Sequences With the Potential to Form Stem-and-Loop Structures Are Associated With Coding-Region Duplications in Animal Mitochondrial DNA

David J. Stanton,[1] L. Lynne Daehler, Craig C. Moritz[2] and Wesley M. Brown

Department of Biology and Museum of Zoology, University of Michigan, Ann Arbor Michigan 48109-1048

Manuscript received March 19, 1993
Accepted for publication January 12, 1994

ABSTRACT

Tandem duplications of gene-encoding regions occur in the mitochondrial DNA (mt DNA) of some individuals belonging to several species of whiptail lizards (genus *Cnemidophorus*). All or part of the duplicated regions of the mtDNAs from five different species were sequenced. In all, the duplication endpoints were within or immediately adjacent to sequences in tRNA, rRNA or protein genes that are capable of forming energetically stable stem-and-loop structures. In two of these mtDNAs, the duplication endpoints were also associated with a direct sequence repeat of 13 bp. The consistent association of stem-and-loop structures with duplication endpoints suggests that these structures may play a role in the duplication process. These data, combined with the absence of direct or palindromic repeats at three of the pairs of duplication endpoints, also suggest the existence of a mechanism for generating *de novo* duplications that is qualitatively different from those previously modeled.

SIZE variations in animal mitochondrial (mt) DNA, although most often resulting from copy number variation of tandemly repeated sequences in non-coding regions (BROWN 1985; MORITZ *et al.* 1987; RAND and HARRISON 1989), sometimes result from direct tandem duplications of coding regions (MORITZ and BROWN 1986, 1987; WALLIS 1987; MORITZ 1991; ZEVERING *et al.* 1991). Ten such duplications in mtDNAs from lizards in the genus *Cnemidophorus* were characterized by cleavage mapping (MORITZ and BROWN 1986, 1987). From those data, 15 of 20 duplication endpoints appeared to be located within or adjacent to tRNA genes, supporting the hypothesis that tRNA genes can mediate mtDNA rearrangements (BROWN 1985). Further, 8 of the 10 duplications had one end in or near a region consisting of the adjacent $tRNA^{Pro}$ and $tRNA^{Thr}$ genes, identifying it as a hot spot for duplication. However, these inferences were speculative, because they depended upon the assumption that mtDNA gene order in Cnemidophorus is the same as in most vertebrates and because the exact duplication endpoints had not been determined by sequence analysis. In particular, it remained a possibility that the sequences mediating the duplications were near to, but not within, the tRNA genes. Also, MORITZ and BROWN (1987) had noted that five of the 20 endpoints appeared to be in genes encoding proteins or rRNAs, rather than in or near tRNA genes. We have resolved all of these ambiguities by sequence analysis.

Five of the Cnemidophorus duplications that would provide maximum information on junction sequences, and thus on mechanisms of duplication, were chosen for sequence analysis. Attention was initally focussed on a 1.5-kb mtDNA duplication in order to test the hypothesized association of one of its endpoints with a tRNA gene and to characterize the other, which appeared not to be associated with a tRNA gene. This duplication had been found in mtDNAs from 13 of 32 individual *Cnemidophorus uniparens* (MORITZ and BROWN 1987). To locate precisely the endpoints and to confirm the gene order in this region, sequences from two individuals of *C. uniparens*, one with a long (L), duplication-containing mtDNA and one with a standard length (S) mtDNA, were compared. Using primers designed from the *C. uniparens* sequence, mtDNA duplication junctions from four other species that had been mapped by restriction analysis (MORITZ and BROWN 1987) were also sequenced, thus providing identification of additional duplication endpoints. The endpoint-containing regions were examined for the presence of common sequence features that might play a role in the duplication process.

MATERIALS AND METHODS

Details of specimen collection, mtDNA purification, and cleavage mapping are published (MORITZ and BROWN 1986, 1987). Purified mtDNA and plasmid DNA vectors (Bluescript; Stratagene, Inc.) were digested with appropriate restriction enzymes, ligated using T4 DNA ligase, and used to transform *Escherichia coli* XL-1 host cells (Stratagene, Inc.) made transformation competent by the method of HANAHAN (1983). Transformed colonies were screened for recombinant plasmids using the X-Gal color system. Plasmid "minipreps" were prepared by alkaline lysis (KLEIN *et al.* 1980) in order to confirm insert identities. *E. coli* cells containing the desired recombinant plasmids were stored as glycerinated stocks (MANIATIS *et al.* 1982). Plasmid DNA was extracted from cells

[1] Present address: Department of Biology, Eastern Michigan University, Ypsilanti, Michigan 48197.
[2] Present address: Department of Zoology, University of Queensland, St. Lucia, Queensland 4067, Australia.

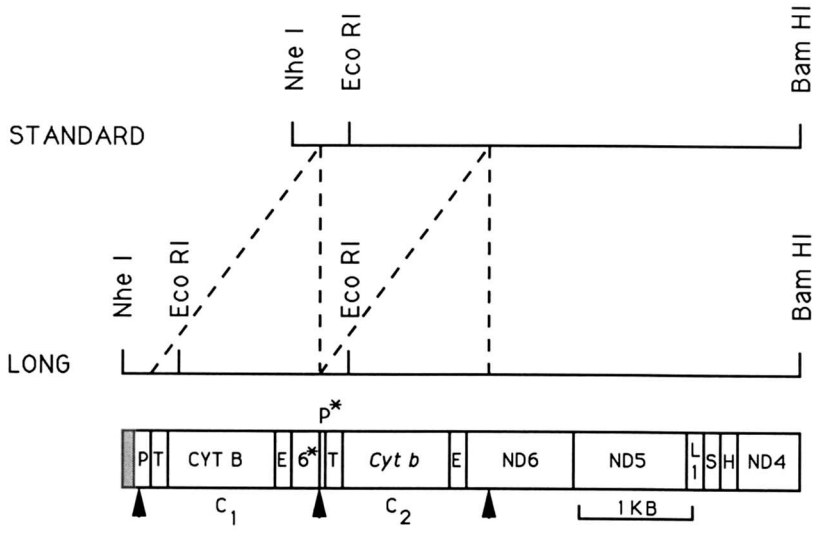


FIGURE 1.—Cleavage maps of homologous regions in the standard and long *C. uniparens* mtDNAs, and the genetic map of the region. Dashed connecting lines indicate the portion of the standard mtDNA that is duplicated in long; the arrowheads indicate the endpoints of the duplicate copies C_1 and C_2. Gene abbreviations: *Cyt b*, cytochrome *b* oxidase apoenzyme; ND4–6, NADH dehydrogenase subunits 4–6; P, T, E, L_1, S, and H, respectively, proline, threonine, glutamate, leucine (CUN), serine, and histidine tRNA; 6* and P*, partial duplications of ND6 and $tRNA^{Pro}$. All genes are transcribed right to left except P, *Cyt b*, E, and ND6. The shaded region at the left end of the genetic map is one end of the control (D-loop) region.

according to MANIATIS *et al.* (1982) and purified by sedimentation equilibrium in CsCl gradients (BROWN 1981; WRIGHT *et al.* 1983) prior to sequencing.

Both strands of each cloned mtDNA were completely sequenced using HATTORI and SAKAKI's (1986) modification of the procedure of SANGER *et al.* (1977), in which double stranded template DNA is alkali denatured and annealed to single primers. The U.S. Biochemical Corp. Sequenase kit was used as the source of sequencing reagents. When necessary, 7-deaza-dGTP was substituted for dGTP in order to resolve band compressions.

Primers complementary to conserved regions flanking the recombinant junctions were made with an ABI model 380-B DNA synthesizer; the conserved regions were identified by comparisons among published mtDNA sequences. The primer sequences appear in the legend to Figure 5. Sequencing reactions were performed according to the supplier's instructions, using the Sequenase kit (U.S. Biochemical Corp.), and products were resolved by electrophoresis in 6 and 4% polyacrylamide gels. Gels were dried, then exposed to X-ray film. For the *C. uniparens* sequence, each nucleotide was independently determined a minimum of six times from the three different copies indicated in Figure 1.

DNA sequences were amplified using the polymerase chain reaction (PCR) (Saiki *et al.* 1988). Each reaction contained approximately 10 ng of purified mtDNA in 100 μl of the following mixture: 10 mM Tris-HCl (pH 8.3); 50 mM KCl; 2.0 mM $MgCl_2$; 0.01% gelatin; 2.0 mM each of the four dNTPs; 1.0 μM each of two primers; and 1.0 unit of AmpliTaq DNA polymerase (Perkin-Elmer Cetus). A thermal cycler (Perkin-Elmer Cetus) was used to carry out 30 cycles consisting of: 1 min at 94° for denaturation; 2 min at 37° for annealing; and 3 min at 72° for elongation. Products were phenol extracted and ethanol precipitated, then treated with 5 units of the large (Klenow) fragment of *E. coli* DNA polymerase I for 15 min at 30° to ensure the presence of blunt ends for ligation, after which they were cloned into plasmid vectors that had been digested with *Eco*RV, and sequenced. In the case of the *Cnemidophorus opatae* and *Cnemidophorus exsanguis* duplication junctions (*e.g.*, Figure 3, D and E), independent clones gave identical junction sequences and no errors due to infidelity during PCR amplification were detected.

Sequence analysis was performed using resident programs in Eugene (MBIR) run on a Sun computer. Gene identities were inferred by comparisons with published sequences for house mouse (BIBB *et al.* 1981) and *Xenopus laevis* (South African clawed toad) (ROE *et al.* 1985) mtDNAs. Duplication endpoints were identified by alignment of the sequences flanking the internal duplication junction with corresponding sequences from the *C. uniparens* S genome, or (for ND1 and rRNA) with appropriate regions of the *Cnemidophorus maslini* and *C. exsanguis* mtDNAs, respectively. Secondary structures for tRNA and rRNA gene sequences were inferred by comparison to published secondary structure models (ROE *et al.* 1985; GLOTZ *et al.* 1981). Putative secondary structures within protein coding regions were inferred using a Macintosh version (supplied by D. GILBERT) of the program MulFold (version 2.0; JAEGER *et al.* 1989a,b; ZUCKER 1989). The free energies of DNA secondary structures were calculated using unpublished, empirically determined values generously provided by HEATHER D. MAYOR (Department of Microbiology, Baylor College of Medicine, Houston, Texas).

RESULTS

Gene arrangement in Cnemidophorus mtDNA: The partial gene arrangement in Cnemidophorus mtDNA (Figure 1), was determined from the sequence shown in Figure 2 and from additional sequence data (D. J. STANTON and W. M. BROWN, unpublished) that will be reported separately. The gene arrangement in this region is identical to those in other vertebrate mtDNAs (BIBB *et al.* 1981; ANDERSON *et al.* 1982; ROE *et al.* 1985; JOHANSEN *et al.* 1989), exclusive of birds (DESJARDINS and MORAIS 1990; DESJARDINS *et al.* 1990).

Duplication endpoints in *C. uniparens*: Comparison of the *C. uniparens* L and S sequences shows that L contains a tandem duplication extending from the $tRNA^{Pro}$ gene into ND6 (Figure 1). The duplicated sequence is 1520 bp in length (Figure 2) and contains the genes for cytochrome *b*, $tRNA^{Thr}$, $tRNA^{Glu}$, and portions of those for ND6 and $tRNA^{Pro}$ (Figure 1). No direct or inverted repeats greater than 12 bp were found within or immediately adjacent to this sequence, nor were there any cases in which corresponding members of smaller (≤12 bp) repeats were located at each duplication endpoint.

Except for the redundancy resulting from the duplication itself, the L and S sequences are nearly identical.

FIGURE 2.—A portion of the *C. uniparens* standard mtDNA sequence that includes the region duplicated in the *C. uniparens* long mtDNA (see Figure 1). Positions 1–97 and 1620–1719 flank this region. Brackets above the sequence indicate gene boundaries, and arrows indicate the direction of transcription. Genes are abbreviated as in Figure 1. Nucleotides at positions that differ between the standard and long mtDNAs, or between the two copies (C_1 and C_2, Figure 1) in the long mtDNA are in boldface and underlined. The C_2 and standard sequences differ by three substitutions (T at position 23, C at 158, and A at 217) and by the additional nucleotide (A) in C_2 between positions 33 and 34. The C_1 and standard sequences are identical, except that C_2 has 6 instead of 8 T's on the interval 1432–1439; this results in a 2 bp shift of the reading frame in this (presumably inactive) partial copy of ND6. Sequences capable of forming potentially stable stem-and-loop structures (Figure 4) are underlined.

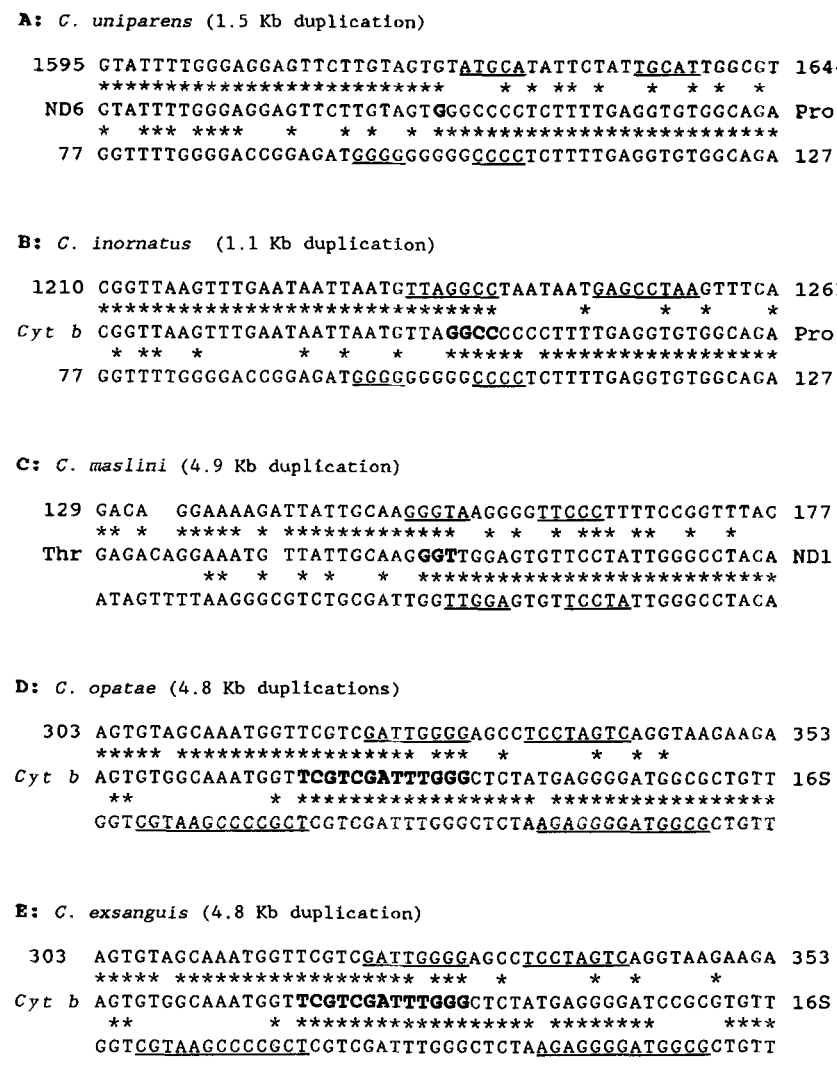


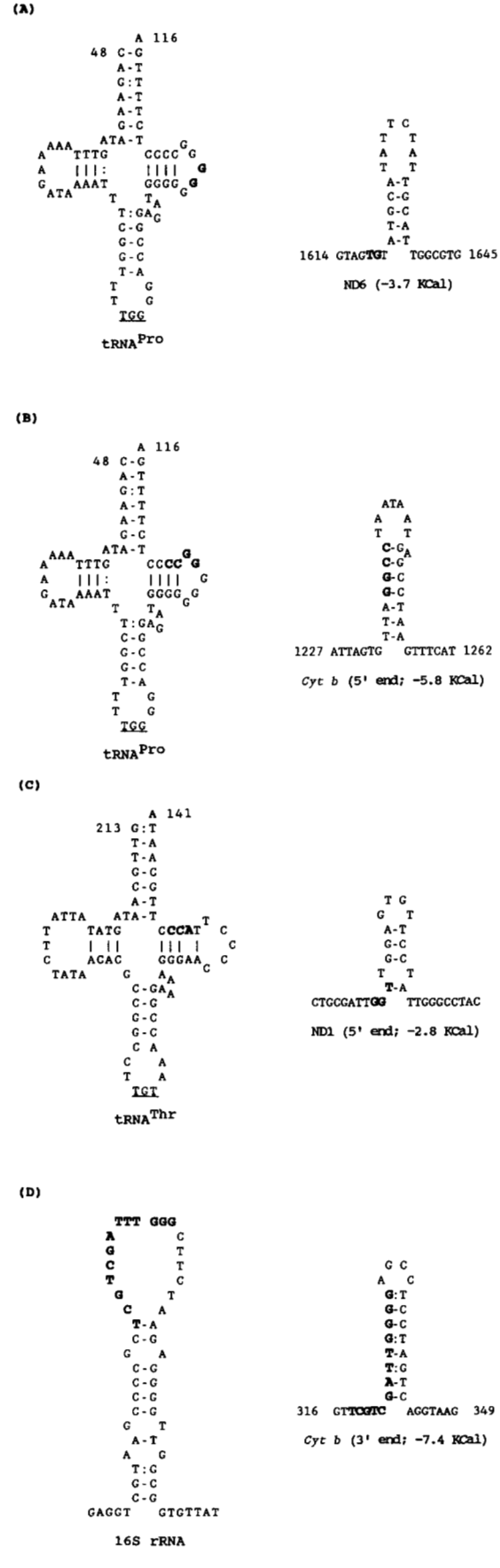
FIGURE 3.—Identification of duplication endpoints. For each of the five duplications (A–E), the sequence containing the junction between duplicate copies is shown between portions of either the *C. uniparens* standard mtDNA sequence (A, B), numbered as in Figure 2, or between a portion of that sequence and a portion of the *C. maslini* (C) or *C. exsanguis* (D, E) mtDNA sequence. Unnumbered sequences for *C. maslini* and *C. exsanguis* were obtained by sequencing PCR amplified regions of mtDNAs from these species. Gene identities for the upper sequences are shown to the left of the middle sequences, and those for the lower sequences are shown to their right. Genes are abbreviated as in Figure 1. Asterisks are placed between identical aligned nucleotides. Sequences that are hypothesized to form stable stem-and-loop structures (see Figure 4) are underlined. Nucleotides at duplication endpoints are in boldface; when an exact endpoint could not be determined, all nucleotides on the interval that must contain the endpoint are in boldface. The two-nucleotide gaps near the left ends of the upper and middle sequences in (C) have been inserted for alignment purposes. The gaps are in an intergenic region that is highly variable in length among vertebrates.

Within L, the duplicate copies (C_1 and C_2, Figure 1) are also nearly identical. The most notable difference is the deletion in C_1 of two T's from the run of eight T's found in the interval 1432–1439 in both S and C_2 (Figure 2). This deletion disrupts the ND6 reading frame in C_1.

Comparison of the junction between the tandem repeats in the L sequence with S sequences corresponding to the ends of the duplicated region identifies the positions of the endpoints to within ±1 bp (Figure 3A), at positions 98 or 99 and 1618 or 1619 (Figure 2). One is in the "TC" loop of the $tRNA^{Pro}$ gene and the other is in the ND6 gene, 193 ± 1 bp from its 5′ end (Figure 4A). No similarities were found between the sequences at or near the ends of this duplication. However, the region just beyond the endpoint in ND6 contains a 5-bp sequence separated from its inverse repeat by 8 bp (Figure 2). These complimentary sequences can potentially interact to form a 5 bp stem with an 8 bp loop (Figure 4A), a structure whose formation is energetically favorable (free energy = −3.7 kcal). Thus, although only one endpoint is within a tRNA gene, both are associated with potentially stable stem-and-loop structures.

Duplication endpoints in other species: Sequences through four other duplication junctions (*i.e.*, those for duplications of 1.1 kb in *C. inornatus*, 4.9 kb in *C. maslini*, and 4.8 kb in *C. opatae* and *C. exsanguis*; MORITZ and BROWN 1987) were also obtained. The sequences of those junctions, the potential secondary structures at or adjacent to their respective endpoints, and the locations of the duplicated regions on the mitochondrial genetic map are shown in Figures 3, 4 and 5.

For the 1.1-kb duplication in *C. inornatus* (Figure 3B), one endpoint is in the "TC" loop of the $tRNA^{Pro}$ gene, at a position on the interval 103–106 (Figure 2) that is from 1 to 5 bp from the corresponding *C. uniparens* duplication endpoint (Figures 3B and 4B). The other endpoint is near the 5′ end of the cytochrome *b* gene, on the interval 1236–1239 (Figure 2); like the *C. uniparens* duplication endpoint in the ND6 gene, it is associated with a potentially stable stem-and-loop structure (Figure 4B).



The endpoints of the 4.9-kb duplication in *C. maslini* are located within the $tRNA^{Thr}$ and ND1 genes (Figure 3C). One endpoint is within the "TC" stem of the $tRNA^{Thr}$ gene and the other is near the 5′ end of ND1, immediately adjacent to a sequence capable of forming a stable stem-and-loop structure (Figure 4C).

The endpoints of the *C. opatae* and *C. exsanguis* duplications could not be determined precisely, due to the presence of a nearly identical (12/13) 13-bp sequence found at the endpoint regions in both the cytochrome *b* and 16S rRNA genes. However, the endpoints for both species must be within or immediately adjacent to this 13-bp sequence (Figure 3, D and E). One endpoint is located near position 330 (Figure 2), within or near a stem-and-loop structure near the 3′ end of the *Cyt b* gene (Figure 4D). The other is located within the 16S rRNA gene in a sequence that forms a stem-and-loop structure (Figure 4D); this structure is conserved in large subunit rRNAs from *E. coli* to man (GLOTZ *et al.* 1981).

Of the 10 duplication endpoints investigated, all are unambiguously associated with potentially stable stem-and-loop structures (Figure 4). Three of the endpoints are associated with the "TC" stem-and-loop of tRNA genes. However, their exact positions within this structure are not constant; one is in the loop (Figure 4A), one is in the stem (Figure 4C), and one is in either the stem or the loop (Figure 4B). The most consistent feature shared by all five duplications is the presence of one endpoint immediately adjacent to the 3′ end of the D-loop region (Figure 5). It may also be noteworthy that there is a consistent association of the endpoints with runs of G and/or C nucleotides (Figure 3). Finally, in two cases the endpoints were associated with stem-and-loop structures and a 13-bp direct sequence repeat (Figure 4D).

FIGURE 4.—Potential secondary structures associated with duplication endpoints. All possible endpoint positions are in boldface (see Figure 3). The sequences from which the pairs of structures in A–C are formed correspond, respectively, to those found at the endpoints of the duplications whose recombinant junctions are shown in A–C of Figure 3; those in D correspond to both D and E of Figure 3, since those duplication endpoints cannot be distinguished. Sequences are numbered as in Figure 2. Values for the minimum energy structures shown were calculated by the MulFold program (JAEGER *et al.* 1989a,b; ZUCKER 1989), using empirically determined energy values for DNA (H. MAYOR, unpublished data); no values were calculated for tRNA or rRNA structures. Dashes represent hydrogen bonding between nucleotides; colons represent pairings with zero free energy. The structures were inferred by inspection and alignment with published vertebrate mitochondrial tRNA and rRNA sequences (see, *e.g.*, GLOTZ *et al.* 1981). All structures are based on *C. uniparens* standard mtDNA sequences except those in the ND1 and 16S rRNA genes, which are based on *C. maslini* and *C. exsanguis* mtDNA sequences, respectively.

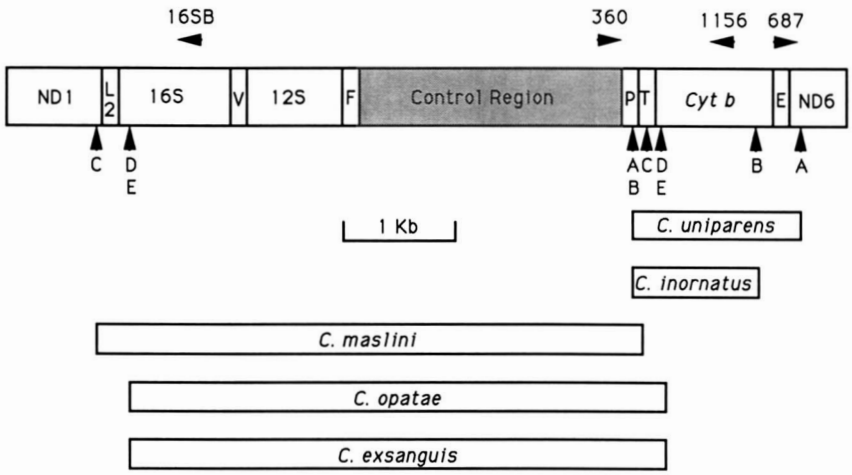


FIGURE 5.—Duplication endpoints for Cnemidophorus mtDNA. Rectangles indicate the regions duplicated in the species indicated. The duplication endpoints for each of the five species (A–E) are indicated by vertical arrowheads; species designations correspond to those in Figure 3. Genes are abbreviated as in Figure 1 and as follows: ND1, NADH dehydrogenase subunit 1; 16S and12S, respectively, large and small ribosomal subunit RNA; F, V, L_2, respectively, phenylalanine, valine and leucine (UUR) tRNA. All genes are transcribed from left to right except P, E, and ND6. Locations of sequences correponding to primers are indicated by horizontal arrowheads. Sequences (5′ to 3′) of the primer pairs that were used to PCR-amplify the recombinant junctions of duplications were: for B, 687 (AGCTAGAAATAGTCCGG) and 1156 (GGATTAAACTCCAACAC); for C, D and E, 360 (GGAAAAGAGATGAATGTAC) and 16SB (GCCTGTTTACCAAAAACAT). The 16SB sequence was a consensus of aligned vertebrate mtDNA sequences. Note that primer orientation precludes amplification in the absence of a duplication.

DISCUSSION

This analysis of sequences both confirms and extends the general conclusions drawn from the previous restriction analysis of these duplications (MORITZ and BROWN 1987). It confirms that the assumed gene arrangement in this region of Cnemidophorus mtDNA and the gene content of the duplications are correct, and that three of the duplication endpoints lie within the $tRNA^{Pro}$ and $tRNA^{Thr}$ genes. However, the sequence analyses also show that two fewer endpoints actually lie within tRNA genes than had been inferred in the previous study (3/10 instead of 5/10). The most significant extension of the previous conclusions is that all 10 endpoints, whether in tRNA, rRNA or protein genes, are either contained within or less than 5 nucleotides from potential stem-and-loop structures (Figure 4). In two cases, duplications that have arisen independently have identical or nearly identical endpoints, which indicates that the duplication mechanism(s) must act with a fair amount of precision.

Two assumptions underlying the conclusions about endpoints are that the five duplications analyzed represent five independent events and that the original endpoints have not been obscured by base substitutions, additions or deletions subsequent to the duplication event. Phylogenetic evidence addresses the issue of independent origins. The duplications most likely to be related by common ancestry are those in *C. uniparens* (1.5 kb) and *C. inornatus* (1.1 kb), which share a common endpoint in the $tRNA^{Pro}$ gene (Figure 5). *C. uniparens* is a hybrid parthenogenetic species whose mtDNA was inherited from *C. inornatus* (DENSMORE *et al.* 1989). It is conceivable that the 1.5-kb duplication arose in *C. inornatus*, was inherited by *C. uniparens*, then subsequently reduced by deletion to 1.1 kb. This is unlikely, however, because the *C. inornatus* populations to which the maternal parent of *C. uniparens* most likely belonged (see DENSMORE *et al.* 1989) lack duplications, as do most *C. uniparens*.

A more dramatic example of shared endpoints is provided by the 4.8-kb duplications in two other species, *C. opatae* and *C. exsanguis*, both of which are parthenogenetic hybrids (Figure 5). Despite having identical or nearly identical endpoints, these two duplications occur only rarely in these distantly related (>10% sequence divergence) mtDNAs (MORITZ *et al.* 1992). The maternal parent species of *C. opatae* is *C. inornatus* (DENSMORE *et al.* 1989) and that of *C. exsanguis* is either *C. costatus* or a closely related species, *C. burti* (MORITZ *et al.* 1989). The maternal parent species are not closely related to each other, and this duplication has not been found in other species closely related to either of them. Thus, it seems most unlikely that the two duplications arose in and were inherited from a common ancestor; that the two duplications arose independently is a far more parsimonious explanation. Although both duplications must end within or immediately adjacent to the same 13-bp sequence (Figure 3, D and E), a more precise determination of their endpoints is impossible.

The possibility that the junction sequences have been modified by deletions cannot be evaluated with these data. Deletions within duplicated segments of mtDNAs have previously been reported in Cnemidophorus (MORITZ and BROWN 1986) and Heteronotia (MORITZ

1991; ZEVERING *et al.* 1991). Such deletions would be impossible to detect, however, if they were immediately adjacent to the internal junction of the duplication.

Sequence elements and endpoints: Directly repeated sequences are associated with both deletions (QUIGLEY and WEIL 1985) and inversions (HOWE *et al.* 1988) in chloroplast DNA, and with several large deletions in human mtDNA (HOFFNER *et al.* 1989; SCHON *et al.* 1989; DEGOUL *et al.* 1991). In theory, such sequences could also promote duplication, but none of those found in or near the *C. uniparens* duplication (Figure 2) or at the junctions between duplicated regions in *C. inornatus* and *C. maslini* (Figure 3, B and C) suggest an involvement in the duplication process. However, the 13-bp direct repeat that is present at the endpoints in both *C. opatae* and *C. exsanguis* may have played a role in the formation of those two duplications.

Pairs of palindromic and quasipalindromic sequences capable of complementary base pairing often occur near duplication and deletion endpoints. Slipped-strand mispairing of such sequences during DNA replication has been postulated as the initiating mechanism for such events (RIPLEY and GLICKMAN 1983; GLICKMAN and RIPLEY 1984). However, no such sequences were found at or near the duplication endpoints in *C. uniparens* mtDNA (Figure 2) or associated with the duplication junctions in the mtDNAs of the four other species examined.

Based on the analysis of the sequences presented and on our previous cleavage mapping data (MORITZ and BROWN 1987), the most reasonable overall hypothesis is that there is a duplication process in Cnemidophorus mtDNA that is mediated by stem-and-loop structures. The stable formation of these stem-and-loop structures in duplex DNA (*e.g.*, by transformation of linear helices to cruciform structures) is improbable, based on free energy considerations. Their stable formation in single-stranded DNA is much more probable. In animal mtDNA, a portion of the molecule is single-stranded for a significant period during replication (reviews by CLAYTON 1982; BROWN 1985). As previously noted, all duplications either begin near one end of or include the control (D-loop) region, where mtDNA replication is initiated (MORITZ and BROWN 1987). Thus, one mtDNA strand in the region within which the duplications occur is single-stranded for a considerable portion of the replication cycle. This provides an opportunity for structures with intrastrand base-pairing to form and to persist for a relatively long period.

The ability of tRNA genes to mediate rearrangement events in DNA is well documented. In bacteria (REITER *et al.* 1989), slime molds (MARSCHALEK *et al.* 1989), and yeast (SANDMEYER *et al.* 1988; CHALKER and SANDMEYER 1990), tRNA genes appear to function as genomic landmarks for the site-specific integration of many different types of genetic elements. In chloroplasts, tRNA genes are associated with rearrangements and inversions (QUIGLEY and WEIL 1985; HOWE *et al.* 1988). In animal mtDNA, gene order comparisons suggest that tRNA genes rearrange at a higher frequency than protein or rRNA genes (BROWN 1985; WOLSTENHOLME and CLARY 1985; CANTATORE *et al.* 1987; MORITZ *et al.* 1987). Thus, the association of tRNA genes and, by extension, stem-and-loop structures with a rearrangement-like process in animal mtDNA is not without precident.

Duplication mechanism: Although these results suggest that stem-and-loop structures might mediate sequence duplication in mtDNA, the duplication mechanism remains obscure. Duplications could arise by intermolecular recombination, transposition, or slipped-strand mispairing during replication. Although not thoroughly studied, recombination appears to be at most, infrequent in animal mtDNA (BROWN 1983); if present, however, even a low frequency of recombination might suffice to produce duplications. Transposition is an unlikely mechanism, because it is improbable that transpositions would produce tandem duplications exclusively, and because the duplication sequences examined lack the terminal direct repeats that are characteristic of most transposable sequences (CALOS and MILLER 1980). Slipped-strand mispairing (LEVINSON and GUTMAN 1987) has been invoked to account for size variation in mtDNA [see references in MORITZ *et al.* (1987) and MORITZ (1991)]. RAND and HARRISON (1989) found seven mtDNA size classes resulting from differences in copy number of a tandemly repeated sequence in crickets. They postulated that these differences could have resulted from slipped-strand mispairing, or from an unknown process mediated by a 14-bp sequence with dyadic symmetry that was found at the repeat termini. The 13-bp direct repeat found at the duplication endpoints in *C. opatae* and *C. exsanguis* is compatible with a slipped-strand mispairing hypothesis. However, because no such sequences are present at the duplication termini in the other Cnemidophorus mtDNAs, and because higher order variation in copy number has not been observed in any Cnemidophorus mtDNAs (MORITZ and BROWN 1987), the present data do not support slipped-strand mispairing as a universal mechanism.

A stem-and-loop structure appears to serve as the signal for the enzyme that initiates L-strand synthesis during mtDNA replication (WONG and CLAYTON 1985, 1986; HIXSON *et al.* 1986). These structures also appear to serve as the signals that mediate processing of the polygenic mtRNA transcript into its components (BATTEY and CLAYTON 1980; OJALA *et al.* 1981). Thus, similar structures, even those arising by chance, might be accidentally recognized by these or other enzymes and initiate a process that occasionally leads to tandem sequence duplication. The structures might also function in combination with other sequences or sequence characteristics, such as the runs of G and/or C that are associated with all ten of the duplication endpoints investigated (Figure 3). Such a

structure/sequence complex might be mistakenly recognized by an enzyme such as endonuclease G, which appears to be capable of generating the primers required to initiate replication of mtDNA. This enzyme is known to cleave DNA at double stranded $(dG)_n \cdot (dC)_n$ and at single stranded $(dC)_n$ sequences (COTE and RUIZ-CARRILLO 1993), like those associated with the duplication endpoints.

These results have relevance that goes beyond duplication in mtDNA alone. The presence of small palindromic or repeated sequence elements (either direct or inverted) has been a common feature of previous models that have been invoked to explain *de novo* sequence duplication. The evidence presented here indicates that there must also be a second, previously undetected duplication mechanism that does not require those elements, but requires stem-and-loop structures instead.

Finally, the presence of both unduplicated and duplicated mtDNA sequences in individuals of the same species also provides opportunities for understanding processes of genetic change within the mitochondrion and, perhaps, some of the events that may be involved in the evolution of mitochondrial gene rearrangements (BROWN 1985).

We thank H. MAYOR for supplying unpublished energy values for DNA secondary structures, and J. BOORE, T. COLLINS, S. HOFFMAN, E. T. SNOW and W. K. THOMAS for comments on the manuscript. This work was supported, in part, by grants from the U.S. Public Health Service (GM30144) and the National Science Foundation (BSR-8517830).

LITERATURE CITED

ANDERSON, S., M. H. L. DEBRUIJN, A. R. COULSON, I. C. EPERON, F. SANGER *et al.*, 1982 Complete sequence of bovine mitochondrial DNA. J. Mol. Biol. **156:** 683–717.

BATTEY, J., and D. A. CLAYTON, 1980 The transcription map of human mitochondrial DNA implicates transfer RNA excision as a major processing event. J. Biol. Chem. **255:** 11599–11606.

BIBB, M. J., R. A. VAN ETTEN, C. T. WRIGHT, M. W. WALBERG and D. A. CLAYTON, 1981 Sequence and gene organization of mouse mitochondrial DNA. Cell **26:** 167–180.

BROWN, W. M., 1981 Mechanisms of evolution in animal mitochondrial DNA. Ann. N.Y. Acad. Sci. **361:** 119–134.

BROWN, W. M., 1983 Evolution of animal mitochondrial DNA, pp. 62–88 in *Evolution of Genes and Proteins*, edited by M. NEI and R. K. KOEHN. Sinauer, Sunderland Mass.

BROWN, W. M., 1985 Evolution of the animal mitochondrial DNA genome, pp. 95–130 in *Molecular Evolutionary Genetics*, edited by R. MACINTYRE. Plenum Press, New York.

CALOS, M. P., and J. M. MILLER, 1980 Transposable elements. Cell **20:** 579–595.

CANTATORE, P., M. N. GADALETA, M. ROBERTI, C. SACCONE and A. C. WILSON, 1987 Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature **329:** 853–855.

CHALKER, D. L., and S. B. SANDMEYER, 1990 Transfer RNA genes are genomic targets for *de novo* transposition of the yeast retrotransposon Ty3. Genetics **126:** 837–845.

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

COTE, J., and A. RUIZ-CARRILLO, 1993 Primers for mitochondrial DNA replication generated by endonuclease G. Science **261:** 765–769.

DEGOUL, F., I. NELSON, S. AMSELEM, N. ROMERO, B. OBERMAIER-KUSSER *et al.*, 1991 Different mechanisms inferred from sequences of human mitochondrial DNA deletions in ocular myopathies. Nucleic Acids Res. **19:** 493–496.

DENSMORE, L. D., C. MORITZ, J. W. WRIGHT and W. M. BROWN, 1989 Mitochondrial DNA analyses and the origin and relative age of parthenogenetic lizards (genus *Cnemidophorus*). IV. Nine *Sexlineatus*-group unisexuals. Evolution **43:** 969–983.

DESJARDINS, P., and R. MORAIS, 1990 Sequence and gene organization of the chicken mitochondrial genome. J. Mol. Biol. **212:** 599–634.

DESJARDINS, P., V. RAMIREZ and R. MORAIS, 1990 Gene organization of the Peking duck mitochondrial genome. Curr. Genet. **17:** 515–518.

GLICKMAN, B. W., and L. S. RIPLEY, 1984 Structural intermediates of deletion mutagenesis: A role for palindromic DNA. Proc. Natl. Acad. Sci. USA **81:** 512–516.

GLOTZ, C., C. ZWIEB and R. BRIMACOMBE, 1981 Secondary structure of the large subunit ribosomal RNA from *Escherichia coli, Zea mays* chloroplast, and human and mouse mitochondrial ribosomes. Nucleic Acids Res. **9:** 3287–3306.

HANAHAN, D., 1983 Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. **166:** 557–580.

HATTORI, M., and Y. SAKAKI, 1986 Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152:** 232–238.

HIXSON, J. E., T. W. WONG and D. A. CLAYTON, 1986 Both the conserved and divergent 5′-flanking sequences are required for initiation at the human mitochondrial origin of light strand replication. J. Biol. Chem. **261:** 2384–2390.

HOFFNER, J. M., M. T. LOTT, A. S. VOLJAVEC, S. A. SOUEIDAN, D. A. COSTIGAN *et al.*, 1989 Spontaneous Kearns-Sayre/chronic external opthaloplegia plus syndrome associated with a mitochondrial DNA deletion: a slip-replication model and metabolic therapy. Proc. Natl. Acad. Sci. USA **86:** 7952–7956.

HOWE, C. J., R. F. BARKER, C. M. BOWMAN and T. A. DYER, 1988 Common features of three inversions in wheat chloroplast DNA. Curr. Genet. **13:** 343–349.

JAEGER, J. A., D. H. TURNER and M. ZUCKER, 1989a Improved predictions of secondary structures for RNA. Proc. Natl. Acad. Sci. USA **86:** 7706–7710.

JAEGER, J. A., D. H. TURNER and M. ZUCKER, 1989b Predicting optimal and suboptimal secondary structure for RNA. Methods Enzymol. **183:** 281–306.

JOHANSEN, S., P. H. GUDDAL and T. JOHANSEN, 1989 Organization of the mitochondrial genome of Atlantic cod, Gadus morhua. Nucleic Acids Res. **18:** 411–419.

KLEIN, R. D., E. SELSING and R. D. WELLS, 1980 A rapid microscale technique for isolation of recombinant plasmid DNA suitable for restriction enzyme analysis. Plasmid **3:** 88–91.

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

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MARSCHALEK, R., T. BRECHNER, E. AMON-BOHM and T. DINGERMANN, 1989 Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum.* Science **244:** 1493–1496.

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. E. DOWLING and W. M. BROWN, 1987 Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. Annu. Rev. Ecol. Syst. **18:** 269–292.

MORITZ, C., J. W. WRIGHT and W. M. BROWN, 1989 Mitochondrial DNA analyses and the origin and relative age of parthenogenetic lizards (genus *Cnemidophorus*). III. *C. velox* and *C. exsanguis.* Evolution **43:** 958–968.

MORITZ, C., J. W. WRIGHT and W. M. BROWN, 1992 Mitochondrial DNA analyses and the origin and relative age of parthenogenetic *Cnemidophorus*: phylogenetic constraints on hybrid origins. Evolution **46:** 184–192.

OJALA, D., J. MONTOYA and G. ATTARDI, 1981 tRNA punctuation model of RNA processing in human mitochondria. Nature **290:** 470–474.

QUIGLEY, F., and J. H. WEIL, 1985 Organization and sequence of five tRNA genes and of an unidentified reading frame in the wheat chloroplast genome: evidence for gene rearrangements during the evolution of chloroplast genomes. Curr. Genet. **9:** 495–503.

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

REITER, W., P. PALM and S. YEATS, 1989 Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. **17:** 1907–1914.

RIPLEY, L. S., and B. W. GLICKMAN, 1983 Unique self-complementarity of palindromic sequences provides DNA structural intermediates for mutation. Cold Spring Harbor Symp. Quant. Biol. **47:** 851–861.

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

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. HIGUCHI *et al.*, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SANDMEYER, S. B., V. W. BILANCHONE, D. J. CLARK, P. MORCOS G. F. CARLE *et al.*, 1988 Sigma elements are position-specific for many different yeast tRNA genes. Nucleic Acids Res. **16:** 1499–1515.

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

SCHON, E. A., R. RIZZUTO, C. T. MORAES, H. NAKASE, M. ZEVIANI *et al.*, 1989 A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. Science **244:** 346–349.

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

WOLSTENHOLME, D. R., and D. O. CLARY, 1985 Sequence evolution of *Drosophila* mitochondrial DNA. Genetics **109:** 725–744.

WONG, T. W., and D. A. CLAYTON, 1985 *In vitro* replication of human mitochondrial DNA: Accurate initiation at the origin of light strand synthesis. Cell **42:** 951–958.

WONG, T. W., and D. A. CLAYTON, 1986 DNA primase of human mitochondria is associated with structural RNA that is essential for enzymatic activity. Cell **45:** 817–825.

WRIGHT, J. W., C. SPOLSKY and W. M. BROWN, 1983 The origin of the parthenogenetic lizard *Cnemidophorus laredoensis* inferred from mitochondrial DNA analysis. Herpetologica **39:** 410–416.

ZEVERING, C. E., C. MORITZ, A. HEIDEMAN and R. A. STRUM, 1991 Parallel origins of duplications and the formation of pseudogenes in mitochondrial DNA from parthenogenetic lizards (*Heteronotia binoei*; Gekkonidae). J. Mol. Evol. **33:** 431–441.

ZUCKER, M., 1989 On finding all suboptimal foldings of an RNA molecule. Science **244:** 48–52.

Communicating editor: J. E. BOYNTON