

Complete Mitochondrial DNA Sequences of Six Snakes: Phylogenetic Relationships and Molecular Evolution of Genomic Features

Songyu Dong,¹ Yoshinori Kumazawa²

¹ Department of Biological Science, School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan

² Division of Material Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan

Received: 22 June 2004 / Accepted: 5 February 2005 [Reviewing Editor: Dr. Bill Ballard]

Abstract. Complete mitochondrial DNA (mtDNA) sequences were determined for representative species from six snake families: the acrochordid little file snake, the bold boa constrictor, the cylindrophiid red pipe snake, the viperid himehabu, the pythonid ball python, and the xenopeltid sunbeam snake. Thirteen protein-coding genes, 22 tRNA genes, 2 rRNA genes, and 2 control regions were identified in these mtDNAs. Duplication of the control region and translocation of the tRNA^{Leu} gene were two notable features of the snake mtDNAs. The duplicate control regions had nearly identical nucleotide sequences within species but they were divergent among species, suggesting concerted sequence evolution of the two control regions. In addition, the duplicate control regions appear to have facilitated an interchange of some flanking tRNA genes in the viperid lineage. Phylogenetic analyses were conducted using a large number of sites (9570 sites in total) derived from the complete mtDNA sequences. Our data strongly suggested a new phylogenetic relationship among the major families of snakes: (((Viperidae, Colubridae), Acrochordidae), ((Pythonidae, Xenopeltidae), Cylindrophiidae), Boidae)), Leptotyphlopidae). This conclusion was distinct from a widely accepted view based on morphological characters in denying the sister-group relationship of boids and pythonids, as well as the basal divergence of nonmacrostomatan cylindrophiids. These results imply the significance to

reconstruct the snake phylogeny with ample molecular data, such as those from complete mtDNA sequences.

Key words: Reptilia — Squamata — Serpentes — Mitochondrial genome — Polymerase chain reaction — Molecular phylogeny — Control region — Concerted evolution — Gene rearrangement — tRNA gene

Introduction

Snakes (Squamata; Serpentes) are a group of reptiles with many derived morphological characters related to their peculiar behavior and life style, e.g., loss or reduction of limbs and girdles, loss of external and middle ears, lidless eyes protected by spectacles, reduction or loss of the left lung, and enlarged numbers of vertebrae (see, e.g., Zug et al. 2001). Extant snakes comprise an ecologically diversified group of over 2900 known species distributed worldwide except Antarctica (Uetz 2003). Snakes have a long evolutionary history and their fossil records are known from mid-Cretaceous (Rage 1987). It has been hypothesized that snakes evolved from a lineage of lizards through fossorial or marine ancestors (reviewed in Rieppel et al. 2003). However, information obtainable from fossil snakes is often fragmentary and it seems difficult to reveal the evolutionary history of snakes solely based on the fossil records (Rage 1987).

Correspondence to: Dr. Yoshinori Kumazawa; email: h44858a@nucc.cc.nagoya-u.ac.jp

The traditional classification of snakes based on morphology (Hoffstetter 1955; Underwood 1967) divided snakes into three major groups: the Scolecophidia (blindsnakes), the Henophidia (primitive snakes) and the Caenophidia (advanced snakes). The scolecophidians are small snakes highly specialized in the fossorial life style, while the latter two groups, often referred to as the alethinophidians or typical snakes, include more ecologically diverse species. The currently widely accepted view of snake phylogeny based on extensive morphological characters (see, e.g., Lee and Scanlon 2002) suggests the scolecophidians to be the earliest shoot-off of extant snakes. Within alethinophidians, nonmacrostomatan (small gape) anilioids (pipesnakes) are supposed to have diverged from a basal position, allowing other macrostomatan (large gape) snakes to have diverged into booids and booid-like groups (boas and pythons), acrochordids (file-snakes), and colubroids (vipers, colubrids, cobras, and sea snakes). However, snakes possess specialized or adaptational characters in various respects of their morphology, which may mislead the morphological characters to reconstruct incorrect or unresolved phylogeny (Heise et al. 1995; Zug et al. 2001).

Due to the potential problems described above, molecular phylogeny has been expected to provide independent information to evaluate the deep-branch phylogeny of snake lineages. There are several molecular phylogenetic studies for snakes (see, e.g., Cadle 1988; Heise et al. 1995; Slowinski and Lawson 2002; Vidal and Hedges 2002; Wilcox et al. 2002; Vidal and David 2004). These previous studies based on a few mitochondrial or nuclear genes, as well as immunological data partly confirmed the morphological view on snake phylogeny, but disagreed with it or with each other in several respects. There still remain unsettled relationships that should be addressed with more molecular data.

In the present study, we therefore attempted to resolve phylogenetic relationships among several major snake families by collecting and analyzing complete mitochondrial DNA (mtDNA) sequences. There are well-recognized advantages of using mtDNA sequences for phylogenetic studies, i.e., relative ease of collecting sequence data due to the compact and conservative gene organization without introns in mitochondrial genes, and clear orthology of homologous gene sequences due to the rare, if any, DNA recombination and gene duplication in mtDNAs (Wolstenholme 1992; Boore 1999; Saccone et al 2002). To the best of our knowledge, complete mtDNA sequence of snakes has been reported only from the colubrid akamata (Kumazawa et al. 1998) and the leptotyphlopoid Texas blind snake (Kumazawa 2004).

In addition, previous work (Kumazawa et al. 1996, 1998) showed duplication and possible concerted evolution of the control regions (CRs) in mtDNAs of

several snake species from Viperidae, Colubridae, and Pythonidae. The CR is a major noncoding region that contains controlling signals for replication and transcription of mtDNAs (reviewed in Clayton 1992). Thus, the collection of complete mtDNA sequences from various snake groups would also help to characterize this interesting phenomenon in more details, especially on the timing of the CR duplication in ancestral lineages of snakes, as well as on the mode of the concerted evolution in diverse snake lineages.

Materials and Methods

Samples and Sequencing

Table 1 lists eight snake species for which complete mtDNA sequences were analyzed. Since mtDNA sequences for two of them (akamata and Texas blind snake) have already been published, we sequenced mitochondrial genomes for the remaining 6 species in this study. Snake samples were obtained from other researchers or a local shop. DNA extraction was performed with a DNeasy Tissue Kit (QIAGEN). Extracted DNA was diluted 10 times and stored at -30°C until use as a template for polymerase chain reaction (PCR).

Complete mtDNA sequencing for the six snakes was carried out essentially as described by Kumazawa and Endo (2004). Supplementary information for the sequencing procedure is provided in the Appendix. The determined nucleotide sequences will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers shown in Table 1.

Gene Identification and Sequence Alignment

For identification of tRNA genes, we searched the mtDNA sequences for regions that can form characteristic secondary structures for mitochondrial tRNA genes (Kumazawa and Nishida 1993). Since rRNA genes and CRs were surrounded by tRNA genes, their boundaries were tentatively recognized by the range of adjacent tRNA genes. For final confirmation of protein-coding genes and RNA genes, sequence similarity to the corresponding genes of other taxa was also referred to.

Nucleotide sequences of all 37 genes of the eight snakes were aligned with counterparts of 24 nonserpentine taxa. These out-group sequences were taken from the database with the following accession numbers: green iguana (AJ278511), fence lizard (AB079242), crocodile lizard (AB080274), alligator lizard (AB080273), Komodo dragon (AB080275 and AB080276), mole skink (AB016606), spinytail lizard (AB079613), chicken (X52392), ostrich (Y12025), indigo bird (AF090341), alligator (Y13113), caiman (AJ404872), green turtle (AB012104), painted turtle (AF069423), helmeted turtle (AF039066), cow (J01394), opossum (Z29573), platypus (X83427), toad (M10217), caecilian (AF154051), lungfish (L42813), coelacanth (U82228), loach (M91245), and trout (L29771). The alignment was carried out with the aid of ClustalX (Thompson et al. 1997) and, finally, corrected by eye using DNASIS 3.5 (Hitachi) and MacClade 4.03 (Sinauer Associates, Inc.).

Phylogenetic Analysis

Two data sets were prepared for concatenated amino acid sequences of the 12 proteins encoded by the heavy strand (3381 sites in total) and for concatenated light-strand nucleotide sequences of

Table 1. Snake taxa analyzed in this study for their complete mtDNA sequences

Scientific name	Common name	Family	Museum voucher No.	mtDNA sequence	
				Length (bp)	Accession No.
<i>Acrochordus granulatus</i>	Little file snake	Acrochordidae	NUM-Az0375	17604	AB177879
<i>Boa constrictor</i>	Boa constrictor	Boidae	MVZ13188	18905	AB177354
<i>Cylindrophis ruffus</i>	Red pipe snake	Cylindrophidae	NUM-Az0376	17499	AB179619
<i>Dinodon semicarinatus</i>	Akamata	Colubridae	—	17191	AB008539
<i>Leptotyphlops dulcis</i>	Texas blind snake	Leptotyphlopidae	MVZ14037	16218	AB079597
<i>Ovophis okinavensis</i>	Himehabu	Viperidae	—	17388	AB175670
<i>Python regius</i>	Ball python	Pythonidae	NUM-Az0377	17245	AB177878
<i>Xenopeltis unicolor</i>	Sunbeam snake	Xenopeltidae	NUM-Az0378	18872	AB179620

Note. The complete mtDNA nucleotide sequences for *D. semicarinatus* and *L. dulcis* were originally reported by Kumazawa et al. (1998) and Kumazawa (2004), respectively. Abbreviations in the museum voucher numbers are NUM for Nagoya University Museum and MVZ for Museum of Vertebrate Zoology of the Uni-

versity of California at Berkeley. Some taxonomists considered boas and pythons to represent subfamilies of the common family Boidae, while *Cylindrophis* has until recently been considered to belong to Aniliidae or Uropeltidae (see Zug et al. [2001] and references therein for details).

the 12 protein genes, as well as 22 tRNA and 2 rRNA genes (9570 sites in total). The third codon positions of the protein genes were removed from the nucleotide data set because of high substitutional rates and consequent multiple substitutions as a source of noise in phylogenetic analyses (Brown et al, 1982; Kumazawa and Nishida 1993).

The only protein gene encoded by the light strand for ND6 (see the legend to Fig. 1 for abbreviations of mitochondrial genes) has an increased proportion of T and G in all codon positions due to the strand-specific base composition bias of mtDNAs, which influences replacement patterns at the amino acid sequence level (Asakawa et al. 1991). For this reason, the ND6 gene has been excluded from most recent phylogenetic analyses using the complete mtDNA sequences (see, e.g., Zardoya and Meyer 2000; Janke et al. 2001; Kumazawa 2004). Although we basically followed this tradition in this study, we found that a data set including the ND6 gene (9812 sites) provided the same phylogenetic conclusions (data not shown).

We also excluded the CRs from phylogenetic analyses because the CRs of animal mtDNAs usually accumulate base substitutions and indels at high rates and thus are not considered for use in deep-branch phylogenetic studies (see, e.g., Hillis and Moritz 1990). All unalignable and gap-containing sites were excluded from phylogenetic analyses. The aligned data files used for phylogenetic analyses can be obtained from YK upon request.

Phylogenetic analyses were carried out using maximum-likelihood (ML) method (Felsenstein 1981), Bayesian method (Huelsenbeck and Ronquist 2001), and neighbor-joining (NJ) method (Saitou and Nei 1987). The ML analyses with the nucleotide data set were conducted with PAUP*4.0b10 (Swofford 2003) by a heuristic search with TBR branch swapping with 10 random taxon additions. The general reversible model (GTR+I+G) and parameters optimized by modeltest 3.0 (Posada and Crandall 1998) were used. Bayesian analyses of the nucleotide sequences were conducted with MrBayes v3.0 (Huelsenbeck and Ronquist 2001) using the GTR+I+G model and separating the data into three partitions (codon first positions, codon second positions, and positions for rRNA and tRNA genes). The Markov chain Monte Carlo process was set to run four chains simultaneously. After the log-likelihood value appeared to become stationary, 10,000 trees were sampled in every 100 generation to provide a 50% majority-rule consensus tree together with posterior probability values in which the frequency of a specific nodal relationship in the sampled tree population is shown as a percentage. The NJ analyses with the nucleotide data set were conducted with PAUP*4.0b10 using the ML-distances estimated with the GTR+I+G model and optimized parameters.

The ML analyses with the amino acid data set were conducted using MOLPHY 2.3 (Adachi and Hasegawa 1996) with the

mtREV24 substitution matrix and amino acid frequency estimated from the data set. As recommended by the authors of MOLPHY, we first obtained a tree by the star decomposition search, from which the local rearrangement search was conducted to provide the ML tree and the local bootstrap values from 1000 replications. The Bayesian analyses of the amino acid data were conducted with MrBayes v3.0 using the mtREV24+I+G model and an empirical amino acid frequency. The Bayesian tree and posterior probability values were obtained using the same procedures described above.

Results and Discussion

Characteristics of the Snake Mitochondrial Genomes

Length, Base Composition, and Gene Organization. The complete nucleotide sequences of mtDNA were determined for the six snake species in this study. The total lengths of the mtDNA molecules were constant at 17.2–17.6 kbp, except for the boa constrictor (18905 bp) and the sunbeam snake (18872 bp) in which CRs contained long tandem repeats (Table 1). When the comparison included two more mtDNA sequences for the akamata and the Texas blind snake, mtDNA sequence for the blind snake is considerably smaller due to the absence of the CR duplication (Table 1). The overall base compositions of the light-strand sequences were A: 35.7%, T: 28.0%, and C: 25.0%, and G: 11.3% for the little file snake, A: 37.1% T: 24.4%, C: 26.0%, and G: 12.5% for the boa constrictor, A: 35.2%, T: 25.3%, C: 27.4%, and G: 12.1% for the red pipe snake, A: 32.6%, T: 25.6%, C: 29.3%, and G: 12.6% for the himehabu, A: 33.6%, T: 24.5%, C: 29.4%, and G: 12.6% for the ball python, and A: 35.6%, T: 26.3%, C: 26.0%, and G: 12.1% for the sunbeam snake. The base compositions in these mtDNAs were skewed similarly to those of other vertebrate mtDNAs (Asakawa et al. 1991), with more A–T base pairs than

listed in Table 1. Only five tRNA genes, i.e., those for tRNA^{Asn}, tRNA^{Gln}, tRNA^{Ser} (UCN), tRNA^{Leu}(UUR), and tRNA^{Leu}(CUN), conserved a complete T stem with 5 bp. The other tRNA genes permitted less than 5 bp for the T stem and many of them (especially tRNA^{Pro}, tRNA^{Gly}, and tRNA^{Met} genes) even allowed 2–3 bp in the T stem (data not shown). These results imply that the evolutionary pressure to simplify the T arm of mitochondrial tRNAs has been operating for almost all extant snake lineages.

Control Regions. Table 2 shows the length of CR sequences for each snake species. In snakes other than the boa constrictor and the sunbeam snake, the CR was 950–1200 bp in length, whereas these two snakes had considerably longer CR sequences due to tandem repeats. The boa constrictor had 11 and 10 arrays of a 98-bp repeat unit in the CR1 and CR2, respectively. The sunbeam snake had 8 and 10 arrays of a 97-bp repeat unit in the CR1 and CR2, respectively. These repeat units could be folded into a stable secondary structure (data not shown), suggesting a role of the slipped-strand mispairing in creating the tandem duplications during replication of mtDNA (Levinson and Gutman 1987).

In mammals, the CR was shown to include information necessary for replication and transcription of mtDNAs (i.e., origin of heavy-strand replication and initiation of both heavy- and light-strand transcriptions) and conserved sequence blocks (CSBs) I–III were identified as conserved sequence elements in CRs of mammalian mtDNAs (reviewed in Clayton 1992). A sequence motif equivalent to the mammalian CSB I was clearly identified in CR sequences of the six snakes and the CSB III motif was also identifiable for five snakes other than the boa constrictor. However, unlike mammals, none of the six snakes had the CSB II motif between CSBs I and III. Instead, a characteristic motif consisting of two tandem arrays of cytidines (C₈ or C₉) interspaced by a TA or TTA sequence was found in the 5' end portion of the CR sequences. This motif, which is similar in sequence to the mammalian CSB-II to some extent, was termed the C-rich sequence and was considered as a structural barrier for DNA polymerase in light of an observation that sequencing reactions for both strands stalled in that region (Kumazawa et al, 1998). The C-rich sequence commonly occurs in all the duplicate CRs for the seven snakes but is not clearly found in the corresponding locations of the blind snake, lizards, mammals, frog, or fish.

For all the seven snake taxa in which the CR was duplicated, nucleotide sequences of the two CRs were identical or quite similar within species, with only zero to seven base differences between them (Table 2). The boa constrictor had the largest base differences between the two CR copies, but still

retaining 99.6% sequence similarity. On the other hand, the CR sequences were divergent to a large extent among species (e.g., approximately 24% divergent between the himehabu and the akamata). If the paralogous copies of CRs evolved independently, much more substantial base differences would be expected between them. These observations therefore suggest concerted sequence evolution of the two CRs located more than 3 kbp away from each other within each snake lineage.

Table 2 shows the distribution of base changes in the CR sequences within individual taxa. Although many changes appeared to be located in the 5' or 3' end portion of the CR rather than in the central portion of the CR, there was no clear rule in the type and location of the base changes. Thirteen changes in total consisted of six transitions between purine bases, one transition between pyrimidine bases, five transversions, and one indel. Among the seven base changes for the boa constrictor, A/G (1173) and A/G (1271) were in the same position of neighboring 98-bp repeat units. A/G (1054) and A/G (1446) also corresponded to the same position of other (though non-neighboring) 98-bp repeat units. It is therefore likely that tandem duplications recently spread the base changes in the boa CRs and that these differences have not yet been homogenized by an unknown mechanism for the concerted evolution.

Phylogenetic Analyses

Figure 2 depicts an ML tree based on the nucleotide sequences of the protein and RNA genes (9570 sites in total). The same tree topology was obtained by the Bayesian method with strong posterior probability values for each nodal relationship (Fig. 2). The phylogenetic relationships obtained for nonserpentine outgroup taxa, including the archosaurian affinity of turtles, were in agreement with previous studies using similar mitochondrial data sets (see, e.g., Zardoya and Meyer 2000; Janke et al. 2001; Rest et al. 2003; Kumazawa 2004) and nuclear data sets (Iwabe et al. 2005).

In the ML tree in Fig. 2, snakes made a sister clade relative to all the lizard taxa from six different families, indicating that our previous conclusion for independent origins of snakes and lizards (Kumazawa 2004) still holds if snake taxa are expanded from two to eight. Phylogenetic affiliation of snakes has been extensively debated in relation to the life style (i.e., fossorial, terrestrial or marine) of ancestral snakes (reviewed in Rieppel et al. 2003). Our results based on complete mtDNA sequences suggest that snakes did not originate from a lineage of varanoids (monitor lizards and their relatives including extinct marine mosasaurs) against a major evolutionary hypothesis supported by

Table 2. Sequence comparison of two control regions in snake mtDNAs

Taxon	Length (bp)		Sequence comparison between two CRs	
	CR1	CR2	Similarity	Sequence differences
Little file snake	1187	1177	99.8% (2/1171) ^a	T/A (3) ^b and C/T (19)
Boa constrictor	1878	1754	99.6% (7/1754)	G/C (47), G/C (50), A/G (1054), A/G (1173), A/G (1271), T/G (1311), and A/G (1446)
Red pipe snake	1134	1133	99.9% (1/1133)	T/gap (741)
Akamata	1019	1018	100% (0/1018)	—
Texas blind snake	1032	—	—	—
Himehabu	1032	1032	100% (0/1032)	—
Ball python	1029	961	100% (0/961)	—
Sunbeam snake	1706	1866	99.8% (3/1683)	A/G (18), A/T (1494), and A/G (1682)

^aNumbers of base differences vs. alignable sites between two control regions (left and right of the slash, respectively).

^bBase changes in CR1 vs. CR2 (left and right of the slash, respectively) at the indicated base position of the aligned CR sequences (in parentheses).

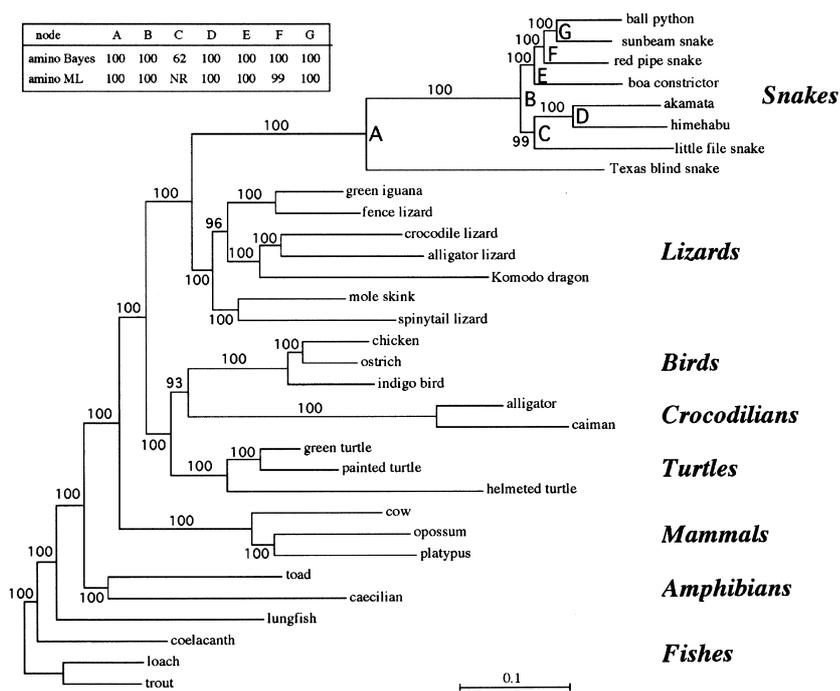


Fig. 2. A maximum-likelihood tree constructed using nucleotide sequences of all the mitochondrial genes except for ND6 (9570 sites). The ML tree was built with PAUP*4.0b10 (Swofford 2003) using the GTR+I+G model and parameters optimized by modeltest 3.0 (Posada and Crandall 1998). The same tree topology was obtained by the Bayesian method (MrBayes v3.0: Huelsenbeck and Ronquist 2001) and resultant Bayesian posterior probability values are shown on the corresponding branches. *Inset:* Bootstrap probabilities for nodes A–G estimated using amino acid sequences of 12 protein genes (3381 sites). The first row shows Bayesian posterior probability values obtained using the amino acid data set. The second row shows local bootstrap values from ML analyses (the star decomposition search + local rearrangement search) using the amino acid data set. NR means that the corresponding nodal relationship was not reconstructed in the best tree topology. Refer to Materials and Methods for the details of these phylogenetic analyses.

some morphological (see, e.g., McDowell and Bogert 1954; Lee 2000) and molecular (see, e.g., Forstner et al. 1995; Rest et al. 2003) studies. Apparent morphological similarities between snakes and varanoids (e.g., long retractile tongue) may therefore be regarded as convergent characters (Kumazawa 2004).

With respect to the phylogenetic relationships within snakes, different tree-building methods consistently supported topological relationships at nodes A, B, and D–G, although nodal C relationship was not reconstructed by some analyses; e.g., an NJ analysis based on the ML distances estimated with the GTR+I+G model and optimized parameters (data not shown) and an ML analysis using the amino acid data set in which the gamma correction

for the site-heterogeneity could not be used for a technical reason (see the inset in Fig. 2). The tree in Fig. 2 suggests that the blind snakes (scolecophidians) diverged most basally among living snakes, followed by two monophyletic groups: one for caenophidians (the little file snake and the akamata–himehabu clade) and the other for henophidians (the ball python, the sunbeam snake, the red pipe snake, and the boa constrictor). Within the henophidian clade, the ball python and the sunbeam snake comprised a sister group, which clustered with the red pipe snake, preceded by the boa constrictor. Snakes appear to possess increased rates of mtDNA sequence evolution (Kumazawa et al. 1998; see also relatively long branch lengths for snakes in Fig. 2).

However, our phylogeny did not change by taxonomic representation for the nonserpentine outgroup and the same phylogenetic relationships among alethinophidians were obtained even if the Texas blind snake was used as the only outgroup taxon (data not shown; see Kumazawa 2004 for discussions on the long branch attraction).

The inferred phylogenetic relationships for snakes (Fig. 3A) are consistent with a widely accepted phylogeny based on morphology (Fig. 3B) with respect to the basal divergence of scoleophidians and monophyly of caenophidians. However, our tree is not in agreement with the morphological tree in denying the sister-group relationship of boids and pythonids, as well as the basal divergence of cylindrophids and xenopeltids. The close phylogenetic relationship between xenopeltids and pythonids has been supported by some morphological studies (Underwood 1976; Rieppel 1979) but there still remains an uncompromisable gap between the molecular and morphological phylogenies.

Previous molecular phylogenetic studies using a few mitochondrial or nuclear gene sequences (see, e.g., Heise et al. 1995; Slowinski and Lawson 2002; Vidal and Hedges 2002; Wilcox et al. 2002; Vidal and David 2004) generally supported the sister relationship of xenopeltids and pythonids, from which boids are phylogenetically isolated. Some studies (Wilcox et al. 2002; Vidal and David 2004) even suggested a derived macrostomatan origin of cylindrophids. However, in these molecular studies, topological relationships among these henophidian taxa and the Caenophidia considerably varied from study to study with low statistical support values in general. It should be noted that none of the topological relationships in these studies exactly corresponded with our conclusion (Fig. 3A).

We then evaluated the statistical basis of our conclusion in comparisons with several alternative hypotheses using the Kishino–Hasegawa (KH) test (Kishino and Hasegawa 1989) and the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa 1999) (see Table 3). The SH test is supposed to be a statistically more legitimate measure than the KH test, although the SH test may be somewhat conservative in rejecting competing hypotheses (Shimodaira and Hasegawa 1999). Whether the KH or SH test was used and whether the amino acid or nucleotide data set was used, tree 1 (Fig. 3A) was significantly ($p < 0.01$) better than tree 2 (Fig. 3B). In other words, our molecular data set strongly rejected the widely accepted phylogeny based on morphology.

Other hypotheses having alternative positions for cylindrophids (Fig. 3C) were mostly rejectable, at least in the 5% significance level by the KH test using the amino acid data set (Table 3). Tree 3, having a basal (potentially nonmacrostomatan) divergence

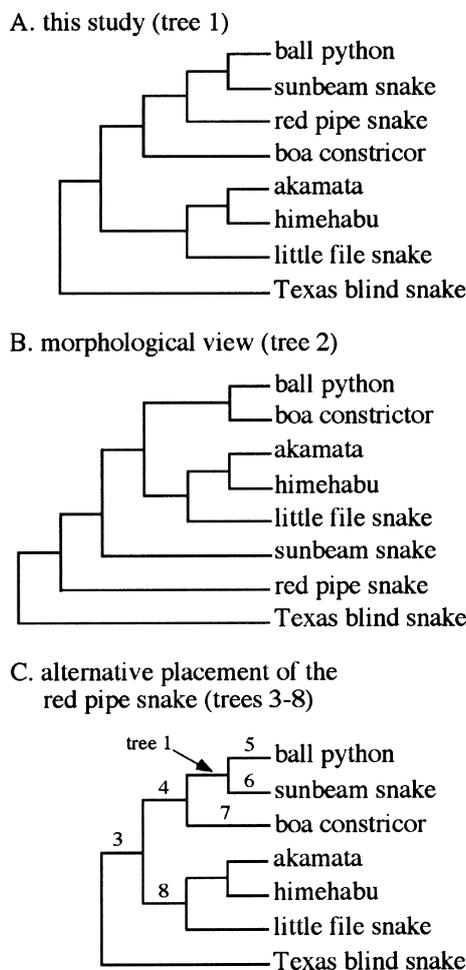


Fig. 3. Hypotheses on the phylogenetic relationships of snakes. **A** A molecular phylogeny based on this study. **B** A phylogeny based on cladistic analyses of morphological characters by Lee and Scanlon (2002). **C** Alternative placements of the red pipe snake on the molecular tree shown in **A**.

point for cylindrophids, was rejectable ($p < 0.05$) also by the SH test using the amino acid data set. Cylindrophids may therefore have secondarily lost the macrostomatan condition in the course of adaptation to the burrowing lifestyle. Taken together with other strong support measures shown in Fig. 2, these results show the robustness of our molecular phylogeny.

We consider that the gap between molecular and morphological phylogenies implies unrecognized homoplasious character changes in the morphological data. In this regard, construction of an independent and reliable phylogenetic framework among extant snake groups with ample molecular data, such as those from complete mtDNA sequences, should receive more expectation than ever.

Molecular Evolution of Snake Mitochondrial Genomes

Based on the phylogenetic framework of snakes thus established, the timing of characteristic changes that

Table 3. Evaluation of alternative hypotheses on the phylogenetic relationship of snakes

Tree	12 proteins (amino acid seq.)			36 genes (nucleotide seq.)		
	$\Delta \ln L/SE$	<i>p</i> -KH	<i>p</i> -SH	$\Delta \ln L/SE$	<i>p</i> -KH	<i>p</i> -SH
1	ML	1.000	1.000	ML	1.000	1.000
2	4.43	0.000**	0.000**	4.43	0.000**	0.000**
3	2.91	0.004**	0.023*	2.39	0.017*	0.073
4	2.07	0.039*	0.161	1.40	0.162	0.414
5	2.96	0.003**	0.128	2.54	0.011*	0.092
6	2.24	0.025*	0.176	2.39	0.017*	0.111
7	2.79	0.005**	0.079	1.44	0.149	0.425
8	3.01	0.003**	0.017*	2.58	0.010**	0.066

Note. The two-sided Kishino-Hasegawa (KH) test (1989) and the Shimodaira-Hasegawa (SH) test (1999) were conducted with TREE-PUZZLE 5.2 (Strimmer and von Haeseler 1996) using either amino acid sequences of 12 protein genes (3381 sites) or nucleotide sequences of 36 mitochondrial genes (9570 sites). The GTR+I+G model and its parameters optimized by modeltest 3.0 were used for the nucleotide data, whereas the mtREV24+I+G model was used for the amino acid data using parameters optimized with TREE-PUZZLE 5.2. ML stands for the ML tree topology (see Fig. 3A).

Alternative hypotheses from the morphological study (tree 2) and for the placement of the red pipe snake (trees 3–8) were assumed as shown in Fig. 3. Results are shown as probability values for each hypothesis (*p*-KH and *p*-SH from the KH and SH tests, respectively). Asterisks and double asterisks mean rejection of the corresponding hypothesis at the 5% ($p < 0.05$) and 1% ($p < 0.01$) significance levels, respectively. $\Delta \ln L/SE$ stands for the difference in log-likelihood from the ML tree topology divided by the standard error.

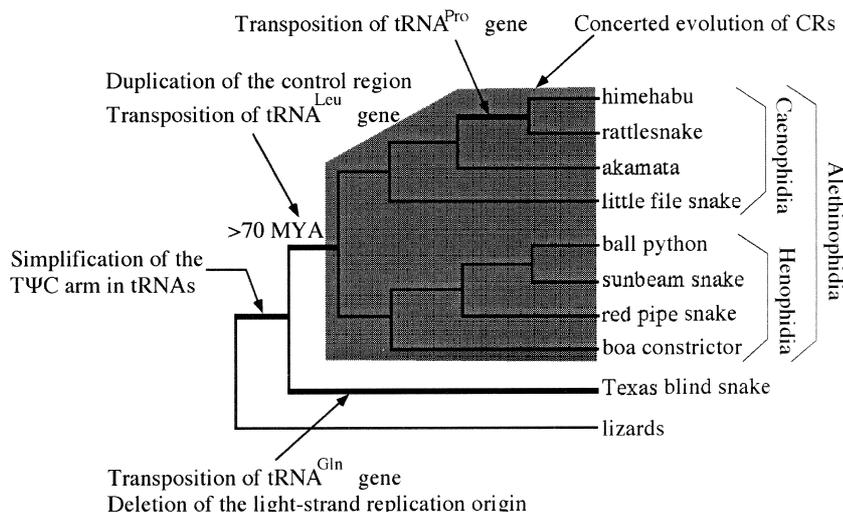


Fig. 4. Occurrence of characteristic features of snake mitochondrial genomes. The phylogeny shown here is based on the results of our phylogenetic analyses (see Fig. 2). Thick lines highlight branches on which changes in the genomic features presumably occurred.

occurred in snake mitochondrial genomes was estimated by the parsimony criterion (Fig. 4). Presumably due to the rarity of the character changes, the parsimony served to map them to specific single positions of the tree without assuming parallel changes or reversals. Mitochondrial genomes of all alethinophidian taxa examined to date possess the second CR located between the $tRNA^{Ile}$ gene and the transposed $tRNA^{Leu}$ gene (Fig. 4), whereas the Texas blind snake and several lizards do not have this feature (Kumazawa and Nishida 1999; Janke et al. 2001; Kumazawa and Endo 2004; Kumazawa 2004). It should be noted that the second CR in the Komodo dragon mtDNA was found in a distinct location and may thus have arisen independently (Kumazawa and Endo 2004). It is therefore likely that the original duplication of CR and possibly concomitant transposition of the $tRNA^{Leu}$ gene took place in an

ancestral lineage of the alethinophidians at least 70 million years ago based on the fossil records (Rage 1987). Detection of the concerted sequence evolution of CRs for all the alethinophidian snakes tested (Table 2) suggests that the duplicate state of CR and the concerted mode of its evolution have persisted more than 70 million years in snake mtDNAs.

As outlined earlier, the viperid himehabu gained another gene rearrangement in which a functional $tRNA^{Pro}$ gene moved from the 5' proximity of CR1 to that of CR2 and a new $tRNA^{Phe}$ gene or pseudogene was created in the 3' proximity of CR2. The positional switch of the $tRNA^{Pro}$ gene was found for another viperid rattlesnake (Kumazawa et al. 1996) but was not found for the nonviperid snakes examined in this study. These observations suggest occurrence of the $tRNA^{Pro}$ gene transposition in an early stage of the viperid radiations (Fig. 4). In light of relative loca-

tions of the tRNA^{Pro} and tRNA^{Phe} genes and two CRs, the duplicate CRs appear to have facilitated an interchange of these flanking tRNA genes between the two CRs. Kumazawa et al. (1998) proposed two possible mechanisms (i.e., the tandem duplication model and the gene conversion model) for the concerted evolution of CRs. While such mechanism operates on snake mtDNAs, the flanking tRNA genes may be copied together with the CR sequence to be pasted in the other homologous region of mtDNA.

In future, sequencing and analyzing mtDNAs from more diverse snake groups will contribute to elucidating intriguing phylogenetic questions, as well as the tempo and mode of molecular evolution in mitochondrial genomic features. We consider that the present study provides a solid basis for that goal. Of course, in spite of several advantages of mtDNA sequences for phylogenetic studies (see **Introduction**), one can trace only a single genealogy by mtDNAs and the use of nuclear genes for this purpose should also be encouraged.

Acknowledgments. We thank Mr. K. Yagi, Drs. M. Nishida and D. Wake, Remix Peponi Co., and the Museum of Vertebrate Zoology, University of California at Berkeley, for providing animal samples. We also thank Dr. T. Nishikawa and Nagoya University Museum for the curation of our specimens and Ms. C. Aoki for her excellent experimental assistance. Gratitude is extended to Dr. H. Shimodaira, two anonymous reviewers, and the Associate Editor for valuable comments on the phylogenetic tests. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan (Nos. 12640680 and 14540641).

Appendix

Complete mtDNA Sequencing for Six Snakes

We employed an efficient and accurate method for sequencing reptile mtDNAs using the long-and-accurate polymerase chain reaction (LA-PCR) amplification of mtDNAs followed by sequencing and assembling nested PCR products with a number of reptile-oriented primers (Kumazawa and Endo 2004). Using the LA-PCR technology, we amplified mtDNA segments for each taxon with the combinations of primers shown in the note to Table A1. Because snake mtDNAs typically contain two CRs, inclusion of two CRs in the same PCR target may cause the jumping PCR (Pääbo et al. 1990) to give rise to artifactually shorter products. This problem was carefully avoided in this study by separating mtDNA sequences into three to five portions for the amplification targets.

More than 85% of the whole mtDNA sequences for the six snakes were successfully amplified and sequenced with the reptile-oriented primers (Kumazawa and Endo 2004). In general, primers designed for some extremely variable regions (e.g., CRs and 3' end portion of ND5 gene to ND6 gene) were less effective in the PCR amplification. These gaps could be readily filled by the primer walking strategy. The boa constrictor and the sunbeam snake had long tandem repeats within CRs and sequencing from surrounding primers did not extend over the repeated region. We estimated

Table A1. Taxon-specific and common primers designed for initial amplification of mtDNA segments

Primer name [length]	Matching gene	Sequence (5' to 3')
Little file snake		
Agra-L1 [28mer]	ND1	GGTCCCATAGGCCTACTACAACCAATCG
Agra-L2 [29mer]	ND4	TGAGGCCACCAATCTGAACGACTTACCGC
Agra-H1 [27mer]	ND1	CGTTCTAGTAGGGTTATAAAGGCTACT
Agra-H2 [29mer]	ND4	GCGGTAAGTCGTTTCAGATTGGTGGCCTCA
Boa constrictor		
Bcon-L1 [19mer]	Gln	AGACTCGAAGCTTGGACCAG
Bcon-L2 [23mer]	12S	CTCTGATTATACATGCAAGACTC
Bcon-H1 [19mer]	Ile	GCTATCAAGGTAGTCCCTG
Bcon-H2 [24mer]	cytb	CGTCTCGTATGATATGTACGATGG
Red pipe snake		
Cruf-L2 [28mer]	cytb	ATGTAATCCTCTACACGAAGAAGGCTC
Cruf-H1 [30mer]	16S	AGGGTACCGCGCCGTTGAATTAGTTTCAC
Cruf-H2 [29mer]	cytb	AGTATTGAGCCGAAGTTTCATCATGTTGA
Cruf-H3 [17mer]	ND1	ATGGCTAGGGTGAGTGG
Himehabu		
Ooki-L1 [26mer]	12S	TCACACCATGACACCAAGCAATAGCC
Ooki-H1 [21mer]	cytb	TGGCCAATTTCTGTGAATGGC
Ooki-H2 [24mer]	Ile	GTTGTGGTTCCTCGTCTACTCTA
Ball python		
Preg-L1 [29mer]	COI	GAGGAGACCCCGTACTATTCCAACACCTG
Preg-L2 [29mer]	cytb	ATAACCACCTGATTATGAGGAGGGTTCGC
Preg-L3 [27mer]	12S	AGACATGAAACCATCCTTTAAAGGCGG
Preg-L4 [17mer]	ND3	CTATAATCATTAAGCC

Continued

Table A1. Continued

Primer name [length]	Matching gene	Sequence (5' to 3')
Preg-H1 [27mer]	COI	AGTGGGATAAGTCAGTTTCCAAAGCCT
Preg-H2 [17mer]	cytb	GTTAGTAGTATTGAGCC
Preg-H3 [29mer]	12S	TCTAGGTTGGGTGAAGTACCGTCAAGTCC
Preg-H4 [30mer]	COIII	ACGAAAAACAGAGAGCCATAAACGCTGTCC
Sunbeam snake		
Xuni-L2 [17mer]	16S	CTATCTTCTTCAAGAGC
Xuni-L3 [26mer]	AT6	TGCGACTCACAGCCAACATTACAGCT
Xuni-H1 [26mer]	16S	CTACCTTTGCACGGTTAGGGTACCGC
Xuni-H2 [17mer]	ND2	TTCATGCCATGATTCAG
Xuni-H3 [26mer]	AT8	CATGGTAAGATTACAGGCTAGTTCTGC
Common primers		
uPhe-1H [20mer]	Phe	GCATCTTCAGTGCYATGCTT
uThr-1L[21mer]	Thr	AAAGCATTGGTCTTGTAARCC
rThr-1H[20mer]	Thr	TCCCATCTTTGGYTTACAAG

Note. Combinations of primers used to amplify mtDNA segments for each taxon were: rPhe-1L and Agra-H1 (2.7 kbp), Agra-L1 and H4433 (2.3 kbp), rND2-1L and Agra-H2 (6.5 kbp), Agra-L2 and rThr-1H (4.7 kbp), and uThr-1L and uPhe-1H (1.3 kbp) for the little file snake; Bcon-L2 and Bcon-H1 (3.5 kbp), L4160m and H4433 (2.1 kbp), Bcon-L1 and Bcon-H2 (10.5 kbp), and L14940 and H690 (3.1 kbp) for the boa constrictor; 16sar-L and Cruf-H3 (1.6 kbp), L4160m and Cruf-H2 (12.0 kbp), and Cruf-L2 and Cruf-H1 (3.8 kbp) for the red pipe snake; Ooki-L1 and Ooki-H2 (3.4 kbp), L4160m and H4433 (1.6 kbp), L4437b and Ooki-H1 (11.2 kbp), and L14940 and H690 (2.2 kbp) for the himedabu; Preg-L3 and Preg-H1 (5.6 kbp), Preg-L1 and Preg-H4 (3.2 kbp), Preg-L4 and Preg-H2 (4.7 kbp), and Preg-L2 and Preg-H3 (2.4 kbp) for the ball python; and Xuni-L2 and Xuni-H2 (3.5 kbp), L5038 and Xuni-

H3 (3.3 kbp), and Xuni-L3 and Xuni-H1 (10.7 kbp) for the sunbeam snake. L and H in the taxon-specific primers mean that they correspond to sequences of the light and heavy strands, respectively. Common primers starting with r and u were originally designed to be oriented for reptiles and universal for wider vertebrate groups, respectively. Names of previously published primers are: 16sar-L by Palumbi et al. (1991); L4160m by Kumazawa and Nishida (1993); H4433 and L5038 by Kumazawa and Nishida (1995); H690, L4437b, and L14940 by Kumazawa et al. (1996); and rPhe-1L and rND2-1L by Kumazawa and Endo (2004). See the legend to Fig. 1 for the abbreviations of genes. Transfer RNA genes are abbreviated in this table with three-letter codes for the corresponding amino acids.

the length of the repeated region from the size of an amplified product using these primers and then carefully assembled sequences obtained from these primers in light of the length of a repeat unit.

References

- Adachi J, Hasegawa M (1996) MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. Computer Science Monographs, No 28. Institute of Statistical Mathematics, Tokyo
- Asakawa S, Kumazawa Y, Araki T, Himeno H, Miura K, Watanabe K (1991) Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. *J Mol Evol* 32:511–520
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27:1767–1780
- Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J Mol Evol* 18:225–239
- Cadle JE (1988) Phylogenetic relationship among advanced snakes: a molecular perspective. *Univ Calif Publ Zool* 119:1–77
- Clayton DA (1992) Transcription and replication of animal mitochondrial DNAs. *Int Rev Cytol* 141:217–232
- Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17:368–376
- Forstner MRJ, Davis SK, Arévalo E (1995) Support for the hypothesis of anguimorph ancestry for the suborder serpentes from phylogenetic analysis of mitochondrial DNA sequences. *Mol Phylogenet Evol* 4:93–102
- Heise PJ, Maxson LR, Dowling HG, Hedges SB (1995) Higher-level snake phylogeny inferred from mitochondrial DNA sequences of 12S rRNA and 16S rRNA genes. *Mol Biol Evol* 12:259–265
- Hillis DM, Moritz C (1990) An overview of applications of molecular systematics. In: Hillis DM, Moritz C (eds) *Molecular systematics*. Sinauer Associates, Sunderland, MA
- Hoffstetter R (1955) *Squamates de type moderne*. In: Piveteau J (ed) *Traité de Paléontologie* 5. Masson, Paris, pp 606–662
- Huelsenbeck JP, Ronquist FR (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755
- Iwabe N, Hara Y, Kumazawa Y, Shibamoto K, Saito Y, Miyata T, Katoh K (2005) Sister group relationship of turtles to the bird-crocodylian clade revealed by nuclear DNA-coded proteins. *Mol Biol Evol* 22:810–813
- Janke A, Erpenbeck D, Nilsson M, Arnason U (2001) The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. *Proc R Soc Lond B* 268:623–631
- Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J Mol Evol* 29:170–179
- Kumazawa Y (2004) Mitochondrial DNA sequences of five squamates: phylogenetic affiliation of snakes. *DNA Res* 11:137–144
- Kumazawa Y, Endo H (2004) Mitochondrial genome of the Komodo dragon: efficient sequencing method with reptile-oriented primers and novel gene rearrangements. *DNA Res* 11:115–125

- Kumazawa Y, Nishida M (1993) Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *J Mol Evol* 37:380–398
- Kumazawa Y, Nishida M (1995) Variations in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. *Mol Biol Evol* 12:759–772
- Kumazawa Y, Ota H, Nishida M, Ozawa T (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
- Kumazawa Y, Ota H, Nishida M, Ozawa T (1998) The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. *Genetics* 150:313–329
- Lee MSY (2000) Soft anatomy, diffuse homoplasy, and the relationship of lizards and snakes. *Zool Scripta* 29:101–130
- Lee MSY, Scanlon JD (2002) Snake phylogeny based on osteology, soft anatomy and ecology. *Biol Rev* 77:333–401
- Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4:203–221
- McDowell SB, Bogert CM (1954) The systematic position of *Lantlanotus* and the affinities of the anguimorph lizard. *Bull Am Mus Nat Hist* 105:1–142
- Pääbo S, Irwin DM, Wilson AC (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem* 265:4718–4721
- Palumbi S, Martin A, Romano S, McMillan WO, Stice L, Grabowski G (1991) The simple fool's guide to PCR. Version 2.0. Department of Zoology and Kewalo Marine Laboratory, University of Hawaii, Honolulu
- Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–818
- Rage JC (1987) Fossil history. In: Seigel RA, Collins JT, Novak SS (eds) *Snakes: Ecology and evolutionary biology*. Macmillan, New York, pp 51–76
- Rest JS, Ast JC, Austin CC, Waddell PJ, Tibbetts EA, Hay JM, Mindell DP (2003) Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol Phylogenet Evol* 29:289–297
- Rieppel O (1979) A cladistic classification of primitive snakes based on skull structure. *Z Zool Syst Evolforsch* 17:140–150
- Rieppel O, Zaher H, Tchernov E, Polcyn MJ (2003) The anatomy and relationships of *Haastophis terrasanctus*, a fossil snake with well-developed hind limbs from the mid-Cretaceous of the Middle East. *J Paleont* 77:536–558
- Saccone C, Gissi C, Reyes A, Larizza A, Sbisà E, Pesole G (2002) Mitochondrial DNA in metazoa: degree of freedom in a frozen event. *Gene* 286:3–12
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114–1116
- Slowinski JB, Lawson R (2002) Snake phylogeny: evidence from nuclear and mitochondrial genes. *Mol Phylogenet Evol* 24:194–202
- Strimmer K, von Haeseler A (1996) Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol* 13:964–969
- Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Uetz P (2003) The EMBL reptile database: [www.embl-heidelberg.de/uetz/Living Reptiles.html](http://www.embl-heidelberg.de/uetz/Living%20Reptiles.html)
- Underwood G (1967) A contribution to the classification of snakes. British Museum of Natural History, London
- Underwood G (1976) A systematic analysis of boid snakes. In: Bellairs Ad'A, Cox CB (eds) *Morphology and biology of reptiles*. Linn Soc Symp Ser No. 3. Academic Press, London, pp 151–175
- Vidal N, David P (2004) New insights into the early history of snakes inferred from two nuclear genes. *Mol Phylogenet Evol* 31:783–787
- Vidal N, Hedges SB (2002) Higher-level relationships of snakes inferred from four nuclear and mitochondrial genes. *CR Biologies* 325:977–985
- Wilcox TP, Zwickl DJ, Heath TA, Hillis DM (2002) Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol Phylogenet Evol* 25:361–371
- Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* 141:173–216
- Zardoya R, Meyer A (2000) Mitochondrial evidence on the phylogenetic position of caecilians (Amphibia: Gymnophiona). *Genetics* 155:765–775
- Zug GR, Vitt LJ, Caldwell JP (2001) *Herpetology*, Academic Press, San Diego