Evolution of the WANCY Region in Amniote Mitochondrial DNA

Gilles Seutin,* B. Franz Lang,* David P. Mindell,† and Réjean Morais*

*Département de Biochimie, Université de Montréal; and †Department of Biological Sciences, University of Cincinnati

In most vertebrate mitochondrial genomes, the site for initiation of light-strand replication, \( O_L \), is found within a cluster of five transfer RNA (tRNA) genes (tRNA\(^{Trp}\), tRNA\(^{Ala}\), tRNA\(^{Asn}\), tRNA\(^{Cys}\), and tRNA\(^{Tyr}\)). This region and part of the adjacent cytochrome \( c \) oxidase subunit I (COI) gene were sequenced for two crocodilian, two turtle, and one snake species and for \( S'phenodon punctatus; \) part of the adjacent nicotinamide adenine dinucleotide dehydrogenase subunit 2 (ND2) gene was also sequenced for the crocodilian and turtle species. All had the typical vertebrate gene order. The turtles and the snake have a lengthy noncoding sequence between the tRNA\(^{Asn}\) and tRNA\(^{Cys}\) genes that we assumed to be homologous to the mammalian \( O_L \). The crocodilians and \( Sphenodon \) lack such a sequence, a condition they share with birds. Most proposed phylogenies for the amniotes require that \( O_L \) at this position was lost at least twice during their diversification or was evolved independently more than once. Within the five tRNA genes, frequencies of substitutions are much higher in loops than in stems. Many loops vary dramatically in size among the species; in the most extreme case, the D-arm of the \( Sphenodon \) tRNA\(^{Cys}\) is a “D-arm replacement” loop of seven nucleotides. Frequency of transitions in stems is relatively uniform across tRNAs, but frequency of transversions varies greatly. Mismatches in stems are infrequent, and their relative frequency in a specific tRNA is unrelated to the frequency of substitution in the corresponding gene. Several features of mammalian mitochondrial tRNAs are conserved in WANCY tRNAs throughout amniotes. The inferred initiation codon for COI is GTG in crocodilians, turtles, and the snake, a condition they share with fishes, certain amphibians, and birds. TTG appears to be the initiation codon for COI in \( Sphenodon; \) if correct, this would be a novel initiation codon for vertebrate mitochondrial DNA. Phylogenetic analyses of the inferred amino acid sequences of ND2 and COI support the sister-group relationship of birds and crocodilians and suggest that mammals are an early derived lineage within the amniotes.

Introduction

In the two vertebrate species where it has been well studied, replication of each strand of the mitochondrial DNA (mtDNA) molecule is initiated from a distinct site (Clayton 1982). Replication starts with the synthesis of a daughter heavy strand (H-strand), beginning at the origin of H-strand replication (\( O_H \)) in the control region. Synthesis proceeds unidirectionally, resulting in the displacement of the parental H-strand. After approximately two-thirds of the H-strand has been replicated, the origin of light strand (L-strand) replication (\( O_L \)) becomes exposed, and L-strand replication is initiated, proceeding in a direction opposite to that of the H-strand. The \( O_L \) sequence, a noncoding DNA segment of \(~30\) nt, is located within the WANCY region, a region coding for five mitochondrial transfer RNAs (tRNAs) (tryptophan, alanine, asparagine, \( O_L \), cysteine, and tyrosine; fig. 1). In its single-stranded form, the \( O_L \) sequence has the potential for forming a stable stem-loop structure which is necessary for the initiation of replication (Hixson et al. 1986).

Sequences with high nucleotide similarity and the same characteristics as the \( O_L \), and thus presumably serving the same function, have been identified at the same location in most vertebrate mitochondrial genomes investigated (see, e.g., Wong et al. 1983). In marsupials, where the WANCY region has been rearranged (Pääbo et al. 1991), a presumptive \( O_L \) sequence was identified in the reorganized tRNA gene cluster in the three species studied. Of all the vertebrates studied so far, only birds do not show an \( O_L \) sequence in the WANCY region (Desjardins and Morais 1990, 1991; Ramirez et al. 1993; G. Seutin and R. Morais, unpublished data).

In this paper, we report on the presence or absence of an \( O_L \) sequence and on the gene order in the WANCY
region of representatives of all major extant amniote lineages. We amplified, cloned, and sequenced the complete WANCY region of species in the taxa Chelonia (turtles), Squamata (lizards and snakes), Rhyncocephalia (Sphenodon), and Crocodilia. We present here the first DNA sequences for the WANCY region in these four groups, and we compare them with published sequences for Aves (birds) and Mammalia.

Although there is no doubt about the monophyly of amniotes (Romer 1966, p. 163; Carroll 1988, pp. 192–201; Hedges et al. 1990), the evolutionary relationships among its member lineages are quite controversial. In particular, the position of turtles and mammals as basal or highly derived lineages and the sister-taxa relationship of crocodilians and birds are strongly debated (see, e.g., Gardiner 1982; Bishop and Friday 1988; Gauthier et al. 1988; Hedges et al. 1990; Hedges and Maxson 1991). Thus, we used our tRNA and flanking protein sequences to infer a phylogeny of the amniotes and used this and other published phylogenetic hypotheses to interpret the pattern of presence-absence of the \( OL \) region in vertebrates.

**Material and Methods**

**DNA Samples**

Purified mtDNA samples were obtained through ultracentrifugation in cesium chloride-ethidium bromide gradients (Dowling et al. 1990) from single individuals of one chelonian, *Malaclemys terrapin*; one squamate, *Epicrates subflavus*; one rhyncocephalian, *Sphenodon punctatus*; and two crocodilians, *Alligator mississippiensis* and *Crocodylus porosus*. Total genomic DNA was extracted from muscle tissue of a second chelonian, *Caretta caretta*, following the procedure of Kocher et al. (1989).

**Polymerase Chain Reaction (PCR) Amplification, Cloning, and Sequencing**

A fragment of DNA encompassing the whole WANCY region and parts of the nicotinamide adenine dinucleotide dehydrogenase subunit 2 (ND2) and cytochrome c oxidase subunit I (COI) genes was successfully amplified by the PCR in the chelonians and crocodilians by using the primers (1613-ND2) 5'-CTAAGCCTATTCTTCTA-3' and (912-COI) 5'-GTGGTTGGTTGAGAATAATCA-3'. In *Epicrates* and *Sphenodon*, these primers failed to produce a specific amplification product, and a smaller fragment was amplified, going from the central part of the tryptophan tRNA gene (primer [1769-TRP] 5'-AACCRAGGGCTTCTTCT-3') to the CO1 gene (primer 912-COI). Amplifications were carried out for 30 cycles with the thermostable Vent DNA polymerase. In a few cases, the desired amplification product was at low concentration, usually because a nonspecific product was coamplified (see Sequence Availability). For those samples, a first round of amplification was conducted (17–20 cycles), and the product of desired size was eluted from a low-melting-point agarose gel through the GeneClean procedure and was used as template for a second round of amplification (25 cycles).

Amplification products were made blunt-ended by using 2 U of T7 DNA polymerase and were phosphorlated by using 5 U of T4 polynucleotide kinase. Those reactions were conducted simultaneously on amplification products at 37°C for 30 min. The resulting products were run on low-melting-point agarose gels, and amplification products of the desired size were recovered in 20 ~1 1 X TE (10 mM Tris, 1 mM ethylenediaminetetraacetate, pH 8.0) by using the GeneClean procedure. The resulting amplification fragments were cloned in the \( Smal \) site of pBlueScript by following conventional protocols (Sanmbruuk et al. 1989, pp. 1.60–1.73). Two clones from each species were completely sequenced in both directions by using the dideoxynucleotide chain-termination method (Sanger et al. 1977). The universal and reverse primers of pBlueScript were used for sequencing, as well as the custom-made primers 1796-TRP (described above) and (1753-CYS) 5'-CATGATGAGTTGCCA-3'.
Sequence Analyses

All sequences were read by two or more independent scorers. Sequences were compared among themselves and with published data by using the program FASTA (Pearson and Lipman 1988); final alignments were made by eye. The tree-skewness approach of Hulsenbeck (1991) was used to evaluate whether phylogenetic information was present in the data matrix. Because our goal was to infer relationships among major lineages, we deleted at random one member of each closely related species pair in these analyses. One thousand random trees were produced using PAUP (Swofford 1990), and Hillis and Huselsenbeck's (1992) table 2 was used to judge the statistical significance of skewness statistics. For phylogenetic reconstructions, the computer packages PAUP and PHYLIP (Felsenstein 1991) were used. Phylogenetic analyses of the ND2 and COI sequences were based on the inferred amino acid sequences and on the matrix of minimum number of nucleotide substitutions for amino acid replacements. For phylogenetic analyses using tRNA genes, only stem sequences were used, because loops were typically highly variable and several could not be aligned across species. The rationale for using sequences involved in secondary structure in phylogenetic reconstructions is discussed by Woese et al. (1980) and Mindell and Honeycutt (1990). Because all phylogenetic inference methods require evolutionary independence of the nucleotide sites analyzed, a condition that is clearly violated in tRNA paired regions, only one side of each stem (selected by the flip of a coin) was considered in the analyses. Because of uncertainties about the ratio of transition events (TS) to transversion events (TV) in the generative process of the mammalian O_i (fig. 3): (i) they can form a stable stem-loop secondary structure when single stranded; (ii) there is a strong asymmetry in nucleotide usage in the stems, with sequences on the 5' side being almost exclusively made up of pyrimidines; (iii) the 5' end of the loops are very rich in pyrimidines, especially thymidine; and (iv) at least in the snake, a sequence 3'-GGCCG-5' is present a few nucleotides upstream from the 5' end of the stem, which is necessary for in vitro replication of the L-strand in humans (Hixson et al. 1986); the turtles have related sequences, 3'-GGCTG-5' and 3'-GTCTG-5'. Because of these similarities, we propose that the snake and turtle sequences are homologous to the mammalian O_i and serve the same function. Analysis of the putative snake O_i (fig. 3), as well as marsupial sequences (Pääbo et al. 1991; fig. 3), provides support for the suggestion of Johansen et al. (1990) that the RNA primers found at the 5' end of nascent mtDNA L-strands are initiated at a run of pyrimidines in the O_i loop and not necessarily at a run of thymidine as suggested by Wong and Clayton (1985).

Desjardins and Morais (1990, 1991) have shown that an O_i homologue with recognizable sequence similarity to other vertebrate O_i's is absent in galliform birds, a finding that was recently extended to a variety of other birds (Ramirez et al. 1993; G. Seutin and R. Morais, unpublished data). Sequences presented here (fig. 2) indicate that neither crocodilians nor Sphenodon have an O_i homologue between tRNA^Asn and tRNA^Cys genes or between other adjacent WANCY tRNA genes. In view of the probable close evolutionary relationships of crocodilians and birds (see, e.g., Romer 1966, pp. 257-263; Carroll 1988, pp. 338-342; Gauthier et al. 1988), the absence of an O_i in crocodilians is not particularly surprising. Because the presence of an O_i in the WANCY region is a primitive condition in amniotes (i.e., it is
Fig. 2.—Nucleotide sequence of the mitochondrial WANCY region, including parts of the ND2 and CO1 genes, of eight amniote species. Asterisk (*) indicates the beginning of a gene; regions that could not be aligned are boxed. Species Latin names are given in the text, except chicken, *Gallus gallus* (sequence from Desjardins and Morais 1990); and human, *Homo sapiens* (sequence from Anderson et al. 1981).
FIG. 3.—Putative secondary structure of the L-strand origin of replication of eight vertebrates. Boldface characters identify sequences that belong to tRNA coding genes; the 3'-GGCCG-5' sequences, or putative homologues, are boxed. Species Latin names are given in the text or in fig. 2, except cod, Gadus morhua (sequence from Johansen et al. 1990); xenopus, Xenopus laevis (sequence from Roe et al. 1985); opossum, Philander opossum (sequence from Pääbo et al. 1991); and mouse, Mus musculus (sequence from Bibb et al. 1981).
identified and shown to be located at the same site in birds and crocodilians. Work is in progress to precisely locate the O_L in birds (B. Ge and R. Morais, unpublished data).

The absence of an O_L homologue in the WANCY region of Sphenodon is more surprising. Rhynchocephalia, of which S. punctatus is the only extant member, is usually regarded as the sister group of Squamata (i.e., snakes+lizards), forming with them the Lepidosauria. The latter taxon is supported by a large number of synapomorphies (Gardiner 1982; Evans 1984; Benton 1985), but its relationships to other taxa are much less clear. Although morphological and molecular data sets exist that support their grouping with almost any other amniote lineage or group of lineages, the most thorough analyses consider them to be the sister group of a Crocodylia-Aves clade (Carroll 1988; Gauthier et al. 1988, preferred tree), of an Aves-Mammalia clade (Hedges et al. 1990, 18S rRNA trees; but see Marshall 1992), or of all other extant amniotes (Gardiner 1982; Gauthier et al. 1988, recent tree). Only under the first hypothesis, i.e., the relationship (Lepidosauria (Crocodylia, Aves)), is it possible that the loss of the O_L sequence was a unique event and did not require independent origins. In this case, the ancestor of these lineages had to be polymorphic for the presence or absence of the O_L and had to have given rise to the Crocodylia-Aves lineage, in which the O_L was absent, and to the Lepidosauria, which maintained the polymorphism until the Squamata and Rhynchocephalia split. This scenario seems unlikely, because it requires the maintenance of an mtDNA organizational polymorphism for a relatively long time. Still, one should remember that the amniote radiation may have occurred quite rapidly (Maeda and Fitch 1981; one should remember that the amniote radiation may have occurred quite rapidly (Maeda and Fitch 1981; Cantatore and Saccone 1987). As expected, V-loops were less variable, with only that of tRNA*Ser differing from that of all other vertebrate mtDNAs. Further, the first tRNA gene in the cluster, for tRNA^Trp, is separated from the end of the ND2 gene by 1 nt in squamates; in turtles these genes are contiguous. At the other end of the WANCY cluster, the tRNA^Tyr and CO1 genes are separated by 1 nt in all the species. Thus, the extreme compactness characteristic of vertebrate mtDNAs (see, e.g., Brown 1985; Cantatore and Saccone 1987) is well illustrated by our results.

**Length Variations and Substitution Patterns**

In the tRNAs we analyzed, loops were far more variable than their stem regions, as is usually the case (see, e.g., Cantatore and Saccone 1987). Most DHU- and TYC-loops varied across species in sequence and size (fig. 2). The most dramatic differences involved the DHU-loops of tRNA^Trp and tRNA^Tyr, which varied by 7–13 and 4–8 nt, respectively, and the TYC-loops of tRNA^Trp and tRNA^Ala, which varied by 3–8 and 2–7 nt, respectively. In the most extreme case, the DHU-arm in the Sphenodon tRNA^Cys gene is reduced to a 7-nt loop (a "D-arm replacement" loop; fig. 4). Among vertebrate mitochondrial genomes, a similar phenomenon has been seen only in tRNA^Ser (AGY), where the replacement loop is 5 nt long (de Bruijn et al. 1980; Cantatore and Saccone 1987). As expected, V-loops were less variable, with only that of tRNA^Ala presenting much variation at the sequence level and with that of tRNA^Trp varying in size by 1 nt across the species considered.

To analyze substitution patterns, we excluded one crocodilian (Crocodylus porosus) and one turtle (Malaclemys terrapin) (both were chosen by the flip of a

**Genes Coding for tRNAs**

**Sequence Organization**

Sequence comparisons and the identification of anticodons through analyses of secondary structures (results not shown) allowed us to unambiguously identify the five tRNA genes present in each of the species studied as those for tRNA^Trp, tRNA^Ala, tRNA^Asn, tRNA^Cys, and tRNA^Tyr. Of these, tRNA^Trp is encoded by the H-strand, whereas the other four are encoded by the L-strand; this organization is typical of vertebrates (see, e.g., Desjardins and Morais 1990). Our results support the contention that the rearrangement of the WANCY region observed in marsupials (Pääbo et al. 1991) is an autapomorphy for that group.

In the species we analyzed, the WANCY tRNA genes formed a tight cluster with very short or no intergenic sequences (0–4 nt; fig. 2). In Sphenodon, tRNA^Cys and tRNA^Tyr gene sequences actually overlap by 1 nt, a phenomenon that has been inferred for several other vertebrate mtDNAs. Further, the first tRNA gene in the cluster, for tRNA^Trp, is separated from the end of the ND2 gene by 1 nt in crocodilians; in turtles these genes are contiguous. At the other end of the WANCY cluster, the tRNA^Tyr and COI genes are separated by 1 nt in all the species. Thus, the extreme compactness characteristic of vertebrate mtDNAs (see, e.g., Brown 1985; Cantatore and Saccone 1987) is well illustrated by our results.

**Length Variations and Substitution Patterns**

In the tRNAs we analyzed, loops were far more variable than their stem regions, as is usually the case (see, e.g., Cantatore and Saccone 1987). Most DHU- and TYC-loops varied across species in sequence and size (fig. 2). The most dramatic differences involved the DHU-loops of tRNA^Trp and tRNA^Tyr, which varied by 7–13 and 4–8 nt, respectively, and the TYC-loops of tRNA^Trp and tRNA^Ala, which varied by 3–8 and 2–7 nt, respectively. In the most extreme case, the DHU-arm in the Sphenodon tRNA^Cys gene is reduced to a 7-nt loop (a "D-arm replacement" loop; fig. 4). Among vertebrate mitochondrial genomes, a similar phenomenon has been seen only in tRNA^Ser (AGY), where the replacement loop is 5 nt long (de Bruijn et al. 1980; Cantatore and Saccone 1987). As expected, V-loops were less variable, with only that of tRNA^Ala presenting much variation at the sequence level and with that of tRNA^Trp varying in size by 1 nt across the species considered.

To analyze substitution patterns, we excluded one crocodilian (Crocodylus porosus) and one turtle (Malaclemys terrapin) (both were chosen by the flip of a
likely that the paired nature of tRNA stems and the need for secondary and tertiary structural stability enforce constraints on, and retard evolution of, the base-paired portions of tRNA genes.

Although the frequency of transitions in stems is quite uniform across tRNAs, that of transversions is not, being close to three times higher in the gene for tRNA$_{Cys}$ than in that for tRNA$_{Tyr}$ (table 1). The latter gene shows little variation within mammals (Pääbo et al. 1991) and birds (G. Seutin and R. Morais, unpublished data; Pääbo et al. 1991) have discussed constraints that may be acting on tRNA$_{Tyr}$ evolution. The gene for tRNA$_{Cys}$ was the most variable gene analyzed, in frequencies of both TS and TV (table 1); it is also the most variable of the WANCY genes in comparisons among mammals (Pääbo et al. 1991) and birds (G. Seutin and R. Morais, unpublished data).

The frequency of each type of TV in tRNA stems is strongly biased against G-T (16.67% of the total number of transversions observed in all pairwise comparisons) and G-C (10.20%). This bias cannot be explained by nucleotide composition alone, since all nucleotides are present at about the same frequency in the regions considered (range = 21.04%–28.91%; expected G-T = 21.54%; expected G-C = 24.52%).

**Mismatches**

Although more than two-thirds (70.0%) of the stem positions we analyzed were variable across taxa, only 16.7% (99/594 pairs analyzed) of the stem pairs were mismatches. Most substitutions in tRNA stems appear to have been compensated to maintain Watson-Crick base pairing. There was no correlation between frequency of substitutions in the stem region of tRNA genes and frequency of stem mismatches in the inferred resulting tRNAs (table 1; Kendall's correlation: $\tau = 0.316, P = 0.439$).

A-C pairs accounted for the large majority (71.7%) of mismatches, followed in frequency by A-A pairs (13.1%) and a number of rarer combinations. G-T pairs, overall the most common mismatches encountered in mitochondrial tRNA genes (Cantatore and Saccone

---

**Table 1**

Substitution Patterns in Stem Regions of Five Mitochondrial tRNAs

<table>
<thead>
<tr>
<th>tRNA</th>
<th>No. of Variable Positions</th>
<th>Substitutions ± SE* (%)</th>
<th>Transitions ± SE* (%)</th>
<th>Transversions ± SE* (%)</th>
<th>TS:TV</th>
<th>No. of Mismatches</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRP</td>
<td>25</td>
<td>35.4 ± 4.4</td>
<td>27.8 ± 4.0</td>
<td>7.6 ± 1.5</td>
<td>3.7</td>
<td>8</td>
</tr>
<tr>
<td>ALA</td>
<td>34</td>
<td>37.8 ± 2.7</td>
<td>25.1 ± 2.4</td>
<td>12.7 ± 1.2</td>
<td>2.0</td>
<td>25</td>
</tr>
<tr>
<td>ASN</td>
<td>26</td>
<td>29.8 ± 2.2</td>
<td>22.5 ± 2.2</td>
<td>7.3 ± 0.8</td>
<td>3.1</td>
<td>18</td>
</tr>
<tr>
<td>CYS</td>
<td>36</td>
<td>46.3 ± 2.6</td>
<td>29.4 ± 2.1</td>
<td>16.9 ± 2.1</td>
<td>1.7</td>
<td>24</td>
</tr>
<tr>
<td>TYR</td>
<td>26</td>
<td>30.8 ± 2.5</td>
<td>25.4 ± 2.1</td>
<td>5.4 ± 0.5</td>
<td>4.7</td>
<td>24</td>
</tr>
</tbody>
</table>

* Average ± SE over all pairwise comparisons.
Conserved Features of Mitochondrial tRNAs

Cantatore and Saccone (1987) have identified three invariant features that mammalian mitochondrial tRNAs share with prokaryotic and cytoplasmic tRNAs. Two of those, the presence of a U preceding the anticodon sequence and of a purine following it, actually appear to be conserved throughout amniote mitochondrial tRNAs, at least as far as the WANCY sequences we studied are concerned. Cantatore and Saccone (1987) also pointed out that the first base in the anticodon loop of all cytoplasmic and prokaryotic tRNAs as well as mammalian mitochondrial tRNAs is a pyrimidine. While this is generally true in our sequences, tRNA\textsubscript{Ala} in the two turtles, in the snake, and in \textit{Sphenodon} has a G at that position, a condition they share with frogs (Roe et al. 1985; Yoneyama 1987).

Cantatore and Saccone (1987) have also identified four relatively conserved features specific to mammalian mitochondrial tRNAs. These show little variation in amniote tRNAs in general. In 59.3% of the sequences presented here, the first base in the DHU-loop was A, and in 51.9%, nucleotides 8 and 9 were UA; these proportions are lower than but similar to those reported by Cantatore and Saccone (1987). In V-loops, a purine was present at the second position in 70.0% of the sequences examined, and a pyrimidine was in the last position in 90.0% of the cases. The above results exemplify the greater evolutionary flexibility of mitochondrial tRNAs compared with other types of tRNAs. It remains to be seen if any of the nonorthodox features of vertebrate mitochondrial tRNAs are compensated for by posttranscriptional editing, as appears to be the case in \textit{Acanthamoeba} (Lonergan and Gray 1993; see also Beier et al. 1992).

Phylogeny Based on tRNAs

To infer a phylogeny of the amniotes from tRNA gene sequences, we used only regions coding for stems; loop regions were too variable to be useful or could not be aligned across taxa (results not shown). We further restricted the data set to sequences corresponding to one side only of each stem, to avoid the problem of correlated evolution of specific nucleotide positions (see Material and Methods). In bootstrap maximum-parsimony analyses of these data (105 bp/species), none of the internal branches were observed in >50% of the replicates, irrespective of the weight attributed to TV versus TS (trees not shown). This could be expected from the tree-skewness statistics (for TV:TS = 1, 2, and 4: \( g_1 > -0.317; P > 0.05 \)), which suggested that little phylogenetic information was available in the data set.

Distance analyses of the same data, using the neighbor-joining approach (Saitou and Nei 1987), produced trees whose topology differed greatly depending on the relative weight given to TS and TV. For instance, with TV:TS = 4, the following relationships were deduced: (\textit{Xenopus} (snake (mammals (crocodilians (turtles (birds, \textit{Sphenodon})))); with TV:TS = 1, (\textit{Xenopus} (snake (crocodilians (mammals (\textit{Sphenodon} (birds, turtles))))))).

The peculiar nature of tRNA evolution, especially the tendency for compensatory substitutions in stem regions and the action of selective constraints that probably result in convergent similarities, may be responsible for the difficulty of inferring phylogenies from tRNA sequence data. Alternatively, inferring within-amniote phylogeny from molecular data may be inherently difficult because the radiation of amniote lineages may have been rapid compared with genetic changes (Maeda and Fitch 1981; Bishop and Friday 1988).

The Protein-coding Regions

ND2

We sequenced 161–185 bp of the carboxy-terminal end of the ND2 gene in the crocodilians and turtles. In the comparison of 158 homologous nucleotide positions sequenced in six species representing three classes (fig. 2), only six positions showed all four nucleotides (2/53 first codon positions and 4/53 third codon positions). This result is partly explained by the strong bias against guanine in this coding region. Overall, 29.1% of the positions sequenced presented three nucleotides (39.6% of the third codon positions), 46.2% presented two nucleotides, and 20.9% were fixed (36.5% of the second codon positions; 5.7% of the third codon positions). The comparison of inferred amino acid sequences (fig. 5) reveals that these sequences are moderately different.

The most striking feature of the ND2 analysis is that crocodilian sequences appear to be 18–24 nt longer at their 3' termini than those of any other vertebrate sequenced to date. Because a complete termination codon (TAA) is present at the very end of the two crocodilian sequences (fig. 2), we suspect that the extra nucleotides are translated. The putative additional amino acids are generally hydrophilic and may have little functional significance. Both crocodilians and birds share an amino acid deletion near the end of the protein (corresponding to nucleotide positions 5442–5444 in the human mtDNA sequence; Anderson et al. 1981). Since all other vertebrates studied have a residue at that position, this amino acid deletion is a synapomorphy supporting a Crocodilia-Aves clade.
ND2

Chicken LRLAYHSTTITPLPPSNNHKMLWRNTKTLNTP-TAILTALSTTLPLS-PLITML
Human  ***I*STS***L*M*N*VKMK*QFEH*KP**FLPT*I**T*L**I**-*FMLMI*
Alligator ***W*N*SS*M**STT*TTR***KSTPQS*NF-*IM*LTMA**T*L**T**MKAI**Q**KQ**YSLC
Crocodile ***W*N*AS*****TI*TQR***KPTQNL*-INSL*MA*LT*I*ATMMKAITQETY
Sea turtle **IS*YA********T*Y~QQ **HKMNQK*YL*L**T***IM**IM-*TLL*IP
Land Turtle **IS*YT******F*TYLQQ **HKTNKKP-YL*M**I****IT-*TLL*IP

FIG. 5.—Amino acid sequence of parts of the mitochondrial ND2 and CO1 genes of eight amniote species. Asterisk (*) indicates amino acid identity. Species Latin names are given in the text or in fig. 2.

Turtle ND2 sequences are similar in length to those of noncrocodilian vertebrates. They end with a single T, as most vertebrate ND2 sequences do, and the termination codon is probably created by polyadenylation of the primary transcript (Ojala et al. 1981).

*Initiation Codon of COI*

Fifteen nucleotides of the 5′ terminal end of the COI gene were sequenced in turtles, crocodilians, Sphenodon, and the snake (fig. 2). In all cases but Sphenodon, the inferred initiation codon is GTG. GTG is also used as initiation codon for the COI gene in a number of bird species (Desjardins and Morais 1990, 1991; Ramirez et al. 1993; G. Seutin and R. Morais, unpublished data), Rana castabeiana (Yoneyama 1987), and fishes (Johansen et al. 1990; Tzeng et al. 1992). Thus, the ATG initiation codons found in mammals and Xenopus appear to have been independently derived. Cantatore et al. (1989) suggested that GTG can only be used as a translation initiation codon in genes that immediately precede the end of the previous gene. This idea is rejected by our results, since we observed in all species one spacer nucleotide between the 5′ end of the tRNA Tyr gene and the putative GTG initiation codon of COI (fig. 2). Johansen et al. (1990), studying the cod sequence, came to the same conclusion.

The most surprising finding that came from our sequencing of COI was that the translation initiation codon in Sphenodon is probably TTG. In metazoans, this triplet normally specifies leucine and had been reported elsewhere as a putative initiation codon only in nematode mtDNAs (Okimoto et al. 1990).

*Phylogenetic Analyses*

The inferred amino acid sequences of ND2 and COI were used jointly in an attempt to establish phylogenetic relationships among mammals, birds, crocodilians, and turtles, using Xenopus as an outgroup. The resulting bootstrapped maximum-parsimony tree is shown in figure 6. The branching pattern is identical to that presented in most of the paleontological literature (see e.g., Carroll 1988, pp. 11 and 193; Gauthier et al. 1988), with birds and crocodilians being sister groups and with mammals being an early derived lineage within the amniotes. Although bootstrap analysis provided only moderate or weak support for internal branches of the tree, and although the tree-skewness statistic ($g_1 = -0.092, P > 0.05$) suggested that limited phylogenetic structure was present in the data set, our data still favor the hypothesis of a bird-crocodilian clade over that of a bird-mammal clade; the former grouping was found in 75.2% of the bootstrap cladograms, the latter in only

![Fig. 6.—Amniote relationships inferred by maximum-parsimony analysis of the partial ND2 and COI amino acid sequences presented in fig. 5. Bootstrap estimates of branch confidence are based on 1,000 replicates.](image-url)
11.3%. The analysis of the same data by using a distance approach and the neighbor-joining clustering algorithm produced a tree with the same topology (result not shown).

**Sequence Availability**

Sequences have been deposited in GenBank under the following accession numbers: L08097, *Crocodylus porosus* (nucleotide sequence of a part of the third subunit of the mitochondrial cytochrome c oxidase gene [COIII]), L08098, *Epicrates subflavescens* (nucleotide sequence of a part of the mitochondrial ND4 gene), and L08099, *Malaclemys terrapin* (nucleotide sequence of a part of the mitochondrial cytochrome b gene).

**Acknowledgments**

We thank John Avise and Barry Campbell for providing DNA samples and Volta Ramirez and Hugo Desjardins for their help in the laboratory. E. Bermingham, W. Brown, and A. Martin made valuable suggestions on the manuscript. This work was supported by a grant from the Medical Research Council of Canada (to R.M.) and by NFS grant BSR9019669 (to D.P.M.). G.S. was supported by an NSERC postdoctoral fellowship.

**LITERATURE CITED**


Wesley M. Brown, reviewing editor

Received August 9, 1993

Accepted October 12, 1993