Copyright © 1998 by the Genetics Society of America

The Complete Nucleotide Sequence of a Snake (*Dinodon semicarinatus*) Mitochondrial Genome With Two Identical Control Regions

Yoshinori Kumazawa,* Hidetoshi Ota,† Mutsumi Nishida‡ and Tomowo Ozawa*

**Department of Earth and Planetary Sciences, Nagoya University, Nagoya 464-8602, Japan, †Tropical Biosphere Research Center, University of the Ryukyus, Nishihara, Okinawa 903-0213, Japan and ‡Department of Marine Bioscience, Fukui Prefectural University, Obama, Fukui 917-0003, Japan*

Manuscript received December 25, 1997
Accepted for publication June 9, 1998

ABSTRACT

The 17,191-bp mitochondrial DNA (mtDNA) of a Japanese colubrid snake, akamata (*Dinodon semicarinatus*), was cloned and sequenced. The snake mtDNA has some peculiar features that were found in our previous study using polymerase chain reaction: duplicate control regions that have completely identical sequences over 1 kbp, translocation of $tRNA^{Leu}$(UUR) gene, shortened TψC arm for most tRNA genes, and a pseudogene for $tRNA^{Pro}$. Phylogenetic analysis of amino acid sequences of protein genes suggested an unusually high rate of molecular evolution in the snake compared to other vertebrates. Southern hybridization experiments using mtDNAs purified from multiple akamata individuals showed that the duplicate state of the control region is not a transient or unstable feature found in a particular individual, but that it stably occurs in mitochondrial genomes of the species. This may, therefore, be regarded as an unprecedented example of stable functional redundancy in animal mtDNA. However, some of the examined individuals contain a rather scanty proportion of heteroplasmic mtDNAs with an organization of genes distinct from that of the major mtDNA. The gene organization of the minor mtDNA is in agreement with one of models that we present to account for the concerted evolution of duplicate control regions.

SINCE the early 1980s, complete mitochondrial DNA (mtDNA) sequences have been determined for a number of vertebrates (Anderson *et al.* 1981; Wolstenholme 1992; Lee and Kocher 1995; Janke and Arnason 1997; Zardoya and Meyer 1997 and references therein). However, these examples are extremely biased toward mammals and fish, with the exception of the chicken (Desjardins and Morais 1990), ostrich (Härlid *et al.* 1997), African clawed toad (Roe *et al.* 1985), and alligator (Janke and Arnason 1997). The vertebrate mtDNAs characterized to date are double-stranded, circular DNAs of 16–17 kbp, they encode genes for 13 proteins, two rRNAs, and 22 tRNAs, and they have a major noncoding or control region that contains signals for replicating the heavy strand of mtDNA and for transcription (*e.g.*, see Clayton 1992; Wolstenholme 1992). These genomes are, therefore, conserved with respect to gene content. No repetition or deletion of any of the genes or the control region has been shown.

In the 1980s, the gene order of the vertebrate mtDNAs was also considered to be conserved. This, however, no longer holds true in light of recent findings of distinct mtDNA gene orders for marsupials (Pääbo *et al.* 1991; Janke *et al.* 1994), birds (*e.g.*, see Desjardins and Morais 1990; Quinn and Wilson 1993; Härlid *et al.* 1997), crocodilians (Seutin *et al.* 1994; Kumazawa and Nishida 1995; Quinn and Mindell 1996; Janke and Arnason 1997), agamas (Macey *et al.* 1997), the Texas blind snake (Kumazawa and Nishida 1995), the bullfrog (Yoneyama 1987), the rice frog (Macey *et al.* 1997), and sea lamprey (Lee and Kocher 1995). In these examples of gene rearrangements, there is no evidence of duplication or deletion of specific genes or the control region element, but rather, simply the reordering of the same 37 genes and the control region. Although some tRNA genes are tandemly duplicated in amphisbaenian mtDNA, one of the redundant gene copies may be a pseudogene (Macey *et al.* 1998).

In our recent work (Kumazawa *et al.* 1996), however, it was found that mtDNAs of several snakes (the viperids himehabu and western rattlesnake, the colubrid akamata, and the boids ball python and boa constrictor) have duplicate control regions. This finding was based on polymerase chain reaction (PCR) amplification and sequencing of two separate mtDNA regions: one surrounded by cytochrome *b* (cytb) and 12S rRNA genes, and the other surrounded by NADH dehydrogenase subunits 1 and 2 (ND1 and ND2, respectively) genes. Notably, in the former three species, the duplicate control regions were found to share nearly the same nucleotide sequence.

This finding raised profound evolutionary questions. We considered that the original duplication of the control region took place in rather ancient lineages of

Corresponding author: Yoshinori Kumazawa, Department of Earth and Planetary Sciences, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan.
E-mail: h44858a@nucc.cc.nagoya-u.ac.jp

snakes because it seems unlikely that the insertion of a control region into a distant and unique location occurred in parallel in multiple snake lineages. Rather, it is more likely that the control region had already been duplicated in mtDNAs of the last common ancestor of the above-mentioned five snakes, which existed at least 70 mya according to fossil records for the minimum divergence time between the Boidae and the Viperidae-Colubridae clade (Rage 1987). If so, it is mysterious that the duplicate state of the control region may have been maintained for such a long period of time in snake mitochondrial genomes because any redundant genes or sequences in mtDNAs, including one of the duplicate control regions, would most likely be deleted rapidly (see Kumazawa *et al.* 1996 for the reasoning).

Another mysterious point is that the control region sequences were nearly identical between the duplicate copies within the species while they were otherwise divergent to a large extent between the species. If the paralogous control regions within an mtDNA molecule had evolved independently since the presumed original insertion event (>70 mya), a much more substantial sequence difference would be expected. It was thus proposed that the duplicate control regions have evolved in a highly concerted fashion (Kumazawa *et al.* 1996). However, unlike the commonness of the concerted evolution in nuclear DNAs, examples of such concerted evolution in mtDNA have not been reported from animals other than the snakes. Currently available information is too scanty to reveal the mechanism of this bizarre phenomenon, making it intriguing to further characterize the concerted evolution in snake mtDNAs.

In this communication, we report an entire mtDNA nucleotide sequence of a Japanese colubrid, akamata. On the basis of the complete gene organization of snake mtDNA, we present models for the concerted evolution of snake control regions. These models are assessed by Southern hybridization experiments that help to detect minor mtDNA components, and by cloning and sequencing of such minor molecules. The Southern hybridization experiments also indicated that the duplicate state of the control region is stable in character and shared by multiple individuals of the species. This may, therefore, be regarded as an unprecedented example of stable functional redundancy in animal mtDNA. Determination of the entire sequence of snake mtDNA also led to an unexpected finding that molecular evolutionary rates of protein genes encoded in snake mtDNA are unusually high.

MATERIALS AND METHODS

Akamata (*Dinodon semicarinatus*) is a nonvenomous colubrid snake that is endemic to the central Ryukyus of Japan. mtDNAs were analyzed for four individuals from Okinawajima Island (designated Okinawa 1–4), two individuals from Tokunoshima Island (Toku 1–2), and two individuals from Amamioshima Island (Amami 1–2). These islands have long been isolated from each other, and it thus seems that there has recently been no gene flow between populations of these islands in nature (Ota 1986). mtDNA was isolated from fresh liver essentially as described by Tapper *et al.* (1983). Briefly, the liver (2–10 g, depending on body size) was homogenized in 250 mm sucrose, 100 mm Tris-HCl, and 10 mm EDTA (pH 7.6), and the mitochondrial fraction was pelleted by the differential centrifugation, followed by lysis with SDS, phenol extraction, and ethanol precipitation. Closed circular mtDNA was recovered by ethidium bromide-CsCl equilibrium density gradient centrifugation. After removal of ethidium bromide, the separate fractions were desalted and concentrated with a Centricon-30 centrifugal microconcentrator (Amicon, Beverly, MA).

Akamata mtDNA was cut with restriction enzymes and cloned into *Escherichia coli* pUC19 or pBR322 vector. Figure 1A shows cloned mtDNA fragments and their arrangements. Clones for three kinds of *Xba*I-digested fragments (3.9, 1.3, and 1.3 kbp) were obtained. Two kinds of *Xba*I/*Eco*RI-digested fragments (4.4 and 1.7 kbp) were also cloned. As described in detail in results and discussion, the remaining portion corresponding to a 4.6-kbp *Eco*RI-cut fragment was refractory to the cloning. Instead, a 5.9-kbp fragment resulting from partial digestion of akamata mtDNA with *Eco*RI and *Bgl*II was cloned successfully. Both strands of these cloned mtDNA fragments were sequenced with a DNA sequencer (model 373A; Applied Biosystems, Foster City, CA) using the primer walking strategy. All overlaps between the clones were confirmed by direct sequencing of PCR products amplified from the purified mtDNA template.

Southern hybridization was done with a chemiluminescence detection system (ECL direct nucleic acid labeling and detection systems; Amersham, Arlington Heights, IL) according to the manufacturer's instructions. Akamata mtDNA (50–150 ng) was cut with appropriate restriction enzymes, electrophoresed on a 0.4% agarose gel, transferred to a nylon membrane (Pall Biodyne A transfer membranes, 0.2-μm pore), and hybridized with a peroxidase-attached DNA probe. After hybridization at 42° overnight in the manufacturer's hybridization buffer (ECL gold buffer), to which 0.5 m NaCl and 5% blocking agent were added, the filter was washed twice at 42° for 20 min with the primary wash buffer containing 6 m urea, 0.4% SDS, and 0.3× SSC, and then washed twice at room temperature for 5 min with 2× SSC. In this study, two kinds of probes were prepared from heat-denatured PCR products: probe A, which corresponds to a sequence containing the $tRNA^{Val}$ gene and parts of flanking rRNA genes (positions 880–1078 in Figure 2), and probe B, which corresponds to a sequence inside the control region (positions 4057–4225 and 16,628–16,796). The chemiluminescent signal was detected with Kodak XAR-5 film. After hybridization with probe A, the blot was stored at room temperature for a few days to allow the chemiluminescent signal to decay completely. The blot was then washed with 2× SSC and reprobed with probe B.

Genes in the determined mtDNA sequences were identified in light of their sequence similarity with the corresponding genes from other vertebrates, as well as structural features of mitochondrial tRNA genes (*e.g.*, Kumazawa and Nishida 1993). The entire nucleotide sequence of the akamata mtDNA will appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases under the accession no. AB008539.

RESULTS AND DISCUSSION

Gene organization: Figure 1A shows the gene organization of akamata mtDNA derived from its complete nucleotide sequence (Figure 2). Akamata mtDNA is

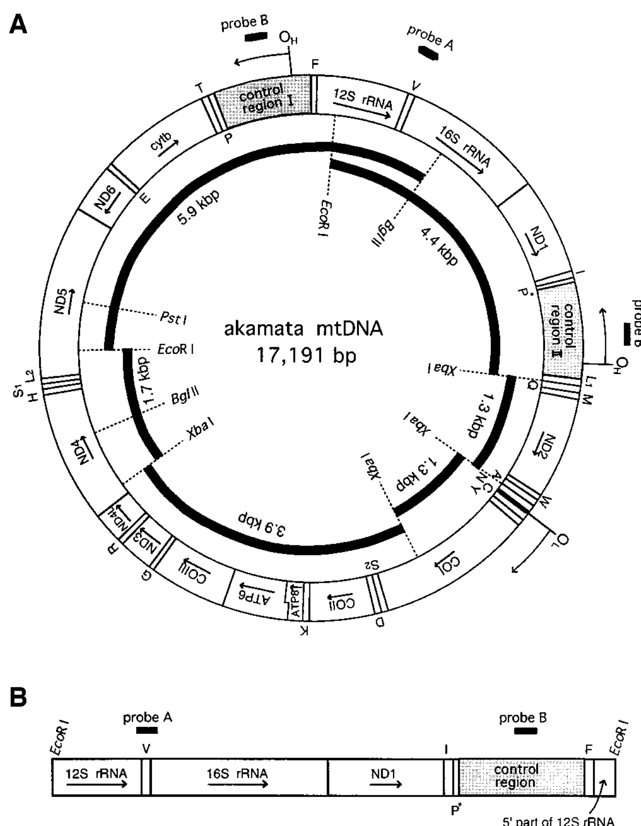

Figure 1.—(A) Gene organization and restriction map of the akamata mtDNA. A 17,191-bp circular mtDNA is depicted as a ring that is divided into individual genes and control regions. Transfer RNA genes are shown as single letters standing for the corresponding amino acid, and those whose sense strand is the light and heavy strand are shown outside and inside the ring, respectively. L1, L2, S1, S2, and P* represent genes for tRNA^{Leu}(UUR), tRNA^{Leu}(CUN), tRNA^{Ser}(AGY), tRNA^{Ser}(UCN), and a pseudogene for tRNA^{Pro}, respectively. Abbreviations for the other genes are as follows: ND1-6, NADH dehydrogenase subunits 1–6; CO I–III, cytochrome oxidase subunits I–III; ATP6, ATPase subunit 6; ATP8, ATPase subunit 8; and cytb, cytochrome *b*. O_H and O_L stand for an origin of heavy- and light-strand replication, respectively. Sites for restriction enzymes *Xba*I, *Eco*RI, *Bgl*II, and *Pst*I are shown, and restriction fragments that were cloned and sequenced are indicated by bold lines. The regions corresponding to the hybridization probes A and B are also shown outside the ring. (B) Gene organization of the 4.6-kbp fragment with *Eco*RI termini, which was unexpectedly cloned from the Okinawa 2 individual and may have been derived from a minor mtDNA component (see text for details).

17,191 bp long and contains 13 protein-coding genes, two rRNA genes, and 22 tRNA genes that are commonly encoded in the mtDNAs from other vertebrates. It also contains two control regions that, in other vertebrates, occur in a single conservative location surrounded by cytb and 12S rRNA genes. The snake genome conserves the control region in this location (control region I) and, in addition, possesses another control region (control region II) within a tRNA gene cluster where genes for tRNA^{Ile}, tRNA^{Gln}, and tRNA^{Met} are typically arrayed in other vertebrates (IQM region). The nucleotide sequences of the two control regions of akamata mtDNA are completely identical over 1 kbp (Figure 2). Because our previous study (Kumazawa *et al.* 1996) used PCR to show the existence of two control regions, we were anxious that we may have analyzed something like PCR artefacts, especially resulting from the jumping PCR phenomenon (Pääbo *et al.* 1990). However, the conventional cloning-and-sequencing strategy was used in the present study, which unambiguously proves that the snake mtDNA has two control regions within a single molecule. This conclusion was additionally confirmed by Southern hybridization experiments (see below).

Besides this anomaly in the control region, akamata mtDNA has two distinctive features in gene organization: the tRNA^{Leu}(UUR) gene, which occurs between genes for 16S rRNA and ND1 in other vertebrates, is translocated to the 3′ vicinity of control region II, and a pseudogene for tRNA^{Pro} exists in the 5′ vicinity of control region II (Figure 1A). Thus, the alteration in gene organization is centered on the IQM region, where insertion of an array of tRNA^{Pro} pseudogene, control region, and tRNA^{Leu}(UUR) gene occurs between genes for tRNA^{Ile} and tRNA^{Gln}. No other distinctive features from the typical vertebrate organization for placental mammals, the African clawed toad, and bony fish are found. A stem-and-loop structure for the putative origin for replicating the light strand (Wong and Clayton 1985) appears to be present between genes for tRNA^{Asn} and tRNA^{Cys} (Figure 2).

The size of akamata mtDNA (17,191 bp) is nearly equal to or slightly larger than those of mtDNAs from other vertebrates reported to date, and the sizes of individual structural genes are mostly common between akamata and other vertebrates (data not shown). The base composition of akamata mtDNA (Table 1) is skewed similarly to those of other metazoan mtDNAs, *i.e.*, generally more A-T base pairs than G-C base pairs, and more A and C contents (accordingly, fewer G and T contents) in the light strand than in the heavy strand (*e.g.*, see Asakawa *et al.* 1991).

Control region: Each of the two control regions found in akamata mtDNA is 1018 bp long. This is the typical length for the control region of vertebrate mtDNAs known to date. Conserved sequence blocks (CSBs) I, II, and III have been known as conserved sequence elements among mammalian control regions (Walberg and Clayton 1981). All or partial members of the CSBs are also identifiable for control regions of vertebrates other than mammals (*e.g.*, Roe *et al.* 1985; Quinn and Wilson 1993; Lee *et al.* 1995; Zardoya and Meyer 1997). The akamata control region contains sequences that show an appreciable level of similarity with CSBs I, II, and III of other vertebrates (Figure 2). However, unlike other vertebrates, CSBs II and III appear to be tightly connected with each other. In light of experimental evidence obtained for mammals (Clayton 1992

TABLE 1

MtDNA base composition of akamata and other vertebrates

Species	Base composition (%)					
	A	C	G	T	A+T	A+C
Akamata	34.7	27.7	12.2	25.3	60.0	62.4
Mouse	34.5	24.4	12.3	28.7	63.2	58.9
Chicken	30.3	32.5	13.5	23.8	54.1	62.8
Frog	33.0	23.5	13.5	29.9	62.9	56.5
Sea lamprey	32.3	23.8	13.5	30.4	62.7	56.1

Base composition percentages of light-strand mtDNA sequences are shown.

and references therein), the displacement loop (D-loop) region of akamata mtDNA may start from 3′ downstream of CSB III and end in the vicinity of termination-associated sequences (TASs; Doda *et al.* 1981), two copies of which are identifiable at the 5′ end of the snake control region (Figure 2). As is often the case with other vertebrates, these TASs are situated in a region where hairpin-like secondary structures and repetitive sequences occur (Figure 2).

5′ upstream of the hairpins and repeats, there exists a C-rich sequence that commonly occurs among several snakes so far examined (Figure 2; Kumazawa *et al.* 1996). Sequencing reactions for both strands tend to stall in the C-rich region (data not shown), suggesting that this is a structural barrier for DNA polymerization. The C-rich sequence may facilitate formation of the D-loop by decelerating the extension of heavy-strand synthesis at this location. The C-rich sequence is not found clearly in the corresponding locations of mammals, frog, and fish. Birds have similar sequences, but they have been interpreted to be involved in a hairpin structure (Quinn and Wilson 1993). As described below, snake mtDNAs possess, just 5′ upstream of the control region, rather stable hairpin-like structures derived from the tRNA^{Pro} gene or pseudogene or an intergenic spacer. This hairpin may also serve as a structural barrier in DNA polymerization and play a role in facilitating the formation of the D-loop concertedly with TASs, the C-rich sequence, and other hairpins at the 5′ end of the control region.

Replication and transcription: In mammals, the control region contains all the template information required for the initiation of nascent heavy-strand synthesis, as well as the essential major *cis*-acting elements necessary for transcriptional initiation (Clayton 1992; and references therein). Because the nucleotide sequences of control regions I and II are completely identical (Figure 2), it is reasonable to infer that heavy-strand synthesis and transcription for both strands can be initiated from two separate locations in the snake mtDNA, whereas light-strand synthesis preceded by heavy-strand synthesis begins from a single location (Figure 1A). This apparently serves to elevate the efficiency of DNA replication and gene expression in the snake mitochondria.

Mammalian mtDNAs maintain a transcriptional attenuator-like, 13-nucleotide sequence at the 5′ end of the tRNA^{Leu}(UUR) gene (Christianson and Clayton 1988). Light-strand RNA transcripts terminate at the end of the 16S rRNA gene much more frequently than they extend beyond this site to make a polycistronic RNA precursor for all the light-strand gene products (Gelfand and Attardi 1981; Ojala *et al.* 1981). This attenuator is thought to regulate expression levels of 12S and 16S rRNAs, which should be much higher than those of tRNAs and mRNAs. Because the snake tRNA^{Leu} (UUR) gene (Figure 3D) is translocated to the 3′ end of control region II, the attenuation mechanism at the 3′ end of the 16S rRNA gene may not be operative in snake mitochondria. It is currently unknown how the high expression levels of rRNAs are maintained in the snake mitochondria. One possibility is that light-strand transcripts may not be able to extend beyond control regions because of the presence of a transcriptional terminator within this region, and that light-strand RNA synthesis from control region I may be regulated to be more efficient than that from control region II.

Transfer RNA genes: Akamata mtDNA encodes 22 tRNA genes and a pseudogene of tRNA^{Pro}. The anticodon triplet sequences of the 22 tRNA genes (Figure 2) are exactly the same as their counterparts for mammals, chicken, and frog. It is therefore likely, though not conclusive, that the snake mitochondria use the same genetic code as mammalian mitochondria (Anderson *et al.* 1981), and that this set of tRNAs is capable of

Figure 2.—Complete nucleotide sequence of the akamata mitochondrial genome. The light-strand sequence of the 17,191-bp akamata mtDNA is presented and numbered from the 5′ end of tRNA^{Phe} gene. The starting and ending positions predicted for each gene are shown with (⌈) or (⌉), together with an arrow that indicates the polarity of the gene. Protein-coding genes are translated into amino acid sequences by the mammalian genetic code. Note that stop codons for ND1, ND3, CO II, CO III, and cytb genes appear by polyadenylation. Anticodons of tRNA genes and recognition sites by major restriction enzymes *Xba*I, *Eco*RI, *Bgl*II, and *Pst*I are underlined. Also underlined are several structural features that exist within the control regions and the origin for light-strand replication. Note that the sequences of two control regions corresponding to 3602–4619 and 16,173–17,190 are completely identical. A guanosine at position 3750, a thymidine at 16,223, and a guanosine at 16,321 were sequenced to be an adenosine, a cytidine, and an adenosine, respectively, for a different individual collected from the same locality of Okinawajima Island (Kumazawa *et al.* 1996). Note that CSB II was not identified clearly in our previous study (Kumazawa *et al.* 1996).

```
┌tRNA(Phe)→                                                          →tRNA(Phe)┐      ┌12S rRNA→
GTTATCATAG CTTACGAATT AAAGCATAGC ACTGAAAATG CTAAGATGGT CTACACCTGA TTTAACACAA GGTCTTGGTC TTAAACCCAA TATTACTTAT   100

ACTCCTTAAT TATACATGCA AGCCTCACCA TAACAGTGAA ACAGCCTACC AATACAGCCT AGAGCAGGTA TAAGGCAATT AAGCCCATAA CACCAAGCCA   200
                                                                                    EcoR I
TATTAAGCCA CACCCCCACG GGCCCGCAGC AGTAGTTAAT ATTAGGCCAT AAGCGAAAGC TTGACCTAGC AAAGGAATTC ACCTTAGGGC CGGTTAATCT   300

CCGTGCCAGC GACCGCGGTT ACACGACAGA CCCAAAATAA TATTAACGGC GTAAAGCACG ACTAAATTAG CCAGTCTACA CATTAGGAAT AAAATCAAGC   400

TGGGCTGTAA AAAGCCATAA GCCACAATAA ACCCCACCCT AACAACAAAC AACTTCCACT CGTGAAAGCT AGGACCCAAA CTAAGATTAG ATACCTTACT   500

ATGCCTAACC GTAACACAAC AATAAAATTA CATGTTGTTC GCCAAATAAC TACGAGTAAA AACTTAAAAT TTAAAAGACT TGACGGTACC CCACAACAAC   600

CTAGAGGAGC CTGTTCAATA ACCGATACTC CACGATTAAT CCAACCTTCC CTAGCCTACC AGTCTATATA CCGCCGTCGC CAGCCTACCT TATGAAAGAA   700

ACAAAGTAAG CCAAACAGTC ACACACTAAC ACGACAGGTC GAGGTGTAAC TCATGGGAAG GACTAGAATG GGCTACATTA TCTAACCCAG ATAACACGGA   800

TAAACTATGA AACAAGAAAC TAAAGGCGGA TTTAGCAGTA CACTATGAAC AATAATACCT AGTCGAAACC AACGCAATGG GGTGCGTACA CACCGCCCGT   900
                                                                                                 →12S rRNA┐ ┌
CATCCCTGTC AACCAACCAA AACACTATAT AAACACAAAT AAAACCTAAA ACAGGGCAAG TCGTAACACG GTAAGCGTAC TGGAAAGTGC GCTTAGAAAC  1000
tRNA(Val)→                                                          →tRNA(Val)┐ ┌16S rRNA→
AAAAAGTAGC TTACAACAAA GCACTCGACC TACACTCGAA CGACATTACC ATAATCTTTT TGAGCCGAAT CAAACAATAC AAACATATTA ACCCCTACCA  1100

AACAAAACAT TTGATAAACC AAGTAGATGT GATCGAACAG TAACAACAAC CACTTAGTAC CGCAAGGGAA TAATCACAAG CAATAAACAG CAAAGATTAA  1200

CCCTTGTACC TTTTGCATCA TGGTCTAGCA AGAAATCAAA GACAAGAAGA ATCATAGCCT CCCCCCCGAA ACCAGATGAG CTACTTTAAA GCAGCCTAAT  1300

GGGCACACCC TTCTCTGTAG CAAAAGAGTG GGAAGACTCA AAAGTAGCGG TGAAACGCCT ACCGAATCTG GAGATAGCTG GCTACCCCAA AAAGAATCTA  1400

AGTTCAACCT TAAGACAAAA CAATACCCCA TTATTGCCTA AGGATATTCA ATAAAGGTAC AGCTTTATTG AAACAGGATA CAACCTGAAT TTGAGAGGAA  1500

AATAAAACAC ATTACCAGTA GGCCTTAAAG CAGCCACCTA ACAAAATATC GTTACAGAAT TGATCTAAAA ATCCAAACAT CAATTAAAAA CTCCAAACCA  1600
         Bgl II
ACTAAAGGTA GATCTATAAC CATAGGTCCT ATTATGCTAG AACTAATAAT AAGACAACCT CTCTTCAATG CACTCGTCCC CTAGACCGGA TACACCACTA  1700

GCCATTAACA GACCAAAACA GGCATAAACG AACCAAAACA CACCTATAAA TCAAACTGTG ACCCCAACAC AGGAGCATTA AAAAGAAAGA TTTAACATTA  1800

TAAAAGGAAC TCGGCAAACA CAGACCCCAA CTGTTTAACA AAAACATAAC CTTTAGCCAA ACAAATATTA AAGGCAACGC CTGCCCAGTG AACAATTAAA  1900

CGGCCGCGGT ATCCTAACCG TGCAAAGGTA GCATAATCAT TTGTCTACTA ATTATAGACC TGTATGAAAG GCAAAATGAG GGTCCTACTG TCTCTTATAA  2000

TCAATCAATT AAACTGATCT ACTAGTAAAA AAGCTAGAAT ACCCACATAA GACCAGAAGA CCCTGTGAAG CTTAAACTAA ACTATTAAAC CATATAATAG  2100

CTACTTTCGG TTGGGGCGAC CTTGGAAAAA AAAAGAACTT CCAACATCAT GACCTTCCTC ATACCATAAC GGCCAACAAG CCAATAAATG ACCCAGCAAA  2200

GCTGACAAAC GAACCAAGTT ACTCCAGGGA TAACAGCGCA ATCTTCTTCA AGAGCCCATA TCAAAAAGAA GGTTTACGAC CTCGATGTTG GATCAGGACA  2300

TCCTAATGGT GCAGCCGCTA TTAAGGGTTC GTTTGTTCAA CGATTAATAG TCCTACGTGA TCTGAGTTCA GACCGGAGTA ATCCAGGTCG GTTTCTATCT  2400

ATGAACCAGC CTTATCCAGT ACGAAAGGAA CGATAAAGCA AAGCCAATAC CACAAGCACG CTTTAACAAA ATACCAAACA CACACCACCA CCACCCAAAT  2500
                                →16S rRNA┐ ┌ND1→
TTAAACCAAG ACAAGGTTAA TTAAGGACCA CCTTAATAAT TTCTATAATC CTACTAAACA TTACCAACTC CCTACTCTAC ATCCTATCAA TTCTTATCGC  2600
                                       M  I   S  M  I   L  L  N  I   T  N  S   L  L  Y   I  L  S  I   L  I  A
CGTAGCATTC CTCACCTTAC TAGAACGAAA ACTCCTTGGA TACATACAAC ACCGCAAAGG GCCAAATCTA GTAGGCCCAA TAGGCCTACT ACAACCAATT  2700
 V  A  F   L  T  L  L   E  R  K   L  L  G   Y  M  Q  H   R  K  G   P  N  L   V  G  P  M   G  L  L   Q  P  I
GCCGACGGAC TAAAACTAAT TACAAAAGAA GCAACAAAAC CAACTATATC CTCACCAATT CTATTCACAC TATCCCCCAT CATAGCACTT ATTCTAGCAC  2800
 A  D  G  L   K  L  I   T  K  E   A  T  K  P   T  M  S   S  P  I   L  F  T  L   S  P  I   M  A  L   I  L  A  L
TAACATCCTG AGCACCAATA CCCATACCAA CCCCCCTAAC CAACATAAAT CTCGGATTAC TCTTCATTAT AGCAATATCA GGTATATTCA CATACGCTAT  2900
  T  S  W   A  P  M   P  M  P  T   P  L  T   N  M  N   L  G  L  L   F  I  M   A  M  S   G  M  F  T   Y  A  I
TCTTTGATCC GGTTGATCCT CTAACTCAAA GTACCCCCTT ATAGGAGCCA TGCGAGCCGT AGCCCAAATT ATTTCATACG AAGTGACCCT AGGGCTAATC  3000
 L  W  S   G  W  S  S   N  S  K   Y  P  L   M  G  A  M   R  A  V   A  Q  I   I  S  Y  E   V  T  L   G  L  I
ATCATCTCCA TAGCCATCTT ATCAGGAGGA TACTCACTAA TACTCTTTAC CGAGACACAA GAACACATAT GACTACTACT ATCATCATGA CCCCTAGCCA  3100
 I  I  S  M   A  I  L   S  G  G   Y  S  L  M   L  F  T   E  T  Q   E  H  M  W   L  L  L   S  S  W   P  L  A  M
TAATATGATT CACATCCACT CTAGCCGAAA CAAACCGATC CCCCTTCGAC CTAACAGAGG GGGAATCAGA ACTAGTCTCA GGATTCAACG TAGAGTTCTC  3200
  M  W  F   T  S  T   L  A  E  T   N  R  S   P  F  D   L  T  E  G   E  S  E   L  V  S   G  F  N  V   E  F  S
AGCCGGACCA TTCGCATTAC TATTCTTGGC AGAATACACC AACATTCTAT TCATAAACAC CCTATCAACT ATAATATTCC TAAACCCAGG AATAACAAAC  3300
 A  G  P   F  A  L  L   F  L  A   E  Y  T   N  I  L  F   M  N  T   L  S  T   M  M  F  L   N  P  G   M  T  N
CCACAACTAT TTACCATTAA CCTTATAACC AAGACAATAA TTCTAACAAC CCTATTCCTA TGAACTCGAG CCTCATACCC TCGTTTCCGC TACGACCAAT  3400
 P  Q  L  F   T  I  N   L  M  T   K  T  M  I   L  T  T   L  F  L   W  T  R  A   S  Y  P   R  F  R   Y  D  Q  L
                                                                                                        →ND1┐ ┌
TAATACACTT ACTATGAAAA CAATATCTTC CACTCACCCT AGCTATATAC CTTCTAAACA CCTCCACTAG TATAGCTTTA TGCGGTACAC CCCCACAATG  3500
  M  H  L   L  W  K   Q  Y  L  P   L  T  L   A  M  Y   L  L  N  T   S  T  S   M  A  L   C  G  T  P   P  Q  *
tRNA(Ile)→                                                          →tRNA(Ile)┐ ┌tRNA(Pro) pseudogene←
GGGGTGTGCC CGAGTAGGGA CTACCTTGAT AGAGTAGACA CGGAACCATA AAACTCCCAC CCCTCCCCAC AAGCCAGTAT TTTAACTTAA ACTACTCTTT  3600
```

```
┐┌control region I             C-rich sequence                                    hairpin 1
GCCAAAAATA AACAATCCTC CTAGGACCCC CCCCCTACCC CCCCCCTTAC ATCTGACCCA ATTCCGACTA TAATGTATCT CTTAGGTTAT AGTCTTTATT  3700
      hairpin 2     ┌───TAS1───┐                      hairpin 3  ┌───TAS2───┐
TCACTATGTA TAATTATACA TTAATGATCT GCCTCACGCC TAATAAACGG GAATTATACT ATAATTATTT GTATAAAAAA GTGGTATATT ACATGCGATT  3800
            └ 20-bp repeat 1    ┘                       └ 20-bp repeat 2    ┘
TACCCCCTCA TTTCCCAAAC GTTCCATGTT ATCTTCGGCC ACCCATTATT AGTAACCATG ACTATCCCGT GCCTAATGGT GTCCCATGAT GTAGCTCAGC  3900
                                                                     ┌                  49-bp repeat 1
CCGTGAAACC CTCTATCCTT CCATTGAAAG CATACAGTCC CGCTTCTCAC GTCCATATAT TGTAACTCCT CCCTTTATGT TCTTTCCAAG GCCGCTGGTT  4000
             ┐                      ┌                  49-bp repeat 2                              ┐     ┌
ACACTCTCAA GAGCATCTCA ATGGTCCGGA ACCACCCCGC CTTACTTGCT TTTTCCAAGG CCTTTGGTCG CACCCTTTAT AGTGGTACAT ATCACCTCAT  4100
     49-bp repeat 3                 ┐
GTTCTTATCA CCTATGCCAG TTCCACCCCT GGTTGTTCTT TTTATCTCTA CCTTTCACCT GACACTCACT TGCCCGTTAC CGTTCCCCTC ACCGGGGCGT  4200
                                                                          CSB I
AGACCATCTA GTCCGGGTGG AGCTATATTC TTGGCTATGC ACATTCCCTA TACGGATACA TTCCTTCATG CTTGTTAGAC ATACTTTTTC CTTGACTGTA  4300

AAATTTCATT ATTCTTTTAT TATAAATTTG CCGCAGTAAA CACTTTTTCA TCACCCCTAT TTTTAAAATC TGAGAAAAAA TAATACGACA AAACTATAAT  4400
     CSB II         CSB III                                                                hairpin 4
AAACCCCCCG CCCCGAAACC CCATATACTT ATTTTCATAC CAGCCTCTCA CCCCGATCAC ACCCCGTGAA AATAGTTTAC TATAAAAATA GCAAACACAT  4500

ATTTTAGAAC CGCATATTTA ACTTCTTTCT CTGAAAAATC ATTTCAAATG CCAAAACACC CCCCACTAAA AAATAGTCAT TTCACCAATT ATCTCAACAA  4600
   control region I┐ ┌tRNA(Leu,UUR)→                                                    →tRNA(Leu,UUR)┐  ┌Xba I
TCTCCCGAAT TTTTATATAA TTAAGGTAGC AAAGCCAGGT TATGCAAAAG GCTTAAAACC TCAACACAGA TGTTCAAATC ATCTCCTTAA TATCTAGAAG  4700
tRNA(Gln)←                                               ←tRNA(Gln)┐ ┌tRNA(Met)→
GCCAAGACTC GAACTTGAAC TAAAAAGCCC AAAACTTTGC ATACTACCAA TATATTACCT TCTACAGTAA GGTCAGCTAA ATAAGCTATC GGGCCCATAC  4800
                  →tRNA(Met)┐ ┌ND2→
CCCGAAAATG CCACAACGGC CCCTACTAAT TAACCCAATA TCCTGATTAA TTATCACAAC AAGCATTGCC CTAAGCACCA CTATAATCAC ATCAACAACA  4900
                              M   N   P   M   S   W   L   I   I   T   T   S   I   A   L   S   T   T   M   I   T   S   T   T
CACTGACTTA TAACATGAGC CTGCCTAGAA ATTAACACCC TATCCATAGT CCCACTAATC TCAAAACCAA ACCACCCCCG AGCAACAGAA GCAGCAACAA  5000
H   W   L   M   T   W   A   C   L   E   I   N   T   L   S   M   V   P   L   I   S   K   P   N   H   P   R   A   T   E   A   A   T   K
AATATTACCT AATCCAAACC ATAGCCTCCA CCTCTATACT ATTTGCAGCC ACAACAAATG CCTTAAACAC CTCAAACTGA GAAACCCATC TCACAACAGA  5100
 Y   Y   L   I   Q   T   M   A   S   T   S   M   L   F   A   A   T   T   N   A   L   N   T   S   N   W   E   T   H   L   T   T   E
ACCAATAGCA ACAACTATTA TTACACTAGC CCTCATAATA AAAATGGCAG CTGCCCCATT CCACTCCTGA CTCCCAAGCG TATCACAAGG CACAACAACC  5200
 P   M   A   T   T   I   I   T   L   A   L   M   M   K   M   A   A   A   P   F   H   S   W   L   P   S   V   S   Q   G   T   T   T
CTAACAACCC TAACCATCCT AACCTGACAA AAAATTGCCC CACTAACAAT TCTATTAACC ACCCATAATA AAACAAATAT TACCCTAATC CTACTATCCG  5300
L   T   T   L   T   I   L   T   W   Q   K   I   A   P   L   T   I   L   L   T   T   H   N   K   T   N   I   T   L   I   L   L   S   A
CAATACTATC CATCACAATA GGAGGCCTTG GAAGCCTTAA CCAAACCCAA CTACGAAAAC TAATAGCCTT CTCATCAATC GCTCACACGG GCTGAATCAT  5400
 M   L   S   I   T   M   G   G   L   G   S   L   N   Q   T   Q   L   R   K   L   M   A   F   S   S   I   A   H   T   G   W   I   M
AGCAACTATC ACAATAGCAC CAAAAATCTC AACCTTAACC TTCACAATTT ACATCATAAC CACCATCCCC ACATTTCTAC TAATCAACAC TACAATATCA  5500
 A   T   I   T   M   A   P   K   I   S   T   L   T   F   T   I   Y   I   M   T   T   I   P   T   F   L   L   I   N   T   T   M   S
ATAACAATCA AAGACCTAGG AACCATGTGA ACCAACTCCC CATACACCAT AACTATCCTA TCTATAACCA TCCTATCTAT AGGAGGACTA CCGCCCCTAT  5600
M   T   I   K   D   L   G   T   M   W   T   N   S   P   Y   T   M   T   I   L   S   M   T   I   L   S   M   G   G   L   P   P   L   S
CAGGATTCAT ACCAAAATGG TTAATCTTAA ATAATCTAAT TTCAATAAAC ATAATTACAG AAGCAACCCT AATAGCCATA GCCTCACTAC TCAGCCTATA  5700
 G   F   M   P   K   W   L   I   L   N   N   L   I   S   M   N   M   I   T   E   A   T   L   M   A   M   A   S   L   L   S   L   Y
CGTATATATG CGACTAACAT ATATATCATC AATAACACTA TCACCCCACA CTACAACCAT ACCATTAAAA TGACGGACAT CAAATAAAAA ACATCCTATA  5800
 V   Y   M   R   L   T   Y   M   S   S   M   T   L   S   P   H   T   T   T   M   P   L   K   W   R   T   S   N   K   K   H   P   M
                                                        →ND2┐ ┌tRNA(Trp)→
GGAACCTCAA TACTAACAAT AATAACAATA CTTCTTCTAC CCCTATCACC AAATATATAG AAGCTTAAGT TATATTAAAC TAGGAACCTT CAAAGTCCCA  5900
G   T   S   M   L   T   M   M   T   M   L   L   L   P   L   S   P   N   M   *
             →tRNA(Trp)┐   ┌tRNA(Ala)←                                                         ←tRNA(Ala)┐  ┌Xba I
AAAAAAGACC ACTTTAGTTT CTGTTTAGAG CTTGCAGAAA CACCACATCT ACTGCTTGCA ACACAGACAT TTTAATTAAA CTAAAGCTCT CTAGATTAGC  6000
tRNA(Asn)←                                             ←tRNA(Asn)┐  origin of L-strand replication     ┌
GGGCCTCGAT CCCACAAAAA CTAATTAACA GCTAGCTGTC AAAACCGGCG GACTTTAATC TAGCTTCTCC GTTTTTGGAC GAAAGAAAAA ACGGAGAAAC  6100
tRNA(Cys)←                                      ←tRNA(Cys)┐ ┌tRNA(Tyr)←
CCGGGGCCGC TGCCGACTTC AGATTTGCAG TCTGACATGT TACACACCTC CGGGTTTGGT AGCAAGGAAT ATCCTATTAA TAATTTTACA GATTACCGCC  6200
       ←tRNA(Tyr)┐  ┌CO I→
TAAATCAGCC ATACTACCCG TGTTCATCAC TCGTTGACTA TTCTCGACAA ATCACAAAGA CATCGGAACC CTATACCTAT TATTTGGTGC CTGATCGGGT  6300
                              M   F   I   T   R   W   L   F   S   T   N   H   K   D   I   G   T   L   Y   L   L   F   G   A   W   S   G
CTAATCGGAG CTTGTCTTAG TATCTTAATA CGAATAGAGC TAACCCAACC CGGATCACTG CTAGGCAGTG ATCAAATCTT CAATGTTTTA GTTACAGCCC  6400
L   I   G   A   C   L   S   I   L   M   R   M   E   L   T   Q   P   G   S   L   L   G   S   D   Q   I   F   N   V   L   V   T   A   H
ATGCGTTTAT CATAATTTTC TTCATAGTTA TACCTATTAT AATCGGCGGC TTTGGCAACT GACTAATTCC ACTAATAATT GGAGCACCTG ACATAGCCTT  6500
 A   F   I   M   I   F   F   M   V   M   P   I   M   I   G   G   F   G   N   W   L   I   P   L   M   I   G   A   P   D   M   A   F
CCCCCGCATA AATAATATAA GCTTCTGACT CCTACCACCC GCTTTACTTT TATTACTGTC ATCCTCTTAT GTAGAAGCAG GGGCAGGTAC AGGGTGAACT  6600
 P   R   M   N   N   M   S   F   W   L   L   P   P   A   L   L   L   L   L   S   S   S   Y   V   E   A   G   A   G   T   G   W   T
GTATACCCAC CCCTGTCAGG AAACCTAGTA CACTCAGGAC CATCAGTAGA CTTAGCGATC TTCTCCCTCC ACCTAGCGGG CGCCTCATCC ATCCTGGGGG  6700
V   Y   P   P   L   S   G   N   L   V   H   S   G   P   S   V   D   L   A   I   F   S   L   H   L   A   G   A   S   S   I   L   G   A
CAATTAATTT TATTACAACA TGCATCAACA TGAAACCAAA ATCTATACCT ATATTTAATA TACCATTATT TGTCTGATCA GTACTAATTA CTGCTATTAT  6800
 I   N   F   I   T   T   C   I   N   M   K   P   K   S   M   P   M   F   N   M   P   L   F   V   W   S   V   L   I   T   A   I   M
ACTTCTTCTA GCCTTGCCCG TACTAGCCGC AGCAATCACC ATACTACTGA CCGACCGAAA TCTTAATACC TCCTTCTTTG ATCCTTGTGG CGGAGGCGAC  6900
 L   L   L   A   L   P   V   L   A   A   A   I   T   M   L   L   T   D   R   N   L   N   T   S   F   F   D   P   C   G   G   G   D
CCCGTATTAT TCCAACACCT GTTCTGATTC TTTGGCCACC CGGAAGTTTA CATTCTCATT CTACCCGGAT TCGGCATTAT TTCAAGCATC ATTACATTCT  7000
P   V   L   F   Q   H   L   F   W   F   F   G   H   P   E   V   Y   I   L   I   L   P   G   F   G   I   I   S   S   I   I   T   F   Y
ATACAGGAAA AAAAAACACA TTTGGATACA CAAGTATAAT TTGAGCAATA ATATCTATCG CCATTCTAGG CTTTGTTGTA TGAGCCCACC ACATATTTAC  7100
 T   G   K   K   N   T   F   G   Y   T   S   M   I   W   A   M   M   S   I   A   I   L   G   F   V   V   W   A   H   H   M   F   T
TGTGGGACTA GACATTGACA GTCGAGCCTA CTTTACAGCA GCAACAATAA TTATTGCTAT CCCAACAGGA ATTAAAGTCT TTGGCTGACT AGCTACCCTA  7200
 V   G   L   D   I   D   S   R   A   Y   F   T   A   A   T   M   I   I   A   I   P   T   G   I   K   V   F   G   W   L   A   T   L
```

Figure 2.—*Continued.*

```
GCAGGTGGCC AAATCAAATG ACAAACCCCC ATCTACTGAG CCCTTGGCTT CATTTTCCTT TTTACTGTCG GCGGCATGAC CGGAATTATC CTAGCAAACT  7300
 A  G  G  Q   I  K  W   Q  T  P   I  Y  W  A   L  G  F   I  F  L   F  T  V  G   G  M  T   G  I  I   L  A  N  S
    Xba I
CCTCTCTAGA TATTGTATTA CACGATACCT ACTACGTAGT AGCACACTTT CACTATGTAC TATCCATAGG GGCGGTATTT GCTATCATAG GTGGACTCAC  7400
  S  L  D   I  V  L   H  D  T  Y   Y  V  V   A  H  F   H  Y  V  L   S  M  G   A  V  F   A  I  M  G   G  L  T
CCATTGATTT CCCCTATTCA CAGGCTACAC CCTAAATCAA ACTATAACAA AAACTCAATT CTGAGTAATA TTTGTAGGGG TTAATATAAC ATTCTTTCCA  7500
 H  W  F   P  L  F  T   G  Y  T   L  N  Q   T  M  T  K   T  Q  F   W  V  M   F  V  G  V   N  M  T   F  F  P
CAACACTTTC TAGGACTATC TGGCATACCA CGACGATACT CAGATTTTCC AGACGCCTTC ACCCTATGGA ACACAATATC ATCTATCGGA TCAACCATTT  7600
 Q  H  F  L   G  L  S   G  M  P   R  R  Y  S   D  F  P   D  A  F   T  L  W  N   T  M  S   S  I  G   S  T  I  S
CTATAGTAGC AGTACTAATA TCCCTATTTA TTGTATGAGA AGCACTAACA TGCAAACGAG AAGTCCAAAT ACCTCTTGGA AAAAAAACAC ACGTAGAGTG  7700
  M  V  A   V  L  M   S  L  F  I   V  W  E   A  L  T   C  K  R  E   V  Q  M   P  L  G   K  K  T  H   V  E  W
ATTCTTTGGT TCTCCCCCAC CATACCACAC ACACACAGAA CCATCATTCA TACTAAACAA CACCTATGCC CCAATTCGAA ACCTTATCTC ATACATAGAG  7800
 F  F  G   S  P  P  P   Y  H  T   H  T  E   P  S  F  M   L  N  N   T  Y  A   P  I  R  N   L  I  S   Y  M  E
  →CO I       ┌          ┐    ←tRNA(Ser,UCN)                     ←tRNA(Ser,UCN)┐ ┌tRNA(Asp)→
TGACCCTGGC CCGAGAAAAG ACGGGACTAA CCGCCATCTG TTAATTTCAA GTTAACCGCA TATTATGCTT TCTTCCCGAG AACCTAGTAA ACACATTACA  7900
 W  P  W  P   E  K  *
                             →tRNA(Asp)┐  ┌CO II→
TGGCTTTGTC GTGGCCAAAT CACAGTCTCT GTGGTCCTCA ATGCCACACG CAAGCCAACT ATCCCTACAA GAAGCCATAG GACCTACAAT AGAAGAAGTA  8000
                                            M  P  H  A   S  Q  L   S  L  Q   E  A  M  G   P  T  M   E  E  V
ATTTTCCTAC ATGACCACGT CCTTTTACTT ACCTGTTTAA TAACCATAGT AATCACAATA TTCACACTAA CCGCAACCAC TACAGCCTTA ACCCACAATG  8100
 I  F  L  H   D  H  V   L  L  L   T  C  L  M   T  M  V   I  T  M   F  T  L  T   A  T  T   T  A  L   T  H  N  D
ACCCCACAGA AGAAGTAGAA CAACTAGAAG CAGCTTGAAC AGTTGCCCCA ATTATAATTT TAATCTTAAC AGCTCTTCCA TCAGTTCGAT CCCTGTATTT  8200
  P  T  E   E  V  E   Q  L  E  A   A  W  T   V  A  P   I  M  I  L   I  L  T   A  L  P   S  V  R  S   L  Y  L
AATAGAAGAA GTGTTCAACC CATACCTAAC CATCAAGGCA ACAGGACATC AATGATACTG AAACTATGAA TACTCAGATG GAGTTAAAAT CTCATTCGAC  8300
  M  E  E   V  F  N  P   Y  L  T   I  K  A   T  G  H  Q   W  Y  W   N  Y  E   Y  S  D  G   V  K  I   S  F  D
TCTTACATAA TCCAAACAAA AGATTTACAA AACGGCTCAC CACGACTACT TGAAGTAGAC CACCGCATAG TCATACCAGC GGGCCTACAA ACCCGTGTTG  8400
 S  Y  M  I   Q  T  K   D  L  Q   N  G  S  P   R  L  L   E  V  D   H  R  M  V   M  P  A   G  L  Q   T  R  V  V
TAGTGACCGC AGAAGACGTT CTTCACTCCT GAACAATCCC ATCACTTGGA GTAAAAGTCG ACGCAGTCCC AGGACGACTA AACCAGCTAC CACTGGCCAC  8500
  V  T  A   E  D  V   L  H  S  W   T  I  P   S  L  G   V  K  V  D   A  V  P   G  R  L   N  Q  L  P   L  A  T
ATCACGAGTA GGGGTCTTTT ACGGCCAATG CTCAGAAATC TGCGGGGCAA ATCACAGCTT TATACCAATT GCAATAGAAG CAACCCCACT ACATCATTTT  8600
  S  R  V   G  V  F  Y   G  Q  C   S  E  I   C  G  A  N   H  S  F   M  P  I   A  M  E  A   T  P  L   H  H  F
                  →CO II┐ ┌tRNA(Lys)→                                                       →tRNA(Lys)┐  ┌ATP8→
GAACAATGAC TAATCTCAGA ACAATCACTG AGAAGCTTTT ATAGCATTAG CCTTTTAAGT TGAAGAAGAA ACACACTTTC CTCAGTGGTA TGCCACAACT  8700
 E  Q  W  L   I  S  E   Q  *                                                                     M  P  Q  L
AGACACAATC TATATTCTTA TAACCCACCT GTGAGCTTGA CTAATGCTAT ACCTAACCAC ACAAAAAATT AAAACCTTCA TTATAACATC ACACCCAATG  8800
  D  T  I   Y  I  L  M   T  H  L   W  A  W   L  M  L  Y   L  T  T   Q  K  I   K  T  F  I   M  T  S   H  P  M
                                 →ATP8     ┌            ┐   ATP6→
ATCTACCATA AACCTAATAA ACAAACACCC ACACCAACAT GACTATAAAC ATATTCGAAC AATTTATAAG CCCAGAACTA TTAATAATTC CAACAGCCTT  8900
 I  Y  H  K   P  N  K   Q  T  P   T  P  T  W   L  *
                                           M   T  M  N   M  F  E  Q   F  M  S   P  E  L   L  M  I  P   T  A  L
ACTATCAATA CTGGTCCCAG TCCTTCTAAT TCACCACAAT CCCAAACTCC TTGGAAACCG CATAACAACA GCAATTGCTT GACTACTAAT GACTATTATA  9000
 L  S  M   L  V  P  V   L  L  I   H  H  N   P  K  L  L   G  N  R   M  T  T   A  I  A  W   L  L  M   T  I  M
TCAAACATAA CCAATCAACT AACCCCCAGT GGACAAAAGT GGTGCCAAGT CCTAACTAGC CTACTTCTTA TAATCCTACT ATCAAACCTA CTAGGCTTAC  9100
 S  N  M  T   N  Q  L   T  P  S   G  Q  K  W   C  Q  V   L  T  S   L  L  L  M   I  L  L   S  N  L   L  G  L  L
TACCATACAC ATTCACATCG ACATCTCAAC TATCTATAAA CATAGCCATA GCCATCCCAC TATGAATAGC CACAATCATC ACAGGAATAA CAAAAAAACC  9200
  P  Y  T   F  T  S   T  S  Q  L   S  M  N   M  A  M   A  I  P  L   W  M  A   T  I  I   T  G  M  T   K  K  P
ATCAATCACA CTAGCCCACA TACTACCAGA AGGATCACCA ACCCCATTAA TCCCCTTCAT GATCATCATT GAAACAATTA GCCTATTAAT ACGACCATTA  9300
  S  I  T   L  A  H  M   L  P  E   G  S  P   T  P  L  I   P  F  M   I  I  I   E  T  I  S   L  L  M   R  P  L
GCGTTAGGAG TACGACTTAC CGCCAACATT ACGGCTGGTC ACCTACTCAT AACAATAGTA AGCACAACTA CATTAAACTT TATTACCTCA CACATCACAC  9400
 A  L  G  V   R  L  T   A  N  I   T  A  G  H   L  L  M   T  M  V   S  T  T  T   L  N  F   I  T  S   H  I  T  L
TTAGCATTAT AACATACCTT CTACTATTCC TACTATGCAT CTTAGAACTA GCAGTAGCCT GCATTCAAGC TTATGTATTT GTATTACTAA TCATCCTATA  9500
  S  I  M   T  Y  L   L  L  F  L   L  C  I   L  E  L   A  V  A  C   I  Q  A   Y  V  F   V  L  L  I   I  L  Y
        →ATP6┌CO III→
CCTTCAAGAA AACACATAAT GACCCACCAA CTCCACCAAT ATCACCTAGT AGACCCCAGC CCATGACCAC TGACAGGAGC CATAGGCTCA TTACTTCTAG  9600
 L  Q  E   N  T  *
                   M   T  H  Q   L  H  Q  Y   H  L  V   D  P  S   P  W  P  L   T  G  A   M  G  S   L  L  L  A
CCTCAGGACT GGCAGTTTGA TTTCATACTA ACAACACCAT ACTACTAAAA TTTGGCCTAT TAACCCTATT ACTAACCATA TTTCAATGAT GACGAGACAT  9700
  S  G  L   A  V  W   F  H  T  N   N  T  M   L  L  K   F  G  L  L   T  L  L   L  T  M   F  Q  W  W   R  D  I
TATTCGAGAA AGCACCTACC AGGGACACCA CACAAGTGGC GTCCAAAAAA ACATACGATA CGGAATAATT CTATTCATCA CATCAGAAGT ATTCTTCTTC  9800
  I  R  E   S  T  Y  Q   G  H  H   T  S  G   V  Q  K  N   M  R  Y   G  M  I   L  F  I  T   S  E  V   F  F  F
CTTGGCTTTT TCTGAGCACT ATACCATGTA AGCCTAGTAC CCACACCAGA ATTAGGAGCA GAATGACCCC CAATCGGAAT TACCCCGCTC AACCCAATAG  9900
 L  G  F  F   W  A  L   Y  H  V   S  L  V  P   T  P  E   L  G  A   E  W  P  P   I  G  I   T  P  L   N  P  M  E
AAGTACCCCT ACTAAACACA GCTGTACTTC TTTCATCAGG AGCAACCATT ACATGATCCC ACCACACAAT AATAAAGGGA AACAAAAAAG AAGCAACCCA 10000
  V  P  L   L  N  T   A  V  L  L   S  S  G   A  T  I   T  W  S  H   H  T  M   M  K  G   N  K  K  E   A  T  H
CGCCCTAATA CTCACTATTA TTCTAGGAGC CTACTTCACA GCCTTACAAC TATCAGAATA TATAGAAACC CCCTTCACCA TTGCAGACAG CGTATACGGA 10100
  A  L  M   L  T  I  I   L  G  A   Y  F  T   A  L  Q  L   S  E  Y   M  E  T   P  F  T  I   A  D  S   V  Y  G
TCATTATTCT TTGTGGCCAC AGGATTCCAC GGACTTCATG TTATAATTGG AACCTCCTTC CTAATAGTCT GCGCACTCCG CCTCGCTAAA CACCACTTTA 10200
 S  L  F  F   V  A  T   G  F  H   G  L  H  V   M  I  G   T  S  F   L  M  V  C   A  L  R   L  A  K   H  H  F  T
                                                                                                      →CO III
CTATCACACA CCACTTCGGA TATGAAGCAG CAATCTGATA CTGACACTTT GTCGATATTG TCTGACTGTT CTTATACATC TCGGTATACT GATGAGGGTC 10300
  I  T  H   H  F  G   Y  E  A  A   I  W  Y   W  H  F   V  D  I  V   W  L  F   L  Y  I   S  V  Y  W   W  G  S
┐ ┌tRNA(Gly)→                                               →tRNA(Gly)┐ ┌ND3→
CTATTTCTTT AGTATACTAG TACAAATGCC TTCCAAGCAT TAAGACCCCC CCGGGAAGAA ATAATTAACC TCATCACCCT TATTATTATG GCCATAGCAA 10400
 *                                                                M  N  L   I  T  L   I  I  M   A  M  A  M
```

Figure 2.—*Continued.*

```
TAACAACCGC CCTCTACACA ATTAACACCT ACACCACCAT AAAACCAGAT ATTAACAAAC TATCCCCCTA CGAATGTGGC TTCGACCCCC TAGGAAACGC 10500
 T  T  A   L  Y  T   I  N  T  Y   T  T  M   K  P  D   I  N  K  L   S  P  Y   E  C  G   F  D  P  L   G  N  A
CCGGACTCCT ATCTCCATTC AATTTTTCCT AGTAGCTATC TTATTCATCC TATTTGACCT AGAAATCGTA CTACTTCTAC CAACCCCTTG AAGCATAAAC 10600
 R  T  P   I  S  I  Q   F  F  L   V  A  I   L  F  I  L   F  D  L   E  I  V   L  L  L  P   T  P  W   S  M  N
ACCAACCCCT CAAACACAAC TATCCTGTTA ATTACCATAC TACTAACAAT TCTAACACTA GGCCTACTAT ACGAGTGATT ACAGGGCGGA CTGGAATGAA 10700
T  N  P  S   N  T  T   I  L  L   I  T  M  L   L  T  I   L  T  L   G  L  L  Y   E  W  L   Q  G  G   L  E  W  T
→ND3┐ ┌tRNA(Arg)→                                                      →tRNA(Arg)┐ ┌ND4L→
CTGAATACCG TGGTAGTCTA ATAGATATTT GGTTTCGACC CAAAAGAACT TACTAACCAT AAGCCACAGT AGTGGAACTA ATAAAAATGA CCCTATACAC 10800
 E  *                                                                         M  E  L   M  K  M  T   L  Y  T
AACCTTCATA ATTACAATCA TCGCCTTATC ACTACAACAA AAACACCTAA TATTAGCCCT AATATGTGTA GAAACAATAA TACTTATCGT ATTCACAATA 10900
 T  F  M   I  T  I  I   A  L  S   L  Q  Q   K  H  L  M   L  A  L   M  C  V   E  T  M  M   L  I  V   F  T  M
CTAGTTATAT TTAACTCTAA CTCACTAACC GTCTCACAAA CCCCAATACC AATCATCCTC CTAACTATCT CAGTGTGTGG GGCAGCTGTA GGCCTAAGCC 11000
L  V  M  F   N  S  N   S  L  T   V  S  Q  T   P  M  P   I  I  L   L  T  I  S   V  C  G   A  A  V   G  L  S  L
                                                           →ND4L┐ ┌ND4→
TAGTAGTCGC AATTACACGA ACTCATGGAA ACGACTTTCT AAAAAATCTA AACCTATTAT AATGCTAAAA ATCACTTTTA TAACCATAAT ACTAATCCCA 11100
 V  V  A   I  T  R   T  H  G  N   D  F  L   K  N  L   N  L  L  *
                                                                    M  L  K   I  T  F  M   T  M  M   L  I  P
                                                                                          Xba I
ACAACTTTAA TTCTAAAACC CAAAATACTT TATCAAACAA CAACCTCATA CTCCTTCATA ATTGCCCTAT TCAGCCTAAG CCTTCTAGAA CCAAACTCAA 11200
T  T  L  I   L  K  P   K  M  L   Y  Q  T  T   T  S  Y   S  F  M   I  A  L  F   S  L  S   L  L  E   P  N  S  N
ACACATATCT CCACCTCGAC TCAACATCAG CCCCACTACT ACTACTATCT TATTGACTCA TACCTATAAC AATAATAGCT AGTCAACATG CAATAAATAA 11300
 T  Y  L   H  L  D   S  T  S  A   P  L  L   L  L  S   Y  W  L  M   P  M  T   M  M  A   S  Q  H  A   M  N  K
AGAACCCCTA CAACGACAAC GAACATTCCT TTCAATCCTC ACACTTCTAC AGTTATTTAT TTCCTTAACA TTCATAGCTT CTAATATCAC CCTAATATAC 11400
 E  P  L   Q  R  Q  R   T  F  L   S  I  L   T  L  L  Q   L  F  I   S  L  T   F  M  A  S   N  I  T   L  M  Y
ATTATATTCG AAGCAACCTT AATCCCAACC CTAATCATTA TTACACGATG AGGACAACAA ACTGAGCGCC TAACCGCAGG CACATATTTC ATAATATATA 11500
I  M  F  E   A  T  L   I  P  T   L  I  I  I   T  R  W   G  Q  Q   T  E  R  L   T  A  G   T  Y  F   M  M  Y  T
CACTAACAAC TTCAATACCC CTACTAACAG CTATCCTATT TATCAATAAC ACAACAAACA CCCCAACTCT TTTTATACAA ATAATACAAA CAACCAGCCC 11600
 L  T  T   S  M  P   L  L  T  A   I  L  F   I  N  N   T  T  N  T   P  T  L   F  M  Q   M  M  Q  T   T  S  P
CTGGACAGAG CTAATACTAT GAATTGCCTG CCTGGGGGCT TTTTTAGCAA AAATACCAGT ATATGGCCTT CATCTCTGAC TACCAAAAGC CCATGTAGAA 11700
 W  T  E   L  M  L  W   I  A  C   L  G  A   F  L  A  K   M  P  V   Y  G  L   H  L  W  L   P  K  A   H  V  E
GCCCCAATCG CAGGCTCAAT AGTACTAGCC GCAATCCTAC TAAAATTAGG AGGCTACGGC ATTATTCGAA TAACACAAAT CCTACCAACA ATAAAGACAG 11800
A  P  I  A   G  S  M   V  L  A   A  I  L  L   K  L  G   G  Y  G   I  I  R  M   T  Q  I   L  P  T   M  K  T  D
                                                                                 Bgl II
ACCTTTTCCT ACCATTCATT GTTCTATCCC TCTGGGGGGC AACACTAGCC AACCTCACCT GCCTCCAACA AACAGATCTA AAATCCCTTA TCGCATACTC 11900
 L  F  L   P  F  I   V  L  S  L   W  G  A   T  L  A   N  L  T  C   L  Q  Q   T  D  L   K  S  L  I   A  Y  S
CTCTGTTAGC CATATGGGGC TGGTTATTGC TGCAACCATA ATTCAAACAC AGTGAAGTTT GTCCGGGGCC ATGGCCCTAA TAATCGCCCA TGGCTTTACC 12000
 S  V  S   H  M  G  L   V  I  A   A  T  M   I  Q  T  Q   W  S  L   S  G  A   M  A  L  M   I  A  H   G  F  T
TCCTCAGCCC TTTTCTGCCT AGCTAATACC ACCTATGAAC GAACCAAAAC CCGTATTATA ATCCTCACGC GGGGATTCCA CAGCATCTTA CCAATAATCA 12100
S  S  A  L   F  C  L   A  N  T   T  Y  E  R   T  K  T   R  I  M   I  L  T  R   G  F  H   S  I  L   P  M  I  T
CAACCTGGTG ACTAATAACC AATCTAATAA ACATTGCAAC TCCACCAAGT ATTAACTTTA CAGGAGAACT ATTAATCGCA TCATCTCTAT TCAACTGATG 12200
 T  W  W   L  M  T   N  L  M  N   I  A  T   P  P  S   I  N  F  T   G  E  L   L  I  A   S  S  L  F   N  W  C
TCCAACAACA ATTATCCTAT TCGGACTATC AATACTAATT ACAGCATCAT ATTCCCTACA CATACTCCTA TCAACACAAA CAGGCACCCC AACATTAAAT 12300
 P  T  T   I  I  L  F   G  L  S   M  L  I   T  A  S  Y   S  L  H   M  L  L   S  T  Q  T   G  T  P   T  L  N
                                                                                                   →ND4┐ ┌
ATAGTAACAT ACCCAACACA CTCACGAGAA CACTTACTTA TAGTATTACA CATCCTCCCA TTAATACTAA TTTCACTAAA ACCAGAACTG ATTATTTAAG 12400
M  V  T  Y   P  T  H   S  R  E   H  L  L  M   V  L  H   I  L  P   L  M  L  I   S  L  K   P  E  L   I  I  *
tRNA(His)→                                                  →tRNA(His)┐ ┌tRNA(Ser,AGY)→
TGTGCGTAAT TTAAAAAAAA TATCAGGCTG TGACCCTGAC AATAGGAATT AACCCCTCAC ACACCGAGGG TGCCATATAG ACCTGCTAAC TCTTTAATCT 12500
   →tRNA(Ser,AGY)   ┌ ┐    tRNA(Leu,CUN)→                                          →tRNA(Leu,CUN)┐ ┌ND5→
GGGAATAATA ACCAGCTCCC TCTATCAAAG GATAATAGTA TTCCACTGGT CTTAGGCACC AAAACCCTTG GTGCAAATCC AAGTGATAGA AATGAACTTA 12600
                                                                                                 M  N  L
ATCACCCCCA CAATTATCCT AACAATTATA CTATCTTTAA TAATATCCAT TATACAACCA CACAACAACA CCAAAAATAA CCTTATAGTA CTCTTTCTAA 12700
I  T  P  T   I  I  L   T  I  M   L  S  L  M   M  S  I   M  Q  P   H  N  N  T   K  N  N   L  M  V   L  F  L  I
TTAGCCTAAT TCCAATCAAC CCGCTATTAA ATAACAACGA ACTAACACTA ACACTCACAC CACTAATTAT CTCACCAACA GAAAATATCA ATATCTCTAT 12800
 S  L  I   P  I  N   P  L  L  N   N  N  E   L  T  L   T  L  T  P   L  I  I   S  P  T   E  N  I  N   I  S  I
                                                                   EcoR I
TACCCTGGAC ACAGCATCAC TACTATTCAC CCCAATCGCC CTATTTATCA CATGATCAAT CACAGAATTC TCTTTATGGT ATATAGCCAC AGACCCCAAC 12900
 T  L  D   T  A  S  L   L  F  T   P  I  A   L  F  I  T   W  S  I   T  E  F   S  L  W  Y   M  A  T   D  P  N
ATCAACAAAT TCATCAAATA CCTGCTAACA TTCTTAATTA CCATACTAGT AATCATTACA GCGAACAACA TATATCAACT TTTCATCGGG TGAGAGGGAG 13000
I  N  K  F   I  K  Y   L  L  T   F  L  I  T   M  L  V   I  I  T   A  N  N  M   Y  Q  L   F  I  G   W  E  G  V
TAGGAATTAT ATCCTTCTTA CTTATCGGGT GATGACACGG CCGCCAAGAC GCCAATACAG CCGCCCTTCA AGCTATTATC TATAATCGTA TCGGAGACAT 13100
 G  I  M   S  F  L   L  I  G  W   W  H  G   R  Q  D   A  N  T  A   A  L  Q   A  I  I   Y  N  R  I   G  D  I
TGGACTTATT ATAACAACCG CATGAATAAT AACCACATCA TCAATCAATA TACAAGAACT TATAATTCAA CACGAAGTAG TTAACATCAT CCCACTACTC 13200
 G  L  I   M  T  T  A   W  M  M   T  T  S   S  I  N  M   Q  E  L   M  I  Q   H  E  V  V   N  I  I   P  L  L
GGACTTGTGG CAGCTGCCAC AGGAAAATCC GCACAATTCA GCCTCCACCC ATGACTTCCC TCAGCTATAG AAGGACCAAC ACCTGTCTCA GCCCTTCTCC 13300
G  L  V  A   A  A  T   G  K  S   A  Q  F  S   L  H  P   W  L  P   S  A  M  E   G  P  T   P  V  S   A  L  L  H
ACTCCAGCAC AATAGTAGTA GCCGGAGTCT TTCTATTAAT TCGACTCCAC CCGATCCTAC ACAATAATAA AATCATACTA ACCTGCTGCC TAATCTTAGG 13400
 S  S  T   M  V  V   A  G  V  F   L  L  I   R  L  H   P  I  L  H   N  N  K   I  M  L   T  C  C  L   I  L  G
GGCAACAACA ACAATATTTG CTGCTGCAGC AGCAACAACC TACTTCGACA TTAAAAAAAT TATCGCACTA TCTACTACAA GCCAACTAGG ACTAATAATA 13500
 A  T  T   T  M  F  A   A  A  A   A  T  T   Y  F  D  I   K  K  I   I  A  L   S  T  T  S   Q  L  G   L  M  M
ACAATAATTG GATTAAACCA ACCAACTCTG GCTTTTCTAC ATATAATTAC CCACTCCTTC TTTAAGGCAA TACTATTCCT ATGCTCAGGA TCCTATATTC 13600
T  M  I  G   L  N  Q   P  T  L   A  F  L  H   M  I  T   H  S  F   F  K  A  M   L  F  L   C  S  G   S  Y  I  H
ACAATCTAAA TAATGAACAG GACATCCGCA TAATAGGAGG ACTACTAAAA ACTTTACCAA TAACCTCATC ATTCCTAACC ATCGCCAACC TGTCGCTAAT 13700
 N  L  N   N  E  Q   D  I  R  M   M  G  G   L  L  K   T  L  P  M   T  S  S   F  L  T   I  A  N  L   S  L  M
```

Figure 2.—*Continued.*

```
AGGAATACCA TTCTTATCAG GGTTCTACTC AAAAGATACT ATTATTGAAA CACTAGCCAA CTCTTACACT AACTCATGAG CCATTATAAT TACAATAATC 13800
 G  M  P    F  L  S  G   F  Y  S    K  D  T    I  I  E  T   L  A  N    S  Y  T    N  S  W  A   I  M  I    T  M  I
GCCACTATCT TATCCGCCTG CTACAGCACA CAAATCATAC TATTCACAAT CATAGAACAC CCACGAACCC ACCACACCAC CCACAAAGAA ACAAAAAACA 13900
A  T  I  L   S  A  C   Y  S  T    Q  I  M  L   F  T  I    M  E  H    P  R  T  H   H  T  T    H  K  E    T  K  N  I
TCACACACCC ACTTGCCCGC CTAATACTCA CATCAATCTT AATAGGAACC ATAACAAAAA TATCAACCCT ACAAACCACA ACAATAGTAA CCATACCAAA 14000
 T  H  P    L  A  R    L  M  L  T   S  I  L    M  G  T    M  T  K  M   S  T  L    Q  T  T    T  M  V  T   M  P  K
AACAATCAAA CTAATAGCAC TCATCTCTAC CATCATTGGA GTTTTACTAT CAAAAGACCT CACTCACATA ACCCACCATA TAAAACCTAA AAAACCTAAC 14100
 T  I  K    L  M  A  L   I  S  T    I  I  G    V  L  L  S   K  D  L    T  H  M    T  H  H  M   K  P  K    K  P  N
AAACAAAATA TATTCTTCAA TCAACTGGCA TTCTTTAATA TCCCCCACCG AACCATTACC ATCAACACCT TAAAAATCAG CCAACAAACC TCCACAGAAC 14200
K  Q  N  M   F  F  N    Q  L  A    F  F  N  I   P  H  R    T  I  T    I  N  T  L   K  I  S    Q  Q  T    S  T  E  L
TAATAGACCT GTGAACCCTA GAAGTCTGAG GACCCAAAGG CCTGTCAAAC ACAATTACCA ACACAATCCA CCTTTTAACA CAACAAAAAA ACATAATCAA 14300
 M  D  L    W  T  L    E  V  W  G   P  K  G    L  S  N    T  I  T  N   T  I  H    L  L  T    Q  Q  K  N   M  I  K
                                                         →ND5    ┌           ┐  ND6←
AAACTATATA GCTATTTTCA CTATGACTAC AATAACTGTA CTCTTATTCA TTATATCTAA AAGGCCGCAA ACCACCCAAT CGAGACCAAC TCAAAATAAC 14400
 N  Y  M    A  I  F  T   M  T  T    M  T  V    L  L  F  I   M  S  K    *
                                                                   *  F  P  R  L   G  G  L    R  S  W    S  L  I  V
TAAAATAGAA AACAATACAA CAAGCAACCC TCAAGAACAC ACAATCAAAC CCCACCCACC ACTAAAATAA AAAACACCTA CCCCATTCAT CTCCAAACAC 14500
 L  I  S    F  L  V    V  L  L  G   W  S  C    V  I  L    G  W  G  G   S  F  Y    F  V  G    V  G  N  M   E  L  C
ACTAAATCCC CTCAATTAAC ATACACCAAT AAACCCCCCA CCTCACGCAA CAAACATAAT AAAACAACAA GCAACAAGGA TACACAAACA ATAAAATACT 14600
 V  L  D    G  W  N  V   Y  V  L    L  G  G    V  E  R  L   L  C  L    L  V  V    L  L  L  S   V  C  V    I  F  Y
TAGTCCCACC AACACTATAA ATACCACTTT CCTTCTCAAC ACTAACACAA TACCCAAAAA CTACAACTAA CCCCCCAAGA TATACAATGT ACATCACCAA 14700
K  T  G  G   V  S  Y    I  G  S    E  K  E  V   S  V  C    Y  G  F    V  V  V  L   G  G  L    Y  V  I    Y  M  V  L
AGCCGCAAAC GTACGACCTA AAAAAACTAT AAAAATACAG CAAAAAAAAG AAACGCCCAT TAAAGCAACA ACCCCTTGAT ATGGGGCAGA AACTACACCT 14800
 A  A  F    T  R  G    L  F  V  M   F  I  C    C  F  F    S  V  G  M   L  A  V    V  G  Q    Y  P  A  S   V  V  G
                                                        ←ND6┐         ┌tRNA(Glu)←
AAAACCACAA CACTCAAAAC TAAAAACACC AAAACTAAAC TAAAAAAATA ATTCATTAAA ACCATAATTC TTGCTCTAAT GAGACCTGTG ATCTGAAAAA 14900
 L  V  V    V  S  L  V   L  F  V    L  V  L    S  F  F  Y   N  M
           ←tRNA(Glu)┐ ┌cytb→
CCACCGTTGT AAATCAACTA CAAAAATCAT GCCCAACCAA CACATACTCA TACTATTTAA TATATTGCCA GTAGGATCAA ACATCTCAAC CTGATGAAAT 15000
                                 M  P  N  Q   H  M  L  M   L  F  N    M  L  P    V  G  S  N   I  S  T    W  W  N
TTCGGTTCCA TACTACTAAC CTGCTCTGCC CTACAAACCA TCACTGGTTT CTTCCTAGCC ATTCACTATA CAGCTAATAT CAATCTTGCT TTCTCATCCG 15100
F  G  S  M   L  L  T    C  S  A    L  Q  T  I   T  G  F    F  L  A    I  H  Y  T   A  N  I    N  L  A    F  S  S  V
TCATCCACAT CACTCGAGAT GTCCCCTACG GATGAATCAT ACAAAACCTC CATGCAATTG GCGCATCCAT ATTCTTCATT TGCATTTACA TCCATATCGC 15200
 I  H  I    T  R  D    V  P  Y  G   W  I  M    Q  N  L    H  A  I  G   A  S  M    F  F  I    C  I  Y  I   H  I  A
ACGCGGACTA TATTACGGGT CCTACCTTAA CAAAAATGTG TGACTCTCAG GAACCATTTT ACTATTTATT CTAATAGCAA CAGCTTTCTT CGGCTATGTT 15300
 R  G  L    Y  Y  G  S   Y  L  N    K  N  V    W  L  S  G   T  I  L    L  F  I    L  M  A  T   A  F  F    G  Y  V
CTACCATGAG GACAAATATC ATTCTGAGCT GCAACAGTAA TCACCAACCT ACTAACAGCA GTACCCTACA TTGGCACAAC ACTAACCAAC TGACTTTGAG 15400
L  P  W  G   Q  M  S    F  W  A    A  T  V  I   T  N  L    L  T  A    V  P  Y  I   G  T  T    L  T  N    W  L  W  G
GGGGATTCTC TATTAATGAC CCAACACTCA CTCGATTTTT TGCCCTACAC TTCATCCTCC CATTCACTAT TATCTCACTA TCTTCAATCC ACATCATACT 15500
 G  F  S    I  N  D    P  T  L  T   R  F  F    A  L  H    F  I  L  P   F  T  I    I  S  L    S  S  I  H   I  M  L
ACTTCACACA GAAGGCTCTA GTAACCCACT AGGAACAAAC TCAGACATTG ATAAAATCCC ATTTCACCCA TACCACACCC ATAAAGACAT TCTAGTATTA 15600
 L  H  T    E  G  S  S   N  P  L    G  T  N    S  D  I  D   K  I  P    F  H  P    Y  H  T  H   K  D  I    L  V  L
ACCATTATAT TAACCACAAT ATTCATTATT ATAACGCTCA CCCCAAATAT CTTCAACTAC CCAGAAAACT TCTCAAAGGC CAACCCACTA GTAACACCAC 15700
T  I  M  L   T  T  M    F  I  I    M  T  L  T   P  N  I    F  N  Y    P  E  N  F   S  K  A    N  P  L    V  T  P  Q
AACACATTAA GCCAGAATGA TACTTCCTAT TCGCCTACGG CATTCTCCGG TCCATCCCCA ATAAACTAGG AGGAACCGTA GCCCTCGTAT TATCCGTAGC 15800
 H  I  K    P  E  W    Y  F  L  F   A  Y  G    I  L  R    S  I  P  N   K  L  G    G  T  V    A  L  V  L   S  V  A
CATTCTACTA ACAACACCAT TTACCCACAC CTCACACATA CGATCCATAA CATTCCGCCC ACTAACACAA CTCATATTTT GAACCCTTGT AGCCACATTT 15900
 I  L  L    T  T  P  F   T  H  T    S  H  M    R  S  M  T   F  R  P    L  T  Q    L  M  F  W   T  L  V    A  T  F
ATTACAATCA CATGAGCAGC TACTAAACCA GTAGAACCCC CATTCACCAT AATCGGCCAA ATAACCTCCC TACTATACTT CTCATTCTTC ATTATAAACC 16000
I  T  I  T   W  A  A    T  K  P    V  E  P  P   F  T  M    I  G  Q    M  T  S  L   L  Y  F    S  F  F    I  M  N  P
                                              →cytb┐ ┌tRNA(Thr)→                                        →tRNA(Thr)
CCCTACTTGG CTGACTAGAA AACAAAATCT CATTCACTAA CACCTGCCCT AGTAGCTTAA CCACTAAAGC ATTGTTCTTG TAAACCAAAG ATGGACCCAA 16100
 L  L  G    W  L  E    N  K  I  S   F  T  N    T  *
      ┐  ┌tRNA(Pro)←                                              ←tRNA(Pro)┐ ┌control region II
CCCTAGAGCA TCAAAGAGAG ACTTCCCATC TCTAGCCCCC AAAGCCAGTA TTTTAACTTA AACTACTCTT TGCCAAAAAT AAACAATCCT CCTAGGACCC 16200
C-rich sequence                                      hairpin 1                          hairpin 2 ┌───TAS1───┐
CCCCCCTACC CCCCCCCTTA CATCTGACCC AATTCCGACT ATAATGTATC TCTTAGGTTA TAGTCTTTAT TTCACTATGT ATAATTATAC ATTAATGATC 16300
                   hairpin 3 ┌───TAS2───┐                                                └ 20-bp repeat 1
TGCCTCACGC CTAATAAACG GGAATTATAC TATAATTATT TGTATAAAAA AGTGGTATAT TACATGCGAT TTACCCCCTC ATTTCCCAAA CGTTCCATGT 16400
┘                           └ 20-bp repeat 2   ┘
TATCTTCGGC CACCCATTAT TAGTAACCAT GACTATCCCG TGCCTAATGG TGTCCCATGA TGTAGCTCAG CCCGTGAAAC CCTCTATCCT TCCATTGAAA 16500
                                          ┌                49-bp repeat 1                   ┐
GCATACAGTC CCGCTTCTCA CGTCCATATA TTGTAACTCC TCCCTTTATG TTCTTTCCAA GGCCGCTGGT TACACTCTCA AGAGCATCTC AATGGTCCGG 16600
   ┌              49-bp repeat 2                        ┐      ┌              49-bp repeat 3
AACCACCCCG CCTTACTTGC TTTTTCCAAG GCCTTTGGTC GCACCCTTTA TAGTGGTACA TATCACCTCA TGTTCTTATC ACCTATGCCA GTTCCACCCC 16700
   ┐
TGGTTGTTCT TTTTATCTCT ACCTTTCACC TGACACTCAC TTGCCCGTTA CCGTTCCCCT CACCGGGGCG TAGACCATCT AGTCCGGGTG GAGCTATATT 16800
                                           CSB I
CTTGGCTATG CACATTCCCT ATACGGATAC ATTCCTTCAT GCTTGTTAGA CATACTTTTT CCTTGACTGT AAAATTTCAT TATTCTTTTA TTATAAATTT 16900
                                                                                   CSB II          CSB III
GCCGCAGTAA ACACTTTTTC ATCACCCCTA TTTTTAAAAT CTGAGAAAAA ATAATACGAC AAAACTATAA TAAACCCCCC GCCCCGAAAC CCCATATACT 17000
                                                         hairpin 4
TATTTTCATA CCAGCCTCTC ACCCCGATCA CACCCCGTGA AAATAGTTTA CTATAAAAAT AGCAAACACA TATTTTAGAA CCGCATATTT AACTCTTTTC 17100
                                                                              control region II┐
TCTGAAAAAT CATTTCAAAT GCCAAAACAC CCCCCACTAA AAAATAGTCA TTTCACCAAT TATCTCAACA ATCTCCCGAA TTTTTATATA A          17191
```

Figure 2.—*Continued.*

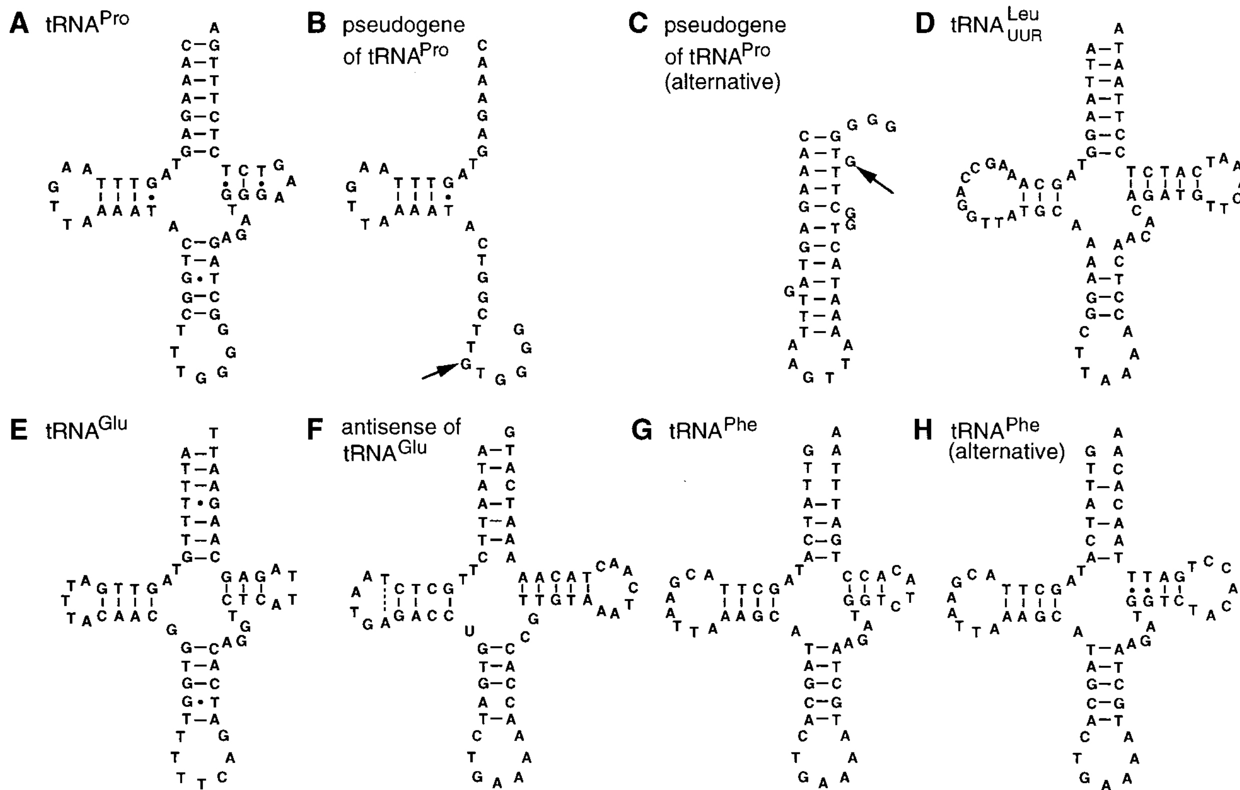

Figure 3.—Several tRNA genes and relevant structural features found in akamata mtDNA. (A) tRNA^{Pro} gene (positions 16,111–16,172 in Figure 2). (B) tRNA^{Pro} pseudogene (3566–3601). (C) An alternative secondary structure for the tRNA^{Pro} pseudogene (3566–3601). (D) tRNA^{Leu}(UUR) gene (4620–4692). (E) tRNA^{Glu} gene (14,866–14,927). (F) An antisense sequence of the tRNA^{Glu} gene (14,864–14,931). (G) tRNA^{Phe} gene (1–65). (H) An alternative secondary structure for the tRNA^{Phe} gene (1–70). Bars in stems represent Watson-Crick basepairs, and dots stand for wobble G-U pairs. Note that the guanosine indicated by an arrow in B and C, which is the only difference from the 5′ half portion of the tRNA^{Pro} gene shown in A, was not found for a different individual from the same locality of Okinawajima Island (Kumazawa *et al.* 1996).

decoding all 60 sense codons of the mitochondrial genetic code without importing nuclear-cytosolic tRNAs.

Considerable truncation of the TψC arm (T arm), previously described for snake mitochondrial tRNAs (Kumazawa *et al.* 1996), is also notable in akamata mitochondrial tRNA genes, most of which have <5 bp in the T stem. For instance, genes for tRNA^{Pro} (Figure 3A), tRNA^{Glu} (Figure 3E), and tRNA^{Phe} (Figure 3G) have only 3 bp in the T stem. Together with small numbers of nucleotides assigned for the T loop region, this results in considerably truncated T arm regions for the snake tRNAs (Figure 4). For 16 tRNAs from tRNA^{Pro} to tRNA^{Ile} in Figure 4, the number of nucleotides in the T arm is smaller in the snake than in other vertebrates and starfish. For the four remaining tRNAs, *i.e.*, leucine tRNAs, tRNA^{Asn}, and tRNA^{Gln}, the figure may be constrained to be 17 because of the need for stabilization of the dihydrouridine loop (D loop)/T loop interaction via conserved sequence elements, such as GG [or AG for tRNA^{Leu}(CUN)] in the D loop and TCR [or GCA for tRNA^{Leu}(CUN)] in the T loop (Figure 2). These results strengthen a previous notion (Kumazawa *et al.* 1996) that snake mitochondrial tRNA genes may have evolved to be somewhat simpler in the T arm structure, which is associated with the loss of the canonical tertiary interactions between the D and T loops.

Because the snake mitochondrial tRNAs do not exactly conform to the canonical secondary structure model, there remains an ambiguity in identifying the tRNA^{Phe} gene. The tRNA^{Phe} gene located in the 3′ vicinity of control region I (Figure 2) can be folded into two alternative secondary structures (Figure 3, G and H), both of which have considerable mismatches in the acceptor stem region. Although a few mismatched pairs may be tolerated in the acceptor stem of mitochondrial tRNAs (Kumazawa and Nishida 1993), mismatches in these cases occur at the acceptor end, in which the structural destabilization is crucial to the function of tRNAs. Another possible candidate for the tRNA^{Phe} gene exists as an antisense sequence of the tRNA^{Glu} gene (Figure 3F). The antisense sequence appears to include a proper anticodon loop for tRNA^{Phe} with more stable acceptor stem structures than the putative tRNA^{Phe} genes described above. One shortcoming in the antisense gene is the shortness of the variable loop (only two nucleotides), which may cause difficulty in tertiary inter-

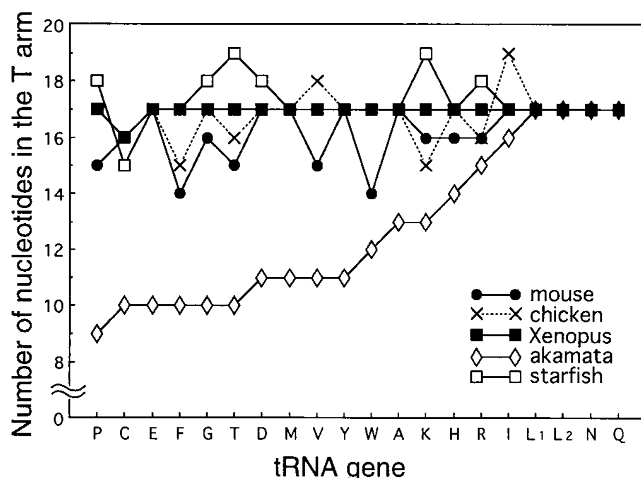

Figure 4.—Simplification of the T arm in snake mitochondrial tRNA genes. The number of nucleotides in the T arm of mitochondrial tRNA genes was compared among the mouse (Bibb *et al.* 1981), chicken (Desjardins and Morais 1990), African clawed toad (Roe *et al.* 1985), akamata (this study), and starfish (Asakawa *et al.* 1995). The comparison was made for all tRNA genes, except for two serine isoacceptors in which a common secondary structure model cannot be defined among vertebrates and starfish (*e.g.*, see Kumazawa and Nishida 1993). The assignment of the T arm region (T stem plus T loop) was done according to the method of Kumazawa and Nishida (1993). Refer to Figure 1 for the abbreviations of the tRNA genes.

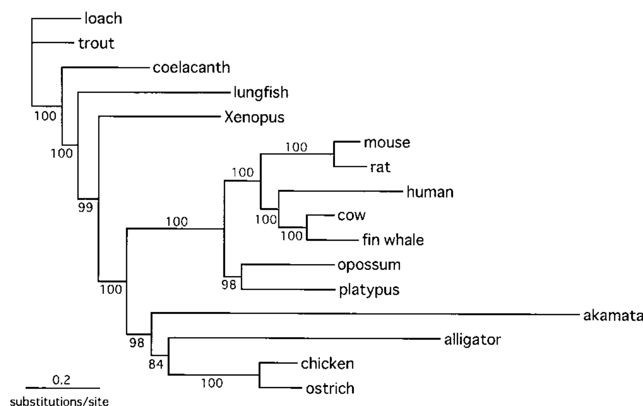

Figure 5.—Maximum likelihood tree constructed from mitochondrial protein sequences. Aligned amino acid sequences for 12 light-strand protein genes from which gap sites were deleted were concatenated (3444 sites in total) and analyzed by PUZZLE version 4.0 (Strimmer and von Haeseler 1996). Shown here is the maximum likelihood tree by the QP algorithm (Strimmer and von Haeseler 1996) assuming the mtREV24 substitution model (Adachi and Hasegawa 1996) and the gamma-distributed rates of substitution over sites (Ota and Nei 1994). The shape parameter of the gamma distribution (0.38) was calculated from the data set. Values along internal branches are QP reliability values in percent obtained from 1000 puzzling steps (Strimmer and von Haeseler 1996). A scale bar indicating 0.2 substitutions per site is shown. Data source: loach, Tzeng *et al.* 1992; trout, Zardoya *et al.* 1995; coelacanth, Zardoya and Meyer 1997; lungfish, Zardoya and Meyer 1996; Xenopus, Roe *et al.* 1985; mouse, Bibb *et al.* 1981; rat, Gadaleta *et al.* 1989; human, Anderson *et al.* 1981; cow, Anderson *et al.* 1982; fin whale, Arnason *et al.* 1991; opossum, Janke *et al.* 1994; platypus, Janke *et al.* 1996; akamata, this study; alligator, Janke and Arnason 1997; chicken, Desjardins and Morais 1990; and ostrich, Härlid *et al.* 1997.

actions involving nucleotides in the variable loop. We thus consider that the most plausible tRNA^{Phe} gene is the one shown in Figure 3G, under the assumption that RNA editing corrects mismatched base pairs in the acceptor end. RNA editing at the acceptor stem is known to occur for mitochondrial tRNAs of pulmonate gastropods and mammals (Yokobori and Pääbo 1995a,b).

Pseudogene for tRNA^{Pro}: The tRNA^{Pro} pseudogene (Figure 3B) consists of a 5′ half portion of the tRNA^{Pro} gene that exists immediately 5′ upstream of control region I (Figure 3A). The insertion of a guanosine into the anticodon loop, as indicated by the arrow in Figure 3B, is the only difference between these two paralogous sequences. It seems improbable that a functional tRNA^{Pro} molecule is expressed from the half-sized gene sequence. Instead, it can be folded into a rather stable hairpin-like structure, as shown in Figure 3C. As discussed above, this structure may facilitate pausing of heavy-strand replication for the formation of the D-loop within control region. In a viperid snake, himehabu, which was also shown to possess duplicate control regions (Kumazawa *et al.* 1996), a functional tRNA^{Pro} gene is found 5′ upstream of control region II, and a similar tRNA^{Pro} pseudogene corresponding to its 5′ half portion exists 5′ upstream of control region I. In the rattlesnake, instead of the half-sized tRNA^{Pro} pseudogene, a long intergenic spacer exists in the 5′ vicinity of control region I (Kumazawa *et al.* 1996). This spacer contains the sequence TTTATAGTAAATACTGTAAA, which may form a stable hairpin structure by use of the complementarity between the underlined parts, and this may substitute for the possible role of the tRNA^{Pro} pseudogene sequence.

Protein genes: Translated amino acid sequences for 12 light-strand protein genes were concatenated and used for phylogenetic analyses among akamata and other vertebrates. A maximum likelihood tree (Figure 5) was obtained by the quartet puzzling (QP) algorithm (Strimmer and von Haeseler 1996, PUZZLE version 4.0), assuming the mtREV24 substitution model (Adachi and Hasegawa 1996) and the gamma-distributed rates of substitution over sites (Ota and Nei 1994). The tree is in agreement with previous studies (*e.g.*, Adachi and Hasegawa 1996; Janke and Arnason 1997; Zardoya and Meyer 1997) with respect to topology and branch length relating to taxa other than akamata. Akamata clusters with archosaurian taxa (*i.e.*, birds and crocodiles) relative to mammals, which is supported with relatively high QP reliability values (>84%) that have the same practical meaning as bootstrap values (Strim-

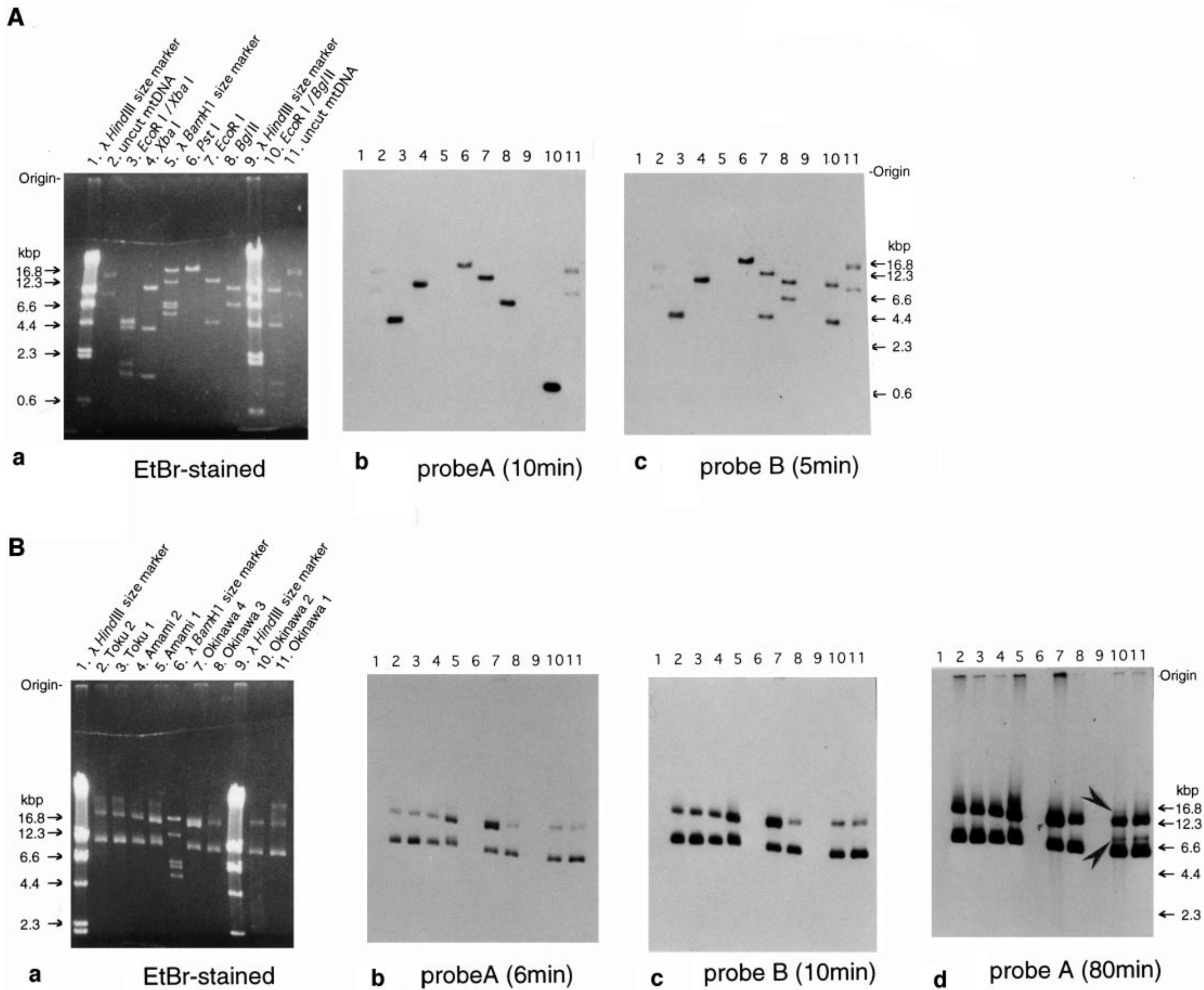

Figure 6.—Southern hybridization of restriction fragments of akamata mtDNA. Samples analyzed are akamata mtDNAs from Okinawa 4, which were digested with various restriction enzymes (A), and akamata mtDNAs from eight individuals, which were uncut (B), digested with *Pst*I (C), and digested with *Eco*RI (D). Hybridization experiments were done as described in materials and methods. Electrophoresed gels were stained with ethidium bromide (a), and their blots were hybridized with probe A (b and d) and then reprobed with probe B (c). The corresponding locations for probes A (complementary to positions 880–1078 in Figure 2) and probe B (positions 4057–4225 and 16,628–16,796) are also indicated in Figure 1. The time in parentheses indicates exposure time to a Kodak film. The arrowheads indicate faint signals discussed in the text. Size markers used are λDNA digested with *Hin*dIII (λ*Hin*dIII) and circularized λDNA digested with *Bam*HI (λ*Bam*HI). The approximate positions corresponding to several marker bands are indicated on the right and left of the figure. Sizes of observed restriction fragments in A-a are as follows: *Pst*I (17.2 kbp), *Xba*I (10.7, 3.9, and a doublet of 1.3 kbp), *Eco*RI (12.6 and 4.6 kbp), *Bgl*II (10.3 and 6.9 kbp), *Eco*RI/*Xba*I (a mixture band of 4.6 and 4.4, 3.9, 1.7, and a doublet of 1.3 kbp), and *Eco*RI/*Bgl*II (10.3, 4.6, 1.3, and 1.0 kbp). Among these restriction fragments, probe A hybridized with the 17.2-kbp fragment produced by *Pst*I, 10.7 kbp by *Xba*I, 12.6 kbp by *Eco*RI, 6.9 kbp by *Bgl*II, 4.6 and/or 4.4 kbp by *Eco*RI/*Xba*I, and 1.3 kbp by *Eco*RI/*Bgl*II (A-b), whereas probe B hybridized with 17.2 kbp by *Pst*I, 10.7 kbp by *Xba*I, 12.6 and 4.6 kbp by *Eco*RI, 10.3 and 6.9 kbp by *Bgl*II, 4.6 and/or 4.4 kbp by *Eco*RI/*Xba*I, and 10.3 and 4.6 kbp by *Eco*RI/*Bgl*II (A-c).

mer and von Haeseler 1996). This phylogenetic relationship is congruent with a traditional one in which snakes are affiliated to Reptilia, Squamata (Rage 1987).

A striking feature in Figure 5 is that a terminal branch leading to akamata is rather long. The branch length measured from the latest common ancestor between snakes and birds is roughly three times larger for akamata than for birds, as averaged between chicken and ostrich. Moreover, pairwise distances among the sequence data computed with various substitution models were substantially larger between akamata and other taxa than among nonserpentine animals (data not shown). This tendency was seen not only for the concatenated sequences, but for individual protein sequences (data not shown). These observations suggest considerably higher rates of molecular evolution in the snake lineage. It has recently been shown that crocodilian mtDNAs evolve faster than mtDNAs of most other verte-

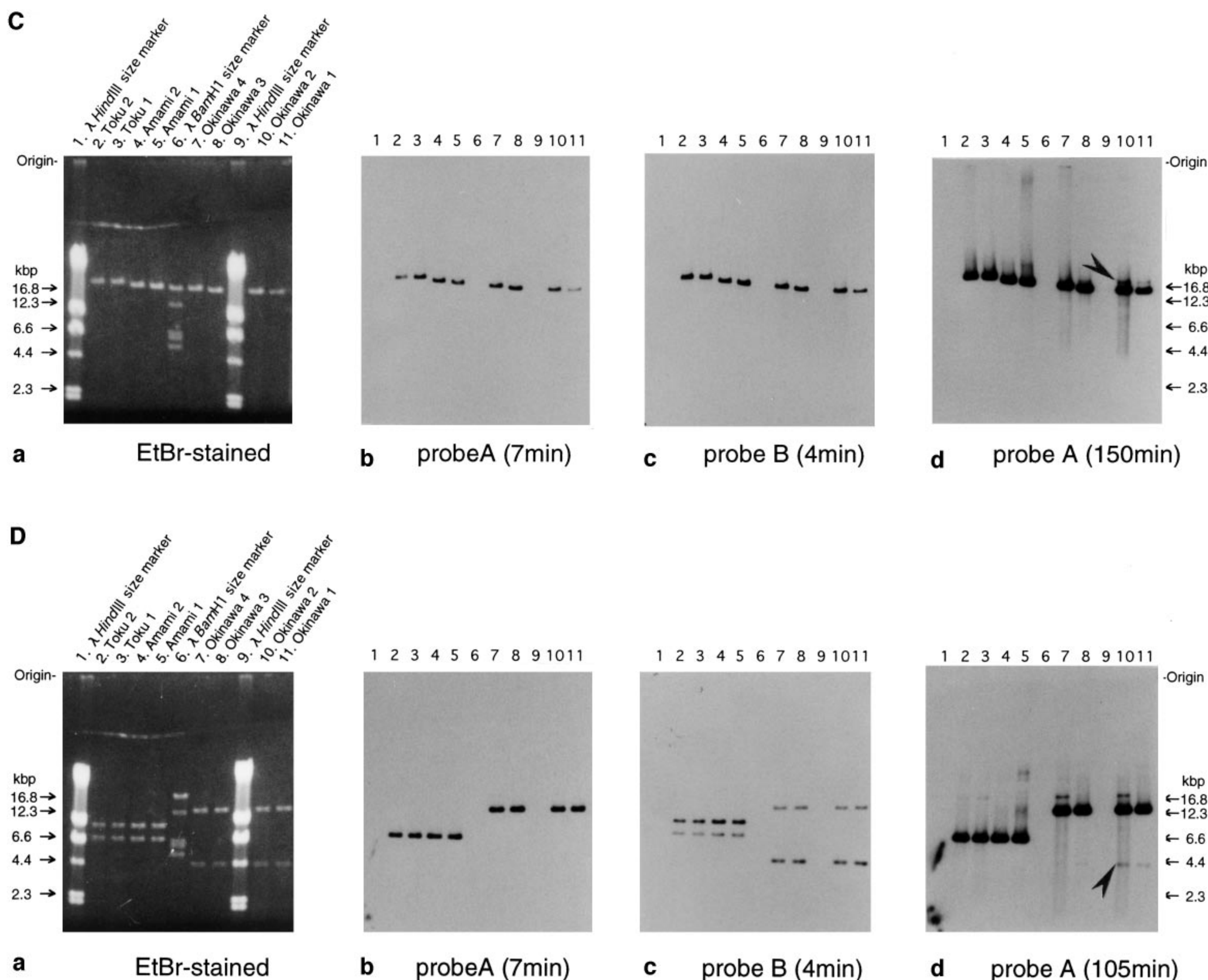

Figure 6.—*Continued.*

brates (Kumazawa and Nishida 1995; Janke and Arnason 1997), but the rate in snakes appears to be even higher than that in crocodilians (Figure 5). Snakes and crocodilians are cold-blooded reptiles and, thus, the cold bloodedness may not necessarily lead to deceleration of molecular evolution despite suggestions by some earlier studies (Thomas and Beckenbach 1989; Adachi *et al.* 1993; Martin and Palumbi 1993). The mechanisms of the rate acceleration in snake mtDNA remain unclear. One possibility is that a replicational mechanism using two control regions (see above) may not be very accurate.

The tree topology shown in Figure 5 was not obtained when the same data set was analyzed on the assumption of uniform rate over sites or two rates (*i.e.*, mixture of variable sites with uniform rate and invariable sites). In these cases, akamata clustered with the alligator relative to birds (data not shown), which most likely represents a false phylogeny resulting from considerable rate fluctuation among lineages, as described above. The Archosauria clade has been supported by various morphological and molecular studies (*e.g.*, Carroll 1988; Hedges 1994; Kumazawa and Nishida 1995; Janke and Arnason 1997), and thus the inclusion of snakes in this clade is very unlikely. It seems reasonable that the gamma-distributed rates were used for analyzing the mitochondrial protein data in Figure 5 because an estimated shape parameter of the gamma distribution for the data set (0.38) points to considerable rate heterogeneity over sites. However, we consider that future phylogenetic studies using akamata mtDNA sequences should be done very carefully because of the unusually high rates of sequence evolution. Simulation studies (Nei 1991) pointed to the difficulty of reconstructing the correct phylogeny by currently used tree-building methods when extreme rate variations exist among lineages.

Cloning of an unexpected fragment: During the course of cloning mtDNA fragments as described in

materials and methods, a 4.6-kbp *Eco*RI-cut fragment covering from ND5 to 12S rRNA genes (Figure 1A) was refractory to cloning using *Eco*RI-cut *E. coli* vectors. Uncloneable mtDNA fragments are known to exist (*e.g.*, see Drouin 1980; Mita *et al.* 1988; Asakawa *et al.* 1995), and suspected reasons for this are inhibitory effects of cloned mitochondrial genes on the growth of *E. coli* as well as profound structural contexts near restriction sites. Repeated rounds of the cloning trial actually generated clones having a 4.6-kbp insert, albeit in a low efficiency (five clones in total after three rounds of the cloning trial using an electroporation apparatus). These clones were initially taken for the wanted ones owing to the reasonable size of the inserts, and one of the clones was sequenced completely.

It turned out that the 4.6-kbp insert consists of an entirely unexpected sequence starting from an *Eco*RI site inside the 12S rRNA gene and continuing to genes for $tRNA^{Val}$, 16S rRNA, ND1, $tRNA^{Ile}$, a pseudogene for $tRNA^{Pro}$, control region, $tRNA^{Phe}$, and the remaining 5′ portion of the 12S rRNA gene (Figure 1B). This sequence corresponds exactly with a concatenated sequence between positions 275–4619 and positions 17191–274 in Figure 2, except for a single base difference: a nucleotide corresponding to a cytidine at position 266 is an adenosine for the 4.6-kbp insert. The control region sequence in this insert is thus surrounded by ND1 and 12S rRNA genes, and this organization is distinct from those around control regions I and II (see Figure 1, A and B). Sequences of the other four clones determined only from *Eco*RI termini were identical to that of the first clone. Moreover, inserts of these five clones were not cut with *Pst*I, suggesting no relation to the 4.6-kbp *Eco*RI fragment ranging from ND5 to 12S rRNA genes.

The recovery of multiple clones having an unexpected sequence necessitated us to confirm the gene organization of akamata mtDNA more carefully, as well as to examine heterogeneity of mtDNA sequences both between and within individuals. These problems were addressed by Southern hybridization experiments, and the results are presented in the next two sections. The origin of the unexpected 4.6-kbp fragment is discussed in the section titled *Minor mtDNA components* (see below).

Confirmation of the gene organization and its stability within species by Southern hybridization: In the Southern hybridization experiments shown in Figure 6, purified mtDNAs from eight akamata individuals (Okinawa 1–4, Amami 1–2, and Toku 1–2) were used, and one of them (Okinawa 2) corresponds to an individual from which the complete mtDNA sequence (Figure 2) was determined. In Figure 6A, mtDNA purified from Okinawa 4 was cut with various restriction enzymes and hybridized with two kinds of probes, as described in materials and methods. The sizes of mtDNA fragments generated by the enzyme digestions, as well as the profile of hybridized bands with the probes A and B, corresponded well with expectations from the physical map (Figure 1A) based on the complete mtDNA sequence (see legend of Figure 6). In addition, this digestion pattern for mtDNA of Okinawa 4 was in agreement with the pattern for mtDNA of Okinawa 2 (data not shown) that had been obtained before its cloning and sequencing. Because probe B matches a part of the control region sequence, its hybridization to two distinct bands is direct evidence for the existence of duplicate control regions within the snake mtDNA.

When purified mtDNAs from eight individuals were electrophoresed on 0.4% agarose gel (Figure 6B-a), they were separated into a closed circular form with a higher mobility on the gel and an open circular form with a lower mobility. The proportion of the two bands varied from sample to sample, presumably because of the variation in intactness of mtDNA samples stored at $-30°$ for different periods. Heterogeneous DNAs with the lowest mobility were apparent in some samples (*e.g.*, lane 11 of Figure 6B-a); however, they do not represent heterogeneity of mtDNA because neither probe A nor probe B hybridize with them (Figure 5, B-b and B-c). They just represent contaminating nuclear DNAs.

All the mtDNAs from eight individuals were cut with a unique *Pst*I site, giving rise to a ~17-kbp fragment (Figure 6C-a). When digested with *Eco*RI, mtDNAs from Okinawa 1–4 were cut into 12.6- and 4.6-kbp fragments (Figure 6D-a). mtDNAs from individuals of the Amami Islands (Toku 1–2 and Amami 1–2) were also cut with *Eco*RI into two major fragments, but of distinct sizes (~9.5 and ~7 kbp) from those of Okinawajima Island (Okinawa 1–4). To account for the difference in size of restriction fragments between the two mtDNA haplotypes by base substitutions, at least two substitutions (a loss and a gain of *Eco*RI site) need to be assumed. This suggests a considerably smaller genetic distance between akamata populations of the two Amami Islands than that between these populations and the Okinawajima population, which is in qualitative agreement with the deduced geohistorical relationship among the three islands (Ota 1986). As shown in Figure 6D-c, probe B hybridizes with both of the *Eco*RI-cut fragments for all eight individuals, irrespective of their geographic origin. This indicates that the duplicate state of the control region is not a transient or unstable feature found in a particular individual, but that it stably occurs in the mitochondrial genomes of the species.

These results validate the gene organization of akamata mtDNA presented in Figure 1A. No distinct length polymorphism was apparent in major mtDNA molecules among the eight individuals tested, although extensive length polymorphism has been known to occur in mtDNAs of some lizards, amphibians, and fish at both the intra- and interindividual levels (Moritz and Brown 1986, 1987; Wallis 1987; Zevering *et al.* 1991; Stanton *et al.* 1994; Gach and Brown 1997). The duplicate state of the control region was previously shown

to be maintained among representative individuals of several snake species (Kumazawa *et al.* 1996). On the basis of this solid evidence, the duplication of control regions in snake mtDNAs may be regarded as an unprecedented example of stable functional redundancy in animal mtDNA.

Minor mtDNA components: As outlined earlier, we unexpectedly cloned a 4.6-kbp *Eco*RI-cut fragment that contains the control region sequence surrounded by ND1 and 12S rRNA genes (Figure 1B). However, the two control regions shown in Figure 1A are surrounded either by cytb and 12S rRNA genes or by ND1 and ND2 genes. If the gene organization of Figure 1A is unambiguous, from where did the 4.6-kbp fragment originate? It is evident that even though the 4.6-kbp fragment is present in mtDNA samples from Okinawa 2, it is not a part of the major 17.2-kbp mtDNA molecule because probe A does not hybridize with a major 4.6-kbp fragment generated by *Eco*RI-digestion of the Okinawa 2 mtDNA (see lane 10 of Figure 6D-b). This 4.6-kbp fragment actually corresponds to a sequence from ND5 to 12S rRNA genes, as judged from its further digestion into smaller pieces (4.0 and 0.6 kbp) with *Pst*I (data not shown).

However, much longer exposure of the same filter to an X-ray film gave rise to a faint but discrete signal of ∼4.6 kbp in size, as indicated by the arrowhead in Figure 6D-d. This signal was clearly detected in Okinawa 1–2, only weakly detected in Okinawa 3, and not detectable in Okinawa 4, Amami 1–2, and Toku 1–2. This signal is not caused by nonspecific hybridization of probe A to the 4.6-kbp major fragment ranging from ND5 to 12S rRNA genes because the size of the faintly hybridized signal (∼4.6 kbp) was unchanged, even when mtDNAs cut with *Eco*RI and *Pst*I, with which the 4.6-kbp major fragment is divided into 4.0- and 0.6-kbp fragments, were used for the hybridization (data not shown). Furthermore, no hybridization with an excessive amount of marker DNAs in Figure 6 supports that nonspecific hybridization is generally negligible under the conditions used. It is, therefore, suggested that mtDNA samples from some individuals of Okinawajima Island origin contain, as a heteroplasmic state, a minor component that gave rise to the unexpected 4.6-kbp fragment cloned into *Eco*RI-cut plasmids.

A much longer exposure time was also applied to filters for uncut (Figure 6B-d) and *Pst*I-cut (Figure 6C-d) mtDNA samples. In lanes 10 (Okinawa 2) and 11 (Okinawa 1) of Figure 6B-d, major signals for both closed and open circular bands are accompanied by faint signals with slightly decreased mobilities, as indicated by arrowheads. These faint signals also appeared when a filter hybridized with probe B (Figure 6B-c) was exposed for an extended period (data not shown). In lanes 10 and 11 of Figure 6C-d, a major signal corresponding to the 17.2-kbp linear form of mtDNA is followed by a faint signal with a lower mobility, which again showed up in prolonged exposure of the filter shown in Figure 6C-c (data not shown). Because the individuals from which faint signals appeared or did not appear are well correlated among Figure 5, B-d, C-d, and D-d, it is reasonable to assume that the heteroplasmic 4.6-kbp *Eco*RI-cut fragment is derived from the larger mtDNA component. Unfortunately, the content of the heteroplasmic mtDNA relative to the major mtDNA molecules is too low (close to the detection limit by the hybridization) for the larger mtDNA molecules to be fully characterized. In this regard, the heteroplasmy in akamata mtDNAs is substantially different from previous examples of heteroplasmic mtDNA length polymorphism with a much higher content of larger mtDNA molecules (*e.g.*, see Moritz and Brown 1987; Gach and Brown 1997).

Molecular mechanisms of the concerted evolution: To account for the mechanism of the concerted evolution of control regions, two possible models are presented in this study (Figure 7). One is a tandem duplication model that operates during replication (Figure 7A). A region surrounded by two control regions is tandemly duplicated via slipped-strand mispairing (Moritz and Brown 1986, 1987) or other mechanisms using two identical control region sequences, giving rise to either a larger mtDNA with three control regions or a normal-sized mtDNA with two homogeneous control region sequences (see legend of Figure 7). In the other model, sequences of two control regions can be homogenized by a mechanism such as frequent gene conversion (Figure 7B). The crossing over of nicked strands between two control regions within a mtDNA molecule leads to formation of a Holliday structure, and a sequence of one control region (and possibly of flanking tRNA genes as well) may be replaced by that of the other via repair of heteroduplex DNA intermediates (Figure 7B). For such a mechanism to occur, DNA recombination activities are necessary in snake mitochondria. Homologous DNA recombination activity, which has been thought to be inactive in animal mitochondria, is now known to exist in mammalian (Thyagarajan *et al.* 1996) and nematode (Lunt and Hyman 1997) mitochondria.

These two models were assessed by Southern hybridization experiments (Figure 6). If the second model were to operate, two minicircle mtDNAs with a single control region (12.6 and 4.6 kbp, respectively) might be generated as alternative products of the homologous DNA recombination (Figure 7B). As shown in Figure 6B-d, however, there is no hybridization signal for the 4.6-kbp minicircle in the reasonable position of a gel; a closed circular DNA being 4.6 kbp in size should appear between two marker bands of 4.4 and 2.3 kbp. This indicates that the unexpectedly cloned 4.6-kbp fragment with *Eco*RI termini did not originate from its circular form. On the other hand, the first model predicts the heteroplasmic existence of larger mtDNA molecules generated by tandem duplication (Figure 7A).

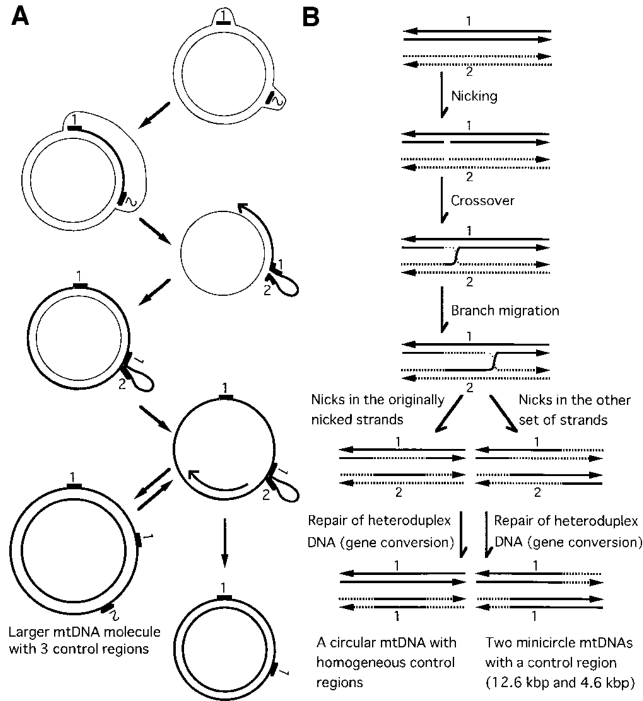

Figure 7.—Models for the concerted evolution or sequence homogenization between two control regions. (A) Tandem duplication model. (B) Gene conversion model. Numerals 1 and 2 stand for control regions I and II, respectively. The model in A is based on the well-elucidated asymmetric replication mechanisms of mammalian mtDNAs (Clayton 1992 and references therein) with consideration to peculiar features of the snake mtDNA. Although initiation of heavy-strand replication and subsequent D-loop formation may be possible at both control regions I and II (see text for discussions), only the replication starting from control region II is considered in this model. A newly synthesized heavy strand pausing at control region I with an extended D-loop structure is unwound to reanneal to control region II for *de novo* initiation of heavy-strand synthesis. Light-strand synthesis in the next round of replication could produce either a larger mtDNA with three control regions or a normal-sized mtDNA with two homogeneous control region sequences. The model in B is based on a well-known mechanism for homologous DNA recombination (*e.g.*, see Lewin 1994). Only parts corresponding to two control regions are shown in B. Note, however, that the two control regions have intramolecular connections within a mtDNA molecule.

As outlined earlier, a scanty proportion of larger mtDNA molecules was actually detected, and the unexpectedly cloned 4.6-kbp fragment was thought to be derived from it (Figure 5, B–D). The 4.6-kbp fragment contained a control region sequence that was surrounded by genes for ND1 and 12S rRNA (Figure 1B), an organization that is predicted to arise by tandem duplication of a region between control regions I and II (see Figure 1A).

These observations are in agreement with the tandem duplication model (Figure 7A), although it is uncertain, and may be difficult to prove because of its extremely low content, that the larger mtDNA molecule detected in Figure 6B-d corresponds with that depicted in Figure 7A or its derivatives. At this point in time, the first model (Figure 7A) is favored by the Southern hybridization experiments, but the second model (Figure 7B) is not excluded. Questions still remain as to the frequency of sequence homogenization between two paralogous control regions and whether the replicational error mechanism that operates for only a minor proportion of mtDNAs in a cell can reasonably account for it (Figure 7A). Biochemical studies for mechanisms of replication and transcription in snake mitochondria, as well as further genetic studies on evolution of snake control region sequences, are expected to provide clues to further unravel the mystery in the concerted evolution.

We thank H. Sato and S. Iwanaga for helping to access part of the akamata samples. We also thank M. Yamaguchi for kind assistance in analyzing a part of the mtDNA sequences. Gratitude is extended to S. Yokobori for providing comments to the original version of manuscript. This work was supported by grants from the Ministry of Education, Science, Sports and Culture of Japan to Y.K. (07740583 and 09214102), H.O. (C-09839024), and M.N. (10660189).

LITERATURE CITED

Adachi, J., and M. Hasegawa, 1996 Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. **42:** 459–468.

Adachi, J., Y. Cao and M. Hasegawa, 1993 Tempo and mode of mitochondrial DNA evolution in vertebrates at the amino acid sequence level: rapid evolution in warm-blooded vertebrates. J. Mol. Evol. **36:** 270–281.

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. de Bruijn, A. R. Coulson *et al.*, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

Anderson, S., M. H. L. de Bruijn, A. R. Coulson, I. C. Eperon, F. Sanger *et al.*, 1982 Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J. Mol. Biol. **156:** 683–717.

Arnason, U., A. Gullberg and B. Widegren, 1991 The complete nucleotide sequence of the mitochondrial DNA of the fin whale, *Balaenoptera physalus.* J. Mol. Evol. **33:** 556–568.

Asakawa, S., Y. Kumazawa, T. Araki, H. Himeno, K. Miura *et al.*, 1991 Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. J. Mol. Evol. **32:** 511–520.

Asakawa, S., H. Himeno, K. Miura and K. Watanabe, 1995 Nucleotide sequence and gene organization of the starfish *Asterina pectinifera* mitochondrial genome. Genetics **140:** 1047–1060.

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.

Carroll, R. L., 1988 Vertebrate paleontology and evolution. W. H. Freeman, New York.

Christianson, T. W., and D. A. Clayton, 1988 A tridecamer DNA sequence supports human mitochondrial RNA 3′-end formation in vitro. Mol. Cell. Biol. **8:** 4502–4509.

Clayton, D. A., 1992 Transcription and replication of animal mitochondrial DNA. Int. Rev. Cyt. **141:** 217–232.

Desjardins, P., and R. Morais, 1990 Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. J. Mol. Biol. **212:** 599–634.

Doda, J. N., C. T. Wright and D. A. Clayton, 1981 Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA **78:** 6116–6120.

Drouin, J., 1980 Cloning of human mitochondrial DNA in *Escherichia coli.* J. Mol. Biol. **140:** 15–34.

Gach, M. H., and W. M. Brown, 1997 Characteristics and distribution of large tandem duplications in brook stickleback (*Culaea inconstans*) mitochondrial DNA. Genetics **145:** 383–394.

Gadaleta, G., G. Pepe, G. de Candia, C. Quagliariello, E. Sbisà *et al.*, 1989 The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. J. Mol. Evol. **28:** 497–516.

Gelfand, R., and G. Attardi, 1981 Synthesis and turnover of mito-

chondrial ribonucleic acid in Hela cells: the mature ribosomal and messenger ribonucleic acid species are metabolically unstable. Mol. Cell. Biol. **1:** 497–511.

Härlid, A., A. Janke and U. Arnason, 1997 The mtDNA sequence of the ostrich and the divergence between paleognathous and neognathous birds. Mol. Biol. Evol. **14:** 754–761.

Hedges, S. B., 1994 Molecular evidence for the origin of birds. Proc. Natl. Acad. Sci. USA **91:** 2621–2624.

Janke, A., and U. Arnason, 1997 The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent Archosauria (birds and crocodiles). Mol. Biol. Evol. **14:** 1266–1272.

Janke, A., G. Feldmaier-Fuchs, W. K. Thomas, A. von Haeseler and S. Pääbo, 1994 The marsupial mitochondrial genome and the evolution of placental mammals. Genetics **137:** 243–256.

Janke, A., N. J. Gemmell, G. Feldmaier-Fuchs, A. von Haeseler and S. Pääbo, 1996 The mitochondrial genome of a monotreme—the platypus (*Ornithorhynchus anatinus*). J. Mol. Evol. **42:** 153–159.

Kumazawa, Y., and M. Nishida, 1993 Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. J. Mol. Evol. **37:** 380–398.

Kumazawa, Y., and M. Nishida, 1995 Variations in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. Mol. Biol. Evol. **12:** 759–772.

Kumazawa, Y., H. Ota, M. Nishida and T. Ozawa, 1996 Gene rearrangements in snake mitochondrial genomes: highly concerted evolution of control-region-like sequences duplicated and inserted into a tRNA gene cluster. Mol. Biol. Evol. **13:** 1242–1254.

Lee, W.-J., and T. D. Kocher, 1995 Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. Genetics **139:** 873–887.

Lee, W.-J., J. Conroy, W. H. Howell and T. D. Kocher, 1995 Structure and evolution of teleost mitochondrial control regions. J. Mol. Evol. **41:** 54–66.

Lewin, B., 1994 *Genes V.* Oxford University Press, New York.

Lunt, D. H., and B. C. Hyman, 1997 Animal mitochondrial DNA recombination. Nature **387:** 247.

Macey, J. R., A. Larson, N. B. Ananjeva, Z. Fang and T. J. Papenfuss, 1997 Two novel gene orders and the role of light-strand replication in rearrangement of the vertebrate mitochondrial genome. Mol. Biol. Evol. **14:** 91–104.

Macey, J. R., J. A. Schulte II, A. Larson and T. J. Papenfuss, 1998 Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol. Biol. Evol. **15:** 71–75.

Martin, A. P., and S. R. Palumbi, 1993 Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. USA **90:** 4087–4091.

Mita, S., R. J. Monnat, Jr. and L. A. Loeb, 1988 Direct selection of mutations in the human mitochondrial $tRNA^{Thr}$ gene: reversion of an 'uncloneable' phenotype. Mutat. Res. **199:** 183–190.

Moritz, C., and W. M. Brown, 1986 Tandem duplications 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.

Nei, M., 1991 Relative efficiencies of different tree-making methods for molecular data, pp. 90–128 in *Phylogenetic Analysis of DNA Sequences*, edited by M. M. Miyamoto and J. Cracraft. Oxford University Press, New York.

Ojala, D., J. Montoya and G. Attardi, 1981 tRNA punctuation model of RNA processing in human mitochondria. Nature **290:** 470–474.

Ota, H., 1986 A review of reptiles and amphibians of the Amami Group, Ryukyu Archipelago. Mem. Fac. Sci., Kyoto Univ. **11:** 57–71.

Ota, T., and M. Nei, 1994 Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. J. Mol. Evol. **38:** 642–643.

Pääbo, S., D. M. Irwin and A. C. Wilson, 1990 DNA damage promotes jumping between templates during enzymatic amplification. J. Biol. Chem. **265:** 4718–4721.

Pääbo, S., W. K. Thomas, K. M. Whitfield, Y. Kumazawa and A. C. Wilson, 1991 Rearrangements of mitochondrial transfer RNA genes in marsupials. J. Mol. Evol. **33:** 426–430.

Quinn, T. W., and D. P. Mindell, 1996 Mitochondrial gene order adjacent to the control region in crocodile, turtle, and tuatara. Mol. Phylogenet. Evol. **5:** 344–351.

Quinn, T. W., and A. C. Wilson, 1993 Sequence evolution in and around the mitochondrial control region in birds. J. Mol. Evol. **37:** 417–425.

Rage, J.-C., 1987 Fossil history, pp. 51–76 in *Snakes: Ecology and Evolutionary Biology*, edited by R. A. Seigel, J. T. Collins and S. S. Novak. Macmillan, New York.

Roe, B. A., D.-P. 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.

Seutin, G., B. F. Lang, D. P. Mindell and R. Morais, 1994 Evolution of the WANCY region in amniote mitochondrial DNA. Mol. Biol. Evol. **11:** 329–340.

Stanton, D. J., L. L. Daehler, C. C. Moritz and W. M. Brown, 1994 Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. Genetics **137:** 233–241.

Strimmer, K., and A. von Haeseler, 1996 Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13:** 964–969.

Tapper, D. P., R. A. Van Etten and D. A. Clayton, 1983 Isolation of mammalian mitochondrial DNA and RNA and cloning of the mitochondrial genome. Methods Enzymol. **97:** 426–434.

Thomas, W. K., and A. T. Beckenbach, 1989 Variation in salmonid mitochondrial DNA: evolutionary constraints and mechanisms of substitution. J. Mol. Evol. **29:** 233–245.

Thyagarajan, B., R. A. Padua and C. Campbell, 1996 Mammalian mitochondria possess homologous DNA recombination activity. J. Biol. Chem. **271:** 27536–27543.

Tzeng, C.-S., C.-F. Hui, S.-C. Shen and P. C. Huang, 1992 The complete nucleotide sequence of the *Crossostoma lacustre* mitochondrial genome: conservation and variations among vertebrates. Nucleic Acids Res. **20:** 4853–4858.

Walberg, M. W., and D. A. Clayton, 1981 Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res. **9:** 5411–5421.

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

Wolstenholme, D. R., 1992 Animal mitochondrial DNA: structure and evolution. Int. Rev. Cyt. **141:** 173–216.

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.

Yokobori, S., and S. Pääbo, 1995a Transfer RNA editing in land snail mitochondria. Proc. Natl. Acad. Sci. USA **92:** 10432–10435.

Yokobori, S., and S. Pääbo, 1995b tRNA editing in metazoans. Nature **377:** 490.

Yoneyama, Y., 1987 The nucleotide sequences of the heavy and light strand replication origins of the *Rana catesbeiana* mitochondrial genome. J. Nippon Med. Sch. **54:** 429–440 (in Japanese).

Zardoya, R., and A. Meyer, 1996 The complete nucleotide sequence of the mitochondrial genome of the lungfish (*Protopterus dolloi*) supports its phylogenetic position as a close relative of land vertebrates. Genetics **142:** 1249–1263.

Zardoya, R., and A. Meyer, 1997 The complete DNA sequence of the mitochondrial genome of a "living fossil," the coelacanth (*Latimeria chalumnae*). Genetics **146:** 995–1010.

Zardoya, R., A. Garrido-Pertierra and J. M. Bautista, 1995 The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss.* J. Mol. Evol. **41:** 942–951.

Zevering, C. E., C. Moritz, A. Heideman and R. A. Sturm, 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.

Communicating editor: N. Takahata