Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 33 (2004) 413–427
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Phylogeny of caecilian amphibians (Gymnophiona) based on complete mitochondrial genomes and nuclear RAG1

Diego San Mauro[a], David J. Gower[b], Oommen V. Oommen[c], Mark Wilkinson[b], Rafael Zardoya[a,*]

[a] *Departamento de Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, José Gutiérrez Abascal, 2, 28006 Madrid, Spain*
[b] *Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK*
[c] *Department of Zoology, University of Kerala, Kariavattom 695 581, Thiruvananthapuram, Kerala, India*

Received 15 January 2004; revised 20 May 2004
Available online 28 July 2004

Abstract

We determined the complete nucleotide sequence of the mitochondrial (mt) genome of five individual caecilians (Amphibia: Gymnophiona) representing five of the six recognized families: *Rhinatrema bivittatum* (Rhinatrematidae), *Ichthyophis glutinosus* (Ichthyophiidae), *Uraeotyphlus* cf. *oxyurus* (Uraeotyphlidae), *Scolecomorphus vittatus* (Scolecomorphidae), and *Gegeneophis ramaswamii* (Caeciliidae). The organization and size of these newly determined mitogenomes are similar to those previously reported for the caecilian *Typhlonectes natans* (Typhlonectidae), and for other vertebrates. Nucleotide sequences of the nuclear RAG1 gene were also determined for these six species of caecilians, and the salamander *Mertensiella luschani atifi*. RAG1 (both at the amino acid and nucleotide level) shows slower rates of evolution than almost all mt protein-coding genes (at the amino acid level). The new mt and nuclear sequences were compared with data for other amphibians and subjected to separate and combined phylogenetic analyses (Maximum Parsimony, Minimum Evolution, Maximum Likelihood, and Bayesian Inference). All analyses strongly support the monophyly of the three amphibian Orders. The Batrachia hypothesis (Gymnophiona, (Anura, Caudata)) receives moderate or good support depending on the method of analysis. Within Gymnophiona, the optimal tree (*Rhinatrema*, ((*Ichthyophis*, *Uraeotyphlus*), (*Scolecomorphus*, (*Gegeneophis Typhlonectes*)))) agrees with the most recent morphological and molecular studies. The sister group relationship between Rhinatrematidae and all other caecilians, that between Ichthyophiidae and Uraeotyphlidae, and the monophyly of the higher caecilians Scolecomorphidae + Caeciliidae + Typhlonectidae, are strongly supported, whereas the relationships among the higher caecilians are less unambiguously resolved. Analysis of RAG1 is affected by a spurious local rooting problem and associated low support that is ameliorated when outgroups are excluded. Comparisons of trees using the non-parametric Templeton, Kishino–Hasegawa, Approximately Unbiased, and Shimodaira–Hasegawa tests suggest that the latter may be too conservative.

Keywords: Gymnophiona; Amphibia; Molecular phylogenetics; Mitochondrial genome; RAG1

1. Introduction

Caecilians (Gymnophiona) are one of the three orders of recent Amphibia. They are readily distinguished from frogs and salamanders by their sensory tentacles and annulated, limbless bodies, and are distinct in many other characters (e.g., Himstedt, 1996; Noble, 1931; Taylor, 1968). Most of the approximately 160 currently recognized species (Frost, 2002; Nussbaum and Wilkinson, 1989) are tropical, soil-dwelling predators for at least their adulthood, but members of the South American family Typhlonectidae are semiaquatic or aquatic (Wilkinson and Nussbaum, 1999). Despite increasing evidence of high local abundance in some species (e.g., Gower et al., 2004; Measey et al., 2003), caecilians

* Corresponding author. Fax: +34-91-564-5078.
E-mail address: rafaz@mncn.csic.es (R. Zardoya).

1055-7903/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2004.05.014

remain probably the least known order of recent tetrapods. Although they are a relatively small group, it is clear that they have a remarkable morphological, ecological, and reproductive diversity (e.g., Gower et al., 2004; Loader et al., 2003; Wake, 1977; Wilkinson and Nussbaum, 1997).

Fossils of potential crown-group caecilians are represented only by isolated vertebrae (Evans and Sigogneau-Russel, 2001) so that inferring phylogenetic relationships among caecilians is essentially a neontological enterprise. Between 1968 and 1979, the previously single caecilian family Caeciliidae was partitioned into the six higher taxa (Nussbaum, 1977, 1979; Taylor, 1968, 1969) that are widely recognized as families today (Duellman and Trueb, 1986; Nussbaum and Wilkinson, 1989; Wilkinson and Nussbaum, 1999). Four caecilian families have relatively restricted distributions—the South American Rhinatrematidae (two genera, eight species) and Typhlonectidae (five genera, 13 species), Indian Uraeotyphlidae (one genus, five species), and African Scolecomorphidae (two genera, six species). Ichthyophiidae (two genera, 30+ species) occurs in South and South East Asia (West of Wallace's line). The more cosmopolitan Caeciliidae (21 genera, ca. 100 species) occurs on all land masses where caecilians are known except South East Asia. The current distribution of extant caecilians is consistent with a Gondwanan origin of the order (Duellman and Trueb, 1986; Hedges et al., 1993; Wilkinson et al., 2002). The four smaller and more local families represent morphologically distinctive caecilian clades (Nussbaum, 1977, 1979, 1985; Wilkinson and Nussbaum, 1999). In contrast, molecular data suggest that ichthyophiids might not be monophyletic (Gower et al., 2002), and morphology and molecules agree that the Caeciliidae, which comprises those caecilians that have not been assigned to the five more recently described families, is paraphyletic (Hedges et al., 1993; Nussbaum, 1979; Wilkinson, 1997; Wilkinson et al., 2003).

Nussbaum (1979) presented the first numerical phylogenetic analysis of caecilians, using morphological characters to investigate the interrelationships of 12 genera. This, and the subsequent analyses of Duellman and Trueb (1986) and Hillis (1991) that used family-level taxa and a subset of Nussbaum's (1979) characters, identified a clade comprising the caeciliids, typhlonectids, and scolecomorphids that Nussbaum (1991) dubbed the "higher" caecilians. The Uraeotyphlidae, Ichthyophiidae, and Rhinatrematidae were successively more distant outgroups to the higher caecilians in these analyses. Diverse morphological evidence that the Rhinatrematidae is the sister group of all other extant caecilians (Nussbaum, 1977; Wilkinson, 1992, 1996a) is considered to provide strong support for this hypothesis, which has been used to root caecilian phylogenetic trees in more recent morphological and molecular analyses (Gower et al., 2002; Wilkinson, 1997; Wilkinson and Nussbaum, 1996; Wilkinson et al., 2002, 2003). Wilkinson and Nussbaum (1996) and Wilkinson (1997) also supported the monophyly of the higher caecilians, but found strong support for an alternative arrangement of more deep-branching families, in which the Ichthyophiidae and Uraeotyphlidae are sister taxa. Whereas earlier family-level phylogenies (Duellman and Trueb, 1986; Hillis, 1991) recovered Caeciliidae and Typhlonectidae as more closely related to each other than to Scolecomorphidae, the most comprehensive morphological study to date (Wilkinson, 1997) was unable to resolve relationships among these higher caecilians.

Previous molecular analyses that have been informative regarding the relationships among caecilian families have used nucleotide sequences of mitochondrial (mt) cytochrome *b* and 12S and 16S rRNA genes (Gower et al., 2002; Hay et al., 1995; Hedges and Maxson, 1993; Hedges et al., 1993; Wilkinson et al., 2002, 2003). These have supported recent morphological analyses by recovering clades comprising Ichthyophiidae + Uraeotyphlidae and Nussbaum's (1991) higher caecilians (caeciliids, scolecomorphids, and typhlonectids) (Wilkinson et al., 2003), and a paraphyletic Caeciliidae (Hedges and Maxson, 1993; Hedges et al., 1993; Wilkinson et al., 2002, 2003). Wilkinson et al. (2003) carried out the only previous molecular analysis to include members of all six currently recognized families. In agreement with the most recent morphological investigation, their study suggested that Caeciliidae is paraphyletic with respect to perhaps Scolecomorphidae as well as Typhlonectidae. However, many relationships within the higher caecilians were not strongly supported, and they suggested that more molecular and morphological data were required to resolve these relationships.

We have determined the complete nucleotide sequences of the mt genomes of five caecilian species, and compared them with the only previously described caecilian mt genome, that of *Typhlonectes natans* (Zardoya and Meyer, 2000). The sampling includes one representative of each of the six currently recognized families. Our mitogenomic (Curole and Kocher, 1999) approach follows several recent studies (Cummings et al., 1995; Russo et al., 1996; Zardoya and Meyer, 1996b) that demonstrated the utility of large sequence data sets for establishing robust high-level phylogenetic inferences. To provide independent data from a different genome, we have also sequenced the nuclear gene RAG1, which has proven useful in inferring relationships among other major vertebrate lineages (e.g., Groth and Barrowclough, 1999; Martin, 1999). Through separate and combined analyses, we explore the utility of these data in establishing a robust higher-level phylogenetic framework for caecilians. The inclusion of comparable data for representatives of frogs, salamanders, and more dis-

tant outgroups, make our analyses relevant to the controversial relationships (Zardoya and Meyer, 2001) among the three recent amphibian orders.

2. Materials and methods

2.1. Taxon sampling

Our sampling includes one species from each of the six currently recognized caecilian families. The typhlonectid *T. natans* was chosen because of the availability of its mt genome sequence (Zardoya and Meyer, 2000). The other five species belong to the type genus (sometimes as the type species) of their respective families (Table 1). Caeciliid paraphyly means that this family is inadequately represented with a single species. The type genus, the Neotropical *Caecilia*, has been shown to be among those caeciliids most closely related to Typhlonectidae (Hedges et al., 1993; Wilkinson et al., 2002, 2003). In contrast, our chosen caeciliid, the Indian *Gegeneophis ramaswamii*, is more distantly related to the Neotropical typhlonectids than is *Caecilia* (Wilkinson et al., 2002), thus providing an opportunity to further explore the nature of caeciliid paraphyly.

For comparisons of mt genomes, we selected the following outgroups (GenBank accession numbers in parentheses): the anuran amphibians *Xenopus laevis* (NC_001573, Roe et al., 1985) and *Rana nigromaculata* (NC_002805, Sumida et al., 2001), the caudate amphibians *Mertensiella luschani* (NC_002756, Zardoya et al., 2003) and *Ranodon sibiricus* (NC_004021, Zhang et al., 2003), and two lobe-finned fishes, a coelacanth, *Latimeria chalumnae* (NC_001804, Zardoya and Meyer, 1997), and an African lungfish, *Protopterus dolloi* (NC_001708, Zardoya and Meyer, 1996a).

Outgroups for examination of RAG1 sequences were one anuran, *X. laevis* (L19324, Greenhalgh et al., 1993), two caudates, *Pleurodeles waltl* (AJ010258, Frippiat et al., 2001) and *Mertensiella luschani atifi* (Table 1), a coelacanth, *Latimeria menadoensis* (AY442925, Brinkmann et al., 2004); and an African lungfish, *P. dolloi* (AY442928, Brinkmann et al., 2004).

2.2. DNA extraction, PCR amplification, cloning, and sequencing

Total DNA was purified from ethanol-preserved liver or muscle, with a standard phenol/chloroform extraction procedure (Sambrook et al., 1989). A suite of 28 primers (Table 2) was used to amplify by PCR contiguous and overlapping fragments that covered the entire mt genome (Fig. 1). PCR amplifications were conducted in 25 μl reactions containing 67 mM Tris–HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.4 mM of each dNTP, 2.5 μM of each primer, template mtDNA (10–100 ng), and *Taq* DNA polymerase (1 U, Biotools), using the following cycling conditions: an initial denaturing step at 94 °C for 5 min; 35 cycles of denaturing at 94 °C for 60 s, annealing at 42–54 °C (see Table 2) for 60 s, and extending at 72 °C for 90 s; and a final extending step of 72 °C for 7 min. PCR products were purified by ethanol precipitation, and sequenced in an automated DNA sequencer (ABI PRISM 3700) using the BigDye Deoxy Terminator cycle-sequencing kit (Applied Biosystems) following manufacturer's instructions. For mtDNA, short amplicons were sequenced directly using the corresponding PCR primers. Long amplicons were cloned into pGEM-T vectors (Promega), and recombinant plasmids were sequenced using the M13 (forward and reverse) universal primers as well as additional walking primers (available from the authors upon request). The sequences obtained averaged 700 base pairs (bp) in length, and each sequence overlapped the next contig by about 150 bp. In no case were differences in sequence observed between the overlapping regions.

Four primers were designed in conserved regions of the RAG1 gene to amplify, by PCR, two contiguous

Table 1
Data for amphibian samples employed in this study

Species	Taxonomic assignment	Voucher No.	Collection locality	GenBank Accession Nos. (mt genomes, RAG1)
Gegeneophis ramaswamii	Gymnophiona: Caeciliidae	MW 331	Thenmalai, India	AY456250, AY456255
Ichthyophis glutinosus[c]	Gymnophiona: Ichthyophiidae	MW 1733	Peradeniya, Sri Lanka	AY456251, AY456256
Rhinatrema bivittatum[c]	Gymnophiona: Rhinatrematidae	BMNH 2002.6	Kaw, French Guyana	AY456252, AY456257
Scolecomorphus vittatus[c]	Gymnophiona: Scolecomorphidae	BMNH 2002.100	Amani, Tanzania	AY456253, AY456258
Uraeotyphlus cf. *oxyurus*[c]	Gymnophiona: Uraeotyphlidae	MW 212	Payyanur, India	AY456254, AY456259
Typhlonectes[c] *natans*	Gymnophiona: Typhlonectidae	BMNH 2000.218	Potrerito, Venezuela	AF154051,[a] AY456260
Mertensiella luschani atifi	Caudata: Salamandridae	—	Fersin, Turkey	NC_002756,[b] AY456261

BMNH, The Natural History Museum, London; MW, field series of the Zoology Department, University of Kerala (India); and the Department of National Museums, Colombo (Sri Lanka).

[a] Zardoya and Meyer (2000).

[b] Zardoya et al. (2003).

[c] Indicates type species/genus of family.

Table 2
Primers used to sequence the complete caecilian mt genomes (see Fig. 1 to trace fragments along the genome)

Fragment name	Primer name	Sequence	Approximate product length (bp)	Annealing temperature (°C) used in the PCR
12S	L1091[a]	5′-AAAAAGCTTCAAACTGGGATTAGATACCCCACTAT-3′	380	51
	H1478[a]	5′-TGACTGCAGAGGGTGACGGGCGGTGTGT-3′		
MID	Amp-12S F	5′-AAGAAATGGGCTACATTTTCT-3′	1200	50
	Amp-16S R	5′-AAGTGATTAYGCTACCTTTGCAC-3′		
16S	16Sar-L[b]	5′-CGCCTGTTTATCAAAAACAT-3′	500	51
	16Sbr-H[b]	5′-CCGGTCTGAACTCAGATCACGT-3′		
P1	MNCN-16S F[c]	5′-GGTTTACGACCTCGATGTTGGATC-3′	1350	42
	Lati-Met R[d]	5′-TCGGGGTATGGGCCCGAAAGCTT-3′		
P2	Amp-P2 F	5′-CAAYTAATRCAYCTAGTATGRAAAA-3′	2500	42
	Amp-P2 R	5′-ATATARCCAAAWGGTTCTTTTTT-3′		
P3	Amp-P3 F	5′-CAATACCAAACCCCCTTRTTYGTWTGATC-3′	900	45
	Amp-P3 R	5′-GCTTCTCARATAATAAATATYAT-3′		
P4	Amp-P4 F	5′-GGMTTTATTCACTGATTYCC-3′	1400	50
	Amp-P4 R	5′-AAATTGGTCAAAKAARCTTAGKRTCATGGTCA-3′		
P5	8.2 L8331[e]	5′-AAAGCRTYRGCCTTTTAAGC-3′	1590	54
	MNCN-COIII R[c]	5′-ACGTCTACRAARTGTCAGTATCA-3′		
P6	Amp-P6 F	5′-ACATGAGCYCAYCACAGYATTAT-3′	1440	50
	Amp-P6 R	5′-CGGGTAATAATAATTAATGTTGG-3′		
P7	Amp-P7 F	5′-AAYCTCCTACAATGYTAAAAAT-3′	1550	48
	Amp-P7 R	5′-CATARCTTTTACATGGATTTGCACC-3′		
P8	MNCN-His F[c]	5′-AAAACATTAGATTGTGATTCTAA-3′	1210	42
	Lati-ND5 R1[d]	5′-CCYATYTTTCKGATRTCYTGYTC-3′		
P9	Amp-P9 F	5′-AGCCARCTYGGCCTAATAATAGT-3′	1630	50
	Amp-P9 R	5′-CAGCCGTARTTTACGTCTCGRCAGAT-3′		
P10	MNCN-Glu F[c]	5′-GAAAAACCACCGTTGTTATTCAACTACA-3′	1170	48
	Amp-P10 R	5′-TTCAGYTTACAAGACYGATGCTTT-3′		
P11	Amp-P11 F	5′-TGRATYGGRGGCCAACCAGTAGAAGA-3′	1550	50
	Amp-12S R	5′-TCGATTATAGAACAGGCTCCTCT-3′		

[a] Kocher et al. (1989).
[b] Palumbi et al. (1991).
[c] Zardoya (Unpublished data).
[d] Zardoya and Meyer (1997).
[e] http://nmg.si.edu/bermlab.htm.

and overlapping fragments that cover a 1500 bp portion of the 3′ end of this gene: Amp-RAG1 F (5′-AGC TGC AGY CAR TAC CAY AAR ATG TA-3′), Amp-RAG1 R1 (5′-AAC TCA GCT GCA TTK CCA ATR TCA CA-3′), Amp-RAG1 F1 (5′-ACA GGA TAT GAT GAR AAG CTT GT-3′), and Amp-RAG1 R (5′-TTR GAT GTG TAG AGC CAG TGG TGY TT-3′). PCR mixtures and cycling conditions were as described above (annealing temperature was 54 °C). PCR products were cloned into pGEM-T vectors and sequenced using the M13 universal primers as described above.

All new nucleotide sequences reported in this paper have been deposited in GenBank under accession numbers given in Table 1.

2.3. Molecular and phylogenetic analyses

Nucleotide sequences (RAG1 gene) and the deduced amino acid sequences of all 13 mt protein-coding genes were aligned separately, using the default parameters of CLUSTAL X version 1.83 (Thompson et al., 1997), and the alignments revised by eye in an effort to maximize

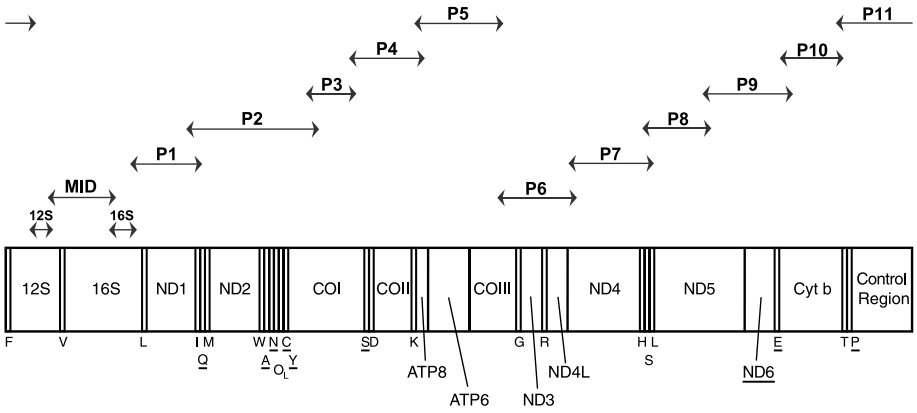


Fig. 1. Gene organization and sequencing strategy for the mt genomes of the caecilians. Genes encoded by the L strand are underlined. Arrow-headed segments denote the location of the fragments amplified by PCR with each pair of primers (see Table 2 for the primer DNA sequence associated with each fragment). *Gegeneophis ramaswamii* departs from this general consensus in lacking the $tRNA^{Phe}$ gene.

positional homology. Alignment ambiguities and gaps were excluded from phylogenetic analyses using GBLOCKS version 0.91b (Castresana, 2000). Alignments and data files are available from the authors upon request.

Four commonly used methods of phylogenetic inference, namely Maximum Parsimony -MP- (Fitch, 1971), Minimum Evolution -ME- (Rzhetsky and Nei, 1992), Maximum Likelihood -ML- (Felsenstein, 1981), and Bayesian Inference -BI- (Huelsenbeck et al., 2001) were applied separately to the RAG1 data and to a concatenated dataset of the amino acid sequences of the mt protein-coding genes. Bayesian and MP analyses of the combined RAG1 (except *Pleurodeles*) and mt amino acid sequence data (except *Ranodon* and *Rana*) were performed. Separate analyses using only the ingroup taxa (caecilian-only data) were also performed for both mt amino acid and RAG1 nucleotide data sets.

Quartet puzzling ML analyses of amino acid sequence data (100,000 puzzling steps) were conducted with TREE-PUZZLE version 5.0 (Strimmer and von Haeseler, 1996). ML analysis of RAG1, and all ME and MP analyses, were performed with PAUP* version 4.0b10 (Swofford, 1998), with 10 random addition sequences and TBR branch swapping. ME analyses of mt amino acid and nuclear DNA sequences used mean character and ML distances, respectively. BIs were made using MrBayes version 3.0b4 (Huelsenbeck and Ronquist, 2001) with four simultaneous chains, each of a million generations, sampled every 100 generations. Trees sampled before the cold chain reached stationarity, as judged by plots of ML scores, were discarded as "burn-in."

Following Yang et al. (1998), we used the mtREV24 model (Adachi and Hasegawa, 1996) in all likelihood and Bayesian analyses of amino acid data, and we employed likelihood ratio tests (LRTs) to select among the following hierarchically nested alternative models: equal rates (eq.) versus proportion of invariant sites (I), versus gamma-distributed rates (Γ), versus gamma-distributed rates and proportion of invariant sites (Γ+I). ML analyses of RAG1 sequences used the best-fit model of nucleotide substitution selected according to the Akaike information criterion (AIC) calculated using Modeltest version 3.4 (Posada and Crandall, 1998). For BI, best-fit models were selected for each RAG1 codon position and model parameters were estimated independently ("unlink" option). For the combined data Bayesian analysis, and for analyses of the caecilian-only data, best-fit models were re-estimated for each partition because of the exclusion of taxa.

Support was evaluated with non-parametric bootstrap proportions (BPs—1000 pseudoreplicates), Bayesian posterior probabilities (BPPs), and quartet puzzling proportions (QPs). Decay indices (*d*) were also calculated using AutoDecay version 5.04 (Eriksson, 2001). Approximately Unbiased -AU- (Shimodaira, 2002), Shimodaira–Hasegawa -SH- (Shimodaira and Hasegawa, 1999), Kishino–Hasegawa -KH- (Kishino and Hasegawa, 1989), and Templeton (Templeton, 1983) tests were used to evaluate the 105 alternative, fully resolved unrooted trees for the caecilian-only data. Templeton test (two tailed) was performed in PAUP*, whereas the other three tests were carried out using CONSEL version 0.1f (Shimodaira and Hasegawa, 2001) with site likelihoods calculated by p4 version 0.79 (Foster, 2003).

Substitution rates and among-site rate heterogeneities were compared among RAG1 (at both nucleotide and amino acid levels), each mt protein (at the amino acid level), and a concatenated data set including all mt proteins (at the amino acid level) using the same subset of taxa used in the combined BI (see above). BI (100,000 generations) was used to estimate substitution rate (measured as tree length -TL-) and among-site rate heterogeneity (α parameter of the gamma distribution).

ANOVA analyses were used to assess variations in substitution rates and among-site rate heterogeneities. Planned comparisons were used to examine contrasts between RAG1 and each mt protein, and between RAG1 and the concatenated mt data set. Statistical analyses were performed using STATISTICA version 6.0 (StatSoft Inc., 2001).

3. Results

3.1. Mitochondrial genome organization and structural features

The complete nucleotide sequences of the L strands of the mt genomes of five caecilians (*G. ramaswamii*, *Ichthyophis glutinosus*, *Rhinatrema bivittatum*, *Scolecomorphus vittatus*, and *Uraeotyphlus* cf. *oxyurus*) were determined. Total length ranged from 15,897 to 16,432 bp. As in *T. natans*, all five newly sequenced caecilian mt genomes encoded for two rRNAs, 22 tRNAs, and 13 protein-coding genes, with the single exception of *G. ramaswamii*'s lack of the $tRNA^{Phe}$ gene. In all cases, the organization (Fig. 1) conforms to the consensus mt gene arrangement for vertebrates (Jameson et al., 2003). Other notable distinct features are only found in non-coding regions.

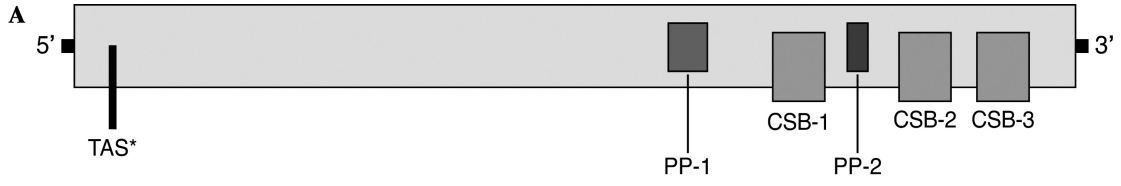
The control regions of the five new caecilian mt genomes are similar in length, ranging from 600 to 682 bp, and are also similar in structure and motifs (Fig. 2A). Three conserved blocks (CSB-1, CSB-2, and CSB-3, Walberg and Clayton, 1981) were identified at the 3′ end of each control region (Fig. 2B). Two polypyrimidine tracts, PP-1 and PP-2, were identified upstream from the CSB-2 and CSB-3 motifs (Fig. 2A). PP-1 consists of a stretch of thymines, and PP-2 is a poly(C) stretch located between CSB-1 and CSB-2 motifs. A putative termination-associated sequence (TAS) was found only in *S. vittatus*, close to the 5′ end of the control region. In contrast to *T. natans* (Zardoya and Meyer, 2000), no tandem repeats were found in the control regions of the newly sequenced caecilian mt genomes.

As in most vertebrates, the putative origin of L-strand replication (O_L) of the five new caecilian mt genomes was located within the WANCY tRNA cluster, between the $tRNA^{Asn}$ and $tRNA^{Cys}$ genes (Fig. 1). The O_L ranges from 30 to 39 bp and, in all five caecilians, has the potential to fold into a stem–loop secondary structure, sharing some nucleotides with the flanking tRNAs (Fig. 3). However, none of them can fold into alternative secondary structures with the adjacent $tRNA^{Cys}$ sequence such as reported for *T. natans* (Zardoya and Meyer, 2000). The 5′-GCCGG-3′ motif that in human mtDNA is involved in the transition from RNA synthesis to DNA synthesis (Hixson et al., 1986), is entirely conserved in the mtDNA of *R. bivittatum* and *U.* cf. *oxyurus*, whereas the remaining caecilian mt genomes show less conserved motifs (Fig. 3).

The mt genomes of *R. bivittatum* and *U.* cf. *oxyurus* have long non-coding regions between $tRNA^{Thr}$ and $tRNA^{Pro}$ genes of 312 and 437 bp, respectively. No secondary structures, tandem repeats, or functional ORFs were found in these intergenic regions, and BLAST searches produced no close matches. The non-coding spacer of *R. bivittatum* exhibits the same base composition as in the L strand of the whole mt genome, whereas in *U.* cf. *oxyurus* there is a much higher frequency of C (33%) and lower frequency of G (8%) than in the L strand.

3.2. RAG1 molecular features

All amphibian RAG1 sequences examined in this study are very conserved and show no indels. Absence of a single codon distinguishes amphibians from lobe-finned fishes. Overall base frequencies of sequences of



B

	CSB-1	CSB-2	CSB-3
Gegeneophis ramaswamii	TATATAAT-GAATGCTTGACGGACATA	TAAACCCCCC-TACCCCCC	CGTCAAACCCCAAAATCCA
Ichthyophis glutinosus	TAATAAAT-GAATGACTTATGGACATA	TAAACCCCCCTTACCCCCC	CGTCAAACCCCAAAATCCG
Rhinatrema bivittatum	T-ATATAT-GAATGATTTATTGACATA	CAAACCCCCCCT-CCCCCC	CGTCAAACCCCTAAATCCG
Scolecomorphus vittatus	TAATAAAT-GAATGTATTGAGGACATA	TAAACCCCCCTTACCCCCC	AGCTAAACCCCTAAATCTG
Uraeotyphlus cf. *oxyurus*	TTATAAAT-GAATGATTTATGGACATA	CAAACCCCCCCTACCCCCC	CGTCAAACCCCAAAATCCG
Typhlonectes natans	TCATAAATTGAATGTTTAATTGATACA	TAAACCCCCC-T--CCCCC	CGCTAAACCCCTAAATTCA

Fig. 2. Main features of the caecilian mtDNA control region. (A) Consensus structure of the control region. All caecilians have three conserved sequence blocks (CSB-1, 2, and 3) and two pyrimidine-rich regions (PP-1 and 2). TAS* is found only in *Scolecomorphus vittatus* and *Typhlonectes natans*. The latter taxon possesses, in addition, seven 109-bp tandem repeats in the right domain, close to the 3′ end. (B) Alignments of the identified CSBs in caecilians. Data for *Typhlonectes natans* are from Zardoya and Meyer (2000).

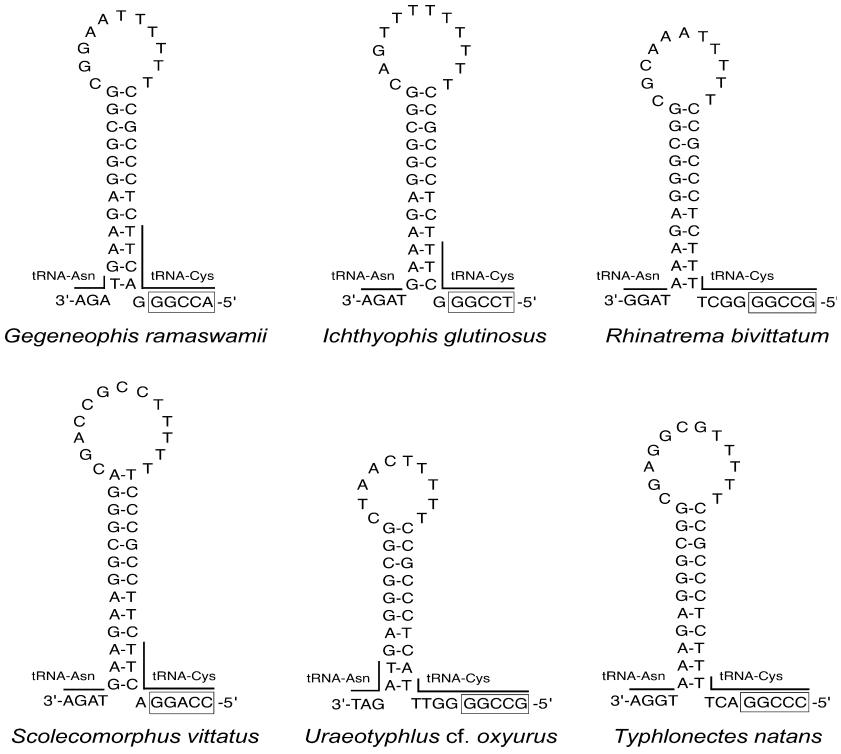


Fig. 3. Proposed secondary structures for the origins of L-strand replication (O_L) in caecilians. The 5′-GCCGG-3′ related motif is indicated by a box. Lines show the nucleotides partially shared with flanking tRNAs. Data for *Typhlonectes natans* are from Zardoya and Meyer (2000).

the frog and salamanders are fairly similar, but sequences of caecilians and lobe-finned fishes have a higher frequency of A (30.2–32.5%) and lower frequency of C (17.7–20.0%). Third codon positions in frogs, caecilians, and lobe-finned fishes have a G+C content of 31.8–47.6%, whereas the salamanders have a moderately strong G+C bias (61.0–62.8%), a difference that is significant in pairwise comparisons ($\chi^2_{5\%(30)} = 202.49$, $P < 0.001$). Pairwise differences among other codon positions across all taxa are not significant.

Average substitution rate of RAG1 is relatively low in comparison with those of most mt proteins (Fig. 4). RAG1 rates (at the nucleotide level) are between those of COII and COIII (at the amino acid level), whereas RAG1 rates at the amino acid level are between rates of COIII and COI (at the amino acid level) (Fig. 4). Relative substitution rates (TL) estimated for the different mt proteins are highly variable (Fig. 4). All statistical contrasts between RAG1 and each mt protein, and between RAG1 and the concatenated mt data set are highly significant ($F_{1,14400}$ values range from 17.79 to 75,809.88; $P < 0.001$ in all cases). For the caecilian-only data, all mt proteins (and the concatenated mt data set) have TL values about half those for the all-taxa data, whereas for RAG1 the differences in TL values are much greater (about 4.6 times at the nucleotide level and 3.6 times at the amino acid level).

Estimated among-site rate heterogeneities (α) are quite similar among the different protein data sets (Fig. 4). With the exception of ATP8 and ND3, α is less than two for all subsets of mt data. RAG1 among-site rate heterogeneity at the nucleotide level is 3.32 ± 0.09, lying between the values for COIII and ND3 (at the amino acid level). RAG1 among-site variation at the amino acid level is 1.17 ± 0.03, lying between the values for ND4 and ND4L (at the amino acid level) (Fig. 4). All statistical contrasts between α values for RAG1 at the nucleotide level and each mt data set are highly significant ($F_{1,14400}$ values ranged from 41.49 to 2579.20; $P < 0.001$ in all cases). At the amino acid level, only contrasts between RAG1 and ATP8, COIII, ND2, ND3, and ND6 are significant ($F_{1,14400}$ values ranged from 5.26 to 3826.08; $P < 0.05$ in all five cases).

3.3. Phylogenetic analyses

The deduced amino acid sequences of all 13 mt protein-coding genes of six caecilians, two salamanders, two frogs, and two lobe-finned fishes were combined into a single data set that produced an alignment of 3857 positions. Of these, 394 were excluded from the analyses because of alignment ambiguities, 1615 are invariant, and 1179 are parsimony informative. Within caecilians, the number of parsimony-informative sites is 518. Mean character distances among caecilians range from 0.15 (*Ichthyophis* vs. *Uraeotyphlus*) to 0.25 (*Ichthyophis* vs. *Typhlonectes*), and among amphibian orders from 0.23 to 0.32. MtREV24+Γ+I was selected as the best-fitting model.

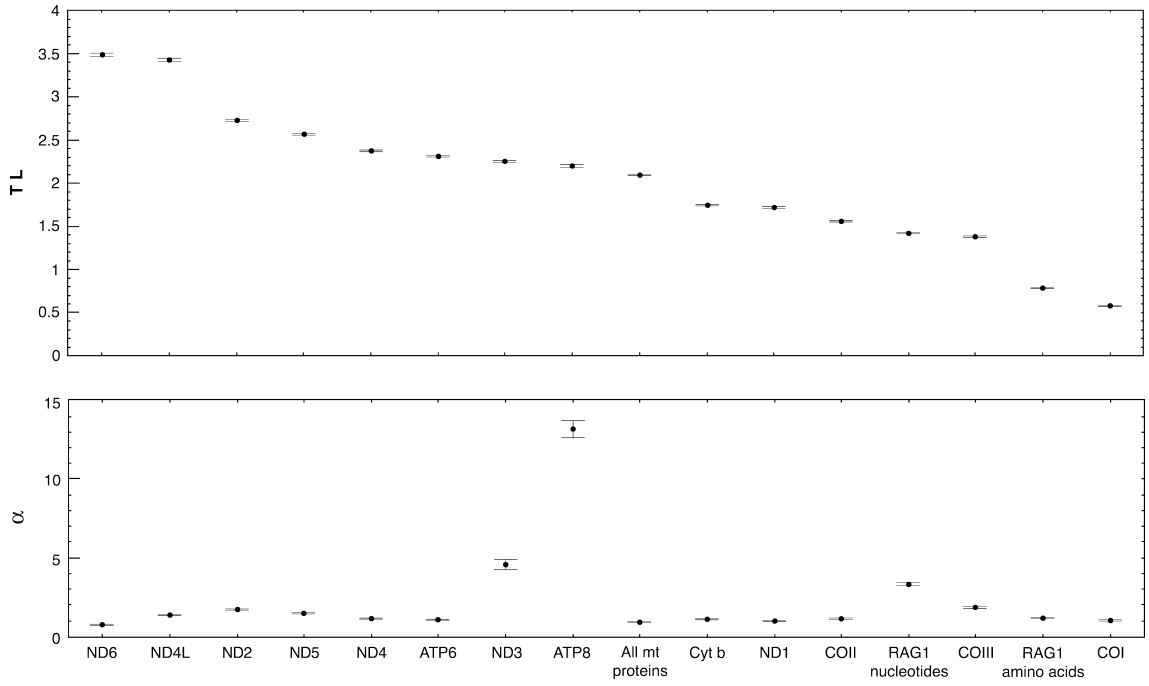


Fig. 4. Substitution rates (measured as Bayesian tree length -TL-) and among-site rate heterogeneities (α parameter of the gamma distribution) of RAG1 (at both nucleotide and amino acid levels), each mt protein (at the amino acid level), and a concatenated data set including all mt proteins (at the amino acid level).

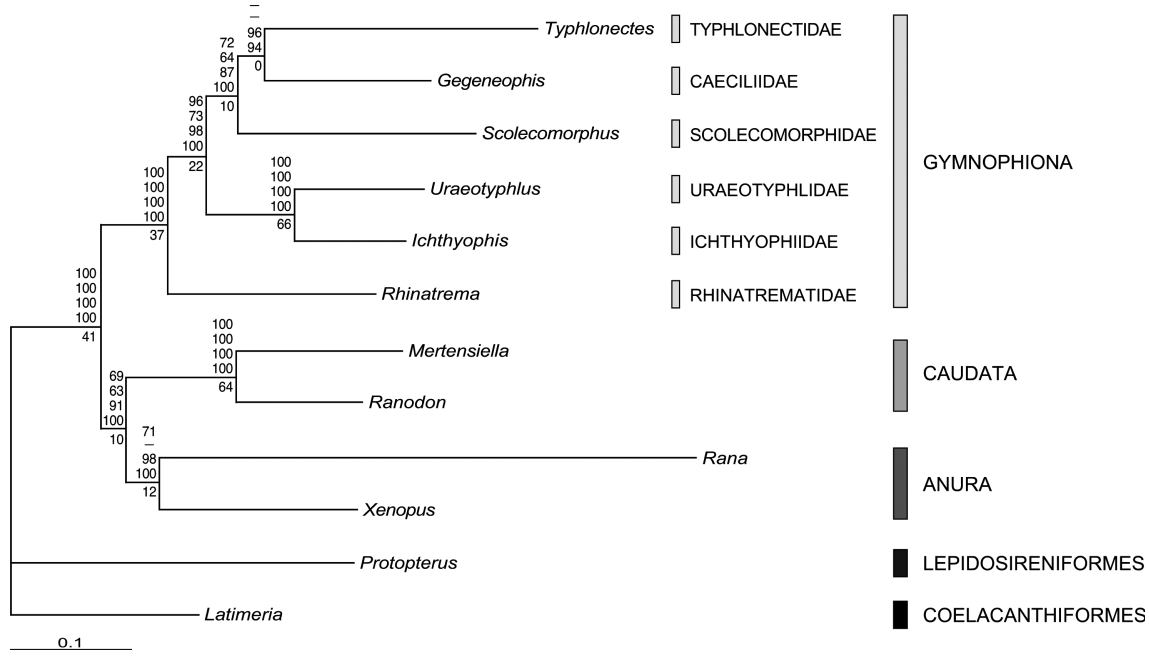


Fig. 5. Phylogenetic relationships (ML phylogram) of caecilians inferred from a single concatenated data set of the deduced amino acid sequences of all 13 mt protein-coding genes. Numbers above branches represent support for MP (BPs; upper value of each quartet), ME (mean character distances; BPs; middle-upper value), ML (mtREV24 + Γ + I model; QPs; middle-lower value), and BI (mtREV24 + Γ + I model; BPPs; lower value). Hyphens indicate support values of less than 50. Numbers below branches represent decay indices. Lobe-finned fishes (*Protopterus* and *Latimeria*) were used as outgroups.

All phylogenetic analyses, MP (6229 steps; CI = 0.71), ME (score = 1.43), ML (−Ln likelihood = 38653.67), and BI (−Ln likelihood = 38676.48) yielded the same inferred relationships with differences only in branch lengths and levels of support (Fig. 5). With all methods and measures, quantitative support for the monophyly of living

amphibians (Lissamphibia), salamanders (Caudata), and caecilians is maximal, whereas for the monophyly of frogs (Anura) and the sister group relationship between Anura and Caudata (the Batrachia hypothesis; Milner, 1988) BPPs are maximal and QPs are high (>90), but BPs are less impressive (63–71). Within caecilians, quantitative support for the sister group pairing of *Ichthyophis* and *Uraeotyphlus* is maximal with all methods. *Rhinatrema* was recovered as the sister group of all other caecilians, with only ME providing support values less than 95. Only BPPs and QPs indicate strong support for a higher-caecilian clade, comprising *Typhlonectes*, *Gegeneophis*, and *Scolecomorphus*, and for a sister group pairing of *Typhlonectes* and *Gegeneophis*.

RAG1 gene nucleotide sequences produced a raw alignment of 1512 positions. One gapped codon was excluded, and of the remaining positions, 795 are invariant and 503 parsimony-informative. Uncorrected "*p*" distances among caecilian taxa ranged from 0.04 (*Ichthyophis* vs. *Uraeotyphlus*) to 0.11 (*Rhinatrema* vs. *Scolecomorphus*), and among amphibian orders from 0.22 to 0.25. Interestingly, only 90 RAG1 positions are parsimony-informative among the sampled caecilians. Using Modeltest, we selected the parameter-rich GTR (Rodríguez et al., 1990) + Γ + I model of substitution for the ML and ME analyses. For the Bayesian analyses, the best-fit models selected were GTR + Γ for the first and the second position partitions and GTR + Γ + I for third positions.

All methods, ML (−Ln likelihood = 8379.63), MP (1507 steps; CI = 0.69), BI (−Ln likelihood = 8071.51), and ME (score = 1.50), produced single trees that differed only in the interrelationships among the caecilian lineages (Fig. 6) and which are otherwise congruent with the single tree inferred from the mt data. As with the mt data, there is maximal or very strong support for the monophyly of Lissamphibia, Caudata, and Gymnophiona, but quantitative support for the Batrachia hypothesis is less impressive and appears strong only with BPPs and MP BPs (Fig. 6A). Within caecilians, only the sister group relationship of *Ichthyophis* and *Uraeotyphlus* was consistently recovered by all methods of analysis. As with the mt data, quantitative support for this relationship is maximal or nearly so. MP, ML, and BI on the one hand, and ME on the other, yielded two different trees for caecilians (Fig. 6B). Of these, the ME tree is most similar to that inferred from the mt data, differing from it only in the resolution of the relationships of the higher caecilians *Gegeneophis*, *Scolecomorphus*, and *Typhlonectes*. MP, ML, and BI yielded a tree (Fig. 6B) that conflicts dramatically with the mt data (Fig. 5). In all analyses, quantitative support values for the non-con-

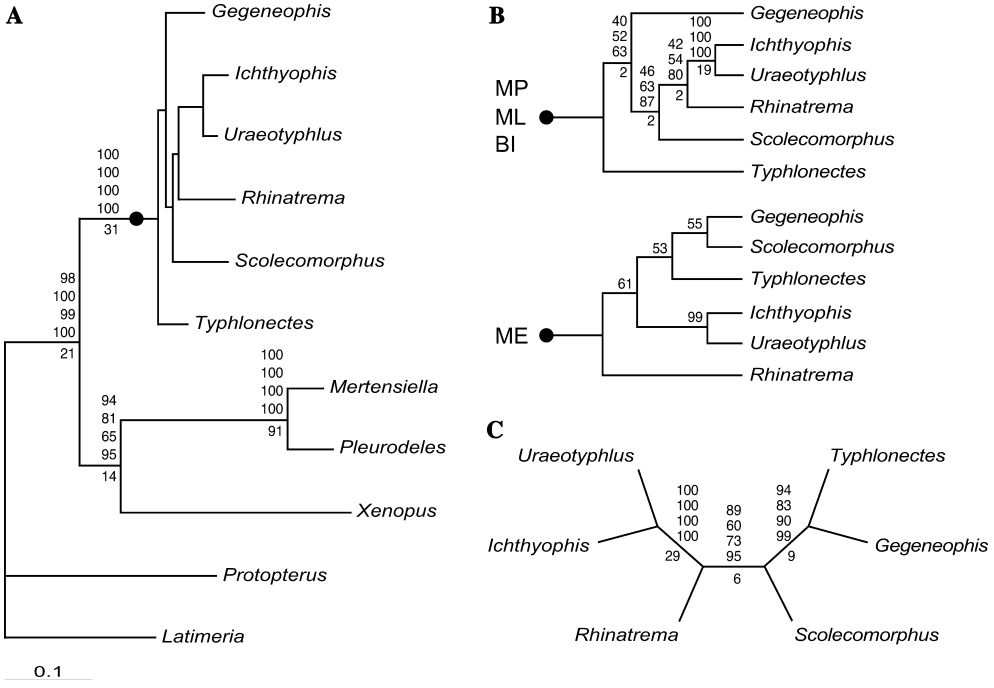


Fig. 6. Phylogenetic relationships of caecilian lineages inferred from RAG1 nucleotide sequence data. (A) ML phylogram showing statistical support for the clades outside Gymnophiona. Numbers above branches indicate support for MP (BPs; upper value of each quartet), ME (GTR + Γ + I distances; BPs; middle-upper value), ML (GTR + Γ + I model; BPs; middle-lower value), and BI (different models according to codon position, see text; BPPs; lower value). Lobe-finned fishes (*Protopterus* and *Latimeria*) were used as outgroups. (B) Inferred relationships and statistical support for each phylogenetic method for the clades within Gymnophiona. (C) Unrooted tree inferred from analysis of caecilian-only RAG1 data, with statistical support for the four methods given in same order as listed for (A). Numbers below branches represent decay indices in all trees.

gruent relationships are unimpressive. Much of the difference between the alternative trees for the RAG1 data is attributable to different rootings of the caecilian clade. To further explore this, we performed an unrooted analysis using the caecilian-only data. All methods yielded an unrooted tree (Fig. 6C) that is fully consistent with the tree supported by the mt data (Fig. 5) and measures of support are considerably enhanced compared to the corresponding splits in the analyses of the full RAG1 data.

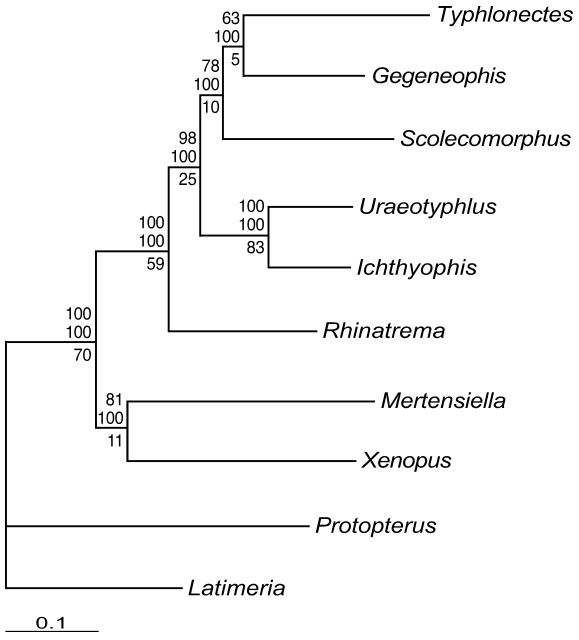


Fig. 7. Phylogenetic relationships (BI phylogram) of caecilians inferred from a combined data set of mt protein amino acids and RAG1 nucleotide sequences. Numbers above branches represent support for MP (BPs; upper value) and BI (mtREV24 + Γ + I model for the mt proteins, and different models according to codon position for the RAG1 gene, see text; BPPs; lower value). Numbers below branches represent decay indices. Lobe-finned fishes (*Protopterus* and *Latimeria*) were used as outgroups.

The combined mt and nuclear data comprised a total of 4991 sites. For BI, the best-fitting model for the mt amino acid partition was mtREV24 + Γ + I, and for the nucleotide RAG1 data were GTR + Γ for the first position, GTR + Γ + I for the second position, and GTR + Γ for the third position. BI and MP yielded the same tree (Fig. 7; −Ln likelihood = 41,461.80; 6551 parsimony steps, CI = 0.74). Relationships among the caecilians are identical to those recovered from the mt data using all methods (Fig. 5) and are fully consistent with the unrooted analysis of the RAG1 caecilian-only data. BPPs are maximal for all relationships in this tree. Parsimony BPs are maximal or nearly so for Lissamphibia, Gymnophiona, the sister group relationship of *Rhinatrema* to all other caecilians and the *Ichthyophis* + *Uraeotyphlus* pairing, substantial (>75) for the Batrachia hypothesis and the "higher" caecilians, but less impressive for the resolution of relationships within the higher caecilians.

Results of AU, SH, KH, and Templeton tests of alternative tree topologies, using caecilian-only data, are summarized in Table 3. Although unrooted, for convenience we describe these trees as if they were rooted on *Rhinatrema*. With either the mt or the RAG1 data, SH tests allow us to reject only trees that do not include the grouping of *Ichthyophis* and *Uraeotyphlus*. AU, KH, and Templeton tests also allow rejection of these trees but, in addition, allow rejection of some of the 15 trees

Table 3
Log likelihood and *p* values of Approximately Unbiased (AU), Shimodaira–Hasegawa (SH), Kishino–Hasegawa (KH), and Templeton tests for each of the 15 unrooted topologies that maintain the *Ichthyophis* + *Uraeotyphlus* pairing for the caecilian-only data

Alternative topologies	mt proteins					RAG1				
	−log L	AU	SH	KH	Templeton	−log L	AU	SH	KH	Templeton
(G,((((I,U),R),S),T))	23748.22	0.900	0.998	0.798	0.697	4094.83	0.902	0.994	0.781	1.000
(G,((((I,U),R),T),S))	23756.37	0.329	0.863	0.202	0.558	4101.83	0.093	0.677	0.077	0.029
(G,(((I,U),R),(S,T)))	23763.21	0.036	0.739	0.044	0.939	4105.61	0.094	0.677	0.077	0.012
(G,(((I,U),(R,S)),T))	23770.63	0.047	0.603	0.032	0.821	4106.19	0.080	0.810	0.126	0.033
(G,((((I,U),S),R),T))	23777.95	<0.001	0.471	0.003	0.125	4101.83	0.360	0.844	0.219	0.134
(G,((((I,U),T),R),S))	23778.06	0.041	0.467	0.040	0.287	4097.87	0.088	0.550	0.068	0.011
(G,(((I,U),(R,T)),S))	23779.09	0.036	0.456	0.033	0.346	4105.77	0.037	0.523	0.052	0.086
(G,(((I,U),T),(R,S)))	23786.58	0.032	0.335	0.015	0.286	4105.61	0.089	0.550	0.068	0.007
(G,((I,U),((R,S),T)))	23790.74	0.007	0.281	0.007	1.000	4107.66	0.004	0.493	0.031	0.007
(G,((I,U),((R,T),S)))	23791.23	0.024	0.275	0.007	0.431	4104.66	0.037	0.522	0.053	0.046
(G,((((I,U),T),S),R))	23793.72	0.002	0.247	0.004	0.158	4098.67	0.087	0.550	0.068	0.016
(G,((I,U),(R,(S,T))))	23794.61	0.002	0.236	0.003	0.305	4105.61	0.004	0.493	0.031	0.007
(G,(((I,U),S),(R,T)))	23796.24	0.002	0.216	0.002	0.127	4107.46	0.144	0.574	0.085	0.071
(G,(((I,U),(S,T)),R))	23799.82	0.005	0.177	0.001	0.031	4106.20	0.003	0.491	0.028	0.004
(G,((((I,U),S),T),R))	23802.56	0.007	0.156	<0.001	0.025	4107.46	0.034	0.541	0.057	0.009

The first topology corresponds to the optimal ML tree. Ninety of the 105 possible topologies are not shown because all four tests reject them at $p < 0.001$. G, *Gegeneophis*; I, *Ichthyophis*; R, *Rhinatrema*; S, *Scolecomorphus*; T, *Typhlonectes*; and U, *Uraeotyphlus*.

that place *Ichthyophis* and *Uraeotyphlus* together. With RAG1, AU tests allow us to reject six of these 15 trees, including all those in which *Gegeneophis* is more closely related to *Ichthyophis* and *Uraeotyphlus* than to any other caecilians. With the mt data, the AU tests allow us to reject all except one suboptimal tree, in which *Gegeneophis* is most closely related to *Scolecomorphus* rather than to *Typhlonectes*. KH test results match closely those of the AU test. Templeton test results match closely those of SH test with the mt data, but allow us to reject 11 of the 15 trees with RAG1.

4. Discussion

4.1. Distinct features of the new caecilian mitochondrial genomes

The new caecilian mt genomes are similar in size and gene arrangement to those of *T. natans* (Zardoya and Meyer, 2000), and thus conform to the vertebrate consensus organization (Jameson et al., 2003). The only exception is the mt genome of *G. ramaswamii*, which lacks the $tRNA^{Phe}$ gene. This presumably derived absence is unique among known vertebrate mt genomes. Absence of other tRNA genes has been previously reported in marsupials (Janke et al., 1997, 2002), and the tuatara (Rest et al., 2003). In marsupials, it has been shown that an alternative tRNA of nuclear origin is imported into mitochondria to participate in the translation process (Dorner et al., 2001). Given that the usage of phenylalanine in the mt proteins of *G. ramaswamii* is comparable to that in the other caecilians (not shown), an analogous importation may be implicated.

All caecilian mt control regions lack tandem repeats with the exception of that of *T. natans* (Zardoya and Meyer, 2000). The newly reported caecilian CSB-1 motifs are not reduced to a truncated pentamotif (5′-GACAT-3′) as in fishes (e.g., Hurst et al., 1999), but share high similarity with the mouse CSB-1 (Walberg and Clayton, 1981). A truncated CSB-1 was tentatively reported for *T. natans* (Zardoya and Meyer, 2000), but the alignment of all caecilian mt control regions allowed us to identify a complete CSB-1 motif in this species. One of the pyrimidine-rich regions, PP-1 (poly(T) stretch), has been previously described for several fishes (Hurst et al., 1999) and might be involved in regulatory aspects of the origin of H-strand replication. A second pyrimidine-rich region, PP-2 (poly(C) stretch), shows a moderately high similarity in most caecilians to the downstream CSB-2 motif, and could be the result of a former duplication. Except for *S. vittatus* and *T. natans*, the general absence of TAS (Doda et al., 1981) at the 5′ end of caecilian mt control regions contrasts with their presence and putative essential role in arresting replication in many other vertebrate mt genomes.

Unusually long intergenic spacers were found between the $tRNA^{Thr}$ and $tRNA^{Pro}$ genes in *R. bivittatum* and *U.* cf. *oxyurus*. Other cases of long intervening noncoding sequences have been reported in salamanders (Zardoya et al., 2003; Zhang et al., 2003), but not in frogs to date, and in all cases sequence similarities are low, suggesting that they are not homologous.

4.2. Comparative analysis of mitochondrial and RAG1 molecular features

Overall base compositions are biased against guanine in all caecilian mt genomes (not shown). This is a typical feature of vertebrate mtDNA, and is mainly due to a strong selection against the use of guanine at third codon positions of protein-coding genes (Zardoya and Meyer, 2000). In contrast, RAG1 gene sequences are homogenous in base composition, similar to what has been reported for this gene in birds and crocodiles (Groth and Barrowclough, 1999). The only exceptions are the salamander RAG1 sequences, which show a moderately strong G+C bias. This may reflect constraints in codon usage in this amphibian group. More amphibian RAG1 gene sequences (especially from anurans and salamanders) need to be determined to further investigate this distinctive condition.

Substitution rate of RAG1 at both nucleotide and amino acid levels was relatively slower than that of almost all mt proteins, being similar to those of the amino acid sequences of the most conservative mt protein-coding genes (cytochrome oxidase subunits, Zardoya and Meyer, 1996b). This makes RAG1 a potentially useful molecular marker for the study of deep vertebrate divergences. Among-site rate heterogeneity of RAG1 is quite similar to those of most mt proteins, being a little higher at the nucleotide level. Only ATP8 shows an unexpectedly high value of among-site rate heterogeneity, which is consistent with the fact that this is the shortest mt protein-coding gene, and shows few conserved positions across vertebrates (Zardoya and Meyer, 1996b).

4.3. Phylogenetics

All relevant analyses provide strong support for four uncontroversial high-level relationships—monophyly of Lissamphibia, Anura, Caudata, and Gymnophiona. Additionally, all relevant analyses are consistent with the Batrachia hypothesis (Gymnophiona, (Anura, Caudata)). This resolution of the Lissamphibia problem is the best hypothesis given the available data, but support is not consistently high in all analyses and, given the limited sampling of anuran and caudate taxa, it cannot be accepted without reservation.

With the exception of RAG1 only, all analyses strongly support the conventional view based on morphology that the Rhinatrematidae is the sister group

of all other caecilians. With outgroups included, RAG1 supports two alternative caecilian trees depending on the method of analysis. Only ME produces the expected sister group relationship between *Rhinatrema* and the other caecilians, with other methods recovering *Typhlonectes* as the sister taxon of other caecilians. In neither arrangement are support values for the basal split high, and low support also characterizes the other intracaecilian relationships except for the pairing of *Uraeotyphlus* and *Ichthyophis*. The latter is extremely well supported in all analyses, in statistical tests, and by previous analyses of morphology and molecules (see below). When outgroups are excluded from RAG1 analyses, all methods yield a single unrooted caecilian tree, fully consistent with relationships inferred from mt amino acid sequence data. This suggests that when outgroups are included there is a local rooting problem in the caecilian tree. Wilkinson (1996b) showed how unstable leaves (taxa) can decrease the bootstrap support for otherwise well-supported relationships. The present example is a special case in which the unstable "leaf" is the root of the caecilian tree. In order to investigate whether low support for the Batrachia hypothesis was a product of the instability of the root of the caecilian tree, we repeated the RAG1 analyses after exclusion of the higher caecilians. This had no substantial impact upon support values (not shown).

The sister group relationship between *Uraeotyphlus* and *Ichthyophis* appears to be the best supported relationship among the sampled caecilians. It is recovered in all analyses with maximal or near maximal support. Monophyly of the higher caecilians (*Gegeneophis*, *Scolecomorphus*, and *Typhlonectes*) is also supported in all analyses that were not affected by local rooting problems but measures of support, though generally high, are not universally high. Most uncertainty remains in the resolution of the higher caecilians. Mt amino acid and combined analyses place *Gegeneophis* with *Typhlonectes*, but with a mixture of high (BPP and QP) and low (MP and ME BPs) support.

There appears to be considerable uncertainty as to which of the various parametric and non-parametric likelihood-based tests are best used to determine whether the difference in fit of two or more trees to the data is significantly greater than expected under the null hypothesis of random sampling error. Although the KH test has been widely used, its validity requires the trees to be specified a priori rather than chosen on the basis of their likelihoods (Goldman et al., 2000). The SH test can be used to evaluate trees chosen a posteriori, but to be valid it requires the inclusion of all "reasonable" trees, and it is unclear how the set of reasonable trees can be selected (Buckley, 2002). On this point, Goldman et al. (2000) note only that selecting all possible trees will always be conservative, but this is an impractical selection for all but the smallest taxon samplings. Empirical comparisons of non-parametric SH tests, and of tests that use parametric bootstrapping, have provided very divergent results and quite different biological conclusions that suggest the SH test is very conservative because of the multiple comparisons, and that parametric bootstrapping may be too liberal as a result of model misspecification (e.g., Buckley, 2002; Goldman et al., 2000; Strimmer and Rambaut, 2001). A further uncertainty arises when the trees to be compared are chosen partly a priori and partly a posteriori, such as when we are interested in a putative monophyletic group but not in the resolution of relationships within that group. The more recently developed AU test is non-parametric and uses a multiscale bootstrap approach. It is less biased than other methods, but is also impractical when the number of trees to be compared is large (Shimodaira, 2002).

We used multiple non-parametric likelihood-based KH, SH, and AU tests and parsimony-based Templeton tests to further evaluate the strength of our inferences on caecilian relationships, and to provide an empirical comparison of the tests. We used the caecilian-only data because for six taxa there are only 105 possible unrooted trees, making the selection of all possible trees practical. Based on previous analyses of morphology and mt DNA sequence data, we expect two splits to be present in the caecilian tree, the pairing of *Uraeotyphlus* and *Ichthyophis*, and the partitioning of the higher caecilians, with particularly strong prior confidence in the former. Thus, based on a priori considerations, we are interested in comparing the three alternative resolutions of the higher caecilians. With RAG1, KH tests do not allow us to reject any of these alternatives, whereas with the more substantial mt amino acid data, KH tests allow the rejection of the grouping of *Scolecomorphus* with *Typhlonectes*, leaving a pairing of *Gegeneophis* with either *Typhlonectes*, as in the optimal tree, or with *Scolecomorphus* as viable alternative hypotheses. Ignoring a priori expectations and examining all 105 possible trees, SH tests are much less discriminatory. Using RAG1 or mt amino acid data, SH tests agree in rejecting only those trees that do not include the pairing of *Uraeotyphlus* and *Ichthyophis*. Using AU tests, these trees are also rejected, but RAG1 allows rejection of six additional trees (those that place *Gegeneophis* in a partition with *Uraeotyphlus* and *Ichthyophis*) and mt amino acid data allow rejection of all except one suboptimal tree (that placing *Gegeneophis* with *Scolecomorphus*).

Although the a posteriori SH tests provide strong support for our a priori confidence in the *Uraeotyphlus* and *Ichthyophis* pairing, the failure to discriminate against other hypotheses is disappointing given the amount of data available and the levels of support indicated by BPs, BPPs, and QPs. This suggests that the conservative SH test is too conservative.

Comparative results from the AU test bear this out. AU tests indicate that the mt amino acid data, in particular, strongly support the higher caecilian grouping also, and fail only to discriminate between the placement of *Gegeneophis* with *Typhlonectes* or with *Scolecomorphus* within the higher caecilians. Despite the concern that KH tests of trees that are not selected a priori are biased (Goldman et al., 2000), our KH test results are very similar to those obtained using the AU test. Good, but less tight correlation between KH and AU test results are reported for other data sets by Shimodaira (2002). The extent to which easily implemented KH tests may be a reasonable proxy for the more computationally demanding AU tests merits further investigation. Templeton test results are quite dissimilar between mt data and RAG1, seeming too conservative with the former and highly discriminative with the latter.

To summarize our phylogenetic investigations (using the initial letters to represent genera), we consider that the mt and nuclear data provide good support for (R,((I,U),(S,(G,T)))), and this is our preferred tree, although we do not discount (R,((I,U),(T,(S,G)))). Assuming that *Caecilia*, the type genus of the Caeciliidae, is more closely related to *Typhlonectes* than is *Gegeneophis* (Wilkinson et al., 2003) the latter tree would indicate that the Caeciliidae is paraphyletic with respect to the Scolecomorphidae as well as with respect to the Typhlonectidae.

Using mt ribosomal DNA sequence data, Wilkinson et al. (2003) were unable to resolve a number of relationships among the sampled caecilians. Our analyses demonstrate the potential of both mt protein gene and nuclear RAG1 data for providing well-supported resolution of caecilian phylogenetic relationships. Thus, expanded taxon sampling for these data is expected to provide much needed additional insights into caecilian phylogeny, particularly with respect to poorly understood relationships among the higher caecilians.

Acknowledgments

Lukas Rüber provided helpful technical advice with laboratory work and phylogenetic analyses. Peter Foster kindly assisted with the AU, KH, and SH tests. In addition to the many people acknowledged by Gower et al. (2002), Wilkinson et al. (2002, 2003), and Gower et al. (2004) for their help with fieldwork, we thank Tito Barros and Simon Loader. Two anonymous reviewers gave insightful comments on an earlier version of the manuscript. D.S.M. was sponsored by a predoctoral fellowship of the Ministerio de Ciencia y Tecnología of Spain. D.J.G and M.W. were granted a BIODIBERIA award to visit the MNCN. This work received financial support from a project of the Ministerio de Ciencia y Tecnología of Spain to R.Z. (CGL2004-00401). Fieldwork was funded in part by NERC GST/02/832.

References

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

Brinkmann, H., Venkatesh, B., Brenner, S., Meyer, A., 2004. Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. Proc. Natl. Acad. Sci. USA 101, 4900–4905.

Buckley, T.R., 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51, 509–523.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Cummings, M.P., Otto, S.P., Wakeley, J., 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. 12, 814–822.

Curole, J.P., Kocher, T.D., 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol. Evol. 14, 394–398.

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

Dorner, M., Altmann, M., Pääbo, S., Morl, M., 2001. Evidence for import of a lysyl-tRNA into marsupial mitochondria. Mol. Biol. Cell 12, 2688–2698.

Duellman, W.E., Trueb, L., 1986. Biology of Amphibians. McGraw-Hill, New York.

Eriksson, T., 2001. AutoDecay. Available from: <http://www.bergianska.se/index_forskning_soft.html>.

Evans, S.E., Sigogneau-Russel, D., 2001. A stem-group caecilian (Lissamphibia: Gymnophiona) from the Lower Cretaceous of North Africa. Palaeontology 44, 259–273.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Fitch, W.M., 1971. Toward defining the course of evolution: minimal change for a specific tree topology. Syst. Zool. 20, 406–416.

Foster, P., 2003. p4. Available form: <http://www.nhm.ac.uk/zoology/external/p4.htm>.

Frippiat, C., Kremarik, P., Ropars, A., Dournon, C., Frippiat, J.P., 2001. The recombination-activating gene 1 of *Pleurodeles waltl* (urodele amphibian) is transcribed in lymphoid tissues and in the central nervous system. Immunogenetics 52, 264–275.

Frost, D.R., 2002. Amphibian species of the World: an online reference. V2.21 (15 July, 2002). Available from: <http://research.amnh.org/herpetology/amphibia/index.html>.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Gower, D.J., Kupfer, A., Oommen, O.V., Himstedt, W., Nussbaum, R.A., Loader, S.P., Presswell, B., Müller, H., Krishna, S.B., Boistel, R., Wilkinson, M., 2002. A molecular phylogeny of ichthyophiid caecilians (Amphibia: Gymnophiona: Ichthyophiidae): out of India or out of South East Asia?. Proc. R. Soc. Lond. B 269, 1563–1569.

Gower, D.J., Loader, S.P., Moncrieff, C.B., Wilkinson, M., 2004. Niche separation and comparative abundance of *Boulengerula boulengeri* and *Scolecomorphus vittatus* (Amphibia: Gymnophiona) in an East Usambara forest, Tanzania. Afr. J. Herpetol. (in press).

Greenhalgh, P., Olesen, C.E., Steiner, L.A., 1993. Characterization and expression of recombination activating genes (RAG-1 and RAG-2) in *Xenopus laevis*. J. Immunol. 151, 3100–3110.

Groth, J.G., Barrowclough, G.F., 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. Mol. Phylogenet. Evol. 12, 115–123.

Hay, J.M., Ruvinsky, I., Hedges, S.B., Maxson, L.R., 1995. Phylogenetic relationships of amphibian families inferred from DNA sequences of mitochondrial 12S and 16S ribosomal RNA genes. Mol. Biol. Evol. 12, 928–937.

Hedges, S.B., Maxson, L.R., 1993. A molecular perspective on lissamphibian phylogeny. Herpetol. Monogr. 7, 27–42.

Hedges, S.B., Nussbaum, R.A., Maxson, L.R., 1993. Caecilian phylogeny and biogeography inferred from mitochondrial DNA sequences of the 12SrRNA and 16S rRNA genes (Amphibia: Gymnophiona). Herpetol. Monogr. 7, 64–76.

Hillis, D.M., 1991. The phylogeny of amphibians: current knowledge and the role of cytogenetics. In: Sessions, S.K., Green, D.M. (Eds.), Amphibian Cytogenetics and Evolution. Academic Press, San Diego, CA, pp. 7–31.

Himstedt, W., 1996. Die Blindwühlen. Westarp Wissenschaften, Magdeburg.

Hixson, J.E., Wong, T.W., Clayton, D.A., 1986. Both the conserved stem-loop and divergent 5′-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J. Biol. Chem. 261, 2384–2390.

Huelsenbeck, J.P., Ronquist, F.R., 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Huelsenbeck, J.P., Ronquist, F.R., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Hurst, C.D., Bartlett, S.E., Davidson, W.S., Bruce, I.J., 1999. The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. Gene 239, 237–242.

Jameson, D., Gibson, A.P., Hudelot, C., Higgs, P.G., 2003. OGRe: a relational database for comparative analyses of mitochondrial genomes. Nucleic Acids Res. 31, 202–206.

Janke, A., Magnell, O., Wieczorek, G., Westerman, M., Arnason, U., 2002. Phylogenetic analyses of 18S rRNA and the mitochondrial genomes of the Wombat, *Vombatus ursinus*, and the spiny anteater, *Tachyglossus aculeatus*: increased support for the Marsupionta hypothesis. J. Mol. Evol. 54, 71–80.

Janke, A., Xu, X., Arnason, U., 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationships among Monotremata, Marsupialia, and Eutheria. Proc. Natl. Acad. Sci. USA 94, 1276–1281.

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.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., Wilson, A.C., 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA 86, 6196–6200.

Loader, S.P., Wilkinson, M., Gower, D.J., Msuya, C.A., 2003. A remarkable young *Scolecomorphus vittatus* (Amphibia: Gymnophiona: Scolecomorphidae) from the North Pare Mountains, Tanzania. J. Zool. 259, 93–101.

Martin, A.P., 1999. Substitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). Mol. Biol. Evol. 16, 996–1002.

Measey, G.J., Gower, D.J., Oommen, O.V., Wilkinson, M., 2003. Quantitative surveying of limbless endogenic vertebrates—a case study of *Gegeneophis ramaswamii* (Amphibia: Gymnophiona) in southern India. Appl. Soil Ecol. 23, 43–53.

Milner, A.R., 1988. The relationships and origin of living amphibians. In: Benton, M.J. (Ed.), The Phylogeny and Classification of the Tetrapods. Clarendon Press, Oxford, pp. 59–102.

Noble, G.K., 1931. The Biology of the Amphibia. McGraw-Hill, New York.

Nussbaum, R.A., 1977. Rhinatrematidae: a new family of caecilians (Amphibia: Gymnophiona). Occ. Pap. Mus. Zool. Univ. Michigan 682, 1–30.

Nussbaum, R.A., 1979. The taxonomic status of the caecilian genus *Uraeotyphlus* Peters. Occ. Pap. Mus. Zool. Univ. Michigan 687, 1–20.

Nussbaum, R.A., 1985. Systematics of the caecilians (Amphibia: Gymnophiona) of the family Scolecomorphidae. Occ. Pap. Mus. Zool. Univ. Michigan 713, 1–49.

Nussbaum, R.A., 1991. Cytotaxonomy of caecilians. In: Sessions, S.K., Green, D.M. (Eds.), Amphibian Cytogenetics and Evolution. Academic Press, San Diego, CA, pp. 22–76.

Nussbaum, R.A., Wilkinson, M., 1989. On the classification and phylogeny of caecilians (Amphibia: Gymnophiona), a critical review. Herpetol. Monogr. 3, 1–42.

Palumbi, S.R., Martin, A., Romano, S., Owen MacMillan, W., Stice, L., Grabowski, G., 1991. The Simple Fool's Guide to PCR. Department of Zoology, University of Hawaii, Honolulu.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rest, J.S., Ast, J.C., Austin, C.C., Waddell, P.J., Tibbetts, E.A., Hay, J.M., Mindell, D.P., 2003. Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. Mol. Phylogenet. Evol. 29, 289–297.

Rodríguez, F., Oliver, J.F., Marín, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.

Roe, B.A., Din-Pow, M., Wilson, R.K., Wong, J.F., 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. J. Biol. Chem. 260, 9759–9774.

Russo, C.A.M., Takezaki, N., Nei, M., 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol. Biol. Evol. 13, 525–536.

Rzhetsky, A., Nei, M., 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. 9, 945–967.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of Log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shimodaira, H., Hasegawa, M., 2001. Consel: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

StatSoft Inc., 2001. STATISTICA (data analysis software system). http://www.statsoft.com.

Strimmer, K., Rambaut, A., 2001. Inferring confidence sets of possible misspecified gene trees. Proc. R. Soc. Lond. B 269, 137–142.

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

Sumida, M., Kanamori, Y., Kaneda, H., kato, Y., Nishioka, M., Hasegawa, M., Yonekawa, H., 2001. Complete nucleotide sequence and gene rearrangement of the mitochondrial genome of the Japanese pond frog *Rana nigromaculata*. Genes Genet. Syst. 76, 311–325.

Swofford, D.L., 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0. Sinauer Associates. Inc, Sunderland, MA, USA.

Taylor, E.H., 1968. The Caecilians of the World: A Taxonomic Analysis. University of Kansas Press, Lawrence, KS.

Taylor, E.H., 1969. A new family of African Gymnophiona. Univ. Kansas Sci. Bull. 48, 297–305.

Templeton, A.R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of human and the apes. Evolution 37, 221–244.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, J., Higgins, D.G., 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Wake, M.H., 1977. The reproductive biology of caecilians: an evolutionary perspective. In: Taylor, E.H., Guttman, S.I. (Eds.), Reproductive Biology of Amphibians. Plenum Press, New York, pp. 73–101.

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

Wilkinson, M., 1992. The phylogenetic position of the Rhinatrematidae (Amphibia: Gymnophiona): evidence from the larval lateral line system. Amphibia–Reptilia 13, 74–79.

Wilkinson, M., 1996a. The heart and aortic arches of rhinatrematid caecilians (Amphibia: Gymnophiona). Zoomorphology 105, 277–295.

Wilkinson, M., 1996b. Majority-rule reduced consensus methods and their use in bootstrapping. Mol. Biol. Evol. 13, 437–444.

Wilkinson, M., 1997. Characters, congruence and quality: a study of neuroanatomical and traditional data in caecilian phylogeny. Biol. Rev. 72, 423–470.

Wilkinson, M., Loader, S.P., Gower, D.J., Sheps, J.A., Cohen, B.L., 2003. Phylogenetic relationships of African caecilians (Amphibia: Gymnophiona): insights from mitochondrial rRNA gene sequences. Afr. J. Herpetol. 52, 83–92.

Wilkinson, M., Nussbaum, R.A., 1996. On the phylogenetic position of the Uraeotyphlidae (Amphibia: Gymnophiona). Copeia 1996, 550–562.

Wilkinson, M., Nussbaum, R.A., 1997. Comparative morphology and evolution of the lungless caecilian *Atretochoana eiselti* (Taylor) (Amphibia: Gymnophiona: Typhlonectidae). Biol. J. Linn. Soc. 62, 39–109.

Wilkinson, M., Nussbaum, R.A., 1999. Evolutionary relationships of the lungless caecilian *Atretochoana eiselti* (Amphibia: Gymnophiona: Typhlonectidae). Zool. J. Linn. Soc. 126, 191–223.

Wilkinson, M., Sheps, J.A., Oommen, O.V., Cohen, B.L., 2002. Phylogenetic relationships of Indian caecilians (Amphibia: Gymnophiona) inferred from mitochondrial rRNA gene sequences. Mol. Phylogenet. Evol. 23, 401–407.

Yang, Z., Nielsen, R., Hasegawa, M., 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol. Biol. Evol. 15, 1600–1611.

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

Zardoya, R., Meyer, A., 1996b. Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. Mol. Biol. Evol. 13, 933–942.

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

Zardoya, R., Meyer, A., 2000. Mitochondrial evidence on the phylogenetic position of Caecilians (Amphibia: Gymnophiona). Genetics 155, 765–775.

Zardoya, R., Meyer, A., 2001. On the origin of and phylogenetic relationships among living amphibians. Proc. Natl. Acad. Sci. USA 98, 7380–7383.

Zardoya, R., Malaga-Trillo, E., Veith, M., Meyer, A., 2003. Complete nucleotide sequence of the mitochondrial genome of a salamander, *Mertensiella luschani*. Gene 317, 17–27.

Zhang, P., Chen, Y.Q., Zhou, H., Wang, X.L., Qu, L.H., 2003. The complete mitochondrial genome of a relic salamander, *Ranodon sibiricus* (Amphibia: Caudata) and implications for amphibian phylogeny. Mol. Phylogenet. Evol. 28, 620–626.