Gene Rearrangements in Snake Mitochondrial Genomes: Highly Concerted Evolution of Control-Region-Like Sequences Duplicated and Inserted into an tRNA Gene Cluster

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

*Department of Earth and Planetary Sciences, Nagoya University; †Tropical Biosphere Research Center, University of the Ryukyus; and ‡Department of Marine Bioscience, Fukui Prefectural University

Mitochondrial DNA (mtDNA) regions corresponding to two major tRNA gene clusters were amplified and sequenced for the Japanese pit viper, himehabu. In one of these clusters, which in most vertebrates characterized to date contains three tightly connected genes for tRNA^met, tRNA^ile, and tRNA^asp, a sequence of approximately 1.3 kb was found to be inserted between the genes for tRNA^ile and tRNA^asp. The insert consists of a control-region-like sequence possessing some conserved sequence blocks, and short flanking sequences which may be folded into RNA^ile and RNA^asc genes. Several other snakes belonging to different families were also found to possess a control-region-like sequence and tRNA^met gene between the tRNA^ile and tRNA^asp genes. We also sequenced a region surrounded by genes for cytochrome b and 12S rRNA, where the control region and genes for tRNA^asp and tRNA^met are normally located in the mtDNAs of most vertebrates. In this region of three examined snakes, a control-region-like sequence exists that is almost completely identical to the one found between the tRNA^ile and tRNA^asc genes. The mtDNAs of these snakes thus possess two nearly identical control-region-like sequences which are otherwise divergent to a large extent between the species. These results suggest that the duplicate state of the control-region-like sequences has long persisted in snake mtDNAs, possibly since the original insertion of the control-region-like sequence and tRNA^met gene into the tRNA gene cluster, which occurred in the early stage of the divergence of snakes. It is also suggested that the duplicated control-region-like sequences at two distant locations of mtDNA have evolved concertedly by a mechanism such as frequent gene conversion. The secondary structures of the determined tRNA genes point to the operation of simplification pressure on the T + C arm of snake mitochondrial tDNAs.

Introduction

Complete mitochondrial DNA (mtDNA) sequences have been determined for a number of vertebrates, representing mammals, birds, amphibians, and fishes (Lee and Kocher 1995; Aronson, Xu, and Gullberg 1996; Ianke et al. 1996 and references therein). Although many partial mtDNA sequences from a variety of animal groups are known, relatively few have been reported from reptiles. The vertebrate mtDNAs characterized to date are double-stranded, circular DNAs of ~16 kb and encode genes for 13 proteins, 2 rRNAs, and 22 tRNAs, as well as having a major noncoding or control region that contains signals for replicating the heavy strand of mtDNA and for transcription (Anderson et al. 1981; Wolstenholme 1992). Some conserved sequence blocks (CSBs) have been identified by comparing control region sequences from a number of vertebrates (see, e.g., Walberg and Clayton 1981; Lee et al. 1995), although some vertebrates lack certain CSB members (Desjardins and Morais 1990; Quinn and Wilson 1993), opossum (Pläibo et al. 1991; Janke et al. 1994), sea lamprey (Lee and Kocher 1995), crocodilians (Seutin et al. 1994; Kumazawa and Nishida 1995; Quinn and Mindell 1996), and the Texas blind snake (Kumazawa and Nishida 1995) indicate that the gene organization of vertebrate mtDNAs is more variable than previously thought. Transfer RNA genes are involved in most cases of such gene rearrangement.

In our previous work (Kumazawa and Nishida 1993, 1995), we thus focused on the tRNA genes which are encoded in the mitochondrial genome as clusters. We designed a set of polymerase chain reaction (PCR) primers based on the conservative regions in each of the protein genes that surround the three major tRNA gene clusters (IQM, WANCY, and HSL)—the tRNA genes are abbreviated by single letters representing the amino acids to be decoded, and the sense strand of the underlined tRNA genes is the heavy strand). These primers proved useful in amplifying the corresponding regions for a wide range of tetrapods (Kumazawa and Nishida 1993, 1995) and enabled us to find tRNA gene rearrangements in crocodilians and the Texas blind snake (Kumazawa and Nishida 1995). The former represents a local rearrangement of tRNA genes within a cluster (HSL to SHL) whereas the latter involves transposition of the tRNA^met gene from the IQM to the WANCY cluster.

During the course of this work, we noticed that much larger products than expected were amplified from the IQM region of several snakes representing the families Viperidae, Colubridae, and Boidae, which suggest-
ed the existence of another novel tRNA gene rearrangement widely present in snakes. This finding prompted us to characterize the large amplified products from these snakes. In this communication, we report on the nucleotide sequences of a long insertion into the IQM cluster of the mtDNAs, together with those of the WANCY and cytochrome b (cytb)-12s rRNA regions. An unexpected finding emerged which revealed that snake mtDNAs possess duplicated control-region-like (CR-like) sequences that are nearly identical to each other within the species. We discuss the implications of our findings as to when and how the original duplication of the CR-like sequences occurred in the ancestral lineage of snakes and how the duplicated CR-like sequences have evolved during the evolution of snakes.

Materials and Methods

Genomic DNAs were extracted from muscle tissues of snakes according to Kocher et al. (1989). The snake species used in this study were the vipers himehabu (Ovophis okinavensis, an endemic pit viper of the Ruykyus Islands of Japan) and western rattle snake (Crotalus viridis), the boids ball python (Python regius) and boa constrictor (Boa constrictor), and the colubrids akamata (Dinodon semicarinatus) and gopher snake (Pituophis melanoleucus). The PCR primers used were mostly derived from previous work: L4160m and H5937m (Kumazawa and Nishida 1993), and H4433, H4614, and L5038 (Kumazawa and Nishida 1995). New PCR primers used for the present study were: L4437b, CAGCTAAAAAAGCTATCGGGCCCATACC (5'→3'); L14940, ATTAACCTAGCTTCTCTCATC; L14973, CACATCACYCGAGATGTCCCCTACGG; H690, GTTAGGCTTGCATAGCACTG; OOK-9, GTCTGTAGCTTAAGCCTAACTAGTATAGC; OOK-12, ATTAGGTGTTTGGCTTGCTTAAGG; and OOK-16, AGAGTAGCTTTGAGTAAATGCTGGC. For the primers L4437b, L14940, L14973, and H690, L or H indicates the primer direction and the accompanying number refers to the 3' end position according to human mtDNA (Anderson et al. 1981). The matching sites of all the primers to the snake mtDNAs are indicated in figure 1. A number of sequencing primers were also synthesized to determine the CR-like sequences by primer walking (sequences not shown).

PCR-aided direct sequencing of the WANCY cluster was carried out as described previously (Kumazawa and Nishida 1995). Briefly, double-stranded PCR products were amplified from the genomic DNA with pairs of primers and Taq polymerase (Takara Shuzo Co.) and subjected to asymmetric PCR. PCR reactions consisted of 30 cycles of denaturation at 92°C for 40 s, annealing at 50–55°C for 1 min, and extension at 72°C for 1–3 min. The single-stranded PCR products were directly sequenced with the amplification primers and appropriate internal primers by the dideoxy chain termination method with [α-35S]dATP and SEQUENASE version 2 (Amersham).

Sequences for the other regions were determined in a different manner. Long (>1.5 kb) DNA fragments were amplified from either the total genomic DNA or the nearly complete (~16 kb) mtDNA template obtained by Long and Accurate PCR (LA PCR) (reviewed in Cheng et al. 1994) using an LA PCR kit version 2 (Takara). The condition for LA PCR was 30 cycles of denaturation at 94°C for 30 s and simultaneous annealing and extension at 68°C for 15 min. The long DNA fragments were cloned into the pUC118 vector of E. coli.
and relevant sequences of multiple independent clones were determined with an Applied Biosystems 373A DNA sequencer using the primer walking strategy. Nucleotide sequences were determined unambiguously by sequencing both strands.

Transfer RNA genes in the determined sequences were searched in the light of their secondary structure (Kumazawa and Nishida 1993). Protein and rRNA genes were identified by the criterion of sequence similarity with the corresponding genes from other vertebrates. All the nucleotide sequences determined will appear in the GSDB, DDBJ, EMBL, and NCBI databases with accession numbers D84255–D84261 and D86118–D86120.
Results

WANCY Cluster

The WANCY region of himehabu was amplified by using primers L5038 and H5937m, whose sequences match the conservative portions of the NADH dehydrogenase subunit 2 (ND2) gene and cytochrome oxidase subunit I (COI) gene, respectively (see fig. 1). PCR amplification gave rise to a discrete product of reasonable size (~900 bp), which was then subjected to asymmetric PCR and sequenced with the help of several of the internal primers used in our previous work (Kumazawa and Nishida 1995). The sequence determined (data not shown) showed that himehabu conserves the usual vertebrate organization of the tRNA genes in the WANCY cluster. The characteristic stem-and-loop structure for the putative light-strand replication origin (Wong and Clayton 1985) located between the tRNAAsn and tRNA^CyS genes, shown to be absent in birds, crocodilians, tuatara, and the Texas blind snake (Desjardins and Morris 1990; Quinn and Wilson 1993; Seutin et al. 1994; Kumazawa and Nishida 1995), can be easily identified at the corresponding position in himehabu (fig. 1). Thus, no feature particularly distinctive from the typical vertebrate organization appears to be present in the WANCY region of himehabu. The secondary structures of these tRNA genes are shown in figure 2 and their structural features are discussed below.

IQM-Related Region

The IQM region was initially amplified by the primers L4T60m and H4433, which respectively match conservative portions of the ND1 and tRNA^Met genes. Although this amplification should give rise to a product approximately 300 bp in length for the normal ND1-IQ-M gene organization, a discrete, long product was consistently amplified from several species belonging to different snake families (fig. 3). The length of the product was 1.4–1.6 kb in himehabu and western rattlesnake (family Viperidae), akamata and gopher snake (Colubridae), and ball python (Boaidae), and longer (~2.2 kb) in boa constrictor (Boaidae). These findings suggested the existence of a novel gene rearrangement involving the IQM region of mtDNAs from diverse phylogenetic groups of snakes.

We decided to characterize the large amplified product from himehabu. Because amplification by primers L4160m and H4614, the latter of which matches a part of the ND2 gene, gave rise to a product approximately 1.8 kb in length (data not shown), this product was selected as the sequencing target. The 1.8-kb DNA fragment was cloned into an E. coli pUC118 vector and multiple independent clones were sequenced by the primer walking strategy. Figure 4 shows the nucleotide sequence determined and figure 1 depicts the inferred gene organization. Flanked by regions that have an appreciable level of sequence similarity with the ND1 and ND2 genes from the Texas blind snake (Kumazawa and Nishida 1995) and other vertebrates, three tRNA genes specifying isoleucine, glutamine, and methionine can be easily identified.

A long insertion of 1,329 bp is found between the tRNA^Ile and tRNA^Gln genes (figs. 1 and 4). A sequence which may be folded into a tRNA^Pro gene exists immediately 3' downstream of the tRNA^Ile gene. In addition, a possible tRNA^Leu(UUR) gene occurs immediately 5' upstream of the tRNA^Gln gene. In the middle of the sequenced 1.8-kb DNA there is also a sequence which may be folded into a tRNA^Phe gene. No protein or RNA gene could be identified in the regions between the tRNA^Pro and tRNA^Phe genes, or between the tRNA^Phe and tRNA^Leu genes. Instead, structural features that are characteristic of a control region occur between the tRNA^Pro and tRNA^Phe genes (fig. 4). This region includes sequences that have clear sequence similarity to CSB-1 and CSB-3 in the control regions of other vertebrates (fig. 5). Although a structure equivalent to CSB-2 cannot be identified between CSB-1 and CSB-3, a C-rich sequence which has high similarity to the putative CSB-2 sequence for teleosts (Lee et al. 1995) is found near the tRNA^Pro gene (fig. 4). Also notable in this noncoding region are hairpin-like structures and repetitive sequences, both of which are often present in the control region. On the basis of these various features, we can regard the noncoding region between the tRNA^Pro and tRNA^Phe genes to be a CR-like sequence. On the other hand, as far as we can deduce, no functional feature exists in the noncoding region between the tRNA^Phe and tRNA^Leu genes, where repetition of two kinds of pentamers seems to occur (see fig. 4).

We also cloned and sequenced the 1.4–1.5-kb PCR products amplified from western rattlesnake (fig. 3, lane
Snake mtDNA Evolution 1247

diamond in A is found to be a guanosine in B; in western rattle snake, the thymidine indicated by an unfilled diamond in A is deleted in B; and in akamata, the thymidine indicated by a filled triangle in A is a cytidine in B. At the left and right junctions of the nearly identical sequence units between the IQM-related and cytb-12s rDNA regions, circled numbers are placed for (1) himehabu, (2) western rattle snake, and (3) akamata. Anticodons of tRNA genes, as well as sequences which may form hairpin structures, are underlined. Repetitive sequence units in noncoding spacer regions are double-underlined. Refer to the nucleotide sequences deposited in the database for more sequence information for parts of ND1, ND2, and cytb genes, as well as for the partially determined IQM-related region sequences of boa constrictor. Species abbreviations: Ook, himehabu (O. okinavensis); Cvi, western rattle snake (Crotalus viridis); Dse, akamata (D. semicarinatus); and Pre, ball python (P. regius).
6), akamata (fig. 3, lane 8), and ball python (fig. 3, lane 3). As shown in figure 4, these products contain a CR-like sequence and tRNA^{Leu(UUR)} gene between the tRNA^{Ile} and tRNA^{Gln} genes. A complete tRNA^{Pro} gene found in himehabu was present only in the western rattle snake, but a partially truncated tRNA^{Pro} gene was found in the corresponding location of akamata (fig. 4). The slightly smaller size of the amplified products from these species as compared with that from himehabu (fig. 3) turned out to be mostly due to the absence of the tRNA^{Phe} gene and its flanking non-coding sequence in this region (figs. 1 and 4). The CR-like sequences for these species conserve CSB-1 and CSB-3 sequences (fig. 5) and CSB-2-like C-rich sequences (fig. 4), although ball python may not have a distinct CSB-3 sequence. These CR-like sequences have clear sequence similarity to that found in himehabu. The pairwise sequence divergence in a clearly alignable portion of the CR-like sequences (i.e., from hairpin 3 to CSB-3 of fig. 4) increased in an order consistent with the phylogenetic relationships among the snakes (e.g., Rage 1987: Zug 1993, pp. 443–467; Heise et al. 1995): 13% between himehabu and the western rattle snake, 24% on average between vipersids and akamata, and 39% on average between vipersids-colubrids and ball python. The PCR product from boa constrictor (~2.2 kb; fig. 3, lane 2) was only partially sequenced to show the existence of genes for tRNA^{Ile}, tRNA^{Leu(UUR)}, and tRNA^{Gln} and the CR-like sequence (data not shown; refer to the sequences deposited in the database with accession numbers D84260 and D84261).

Cytb-12S rRNA Region

The finding of the long insertion in the IQM cluster region of several snakes consisting of a CR-like sequence and some flanking tRNA genes prompted us to sequence the region surrounding the cyt b and 12S rRNA genes, where the control region and three tRNA genes specifying threonine, proline, and phenylalanine are encoded in the mtDNAs of most vertebrates characterized to date (see fig. 1). For this purpose, we first sequenced a short segment of the cyt b gene of each snake species, which can be amplified and sequenced in combination with the well-conserved primers L14841 and H15149 (see Kocher et al. 1989). Primers, named L14940 and L14973, were then synthesized so as to match the determined short cyt b sequences (data not shown) for himehabu and for the western rattle snake and akamata, respectively (see fig. 1). Another primer, H690, was also synthesized so as to match a well-conserved region of the vertebrate 12S rRNA gene in proximity to the 5' end of the gene. Amplifications with L14940 and H690 from himehabu genomic DNA produced a discrete band approximately 2.1 kb long (fig. 6, lane 4). This single band was also seen by LA PCR with the same template and primers (data not shown). The 2.1 kb fragment was cloned into an E. coli plasmid vector and multiple clones were sequenced by primer walking (fig. 4).

The 5' and 3' end portions of the 2.1-kb fragment have clear sequence similarity with the corresponding regions of the cyt b and 12S rRNA genes of other vertebrates, respectively. In the 3' end portion, a well-conserved hairpin structure which is characteristic of 12S
rRNA genes is also found (fig. 4). It is therefore unlikely
that unrelated regions were amplified by the PCR. Im-
mediately 3' downstream of the cytb gene, a sequence
which may be folded into a tRNA^{Thr} gene is found. Next
to the tRNA^{Thr} gene there is a truncated tRNA^{Pro} gene
(fig. 4) which has a sequence identical to the 5' half of
the putative tRNA^{Pro} gene found in the IQM-related
region (fig. 4). Because the truncation disrupts even the
anticodon loop, it is very unlikely that a functional
tRNA molecule can be produced from the truncated
tRNA^{Pro} gene. Rather, it most likely represents a pseudo-
dogene, which may have been generated in association
with gene conversion or duplication-and-deletion events,
as discussed below. A mitochondrial tRNA pseudogene
which may have been generated by the duplication-and-
deletion mechanism has been found in a parthenogenetic
lizard (Zevering et al. 1991).

Immediately 5' upstream of the 12S rRNA gene,
there is a sequence which may be folded into a tRNA^{Phe}
gene (fig. 4). This sequence is nearly identical to an alter-
native tRNA^{Phe} gene found in the IQM-related re-
gion, but lacks the sequence from the T stem to the
acceptor stem of the alternative gene (see fig. 2). As
shown in figure 2, secondary structures for both of the
tRNA^{Phe} genes are similarly unstable in the acceptor-
and T-stems, although truncation of the T-stem is com-
mon among many snake mitochondrial tRNAs (see Dis-
cussion). Comparison of the acceptor-stem sequences
with tRNA^{Phe} genes from the western rattle snake and
akamata (see fig. 4) favors the idea that the gene found
in the cytb-12S rRNA region may be the primary t-
RNA^{Phe} gene for himehabu. However, the possibility
that the alternative gene also encodes a functional tRNA
cannot be excluded. The question of which (or possibly
it is both) of these tRNA^{Phe} genes actually functions in
snake mitochondria can only be answered by biochem-
ical characterization of tRNA^{Phe} expressed in the mito-
chondria.

A long sequence (1,127 bp) between the truncated
tRNA^{Pro} and tRNA^{Phe} genes proved to be nearly iden-
tical to the sequence between the tRNA^{Pro} and tRNA^{Phe}
genes found in the IQM-related region, with the excep-
tion of a single A \rightarrow G base difference at the immediate
5' residue of the tRNA^{Phe} gene (refer to the legend of
fig. 4). This means that almost the same CR-like se-
quence appears to be present in both the IQM-related
and cytb-12S rRNA regions in himehabu (see fig. 1).

The cytb-12S rRNA regions were also amplified
and sequenced with L14973 and H690 from western
rattle snake and akamata. As shown in figures 1 and 4,
these snakes possess a CR-like sequence and genes for
tRNA^{Thr} and tRNA^{Phe} in this region as himehabu does.
The western rattle snake, which has a complete tRNA^{Pro}
gene in the IQM-related region, lacks another tRNA^{Pro}
gene between the tRNA^{Thr} gene and the CR-like se-
quence of the cytb 12S rRNA region, where an appar-
ently noncoding sequence (130 bp) with no particularly
distinctive feature occurs (fig. 4). On the other hand,
akamata, which has a truncated tRNA^{Pro} gene in the
IQM-related region, possesses a complete tRNA^{Pro} gene
in the cytb-12S rRNA region. The CR-like sequence in
this region proved to be almost completely identical to
the one found in the IQM-related region in both the
western rattle snake and akamata. The only differences
in the CR-like sequences between the two regions are
the deletion or insertion of T in the western rattle snake
and a T=C base substitution in akamata (refer to the
legend of fig. 4).

Two CR-like Sequences in MtDNA

In order to obtain further evidence for the existence
of nearly identical CR-like sequences at two different
locations of mtDNA, the following experiments were
carried out for himehabu. Two primers, named OOK-9
and OOK-16, were synthesized to respectively match
the nucleotide sequences determined for the tRNA^{Phe}
and tRNA^{Pro} genes of himehabu (see fig. 1). If the CR-
like sequence flanked by these tRNA genes were pres-
ent at only a single location in a circular mtDNA, LA
PCR amplification with OOK-9 and OOK-16 should
produce a single band covering nearly the entire mtDNA
sequence except for the control region. In fact, LA PCR
gave rise to two clear bands of approximately 4 and 10–
11 kb (fig. 6, lane 8), thus supporting the existence of
duplicate sequences. The annealing temperature used
for LA PCR (68°C) was high enough to make nonspecific
annealing of these primers to unrelated regions unlikely,
though not impossible. Although the sequences between
the IQM-related and cytb-12S rRNA regions have not
been determined, our preliminary PCR amplification ex-
periments using primers designed for the 12S rRNA,
16S rRNA, and ND1 genes suggest that these genes are
encoded in this order in the \sim2-4 kb fragment just as in
other vertebrates (unpublished data).

Two primers, named L4437b and OOK-12, were
then constructed to respectively match the nucleotide
sequences determined for the tRNA^{Met} and tRNA^{Gln} genes
(see fig. 1) for the purpose of obtaining a nearly entire
mtDNA sequence by LA PCR, figure 6 shows a discrete
amplified product of ~16 kb (lane 7). This product was
highly purified by electrophoresis on 0.4% agarose gel
and used as a template for the subsequent amplifications.
The amplification from this template with OOK-9 and
OOK-16 produced a single band (fig. 6, lane 9) which
corresponded to the shorter band (~4 kb) produced from
the total genomic DNA with the same primers (fig. 6,
lane 8), providing strong evidence for the duplication of
the CR-like sequence and flanking tRNAs ~4 kb apart
within the mtDNA molecule. The disappearance of the
10-11 kb band may be reasonably accounted for by the
fact that a circular mtDNA is linearized at the tRNA^{Gln}–
tRNA^{Met} junction for the ~16 kb template. With the lin-
erized ~16-16 kb mtDNA as a template, 1.5- and 2.1-kb
products were able to be amplified with L4160m and
OOK-12, and with L14940 and H690, respectively (fig.
6, lanes 2 and 5). Partial sequencing of these products
from several primer sites revealed no differences from
the sequences that were amplified and sequenced from
the total genomic DNA (data not shown), thus refuting
the possibility that the sequences determined are of nu-
clear origin.
Frequent polymorphic length mutations are known to occur in and around the control region of lizard mtDNAs (see, e.g., Moritz and Brown 1986, 1987; Zervenig et al. 1991). In this respect, it is noteworthy that we have not detected a polymorphic state in himehabu mtDNA; PCR products of the same size were consistently obtained from three individual himehabu with L4160m and H4433, and with L14940 and H690 (data not shown).

**Discussion**

Snake tRNAs with a Truncated TΨC Arm

Metazoan mitochondrial tRNAs are known to frequently lack invariant and seminvariant nucleotides, and even an entire arm (see, e.g., Dirheimer and Martin 1990; Wolstenholme 1992; Kumazawa and Nishida 1993). The most extreme examples are the complete lack of the dihydrouridine arm (D arm) for tRNA^Ser^ specifying AGY codons (Anderson et al. 1981; de Bruijin and Klug 1983), the lack of the position 8 residue and the increased number of anticodon stem pairs for tRNA^Ser^ specifying AGY codons (Yokogawa et al. 1991; Steinberg and Cedergren 1994; Watanabe et al. 1994a), and the apparent lack of the stable stem-and-loop structure in the TΨC arm (T arm) for most tRNAs from nematode mitochondria (Wolstenholme et al. 1987; Watanabe et al. 1994b). Except for these extreme examples, many mitochondrial tRNAs possess a common cloverleaf secondary structure (Kumazawa and Nishida 1993). Figure 2 shows secondary structures of possible tRNA genes found in the himehabu mtDNA sequences. It is striking that many of these tRNAs can have only two or three base pairs in the T stem region. For example, tRNA^Phe^ has only two G-C pairs in the T stem. Judging from the excellent base matching in the acceptor and anticodon stem regions, as well as the shortness of the T loop (only four nucleotides), it is hard to envisage alternative secondary structures for the tRNA^Phe^.

To the best of our knowledge, this type of tRNA with a truncated TΨC arm has not yet accumulated to the simplification pressure. Since it is relatively easy to collect snake tissues in sufficiently large amounts for biochemical characterization, snake mitochondrial tRNAs might provide intriguing subjects for studying the structure-function relationships of RNA molecules, especially as how tRNAs with unusual cloverleaf structures are folded into a functional tertiary structure, and how they have evolved the mechanisms by which they interact with recognition enzymes.

**Hypothetical Gene Rearrangement Mechanism**

As Lee and Kocher (1995) have recently noted, two major mechanisms have been proposed to explain gene rearrangements in mtDNAs. One is the duplication of particular tRNAs. The data imply that as snake mitochondrial tRNAs have evolved, their T arm structure has become somewhat simplified to resemble the nematode tRNAs, but that some snake tRNAs, whose D-loop/T-loop interactions are functionally important, have not yet acceded to the simplification pressure. Since it is relatively

---

**Table 1**

<table>
<thead>
<tr>
<th>tRNA</th>
<th>Snake</th>
<th>Species*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

* Spectra employed: Ook, himehabu (Ophisops ichnolus); Ldu, Texas blind snake (Lampropeltis duclosi); Boo, boa constrictor (Boa constrictor); Mdo, mouse (Mus domesticus); Bta, cow (Bos taurus); Hsa, human (Homo sapiens); Gga, chicken (Gallus gallus); Ala, frog (Xenopus laevis).

**Notes**—For the assignment of the T arm region (T stem plus T loop), refer to figure 4 of this study and Kumazawa and Nishida (1993, 1995). In order to calculate the number of base pairs in the T stem, only Watson-Crick and G·U wobble pairs are considered under the assumption that at least two nucleotides are required to maintain the T loop. Bars mean that the corresponding genes have not been sequenced.

---

**Figure 2**

Secondary structures of possible tRNA genes found in the himehabu mtDNA sequences. The gene rearrangements in birds may well be explained by such a mechanism (Desjardins and Morais 1990; Quinn and Wilson 1993), and its feasibility is also supported by the frequent polymorphic duplications of 1250 Kumazawa et al.
mtDNA sequences found in lizards (Moritz and Brown et al. 1989).

We speculate that the original gene rearrangements, which resulted in the insertion of the CR-like sequence and tRNALeu(UUR) gene into the IQM region, may have occurred in the ancestral lineage of the snakes by the first of the mechanisms outlined above. Assuming that a long DNA fragment ranging from the tRNAPro to t-RNAHc genes of the typical organization (see fig. 1) is duplicated, subsequent deletions of the duplicated portions for tRNAPro-16S rRNA and ND1-tRNAHc would give rise to the rearranged gene organization either represented by boids with additional deletion of the tRNAPro gene or by the western rattle snake without such deletion.

Gene Rearrangement Timing

When did the original gene rearrangements take place? As described above, the insertion of the CR-like sequence and tRNALeu(UUR) gene into the IQM region was commonly found in snakes belonging to the Boidae, Viperidae, and Colubridae families. It is therefore reasonable to infer that the original gene rearrangements that resulted in the insertion of these segments into the IQM cluster took place long ago, i.e., more than 70 Myr as inferred from the minimum divergence time between the Boidae and the Colubridae-Viperidae clade (Rage 1987) (refer to fig. 7). Snakes are phylogenetically derived from a lineage of lizards (see, e.g., Benton 1990; Zug 1993, pp. 438-444), some species of which have been shown to possess the normal IQM cluster like crocodilians, birds, mammals, and fishes (Kumazawa and Nishida 1995). Furthermore, the Texas blind snake, which is considered to belong to the earliest snake lineage (Scolecodophians) (Rage 1987; Zug 1993, pp. 443–467; Heise et al. 1995), has a rearranged organization resulting in an IM cluster (Kumazawa and Nishida 1995). These results suggest that multiple rearrangement events involving the IQM cluster took place early in snake lineages (possibly 70–140 MYA; Rage 1987) (fig. 7).

Duplicate CR-like Sequences

It was an unexpected finding that himehabu, the western rattle snake, and akamata possess almost completely identical CR-like sequences in two locations of their mtDNAs, there being only a single base alteration with respect to their size and gene number, and, as a consequence, repetitive sequences would be likely to be deleted rapidly. This reasoning predicts that the duplicate state of the CR-like sequence may be only transient. In disagreement with this prevalent notion, our data strongly support the long-term persistence of the duplicate state during the evolution of snakes. As described above, the three snake species all conserve the duplicate state of the CR-like sequence. Although we have not sequenced the cytb-12S rRNA region of boids, preliminary PCR amplification experiments using primers L14940 and H690 produced a band (~3.2 kb, data not shown) from boa constrictor, and the presence of a CR-like sequence within this product was suggested by PCR experiments using this ~3.2-kb template and primers synthesized on the basis of the partially determined CR-like sequence in the IQM-related region of boa constrictor (data not shown). However, discrete amplifica-
Frequent Gene Conversion?

Why are the duplicated CR-like sequences so similar to each other, whereas they differ to a large extent among the species? If the paralogous CR-like sequences within a mtDNA molecule had evolved independently since the inferred time of the original insertion event (>70 MYA), a much more substantial sequence difference would be expected. Because the origin of the duplicated state at least predates the divergence between the Boidae and the Viperidae-Colubridae clade (see above), the difference between two duplicated CR-like sequences within a mtDNA can be expected to be greater than the difference between the viperids-akamatsu and ball python (39% on average, see above).

In order to explain this discrepancy, it seems necessary to consider that the duplicated CR-like sequences have evolved in a concerted manner. The control region is thought to contain many nucleotide positions at which base alterations are rapid and selectively neutral (see, e.g., Lee et al. 1995). The concerted evolution between the two CR-like sequences cannot, therefore, simply be explained by directional mutation pressure and/or selection. The most plausible explanation is that frequent gene conversion has persistently homogenized the duplicated CR-like sequences during the evolution of snakes. Gene conversion is a nonreciprocal recombination that results in the complete substitution of a sequence by another relevant sequence. An alternative possibility is that snake mtDNAs have a hitherto unknown replication mechanism that allows duplication of the CR-like sequence from one of two CR-like sequences of a template molecule in each cycle of replication. However, this seems less likely in light of the fact that the duplicated CR-like sequences are not completely identical for each species (fig. 4).

Gene conversion is known to occur commonly in nuclear genomes (see, e.g., Li and Graur 1991, pp. 162-169) but there has been little experimental evidence to show or suggest its occurrence in mitochondrial genomes. On the contrary, any type of recombination has been thought to be absent or very rare in mitochondrial genomes (see, e.g., Wolsenholme 1992). In this respect, the present study may require this prevalent notion to be reevaluated. As shown in figure 4 by circled numbers, the duplicate sequence unit varies from species to species but the left and right junctions are centered around the genes for rRNA^16S and rRNA^12S, respectively. Sequencing the IQM-related and cytb-12S rRNA regions from closely related vipers and multiple individuals within the species would provide a key to elucidating the mechanism underlying the concerted evolution of the CR-like sequences in more detail. It would also be important to reveal the entire gene organization and sequence for himehabu mtDNA in order to gain insights into how snake mtDNA replicates with two CR-like sequences. It would be also intriguing to investigate whether any of the gene arrangement features reported herein could serve as a useful molecular marker for snake phylogenetics.
Acknowledgments

We thank A. Sidow, A. C. Wilson, D. B. Wake, C. J. Schneider, and the Museum of Vertebrate Zoology of the University of California at Berkeley for generously providing animal tissues or extracted DNAs. We also thank Dr. Craig Moritz and anonymous reviewers for valuable comments on the original manuscript. Gratitude is extended to M. Nishimura and T. Kamura for helping with access to part of the himehabu samples. This work was supported by a grant from the Ministry of Education, Science, Sports and Culture of Japan to Y.K. (No. 07740583). An Alfred P. Sloan postdoctoral fellowship to Y.K. was also helpful in the very early stage of this work.

LITERATURE CITED


CRAIG MORITZ, reviewing editor

Accepted August 6, 1996