.

POWERS, T. O., E. G. PLATZER and B. C. HYMAN, 1986b Species-specific restriction site polymorphism in root-knot nematode mitochondrial DNA. J. Nematol. **18:** 288–293.

PRECIOUS, W. Y., and J. BARRETT, 1989 The possible absence of cytochrome P450 linked xenobiotic metabolism in helminths. Biochim. Biophys. Acta **992:** 215–222.

RAND, D. M., and R. G. HARRISON, 1986 Mitochondrial DNA transmission genetics in crickets. Genetics **114:** 955–970.

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

ROE, B. A., D.-P. MA, R. K. WILSON and J. F.-H. WONG, 1985 The complete nucleotide sequence of the *Xenopus* mitochondrial genome. J. Biol. Chem. **260:** 9759–9774.

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., 1984 Structural variation in mitochondrial DNA. Adv. Genet. **22:** 1–108.

SHORT, J. M., J. M. FERNANDEZ, J. A. SORGE and W. D. HUSE, 1988 λ ZAP: a bacteriophage λ expression vector with *in vivo* excision properties. Nucleic Acids Res. **16:** 7583–7600.

SNYDER, M., A. R. FRASER, J. LAROCHE, K. E. GARTNER-KEPKAY and E. ZOUROS, 1987 Atypical mitochondrial DNA from the deep-sea scallop *Placopecten magellanicus*. Proc. Natl. Acad. Sci. USA **84:** 7595–7599.

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

SOLIGNAC, M., J. GENERMONT, M. MONNEROT and J.-C. MOUNOLOU, 1984 Genetics of mitochondria in Drosophila: mtDNA inheritance in heteroplasmic strains of *D. mauritiana*. Mol. Gen. Genet. **197:** 183–188.

SPURR, N. K., A. C. GOUGH, C. A. D. SMITH and C. R. WOLF, 1991 Genetic analysis of cytochrome P450 system. Methods Enzymol. **206:** 149–166.

STREISINGER, G., Y. OKADA, L. 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.

VIEIRA, J., and J. MESSING, 1987 Production of single-stranded plasmid DNA. Methods Enzymol. **153:** 3–11.

WALLACE, D. C., 1982 Structure and evolution of organelle genomes. Microbiol. Rev. **46:** 208–240.

WALLACE, D. C., 1992 Mitochondrial genetics: a paradigm for aging and degenerative diseases? Science **256:** 628–632.

WALLIS, G. P., 1987 Mitochondrial DNA insertion polymorphism and germ line heteroplasmy in the *Triturus cristatus* complex. Heredity **58:** 229–238.

WILKINSON, G. S., and A. M. CHAPMAN, 1991 Length and sequence variation in evening bat D-loop mtDNA. Genetics **128:** 607–617.

WOLSTENHOLME, D. R., R. OKIMOTO, J. L. MACFARLANE, G. A. PONT, H. M. CHAMBERLIN, J. R. GAREY and N. A. OKADA, 1990 Unusual features of lower metazoan mitochondrial genomes, pp. 103–106 in *Structure, Function and Biogenesis of Energy Transfer Systems*, edited by E. QUAGLIARIELLO, S. PAPA, F. PALMIERI and C. SACCONE. Elsevier Science Publishers B.V., The Netherlands.

ZEVERING, C. E., C. MORITZ, A. HEIDEMAN and R. A. STURM, 1991 Parallel origins of duplication and the formation of pseudogenes in mitochondrial DNA from parthenogenetic lizards (*Heteronotia binoei: Gekkonidae*). J. Mol. Evol. **33:** 431–444.

ZURAWSKI, G., M. T. CLEGG and A. H. D. BROWN, 1984 The nature of nucleotide sequence divergence between barley and maize chloroplast DNA. Genetics **106:** 735–749.

Communicating editor: A. G. CLARK