Syst. Biol. 52(3):386–402, 2003

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150390197028

True and False Gharials: A Nuclear Gene Phylogeny of Crocodylia

JOHN HARSHMAN,[1,2] CHRISTOPHER J. HUDDLESTON,[1] JONATHAN P. BOLLBACK,[1,3,4] THOMAS J. PARSONS,[1,5] AND MICHAEL J. BRAUN[1]

[1]*Department of Systematic Biology, National Museum of Natural History, Smithsonian Institution, Suitland, Maryland 20746, USA; E-mail: braun@lab.si.edu (M.J.B)*
[2]*4869 Pepperwood Way, San Jose, California 95124, USA; E-mail: jharshman@pacbell.net*
[3]*Department of Biology, University of Rochester, Rochester, New York 14627, USA*
[4]*Section of Ecology, Behavior and Evolution, Division of Biological Sciences, University of California, San Diego, La Jolla, California 92093, USA*
[5]*U.S. Armed Forces DNA Identification Laboratory, Armed Forces Institute of Pathology, Rockville, Maryland 20850, USA*

Abstract.— The phylogeny of Crocodylia offers an unusual twist on the usual molecules versus morphology story. The true gharial (*Gavialis gangeticus*) and the false gharial (*Tomistoma schlegelii*), as their common names imply, have appeared in all cladistic morphological analyses as distantly related species, convergent upon a similar morphology. In contrast, all previous molecular studies have shown them to be sister taxa. We present the first phylogenetic study of Crocodylia using a nuclear gene. We cloned and sequenced the c-*myc* proto-oncogene from *Alligator mississippiensis* to facilitate primer design and then sequenced an 1,100-base pair fragment that includes both coding and noncoding regions and informative indels for one species in each extant crocodylian genus and six avian outgroups. Phylogenetic analyses using parsimony, maximum likelihood, and Bayesian inference all strongly agreed on the same tree, which is identical to the tree found in previous molecular analyses: *Gavialis* and *Tomistoma* are sister taxa and together are the sister group of Crocodylidae. Kishino–Hasegawa tests rejected the morphological tree in favor of the molecular tree. We excluded long-branch attraction and variation in base composition among taxa as explanations for this topology. To explore the causes of discrepancy between molecular and morphological estimates of crocodylian phylogeny, we examined puzzling features of the morphological data using a priori partitions of the data based on anatomical regions and investigated the effects of different coding schemes for two obvious morphological similarities of the two gharials. [c-*myc*; crocodylia; data conflict; data partitions; partitioned likelihood analysis; phylogeny; rooting.]

The story of molecules versus morphology is a familiar one in phylogenetics. Analysis of a molecular data set can indicate that a group seemingly well supported by morphological data is instead polyphyletic and that its apparent synapomorphies must be explained as functional convergences (e.g., Graur et al., 1991; McCracken et al., 1999). The controversy regarding Crocodylia is different. The true gharial (*Gavialis gangeticus*) and the false gharial (*Tomistoma schlegelii*) are among the most morphologically specialized of living crocodylians, having highly elongated, narrow snouts. However, specializations of this sort have evolved several times within Crocodyliformes (Clark, 1994; Brochu, 2001), and the similarities between the two gharials have traditionally been considered examples of convergence. All cladistic morphological analyses (Norell, 1989; Poe, 1996; Salisbury and Willis, 1996; Brochu, 1997, 1999a; Buscalioni et al., 2001) have confirmed that assessment, with the true gharial the living sister group of all other extant crocodylians and the false gharial the sister group of the two crocodile genera *Crocodylus* and *Osteolaemus* (Fig. 1a). In contrast, all previous molecular analyses (Densmore, 1983; Densmore and Dessauer, 1984; Densmore and White, 1991; Gatesy and Amato; 1992; Hass et al., 1992; Gatesy et al., 1993; Aggarwal et al., 1994; Poe, 1996; White and Densmore, 2001, unpubl.) have agreed in placing the true gharial as the sister of the false gharial and both gharials nested well within Crocodylia (Fig. 1b). The analyses of Gatesy et al. (2003; this issue) also produced the molecular topology.

Unlike many other molecules versus morphology conflicts, this disagreement has survived multiple reconsiderations of the data (Poe, 1996; Brochu, 1997; Brochu and Densmore, 2001). Thus, the conflict is not due to obviously flawed analyses or to indecisive data. Many of the molecular data sets have been criticized for using inherently phenetic data or phenetic analyses (Norell, 1989), but we see no a priori problem with such analyses as long as their assumptions are clear. Poe (1996) used parsimony to reanalyze those data sets that could be turned into discrete characters, and topology did not change. A more serious criticism (Norell, 1989) is the lack of outgroup rooting in some molecular studies. Although all these studies support the molecular ingroup topology, unless some limit on variation in molecular evolutionary rates is assumed, trees from these studies cannot be rooted. In most cases, small deviations from a perfect molecular clock would be sufficient to reroot the tree, producing, for example, root 5 of Figure 1d. The two gharials are still sisters on this tree, however. In all molecular data sets, the distance between *Tomistoma* and *Gavialis* is small compared with the distance between either of those taxa and any other crocodylian. Large differences in evolutionary rates would be required to root the tree such that they are not sister taxa. Still, without an outgroup to test for rate variation, such large differences cannot formally be ruled out.

All studies of mitochondrial DNA (mtDNA) sequence data (Gatesy and Amato, 1992; Gatesy et al., 1993; Aggarwal et al., 1994; Poe, 1996; White and Densmore, 2001, unpubl.) have included outgroups, and all have supported the molecular tree. However, one of White and Densmore's (unpubl.) analyses produced a different result (root 4, Fig. 1c). Rooting is a difficulty in molecular analyses, because the closest extant outgroup, birds, is

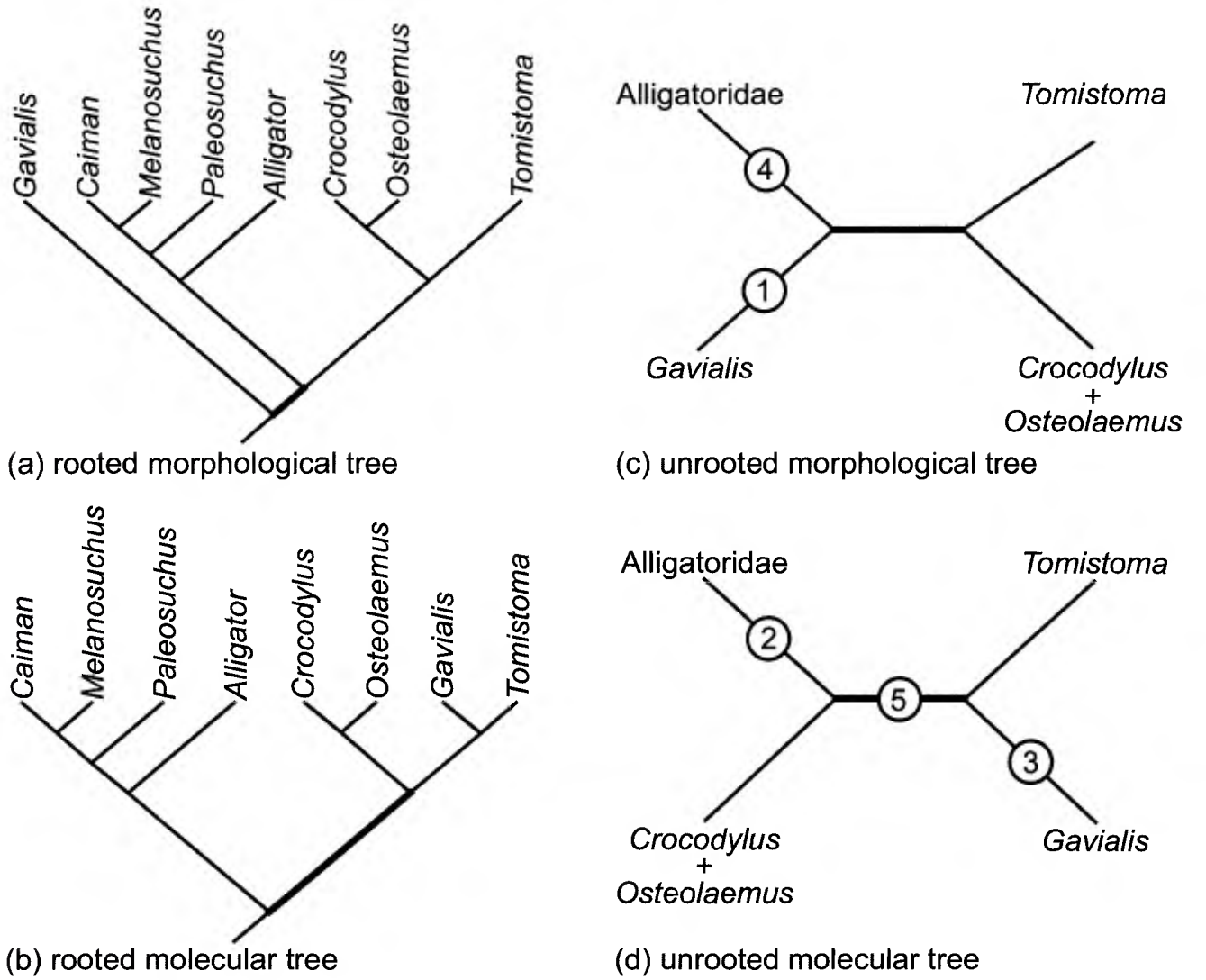

FIGURE 1. Alternative phylogenies of Crocodylia. Rooted (a) and unrooted (c) morphological topologies are compared with rooted (b) and unrooted (d) molecular topologies. Disagreements among topologies concern the existence of the bold branches. Alternative rootings are indicated by circled numbers: 1 = the morphological root (as in a); 2 = the molecular root (as in b); 3 = root from a combined analysis of morphological and molecular data (Brochu, 1997); 4 = root from a maximum likelihood analysis of mitochondrial sequences (White and Densmore, unpubl.); 5 = root from a different combined analysis of morphological and molecular data (Poe, 1996) and from a maximum likelihood analysis of the nuclear gene *RAG-1* (Gatesy et al., 2003).

extremely distant from the ingroup; thus, the branch connecting the ingroup and the outgroup is very long. It might be argued that mtDNA sequences retain relatively little phylogenetic signal at such high divergences. Gatesy et al.'s (2003) analyses of nuclear and mitochondrial data also resulted in the molecular tree, although maximum likelihood analysis of their nuclear data alone rooted the tree differently (root 5, Fig. 1d). Thus, the weight of evidence from all previous molecular studies strongly contradicts the morphological ingroup topology and supports the molecular ingroup topology but is less conclusive about root placement.

Combined analyses of morphological and molecular data (Poe, 1996; Brochu, 1997; Brochu and Densmore, 2001; Gatesy et al., 2003) have produced either the morphological tree or a topological compromise; i.e., the molecular ingroup topology with either a morphological root (root 3, Fig. 1d) or a compromise root (root 5, Fig. 1d). Such compromises are to be expected from a combination of strongly conflicting data sets (e.g., Swofford, 1991; Bull et al., 1993). More conclusive data are needed to help choose among alternative topologies and particularly to root the tree.

Here, we present a new DNA sequence data set based on an 1,100-base pair (bp) fragment of the proto-oncogene c-*myc*, sampling all extant crocodylian genera and several avian outgroup species. The structure, function, and expression of c-*myc* have been intensely studied because of the central role of this gene in cancer biology (Marcu et al., 1992; Ryan and Birnie, 1996; Prendergast,

1997). Its coding sequence is phylogenetically informative for deep nodes in vertebrates (Braun et al.,1985; Graybeal, 1994; Ericson et al., 2000; Miyamoto et al., 2000; Irestedt et al., 2001; Johansson et al., 2001), as are its noncoding sequences (Mohammad-Ali et al., 1995). The noncoding sequences are easily alignable and rapidly evolving, and have most sites free to vary, and thus actually have a much higher information content than an equal length of coding sequence. Additionally, the noncoding regions have many small insertion/deletion (indel) regions, which are much less susceptible to homoplasy than are nucleotide substitutions and are especially valuable indicators of phylogeny (van Dijk et al., 1999; Ericson et al., 2000). The superior performance of noncoding sequences in intermediate-level vertebrate phylogenetics has been noted by others (Prychitko and Moore, 1997) and is probably a general phenomenon.

The c-*myc* data set is highly informative for crocodylian relationships and appears to provide a robust resolution of the position of *Gavialis*. Through analysis of the new data set and reconsideration of the previous data, we addressed several questions. What is the unrooted ingroup topology within Crocodylia? What is the proper placement of the root on this topology? What biases or artifacts would cause molecular or morphological data to give false answers, and can we eliminate all these potential explanations? We consider these latter questions separately for the unrooted tree and for root placement.

Materials and Methods

Sampling, Molecular Methods, and Data Sets

L. Densmore and H. Dessauer kindly provided crocodylian DNA samples (Table 1). Avian DNA or tissue samples for use as outgroups were obtained from genetic resource collections. DNA was isolated from tissues by standard protocols involving proteinase K digestion, phenol–chloroform extraction, and ethanol precipitation (Sambrook et al., 1989).

Cloning.—A genomic library of *Alligator mississippiensis* DNA was prepared by partial digestion of high-molecular-weight DNA with *Sau*3AI and ligation into the *Bam*HI site of the phage lambda vector EMBL3A (Frischauf et al., 1983; Sambrook et al., 1989). After in vitro packaging of the ligated DNA, about 200,000 recombinant phage were plated at high density, transferred to nitrocellulose filter lifts, and screened by hybridization to a ^{32}P-labeled probe derived from a *Sal*I–*Pst*I fragment of chicken (*Gallus gallus*) v-*myc* corresponding to c-*myc* exon 3 (Alitalo et al., 1983). Positive clones were selected and plaque purified by screening at lower density. A 3-kilobase c-*myc*-positive *Bam*HI–*Sal*I fragment was

TABLE 1. Crocodylian and avian taxa examined.

Scientific (common) names	Tissue no.[a]	Source and Voucher no.[a]	Origin	GenBank no.
Alligator mississippiensis (American alligator)	Uncataloged[b]	LSU HSC	Louisiana	AY277494
Caiman yacare (yacare caiman)	LLD 101788-5	Jumbolair, Inc., Ocala, FL, USA	captive	AY277491
Crocodylus cataphractus (African slender-snouted crocodile)	LLD 031787-2	MMZ	captive	AY277487
Gavialis gangeticus (true gharial)	LLD 100187-1	NZP	captive	AY277490
	LLD 100187-4	NZP	captive	AY277490
Melanosuchus niger (black caiman)	LLD 093087-1	NZP	captive	AY277492
Osteolaemus tetraspis (dwarf crocodile)	Y21 (from LLD)	T. Cullen, Milwaukee, WI, USA	captive	AY277488
Paleosuchus palpebrosus (Cuvier's dwarf caiman)	LLD 10987	NZP	captive	AY277493
Tomistoma schlegelii (false gharial)	LLD 42290-1	MMZ	captive	AY277489
Anseranas semipalmata (magpie goose)	USNM B2954	USNM 621019	captive	AY277499
Casuarius casuarius (southern cassowary)	LSU MNS B10202	LSU MNS 136437	captive	AY277495
Coragyps atratus (black vulture)	USNM B1320	USNM 613353	Panama	AY277498
Musophaga violacea (violet turaco)	FMNH 396418	LSU MNS 168419	Ghana	AY277500
Ortalis cinereiceps (grey-headed chachalaca)	USNM B1282	USNM 613360	Panama	AY277497
Struthio camelus (ostrich)	LSU MNS B8609	SDZ	captive	AY277496

[a]LSU HSC = Herbert C. Dessauer, Health Sciences Center, Louisiana State University, New Orleans; LLD = Llewellyn L. Densmore, Texas Tech University, Lubbock; LSU MNS = Museum of Natural Science, Louisiana State University, Baton Rouge; USNM = National Museum of Natural History, Smithsonian Institution, Washington, D.C., FMNH = Field Museum of Natural History, Chicago, IL; NZP = National Zoological Park, Smithsonian Institution, Washington, D.C.; MMZ = Miami Metro Zoo, Miami, FL; SDZ = San Diego Zoo, San Diego, CA.

[b]Pooled blood sample from several individuals for genomic library preparation.

TABLE 2. Oligonucleotide PCR and sequencing primers used for crocodylian (C) and avian (A) sequences.

Primer	Taxon	Sequence
MYC-F-26	C	5′-GACGACTCCACCCCATGCAGGAGC-3′
MYC-F-18	C	5′-CAAGCCCAGAATGAGGTCC-3′
MYC-F-03	A, C	5′-AGAAGAAGAACAAGAGGAAG-3′
MYC-R-27	C	5′-CCATTCACCTCTACTAGCGTGAC-3′
MYC-R-25	C	5′-GACTCTGTGCCGGATTCAAC-3′
MYC-F-19	C	5′-AAAGAGGTTAAAATTGGAC-3′
MYC-R-20	C	5′-GGGGACTTGAGCACTTTCTG-3′
MYC-F-41	C	5′-CCAAAGTCGTCATCCTT-3′
MYC-F-42	C	5′-GAAAGAGCAGTTGAGGAGGC-3′
MYC-R-01	A, C	5′-CCAAAGTATCAATTATGAGGCA-3′
MYC-F-01	A	5′-TAATTAAGGGCAGCTTGAGTC-3′
MYC-F-02	A	5′-TGAGTCTGGGAGCTTTATTG-3′
MYC-F-43	S[a]	5′-TTAAAGGAAAAGCTCTCCAGG-3′
MYC-R-09	A	5′- CTTCYTCTTGTTCTTCYTCT -3′
MYC-R-04	A	5′-GGCTTACTGTGCTCTTCT-3′
MYC-R-22	S	5′-TGGTTCTGTGCAAATAAAGA-3′
MYC-F-04	A	5′-AAAAGGCTAAAGTTGGAC-3′
MYC-R-04	A	5′-CATTTTCGGTTGTTGCTG-3′
MYC-R-21	S	5′-CATTTGCGGTTATTACTG-3′
MYC-F-06	S	5′-AAGGTTGTCATCCTGAAAAAAGC-3′
MYC-F-07	A	5′-AGAGAAAGAGCAGTTGAGG-3′
MYC-F-05	A	5′-CACAAACTYGAGCAGCTAAG-3′
MYC-R-06	A	5′-TTAGCTGCTCAAGTTTGTG-3′
MYC-R-08	A	5′-TTGAGGTTCTGGCCAAACCCTT-3′
MYC-R-02	A	5′-TGAGGCAGTTTTGAGGTTCT-3′

[a]Used only for *Struthio camelus*.

subcloned into the plasmid vector Bluescript SK+ and sequenced on both strands by primer walking. This fragment was truncated by cloning at a *Sau*3A I site near the end of the coding sequence, so the sequence was completed by analysis of a polymerase chain reaction (PCR)-derived fragment.

PCR amplification of crocodylian sequences.—PCR and sequencing primers (Table 2) were designed from the *Alligator* c-*myc* sequence by targeting regions conserved in *Alligator* and chicken. The 1,100-bp focal region was amplified from other crocodylians in two overlapping fragments of 690 bp (MYC-F-26 to MYC-R-20) and 813 bp (MYC-F-03 to MYC-R-01) using standard methods. A 100-μl reaction mixture containing 1.5 mM $MgCl_2$, 0.2 mM each dNTP, 0.3 μM each primer, 6.25 U of *Taq* polymerase (Promega, Madison, WI), and 100 ng of DNA template was used with the following cycling parameters in a Robocycler (Stratagene, La Jolla, CA) with a thermal cover: one cycle of 95°C for 3 min, 35 cycles of 95°C for 45 sec, 48°C or 50°C for 45 sec, and 72°C for 1 min, and a final 72°C extension for 7 min. For a number of species, multiple bands were observed, and target fragments were gel purified prior to sequencing. All other products were PEG/NaCl precipitated prior to sequencing.

PCR amplification of bird sequences.—Avian PCR and sequencing primers were designed by comparison of the published chicken and canary (*Serinus canaria*; Collum et al., 1991) sequences. The 1,100-bp focal region was amplified in a single fragment using primers MYC-F-01 and MYC-R-01 (Table 2). PCR conditions differed slightly among taxa. Some were amplified in a Robocycler with a thermal cover with these parameters: one cycle of 95°C for 3 min, 30 cycles of 95°C for 50 sec, 45°C for 50 sec, and 72°C for 80 sec, and a final 72°C extension for 7 min. Others were amplified on a Perkin-Elmer 480 cycler with one cycle at 95°C for 3 min, 30 cycles of 95°C for 30 sec, 53°C for 30 sec, and 72°C for 1 min, and a final 72°C extension for 7 min. All reaction mixtures contained 0.025 U/μl of *Taq* polymerase, 0.2 mM each dNTP, 0.75 μM each primer, and 0.05 mg/ml bovine serum albumin. Robocycler reactions had 1.5 mM $MgCl_2$, and PE 480 reactions used Hot Beads (Lumitekk, Salt Lake City, UT) to provide a hot start (2.5 mM $MgCl_2$). Initial amplifications resulted in low yield for some taxa; these products were gel purified and reamplified with the same reaction conditions and primers. *Casuarius* and *Struthio* exhibited length polymorphism in a poly-T tract at the 3′ end of the intron. To simplify sequencing of this region, PCR products were cloned using a PCR-Script Amp Electroporation-Competent cell cloning kit (Stratagene), and at least three positive clones were sequenced for each taxon.

DNA sequencing.—PCR products and clones were sequenced using ABI PRISM Dye Terminator Cycle Sequencing, FS kit (Applied BioSystems, Foster City, CA) following the manufacturer's protocol. Sequencing reactions were purified using Centri Sep spin columns (Princeton Separations, Adelphia, NJ) prior to analysis on an ABI 373A or 377 automated sequencer. Both DNA strands were sequenced completely for all taxa. Sequences were edited and aligned using Sequencher 3.1.1 (Gene Codes Corporation, Ann Arbor, MI).

Morphological data.—C. Brochu (pers. comm.) provided a matrix of 164 morphological characters for Crocodylia, previous versions of which have been published elsewhere (Brochu, 1997, 1999a; Brochu and Gingerich, 1999). Versions differ only by successive addition of further fossil taxa, a redefinition (Brochu, 1999a) of two characters, and correction of a few typographical errors. All characters were coded as unordered, as in the original publications.

Combined data matrix.—C-*myc* data were added to the morphological matrix. For all taxa lacking molecular data, C-*myc* characters were coded as missing. The 10 indel characters were included. Sequence and morphological characters were equally weighted. The molecular data were rooted by giving the Glen Rose form, the morphological outgroup (Brochu, 1997), the states reconstructed (ACCTRAN optimization) for the basal crocodylian ingroup node on the molecular tree.

Phylogenetic Analyses

Alignment was by eye. Parsimony and maximum likelihood analyses were performed using PAUP* 4.0 (Swofford, 1998). In parsimony analyses, molecular characters were unordered. Gaps were coded as missing data, but in some analyses, separately coded indel characters were added. There were 10 informative indels, 9 coded as present/absent and one coded as a three-state unordered character. An 11th indel is informative if treated

as a three-state ordered character; it was not included in any analyses but can be mapped onto the resulting trees. In some analyses, sequence characters were divided into a priori partitions: intron, protein-coding region of the exon, and 3′ untranslated region (UTR). Maximum likelihood analyses of c-*myc* sequences were performed using a version of PAUP* 4.0 currently under development for future release. This version supports partitioned likelihood analyses by allowing independent parameter estimates for each subset of the data. For each partition, and for a single-model analysis of the entire data set, the general time reversible model was used, with a gamma distribution and invariable sites (GTR + Γ + I), with parameters estimated from the data. Simpler submodels for each partition provided only a slight decrease in likelihood, but the optimal submodels differed greatly among partitions, and we chose to use a single framework for ease of direct comparison. Parsimony analyses used branch-and-bound searches, and maximum likelihood analyses used heuristic searches, with 10 random-addition sequence replicates. All searches used tree bisection–reconnection branch swapping. Felsenstein's (1981) likelihood ratio test was used to test whether c-*myc* data conform to a molecular clock.

A Bayesian phylogenetic analysis was performed with MrBayes 3.0b3 (Huelsenbeck and Ronquist, 2001), using the same partitioned model as for the maximum likelihood analysis. MrBayes 3.0 allows some or all of the parameters of a phylogenetic model to be estimated separately for each data partition or across all partitions. Two separate analyses were performed, each with 5,000,000 generations of the chain. Topology and model parameters were sampled every 100th generation and used to determine the posterior probabilities of clades and estimates of model parameters. The first 1,000,000 generations (10,000 samples) were discarded to ensure that inferences were based on valid samples from the target distribution. This burn-in period far exceeded the point at which the log probabilities of both chains had reached a plateau, an indicator that the chains have converged to the same state space (Huelsenbeck et al., 2001). In addition to monitoring the log probabilities of the different runs, clade probabilities for the different runs were compared to determine whether both chains had converged on the same target distributions (Huelsenbeck et al., 2001). Model parameters were summarized as the mean of the marginal posterior distribution of model parameters and their respective 95% credibility regions (CRs). A total of 40,000 samples from a single run were used in the estimation of topology and model parameters. (Output files from MrBayes can be obtained from the authors upon request.) For the intron, model parameter values were estimated from a data set excluding the outgroup, as in maximum likelihood analyses.

A nonhomogeneous maximum likelihood (NHML) analysis was performed with the program package NHML 2 (Galtier and Gouy, 1998; http://www.univ-montp2.fr/%7Egenetix/nhml.htm), which allows base composition to vary over the tree. NHML does not support the GTR model, instead using Tamura's (1992) model with a gamma distribution of site rates. The intron was omitted for this analysis, which was unpartitioned. NHML options were set as follows: branch lengths, GC content, transition:transversion ratio, root position, ancestral GC content, and gamma distribution α parameter were all estimated, and initial values were all set at defaults.

Morphological characters were treated as unordered (Brochu, 1999a). The number of most-parsimonious trees for the morphological data was extremely large. In an attempt to estimate the total, we performed a mark–recapture experiment (Seber, 1973:60). Two PAUP* runs of 2,000 random-addition replicates, each limited to five trees per replicate, were used as samples. Some analyses of morphological data were done using the molecular tree as a backbone constraint. For some analyses, morphological characters were divided into a priori partitions under the simplest scheme we could create. Cranial characters were separated from postcranial characters, under the assumption that such broad categories were most likely to be functionally independent. Characters of the cervical vertebrae and hyoid were combined into a third partition because their functional relationships to the first two partitions are ambiguous, and we created a fourth partition for the few soft-anatomy characters.

Fit of trees to data was assessed by KH tests (Kishino and Hasegawa, 1989); both trees (molecular, Fig. 1b; morphological, Fig. 1a) were specified a priori, based on previous studies (e.g., Brochu, 1997, and works cited therein). The null distribution for likelihood-based KH tests was obtained by RELL bootstrapping. Node support in parsimony and maximum likelihood analyses was assessed by bootstrapping (Felsenstein, 1985), with 100 replicates for each analysis. All settings were the same as for the other analyses except that the maximum likelihood bootstrap used a single model for all sites.

Taxonomy

The taxonomy used in this paper follows that of Brochu (2003), which is itself based on the taxonomy proposed by Norell et al. (1994) and Brochu (1999a). Brochu (2003) presented two definitions each for Crocodylidae and Gavialidae, one to be used in the context of the morphological tree and the other in the context of the molecular tree; here, we use the molecular-context definition in both cases.

Given the molecular tree, two more clades need to be named. We define Longirostres (a complement to Brochu's Brevirostres) as a node-based group including the last common ancestor of *Crocodylus niloticus* and *Gavialis gangeticus* and all of its descendants. If the morphological tree were to be adopted, Longirostres would be a junior synonym of Crocodylia; similarly, if the molecular tree were to be adopted, Brevirostres would be a junior synonym of Crocodylia. The second clade, which we do not name here, is a stem-based group including

C. niloticus and all taxa more closely related to it than to *Alligator mississippiensis*. If the morphological topology were to be adopted, this group would be a junior synonym of Crocodyloidea. If the molecular topology were to be adopted, there would be many extinct crocodylians that would fall into this group but would be excluded from Longirostres; we refer to these taxa informally as stem-Longirostres.

One final consequence of basing the taxonomy on the molecular tree is that if it were to be adopted, many extinct taxa considered by Brochu (1999a) to be crocodylians would become instead noncrocodylian eusuchians.

RESULTS

*Sequence and Structure of Crocodylian c-*myc *Genes*

The *Alligator* c-*myc* sequence we determined is 3,038 bp in length and contains elements identifiable by similarity to the chicken sequence (Watson et al., 1983) as exon 2 and exon 3, including the complete coding sequence. Exon 2 (745 bp) is full length, but our exon 3 sequence (776 bp) is truncated at the 3′ end. It does not include a region homologous to the last 228 bp of the chicken 3′ UTR and lacks a polyadenylation signal and polyadenylation site. The structure of the gene is similar to that of other vertebrate c-*myc* genes (Ryan and Birnie, 1996). The open reading frame for the standard *Myc* gene product is 1,317 bp in length, beginning within exon 2 and ending in exon 3. The predicted protein sequence of 439 residues differs from the chicken sequence by seven indels and 52 replacements (11.8%). Most differences occur within five regions of the protein previously identified as prone to variation (Braun et al., 1985). Exon 3, with only one indel and 19 replacements, is less divergent than exon 2 because exon 3 contains the highly conserved nuclear localization signal (Marcu et al., 1992) and the DNA-binding domains of the protein, the basic helix-loop-helix and leucine zipper (Atchley and Fitch, 1997). These are readily recognizable in the alligator sequence as are the three *myc* boxes of exon 2, which are involved in transcriptional activation (Atchley and Fitch, 1995).

The exons are bounded by consensus splice donor and acceptor signals and separated by a 1,186-bp intron (intron B), which lies at exactly the same point in the alligator coding sequence as it does in chicken, human, trout, and other vertebrate sequences. Beginning 288 bp from the 5′ end of this intron, there is an ~80-bp element that appears again as an imperfect inverted repeat 384 bp downstream. The 3′ half of the element appears once more in inverted format between the two full-length repeats. Exon 2 is preceded by a 331-bp sequence element that presumably represents a portion of intron A. Neither of the introns bears significant sequence similarity to the corresponding chicken introns.

We determined c-*myc* sequences for nine individuals of eight crocodylian species, including all extant crocodylian genera. Sequences of six birds from different orders were also determined for use as outgroups. The sequenced fragment ranges in length from 1,118 to 1,130 bases within crocodylians and includes the 3′ portion of intron B (342–353 bp), the complete coding region of exon 3 (591–594 bp), and an adjacent part of the exon 3 3′ UTR (183–188 bp).

The two *Gavialis* sequences are identical, and only one was used in phylogenetic analyses. Sequences for *Crocodylus cataphractus* and *Osteolaemus tetraspis* are nearly identical, differing at two positions that may be polymorphic. There are no other polymorphic or ambiguous sites in the crocodylian sequences. Other uncorrected divergences within Crocodylia range from 1.2% (*Caiman–Palcosuchus*) to 6.4% (*Gavialis–Melanosuchus*). Divergences vary by region, with intron distances 1.2–9.4%, the coding region distances of 0.6–3.9%, and 3′ UTR distances of 1.1–10.5%. Uncorrected distances between crocodylians and birds are much greater, 14.8–19.5% for the coding region and 3′ UTR alone (the intron was not alignable between crocodylians and birds). Within crocodylians, 106 sites are variable (counting gaps as unknown) and 74 are parsimony informative. If the six birds are included, 246 sites are variable and 192 are parsimony informative. In the protein-coding portion of the exon, 33 sites are variable within crocodylians, four of which cause amino acid replacements. Two of the replacements are parsimony informative, but neither can be polarized by outgroup comparison.

Base composition differs between sampled birds and crocodylians, with birds having a slightly lower proportion of T and slightly higher proportion of A; this difference was not significant in a chi-square test ($P = 0.08$). Compositional differences within crocodylians also were not significant ($P = 1.00$). However, when invariant sites were excluded, crocodylians differed significantly among themselves ($P = 0.01$), attributable to differences in base frequency in variable sites only between alligatorids (*Caiman*, *Melanosuchus*, *Palcosuchus*, and *Alligator*) and Longirostres (*Crocodylus*, *Osteolaemus*, *Tomistoma*, and *Gavialis*), with alligatorid variable sites having a higher proportion of A and T and a lower proportion of C and G than Longirostres.

All sequences have been uploaded to GenBank, with accession numbers given in Table 1. The sequence alignment and relevant trees have been uploaded to TreeBase (reference numbers S880 and M1427).

There was one insertion in the coding region, a duplication of a single codon in *Palcosuchus*. This duplication was very near the 5′ end of the exon and involved the insertion of a fourth GAA codon in a string of three GAA's. Indels in the intron and 3′ UTR were more common. There were 13 indel events in the intron, 7 insertions, 5 deletions, and 1 unpolarizable, of which 7 were informative. There were four deletions, of which three were informative, and no insertions in the 3′ UTR. Length ranged from one to six bases. In addition, *Melanosuchus* showed a 29-base inversion in the intron of a sequence with an internal inverted repeat (underlined); 5′-TCAAGCCATGGGCAGGAGCATCCTGTTGC-3′ is the sequence in *Caiman*, of which the sequence in *Melanosuchus* is an exact inversion.

TABLE 3. Consistency indices (CI) and tree lengths for crocodylian c-*myc* data.

Data partition	CI (with/without uninformative characters)	Tree length	No. characters	No. informative characters
Ingroup only				
All sites[a]	0.953/0.938	128	1,191	84
Intron	0.931/0.902	58	390	32
Exon coding region	0.943/0.917	35	594	22
3′ UTR	1.000	24	197	20
Indels	1.000	11	10	10
Ingroup and outgroup				
All sites[a]	0.817/0.786	345	1,191	202
Intron[b]	0.931/0.902	58	390	32
Exon coding region	0.721/0.686	197	594	111
3′ UTR	0.962/0.955	78	197	49
Indels[c]	1.000	11	10	10

[a]Including indels.
[b]Intron was not alignable between ingroup and outgroup, so these values are the same as for the ingroup-only tree.
[c]Only 3 of 10 indels were alignable between ingroup and outgroup.

*Phylogenetic Analysis of c-*myc *Data*

All analyses of the full c-*myc* data set produced a single ingroup topology (Fig. 2) identical to that of the molecular tree of Figure 1b, as did all analyses of the 3′ UTR and maximum likelihood analyses of the other two partitions. The intron sequence was not alignable between ingroup and outgroup and thus when analyzed alone could not address the location of the root. In all trees, *Gavialis* is the sister of *Tomistoma*, and together these are the sister group of Crocodylidae.

Parsimony.—Equally weighted analysis of all sites, including indels, produced two most-parsimonious trees of length 345, differing only in relationships among avian outgroups. All ingroup nodes received strong bootstrap support (Fig. 2), with at least two of the three data partitions providing strong individual support for each node. The 3′ UTR in particular was perfectly consistent, i.e., within crocodylians it had a consistency index (CI) of 1.00, and it had a CI of >0.95 when outgroups were included. Parsimony analyses of the intron and exon coding region partitions both produced the molecular tree, except for failing to resolve a single node each (Alligatoridae for the coding sequence and relationships among caiman genera for the intron). Tree lengths and CIs are given in Table 3.

Like the 3′ UTR, indels were perfectly consistent, with CI of 1.00 on the molecular tree (Fig. 2). One indel supports each of the two branches of the molecular tree in conflict with the morphological tree. Indel 3 (position 75 of the intron alignment) supports the unrooted molecular tree, and indel 10 (positions 55 and 56 of the 3′ UTR) supports the molecular root position. At least one indel supports each of the other branches except for caimans and Alligatoridae. Another indel (intron positions 8 and 9), which was not included in the analyses, probably provides additional support for a *Gavialis–Tomistoma* clade. This indel falls in a poly-G tract, for which *Tomistoma* has one more and *Gavialis* has two more G's than other crocodylians. When this indel is coded as a single ordered character, it supports a sister group relationship between the gharials. This coding is probably the most reasonable one, because polynucleotide tracts often evolve in a stepwise fashion (Weber and Wong, 1993). However, this indel could be coded as unordered (or as two characters), in which case it would be uninformative.

In parsimony-based KH tests (Table 4), all three partitions and the complete data set, with or without indels, were significantly more compatible with the rooted molecular tree than with the rooted morphological tree. In tests of the unrooted ingroup-only trees, parsimony-based tests of all data and of the intron partition significantly rejected the morphological tree.

Maximum likelihood.—Estimated parameters of the single and partitioned GTR + Γ + I models are shown in Table 5. Parameters differed strongly among partitions, and the partitioned model, as might be expected, is a much better fit to the data than is the single model (negative log likelihood [− ln L] = 3324.67 and 3253.65, respectively, a difference of 71.02). A likelihood ratio test of the difference is highly significant (df = 70, $P = 8.08 \times 10^{-7}$). The intron has the least biased base composition,

TABLE 4. Maximum likelihood and Bayesian parameter estimates. Bayesian numbers are means, with 95% CR in parentheses. p = proportion of invariable sites; α = gamma distribution shape parameter.

	Maximum likelihood				Bayesian		
Parameter	All sites	Intron[a]	Exon coding	3′ UTR	Intron[a]	Exon coding	3′ UTR
Freq A	0.299	0.224	0.331	0.334	0.222 (0.184–0.262)	0.327 (0.292–0.362)	0.331 (0.275–0.390)
Freq C	0.230	0.212	0.236	0.231	0.211 (0.174–0.252)	0.236 (0.207–0.267)	0.226 (0.178–0.279)
Freq G	0.247	0.287	0.256	0.165	0.288 (0.246–0.333)	0.250 (0.219–0.282)	0.175 (0.133–0.222)
Freq T	0.223	0.278	0.177	0.270	0.279 (0.236–0.322)	0.187 (0.160–0.215)	0.268 (0.215–0.324)
p	0.496	0.435	0.632	0.470	0.577 (0.174–0.780)	0.62 (0.351–0.706)	0.484 (0.271–0.610)
α	1.072	0.662	2.270	infinite	23.297 (0.844–48.569)	10.612 (0.403–40.121)	30.715 (1.182–49.375)
Rate A-C[b]	1.751	4.072	2.349	0.697	12.103 (3.224–26.202)	3.609 (1.501–6.832)	6.682 (1.435–17.216)
Rate A-G	8.237	17.248	9.229	3.615	43.386 (30.688–49.770)	16.047 (7.740–26.895)	37.092 (19.044–49.281)
Rate A-T	0.674	2.041	0.188	0.304	9.212 (2.232–24.768)	0.594 (0.019–1.986)	5.227 (1.168–13.646)
Rate C-G	1.494	3.381	1.068	0.991	11.731 (3.628–25.522)	2.538 (0.624–5.543)	14.902 (4.917–29.181)
Rate C-T	13.050	9.921	26.554	2.504	30.003 (14.082–48.998)	42.117 (26.112–49.745)	32.127 (14.475–49.214)

[a]Parameters estimated for ingroup taxa only.
[b]Rate of G-T transformations set at 1.

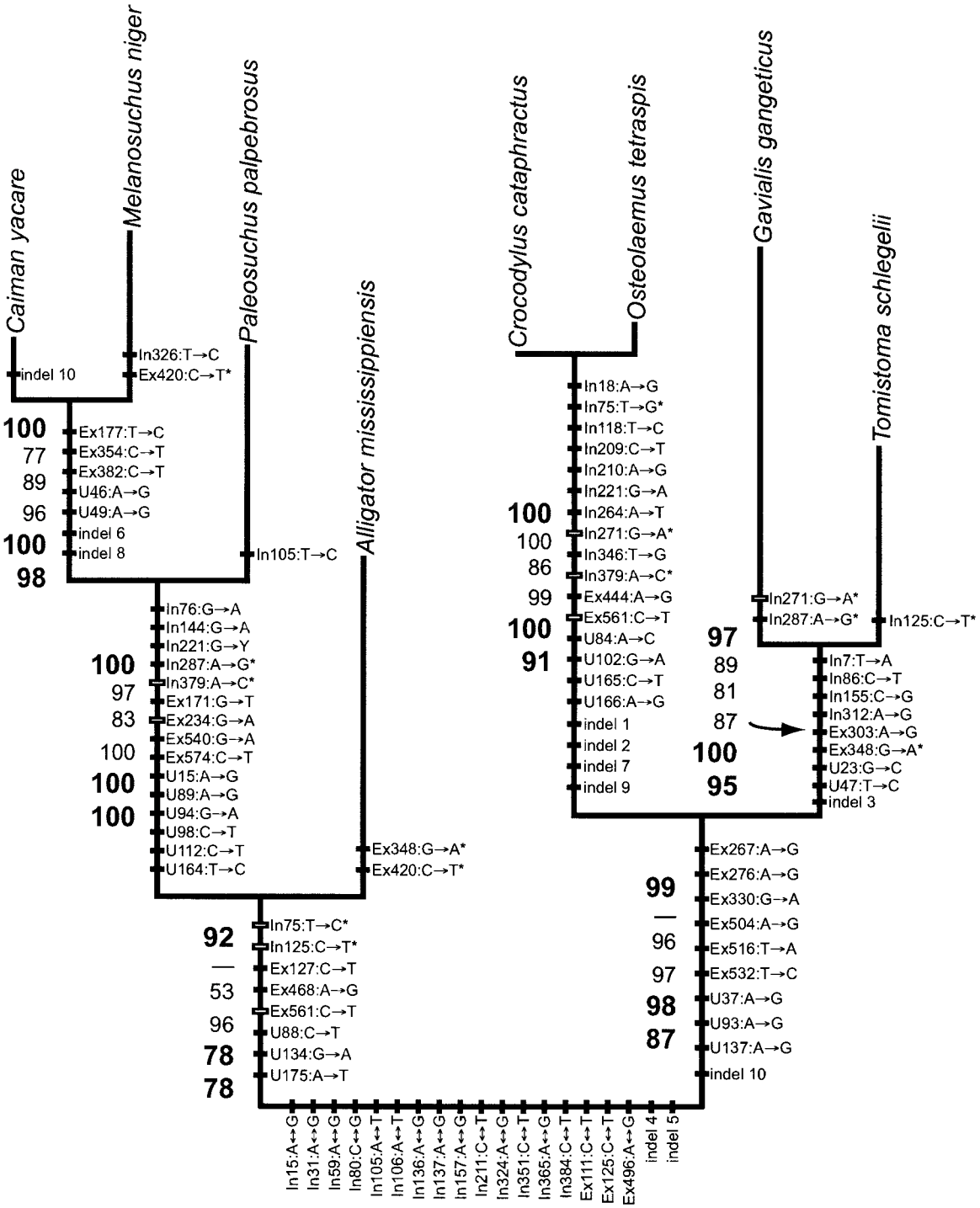

FIGURE 2. Single best tree from all analyses of crocodylian c-*myc* data. The outgroup (six birds) is not shown but roots the tree at the basal branch. Branch lengths were determined using parsimony. Numbers next to branches are support values from various analyses. The top number (larger, bold) is the unweighted parsimony bootstrap value (100 replicates) for the full data set (including indels), and smaller numbers below are parsimony bootstrap support for the three data partitions: intron, exon coding region, and 3′ UTR region (without indel characters). Because the intron could not be aligned between ingroup and outgroup, the intron tree is unrooted, and support could not be assessed for the two basal branches. The bottom two numbers (larger, bold) are the Bayesian posterior probability for an analysis using a GTR + Γ + I model with separate parameters for each of the three data partitions and the maximum likelihood bootstrap values for an analysis using a single GTR+Γ+I model across all sites. Character transformations are mapped onto the tree using unweighted parsimony. Ambiguous optimizations are shown with an open tick mark. Ambiguous transformations are placed so as not to occur on the two branches that conflict with the morphological tree, if possible. Characters showing homoplasy within Crocodylia are indicated with an asterisk. Characters that cannot be rooted by the outgroup (almost all of them intron characters) are placed on the basal horizontal branch; each character supports one of the two basal branches, but there is no way to tell which one. Character numbers correspond to position in each of the three partitions of the data matrix: In = intron; Ex = exon coding region; U = 3′ UTR, e.g., Ex540 is the 540th site in the exon coding region alignment. Only informative characters are shown, but branch lengths include autapomorphies.

TABLE 5. Likelihoods. Numbers are negative log likelihoods (−ln L).

Data partition	Molecular tree	Morphological tree	Difference
All sites, partitioned model	3253.653	3269.060	15.407
Intron	846.119	850.317	4.197
Exon coding region	1748.620	1754.355	5.735
3′ UTR	658.914	664.388	5.475
All sites, single model	3324.669	3338.944	14.274
Difference, single vs. partitioned model	71.016	69.884	

the exon coding region has the highest proportion of invariant sites, and the 3′ UTR has the least variation in rate among sites and the least variation in rate among transformation types.

All maximum likelihood analyses produced a single tree (Fig. 2), identical to the molecular tree of Figure 1b. Likelihoods for the molecular tree are from 4 (intron only) to 15 (partitioned analysis) log-likelihood units better than those for the morphological tree (Table 6). However, in KH tests, only a single-model test using the entire data set significantly rejected the morphological tree (Table 4). Indels were not considered in the likelihood analyses, but they map perfectly onto the tree, thus strengthening support for the molecular topology.

Bayesian analyses.—The individual Bayesian analyses both plateaued at very similar log probabilities (mean log probabilities of −3312.37 and −3310.87) and at nearly the same time (close to generation 5,000), indicating both runs had converged to the same distribution. A comparison of the clade probabilities between runs showed no large deviations. For the ingroup species, identical clade probabilities were observed between runs, which suggests that the chains had converged on the same target distribution. Thus, only one run was used to summarize topologies and model parameters. The Bayesian majority rule topology is identical to that of the molecular tree, with high posterior probabilities for most internal branches (Fig. 2).

Bayesian estimates of the model parameters (Table 5) are very similar to those obtained under the partitioned maximum likelihood analysis with the exception of the gamma shape parameter (α). In general, the Bayesian values for α were larger than those obtained under maximum likelihood (e.g., 10.612 vs. 2.270 in the exon) but had broad confidence intervals (0.4–40.1 for the exon). The explanation for this discrepancy appears to be that the likelihood surface for α is a large plateau on which many different values have similar likelihoods. We plotted this surface for the intron using only ingroup species (Fig. 3a). A small peak, representing the maximum likelihood value, can be seen near zero, but the rest of the distribution is nearly flat. None of the log likelihood values are >1.92 log units from the maximum and therefore cannot be rejected as inadequate by a likelihood ratio test. Furthermore, Bayesian mean posterior estimates of α for each data partition were not significantly different from the corresponding maximum likelihood estimates (Table 6) in likelihood ratio tests (intron: $P = 0.502$; exon: $P = 0.414$; 3′ UTR: $P = 0.314$).

As another approach to the discrepancy in estimates of α, we plotted the posterior distribution of α on the same scale as its prior distribution, a uniform distribution from 0 to 50, for the intron partition (Fig. 3b). There is very little difference between the prior and posterior distributions. As with the likelihood surface plot there is a small peak near zero, but the remainder of the surface closely matches the prior, indicating that there is very little information in the data about the value of α; longer sequences would be needed for a more constrained estimate. This pattern also was observed for the other two partitions (not shown). One advantage of the Bayesian method is that we derive CRs for the model parameters rather than a single value and are able to determine that support for our estimates of some parameters, especially α, is not strong.

NHML analysis.—This method allows base composition to vary among taxa, providing a test of the possibility that the position of *Gavialis* is being influenced by its similarity in base composition to *Tomistoma* and Crocodylidae. We calculated likelihoods over 12 topologies, including the molecular and morphological trees, the various compromise rootings of Figure 1, and other alternative placements of *Gavialis* and *Tomistoma* that maintained a set topology within assumed monophyletic Alligatoridae, Crocodylidae, and Aves. The molecular tree had a better likelihood score (−ln L = 2268.70) than the next best tree, in which the gharials were the sister group of Alligatoridae (− ln L = 2276.17), a slight rerooting of the molecular topology. The morphological tree had the second-worst likelihood of all trees tested (−ln L = 2286.95); the worst (−ln L = 2287.00) left

TABLE 6. Kishino–Hasegawa tests of the crocodylian c-*myc* data set.

Data partition	No. characters informative for the ingroup	No. characters that prefer morphology/molecules	KH test P value: Ingroup only, Parsimony	Ingroup only, Likelihood	All taxa, Parsimony	All taxa, Likelihood
All sites	74	1/17	0.0081**	0.106	0.0002**	0.042* (0.123)[a]
Intron	32	0/4	0.0454*	0.236		
Coding region	22	1/8	0.3177	0.596	0.0195*	0.093
3′ UTR	20	0/5	0.1578	0.124	0.0250*	0.218
All sites plus indels	84	1/19	0.0046**		0.0001**	

[a] First-value is for a single model; value in parentheses is for a partitioned model.
*$P < 0.05$.
**$P < 0.01$.

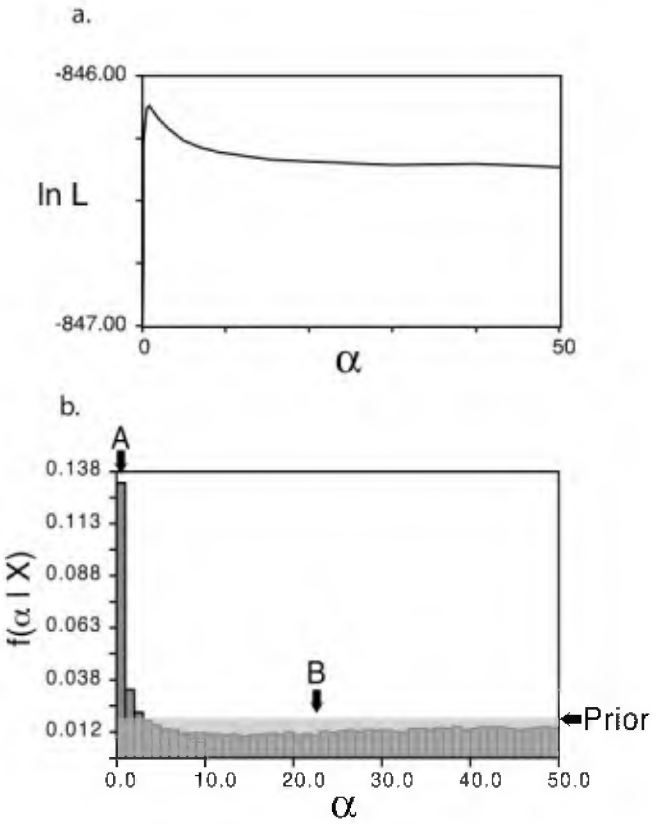

FIGURE 3. Differences between maximum likelihood (mode) and Bayesian (mean) estimates of gamma shape parameter α. (a) Likelihood surface for α from intron data partition. The surface is very flat except for a small spike at the maximum likelihood value. Surfaces for the other three partitions (not shown) are similar. (b) Posterior and prior probability distributions for α from intron data partition. Arrow A is the maximum likelihood value (0.662), and arrow B is the mean of the posterior distribution (22.747). The uniform (0, 50.0) prior is shown by the lightly shaded box, and the shaded bars are the posterior distribution.

Tomistoma as sister to Crocodylidae and made *Gavialis* sister to Alligatoridae.

Tests of the molecular clock.—A likelihood ratio test using all ingroup taxa, all data partitions, and the same partitioned model used in phylogenetic analysis was highly significant and rejected a uniform molecular clock within Crocodylia ($2 \times$ diff. $-\ln L = 27.75$, df $= 6$; $P = 0.0001$).

Phylogenetic Analysis of Morphological and Combined Data

Parsimony analysis of the morphological data produced a very large number of trees, exceeding the memory capacity of our computers. A mark–recapture experiment estimated the total number at several million. However, the large number of trees arises from uncertainty about relationships among a fairly small proportion of the extinct taxa, many of them diagnosed from very incomplete specimens. The strict consensus of the 15,000 trees we found is well resolved (completely so for extant taxa, with the topology of Fig. 1a) and is compatible with other published trees using older versions of the same data set (Brochu, 1997, 1999a). When the molecular tree was used as a backbone constraint, the strict consensus of 15,000 trees out of an estimated several million was again well resolved and compatible with Brochu's (1997) similar exercise (and necessarily having the topology of Fig. 1b for extant taxa). These two strict consensus trees were used as the morphological tree and the molecular tree, respectively, in subsequent analyses involving mapping of morphological characters onto trees (e.g., for the KH tests). The combined data produced at least 15,000 (probably much higher; no mark–recapture experiment was performed) most-parsimonious trees of 663 steps and CI = 0.504 (0.470 excluding uninformative characters), with a strict consensus topology identical to the molecular constrained analysis.

TABLE 7. Kishino–Hasegawa tests of the crocodylian morphological data set.

Partition	No. characters	No. characters that prefer morphology/molecules	KH test *P* value
All data	164	32/17	0.03*
Cranial	116	17/11	0.26
Cervical	21	4/3	0.72
Postcranial	21	10/3	0.05*
Soft anatomy	6	1/0	0.36
All nonpostcranial	143	22/14	0.18

*$P < 0.05$.

In parsimony-based KH tests of the morphological data, using the morphological and molecular-constrained consensus trees (Table 7), the full morphological data set supported the morphological tree significantly better than it supported the molecular tree. Of data partitions, only the postcranial characters significantly rejected the molecular tree. We used chi-square tests to determine whether the morphological characters supporting the morphological and molecular constrained trees were randomly distributed among the data partitions (Table 8). Characters supporting the morphological tree were overrepresented in the postcranial partition and underrepresented in the cranial partition ($P = 0.019$). Characters supporting the molecular tree were randomly distributed among partitions ($P = 0.722$).

Reanalyses of Mitochondrial Sequences

White and Densmore's (unpubl.) maximum likelihood analysis of mtDNA data is the only molecular analysis

TABLE 8. Chi-square tests of the random distribution among partitions of morphological characters supporting the morphological and molecular trees (df = 3). For characters supporting the morphological tree, SS = 66.62, $P = 0.019$. For characters supporting the molecular tree, SS = 2.79, $P = 0.722$.

	Support morphological tree			Support molecular tree		
Partition	Observed	Expected	Observed − expected	Observed	Expected	Observed − expected
Cervical	4	4.10	−0.10	3	2.18	0.82
Postcranial	10	4.10	5.90	3	2.18	0.82
Cranial	17	22.63	−5.63	11	12.02	−1.02
Soft	1	1.17	−0.17	0	0.62	−0.62

that has produced an ingroup topology different from the others. To determine why, we analyzed both the full data set and a reduced data set consisting only of species for which we have c-*myc* sequences plus the outgroup *Gallus gallus*. Their original maximum likelihood analysis using the Hasegawa–Kishino–Yano (1985) model (HKY85) produced the morphological topology with the molecular root (Fig. 1c, root 4). Our attempt to replicate this analysis using both full and pruned data sets recovered that topology ($-\ln L = 3108.08819$ and 1966.21667, respectively). However, adding site-to-site rate variation (HKY85 + Γ) with the pruned data set changed the topology to the molecular topology with the molecular root (identical to the c-*myc* trees) and greatly improved likelihood ($-\ln L = 1912.17433$). Analysis of the full data set using the HKY85 + Γ model produced a molecular ingroup topology rooted inside *Crocodylus* ($-\ln L = 2948.66822$). This model was chosen because more complex alternative models gave only modest increases in likelihood scores.

DISCUSSION

The c-*myc* data provide the strongest evidence for crocodylian phylogeny of any single data set to date. They show very strong support for the molecular tree, in both ingroup topology and root position, and this support is strong in all three data partitions and indel characters. The disagreement between molecular and morphological trees concerns the position of the true gharial, *Gavialis*, and can be divided into two independent issues: ingroup topology and rooting. These issues boil down to the existence (or nonexistence) of two internal branches.

Unrooted Ingroup Tree

The unrooted morphological and molecular trees differ by one branch (Figs. 1c, 1d). The c-*myc* data strongly support the molecular topology, as determined by several different measures, including parsimony and likelihood bootstraps, Bayesian posterior probability, indels (all in Fig. 2), and parsimony KH tests (Table 4).

The reason for the failure of the data to distinguish among topologies in most of the maximum likelihood KH tests is unclear. It may be that these data fit the expectations of the parsimony model extremely well but that the small total number of changes in the tree is insufficient to allow maximum likelihood parameters to be estimated efficiently. This explanation seems plausible considering that the data are split among three regions evolving under quite different parameters, and the wide Bayesian CRs for many parameters (Table 5) support this interpretation. Another hypothesis, that long-branch attraction inflates support under parsimony, is unlikely given the high CIs for the data, particularly the CI of 1.0 for the 3′ UTR, coupled with the very low differences in branch lengths between parsimony and likelihood trees.

All previous molecular studies support the molecular ingroup topology. Although the strength of that support has in most cases not been evaluated for the individual data sets, the concordance of several independent estimates makes a powerful argument for the correctness of their common topology.

Location of the Root

The c-*myc* data strongly support the molecular root, as indicated by many measures. The relevant branch (the right basal branch of Fig. 1b) is supported by parsimony and likelihood bootstraps, Bayesian posterior probability, all parsimony KH tests, and a likelihood KH test using a single GTR + Γ + I model. An unambiguous indel (number 10, positions 55 and 56 of the 3′ UTR alignment) also supports the branch. (Although *Caiman* has an overlapping insertion, all evidence nests it well within Alligatoridae, so the two events must be considered independent.) None of the maximum likelihood KH tests of individual partitions or of the partitioned analysis were significant, possibly because of the relatively small amount of information available for estimating model parameters.

Previous studies using mitochondrial 12S ribosomal DNA (Gatesy and Amato, 1992; Gatesy et al., 1993; Aggarwal et al., 1994; Poe, 1996) and a combination of mitochondrial protein-coding genes and a tRNA (White and Densmore, unpubl.) all recovered the molecular root. Each used a single bird outgroup species. Gatesy et al. (2003), in combined and separate analyses of morphological, mitochondrial, and nuclear data (with two birds used as molecular outgroups), found strong support for the molecular root. In other studies, no outgroup was used, and thus the trees cannot be rigorously rooted. However, in unrooted trees the location of the root can be constrained, depending on how much variation in molecular evolutionary rates we are willing to consider credible. In previous unrooted trees, the distance between *Gavialis* and *Tomistoma* is much less than the distance between either one and any other taxon. To accept a topology in which the two gharials are not sister taxa—particularly to accept the morphological root, on *Gavialis*—we must accept considerable disparity in molecular evolutionary rates between *Gavialis* and whichever other crocodylian is most distant from it.

The best case for the morphological root, i.e., the case requiring the least disparity of rates, is a trichotomy of *Gavialis*, *Tomistoma*, and all other crocodylians; any other situation would require greater disparity of branch lengths and therefore of rates. To estimate rate disparity required by the morphological tree, we calculated the lengths of the branches in a number of published data sets and our c-*myc* data set. When data were distance measures, we calculated the branch lengths from published pairwise distances among three taxa: *Gavialis*, *Tomistoma*, and whichever crocodylian was most distant from *Gavialis*. For DNA sequences, we used the maximum likelihood branch-length estimates on the unrooted maximum likelihood trees. In all cases, the species most distant from *Gavialis* was a caiman, either *Melanosuchus* or (for tryptic peptide distances) *Caiman crocodylus*. The ratio of lineage lengths from the most distant caiman to the morphological root and from *Gavialis*

TABLE 9. Differences in molecular evolutionary rates implied by the morphological root (on *Gavialis*). A is the length of the branch from *Gavialis* to the morphological root, and B is the length of the branch to the morphological root from the most distant crocodylian taxon.

Type of data	A	B	Rate difference (B/A)	Reference
Immunodiffusion distances	0.03	4.58	152	Densmore, 1983
Tryptic peptide fingerprint distances	1	22	22	Densmore, 1983
Microcomplement fixation distances	9.25	114.25	12.35	Hass et al., 1992
mtDNA maximum likelihood branch lengths	0.207	0.950	4.59	White and Densmore, unpubl.
c-*myc* maximum likelihood branch lengths	0.013	0.075	5.77	This study

to the root is the greatest difference in evolutionary rates implied by that rooting. Implied rate differences for different data sets ranged from a high of 152 (i.e., the caiman lineage has evolved 152 times as fast as the *Gavialis* lineage) to a low of 4.59 (Table 9). The high rate differences required by the morphological root in such a variety of data sets strain credibility.

Our data and all previous sequence analyses strongly support the molecular root as correct. We consider these results conclusive and believe that explanations other than phylogeny for the root location are unlikely, but because of the unavoidably large distances to the outgroup (Fig. 4), there remains some slight room for doubt. However, given the molecular ingroup topology and even a slight limitation in evolutionary rate variation, the only alternative to the molecular root is a root between *Tomistoma* + *Gavialis* and other crocodylians (root 5, Fig. 1d). One molecular analysis, of RAG-1 nuclear data using maximum likelihood (Gatesy et al., 2003) did produce this root, but strength of support and model parameters were not stated. This root has the same conflict with stratigraphy and with morphological characters as does the molecular root and thus does not solve the problem of data incongruence.

There is no possibility of shortening the branch leading to the outgroup (no unsampled extant taxa attach to it), but there are three simple strategies to confirm the root that may be more fruitful than simply adding a more sequence data: (1) examine characters that evolve more slowly than c-*myc* (e.g., amino acid characters of a less-conserved gene), (2) examine characters that can be polarized without an outgroup, such as SINEs (Shedlock and Okada, 2000; Shedlock et al., 2000) or pairs of paralogous genes, or (3) examine a slowly evolving outgroup or one whose basal split is much older than the basal split within birds, e.g., turtles (Rieppel, 1994; Hedges and Poling, 1999; Rieppel and Reisz, 1999).

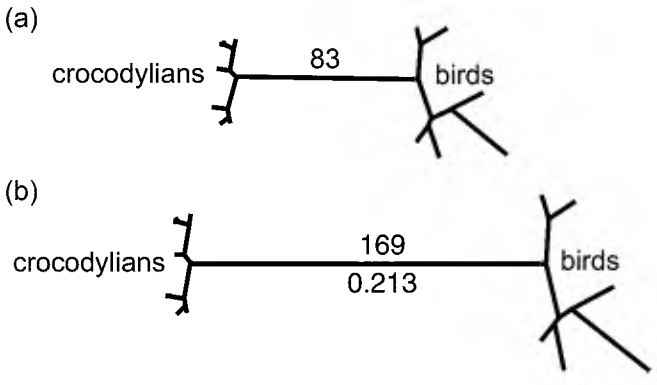

FIGURE 4. Trees based on c-*myc* data showing length of the branch from ingroup to outgroup and lengths of the other branches for comparison. (a) Parsimony tree. (b) Maximum likelihood tree. Branch lengths are to scale. Numbers above the central branches are inferred numbers of changes. The number below the central branch on the ML tree is inferred changes per site.

Functional Convergence?

As conclusive as the molecular data seem, we must consider ways in which they could mislead us. We have rejected long-branch attraction. Functional convergence is the one remaining possibility. However, almost all changes in the exon coding region are silent, so convergence in protein characteristics is ruled out. Convergence in sequence characteristics, e.g., base composition, would be possible, but there is little difference noted among crocodylians or between crocodylians and birds. Differences in base composition in variable characters alone are significant in a chi-square test but not large. Convergence is also an unlikely explanation when it must affect multiple partitions, i.e., exon coding region, 3′ UTR, and indels, evolving according to different parameters. The data partitions differ in base frequency among themselves (within species) to a much greater degree than base frequencies differ among species. Nevertheless, base composition differences have been implicated in some disagreements about phylogeny (e.g., Haddrath and Baker, 2001). Analyses using LogDet distances, which are insensitive to base compositional differences (Lockhart et al., 1994), or maximum likelihood models that allow base composition to vary over the tree (Galtier and Gouy, 1998) still recover the molecular root.

We are unable to think of any explanation for the patterns in the molecular data that fits the evidence better than our preferred explanation: both the molecular ingroup topology and the molecular root reflect the true phylogeny.

Morphological Data

Several features of the morphological data are puzzling if morphological similarities between the two gharials are to be explained by functional convergence. If the two gharials were in fact convergent in morphology, we would expect a strong secondary signal, presumably strongest in the cranial partition, uniting the two (and

their extinct long-snouted relatives) to the exclusion of other crocodylians. It is unreasonable to expect that all homoplasy will have been recognized a priori and coded as separate states, because the most powerful test of homology/homoplasy is congruence/incongruence with other characters.

Trueman (1998) used reverse successive weighting to locate a secondary signal supporting the molecular tree. This method has advantages when looking for any secondary signal rather than a specific one. However, because we have only two trees to consider, there are more precise methods of identifying such a signal; it must consist of those characters that have fewer steps on the molecular tree than on the morphological tree. There are 17 of these characters (Table 10). Trueman found only 12 (although his analysis implies that there are at least 15); unfortunately, he listed only 5 of these, and 1 of them (character 36) is wrong (it prefers only the unrooted molecular tree; if the choice is between rooted trees, it is equally parsimonious on both). Brochu (1999b), in further reverse successive weighting analyses, accepted Trueman's count but did not elaborate. We are thus unable to define the specific reasons for the discrepancy, but reverse successive weighting is probably not the method of choice for identifying secondary signal when the alternative tree is known a priori.

Distribution of signals in data partitions.—We would expect the secondary morphological signal to be concentrated in the cranial partition, and conflicts created by such a signal would potentially explain why the cranial partition does not significantly prefer the morphological tree over the molecular tree in KH tests (Table 7). However, a chi-square test (Table 8) shows that the secondary signal is not concentrated in any partition. In contrast, the primary signal, i.e., support for the morphological tree, is concentrated in the postcranial partition and attenuated in the cranial partition (Table 8). The cranial partition contains more of the secondary signal than do the other partitions (11 of 17 characters) but less than the proportion expected by chance, given that cranial characters make up the bulk of the data set. Rather than containing more conflicting signal, the cranial partition seems merely to contain less signal relevant to the position of *Gavialis*.

Secondary signal in the data set.—We found some characters with the expected pattern, uniting *Gavialis* with *Tomistoma* (and various putative extinct gavialoids), but there are only six of these charaters (2, 9, 30, 60, 61, 62), and they make up a small proportion of the secondary signal. Seven characters (12, 27, 39, 78, 79, 119, 145) unite *Gavialis* and its putative extinct relatives with several more inclusive clades, from Longirostres to Crocodylia. Because most of the united taxa are not long-snouted, an explanation of functional convergence does not fit these characters. Another three characters (84, 118, 130) can be interpreted as convergences among long-snouted taxa but not between the two living species, because *T. schlegelii* lacks the derived states. A final character (103) unites a putative gharial relative, *Thoracosaurus*, with stem-Longirostres, but *G. gangeticus* has an autapomorphic state.

Gatesy et al. (2003), although using the same morphological data set, found a somewhat different list of characters to comprise the secondary signal, as a result of using different methods and a different criterion for secondary signal. Gatesy et al. considered the secondary signal to include all those characters that unambiguously change on the basal branch of Gavialidae. We considered it to include all characters that support the molecular over the morphological topology, whatever the branch on which they are optimized as changing and whether or not that optimization is ambiguous.

TABLE 10. Morphological characters that support the molecular tree (from Brochu 1997, 1999a).

Character number, derived state	Partition[a]	Description	Support for molecular clade
2, 2	V	proatlas massive and block shaped	Gavialidae
9, 1	V	neural spine on first postaxial cervical narrow, dorsal tip acute and less than half the length of the centrum without the cotyle	Gavialidae
12, 0	V	axis neural spine crested	stem-Longirostres
27, 1	P	olecranon process of ulna wide and rounded	Crocodylia
30, 1	P	interclavicle with moderate dorsoventral flexure	Gavialidae
39, 0	P	ventral armor absent	Longirostres
60, 0	C	sulcus between articular and surangular	Gavialidae
61, 0	C	surangular with spur bordering the dentary toothrow lingually for at least one alveolus length	Gavialidae
68, 2	C	dentary linear between 4th and 10th alveoli	Gavialidae
78, 2	C	dentary teeth occlude in line with maxillary toothrow	stem-Longirostres
79, 1	C	naris projects dorsally	stem-Longirostres
84, 1	C	squamosal groove for external ear valve flares anteriorly	Gavialidae minus *T. schlegelii*
103, 1	C	dorsal edges of orbits upturned	Longirostres minus *Gavialis*
118, 1	C	palatine process in form of thin wedge	Gavialidae minus *T. schlegelii*
119, 0	C	basisphenoid not broadly exposed ventral to basioccipital at maturity; pterygoid short ventral to median eustachian opening	stem-Longirostres
130, 0	C	capitate process of laterosphenoid oriented laterally toward midline	Gavialidae minus *T. schlegelii*
145, 1	C	dorsal premaxillary processes long, extending beyond third maxillary alveolus	stem-Longirostres

[a] V = cervical vertebrae; P = postcranial; C = cranial.

Characters (43, 88, 95) counted by Gatesy et al. but not by us are reconstructed as changing unambiguously on the branch uniting *Gavialis* and *Tomistoma*, but they do not differ in number of steps between the molecular and morphological topologies. Most characters counted by us but not by Gatesy et al. change at branches more basal than Gavialidae. Others (9, 30, 84, 118) do not change unambiguously at the base of Gavialidae because of a basal polytomy; however, these characters do unambiguously support a clade that includes *Gavialis* and *Tomistoma* to the exclusion of all other extant species. One character (130) supports a clade within Gavialidae that excludes *Tomistoma* but includes both *Gavialis* and *Paratomistoma*, a species that is close to *Tomistoma* on the morphological tree; thus, we consider it as support for Gavialidae.

Search for missing secondary signal.—In an attempt to locate missing secondary signal (the true signal if we accept the molecular tree as correct) uniting the long-snouted taxa, we examined the anatomical region that would be expected to provide the strongest such signal, the long snout itself. In the lower jaws of both *Gavialis* and *Tomistoma*, the splenial makes up a major part of the mandibular symphysis. In other living (and most extinct) crocodylians, the splenial is excluded from the symphysis. In the upper jaws of the two gharials, but not most other taxa, the nasals are excluded from the external nares by the premaxillae; in *Gavialis*, the nasals are further excluded (by the maxillae) from contact with the premaxillae. Both of these features were included in the morphological data set (Brochu, 1997, 1999a), but their coding did not recognize similarities between the two gharials. Clark (1994), in an analysis of crocodyliform relationships (which did not examine relationships within Crocodylia), also coded both of these features but in a different way that emphasized gharial similarities. Table 11 compares both codings of these features.

Brochu's character 43 coded the conditions of the mandibular symphysis in *Tomistoma* and *Gavialis* as separate states (states 2 and 3, respectively). Because the character was unordered, the two states were interpreted as synapomorphies of separate groups (other members of both groups are extinct) rather than as a putative synapomorphy of the two gharials (and their extinct relatives). Clark's coding of his character 77 would give both taxa state 2. Character 43 thus has the same number of steps on the molecular and morphological trees (Figs. 5a, 5c), whereas character 77 is more parsimoniously optimized on the molecular tree (Figs. 5b, 5d). Brochu's coding contains information that Clark's does not (about lower level relationships), whereas Clark's contains information that Brochu's does not (about higher level relationships). An ordered version of Brochu's character could preserve both sorts of information.

Similarly, Brochu's character 95 coded all differences in bones of the upper jaw as a single multistate, unordered character, and again the states in the two gharials are necessarily interpreted as synapomorphies of more restricted groups and thus are uninformative about their relationships to each other (Figs. 5a, 5c). Clark coded the same information as two binary characters, numbers 13 and 14, equivalent to a three-state ordered character. Character 13 recognizes the similarity of the two gharials, whereas character 14 recognizes their differences; character 13 is more parsimoniously optimized on the molecular tree than on the morphological tree, and character 14 is an autapomorphy on either tree (Figs. 5b, 5d).

Recoding of characters 43 and 95 does not by itself change the tree supported by the entire morphological data matrix. In another analysis of crocodylian phylogeny in which the two characters were coded in a way similar to Clark's characters, Salisbury and Willis (1996) still came up with the same morphological tree. We have not closely examined other morphological characters that might be sources of additional secondary signal if recoded, i.e., those characters that under current codings fail to differentiate between the molecular and morphological trees. We hope that these examples will encourage further examination of the data, with an eye toward the effects of character coding and ordering on phylogenetic analyses.

Difficulties of character coding.—Although we do not claim that molecular data are inherently superior to morphological data, they do have one simplifying

TABLE 11. Coding of crocodyliform rostral and mandibular characters.

Feature	Character number and coding	
	Brochu (1997, 1999a)	Clark (1994)
Bones of the rostrum	95: External naris bisected by nasals (0), nasals contact external naris but do not bisect it (1), nasals excluded, at least externally, from naris and nasals and premaxillae still in contact (2) or nasals and premaxillae not in contact (3).	13: Nasal takes part in narial border (0) or does not (1). 14: Nasal contacts premaxilla (0) or does not (1).
Mandibular symphysis	43: Splenial participates in mandibular symphysis and splenial symphysis adjacent to no more than five dentary alveolae (0), splenial excluded from mandibular symphysis and anterior tip of splenial passes ventral to Meckelian groove (1), splenial excluded from mandibular symphysis and anterior tip of splenial passes dorsal to Meckelian groove (2), deep splenial symphysis, longer than five dental alveoli, and splenial forms wide "V" within symphysis (3), or deep splenial symphysis, longer than five dental alveoli, and splenial constricted within symphysis and forms narrow "V" (4).	77: Splenial not involved in symphysis (0), or involved slightly (1), or involved extensively in symphysis (2) (ordered).

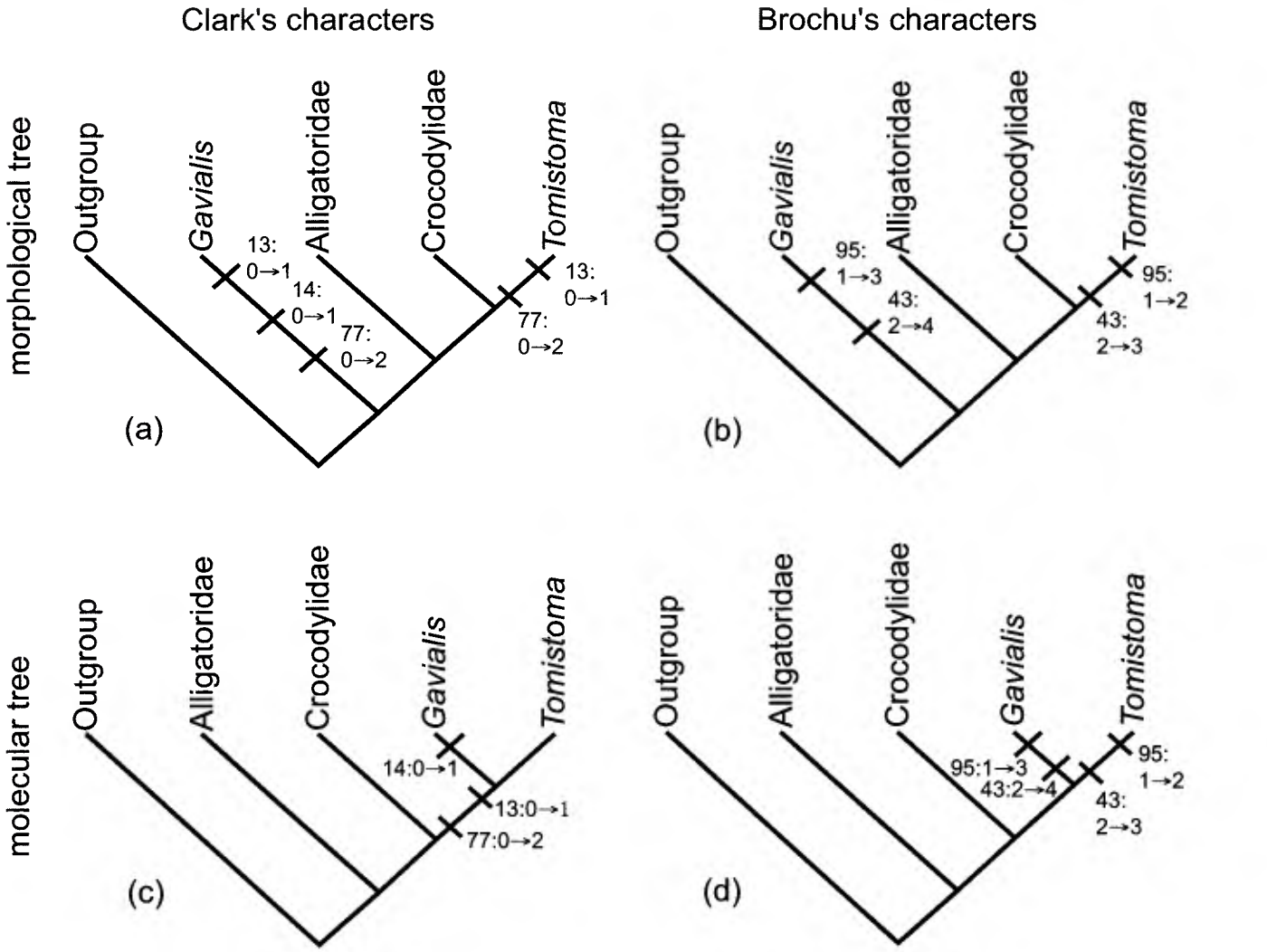

FIGURE 5. Morphological (a, b) and molecular (c, d) trees showing how alternative codings of the same morphological features can change hypotheses of monophyly. Clark's (1994) characters 13, 14, and 77 require five steps on the morphological tree and three on the molecular tree. Brochu's (1997, 1999a) characters 43 and 95, describing the same features, require four steps on either tree. Details irrelevant to differences between trees have been omitted.

advantage: DNA sequence character states really are discrete and objectively scored. Morphological character states, in contrast, are generally abstracted from more variation than is coded: each state is a cluster of points in some multidimensional state space rather than a single point. If the separations in state space among points assigned to each state are all much less than the separation between any two points assigned to different states, the character coding approaches objectivity. If clusters are in any way ambiguous, necessarily subjective judgments must be made.

Choices about relationships among different assigned states, e.g., between ordered and unordered characters, add further issues of judgment. If a character contains hierarchical information at multiple levels (to maintain the state space metaphor, if there are clusters of clusters), simple unordered characters can represent only one of these levels, as in the two examples above. To choose a molecular example, if there is a transition:transversion bias in sequence data, such that A and T are considered closer to each other than either is to C or G, unordered characters can represent only one level of the hierarchy: either the difference between A and T (ACGT coding) or their similarity (purine/pyrimidine coding). For this reason, character transition matrices (transversion parsimony) are often used. Character 43 is analogous.

Consequences for character evolution if the morphological tree is wrong.—Assuming that the molecular tree is correct, what could be causing the morphological data to be misleading? *Gavialis* must undergo a great many reversals to the primitive condition, with postcranial characters disproportionately represented, from that shared among various more or less inclusive groups ranging from all other Crocodylia to all other Longirostres. We have no explanation for this pattern, but it deserves investigation.

Stratigraphic fit.—The stratigraphic fit of the molecular-constrained tree is poor, because putative members of Gavialoidea are known from the Late Cretaceous Campanian stage (Brochu, 1997), earlier than any other Longirostres or any stem-Longirostres. However, if we assume that only one fossil taxon, *Thoracosaurus*, is wrongly assigned to Gavialoidea, the fit is improved significantly, because the next oldest putative gavialoid (*Eogavialis*) is latest Eocene in age, a difference of 40 million years. Do the c-*myc* data fit this younger date? Changes within Crocodylia are too few for accurate assessment of evolutionary distances. More rigorous tests of evolutionary rates must await longer sequences.

Recently, Brochu (2001, 2003) suggested that other extinct taxa younger than *Thoracosaurus* (which itself may not be monophyletic) but older than *Eogavialis* are gavialoids. Pending formal analyses of these taxa, we can only speculate. However, a similar parsimony debt may be incurred if we separate all these taxa (which would then form a clade) from *Gavialis* or if we separate one of them, if the same characters unite all of them.

Combined data.—Combined analysis is not useful in this case. It reveals that the signal in the molecular data is stronger and more consistent (with less implied homoplasy) than the conflicting signal in the morphological data, and for that reason combined analysis produces the molecular tree. However, the conflicting signals are unexplained, and we gain no further insight into which data set is a better guide to the true topology. Similar opinions have been expressed by previous workers on the problem (Brochu, 1997; Brochu and Densmore, 2001; see however Gatesy et al., 2003).

Partitions and character coding.—Molecules versus morphology is not a useful view of data conflict. It is more productive to look at conflicting and congruent signals within both molecular and morphological data sets by subdividing each into natural (as far as we can make them) a priori partitions (McCracken et al., 1999; Naylor and Adams, 2001). In the present case, examination suggests that morphological support for the molecular tree may have been reduced by the character codings used. The crocodylian morphological data need yet another examination, with special attention to the effects of character state assignment and state ordering on phylogenetic hypotheses.

ACKNOWLEDGMENTS

We thank Lou Densmore and Herb Dessauer for crocodylian tissue samples. Dave Swofford, Jim Wilgenbusch, and Kevin de Queiroz gave us much helpful advice. Dave also allowed us to use an experimental version of PAUP* with partitioned likelihood, and Jim also provided programs to make possible partitioned model KH tests. Chris Brochu and Lou Densmore sent us preprints of their papers in press, and Chris provided an unpublished version of his morphological data set. Allan Baker, Lou Densmore, and an anonymous reviewer provided useful comments on the manuscript. We especially wish to acknowledge Chris Brochu's help; although we remain in disagreement on many points, his comments on several previous drafts have greatly improved this paper.

REFERENCES

AGGARWAL, R. K., K. C. MAJUMDAR, J. W. LANG, AND L. SINGH. 1994. Genetic affinities among crocodilians as revealed by DNA fingerprinting with a Bkm-derived probe. Proc. Natl. Acad. Sci. USA 91:10601–10605.

ALITALO, K., J. M. BISHOP, D. H. SMITH, E. Y. CHEN, W. W. COLBY, AND A. D. LEVINSON. 1983. Nucleotide sequence of the v-*myc* oncogene of avian retrovirus MC-29. Proc. Natl. Acad. Sci. USA 80:100–104.

ATCHLEY, W. R., AND W. M. FITCH. 1995. Myc and Max: Molecular evolution of a family of proto-oncogene products and their dimerization partner. Proc. Natl. Acad. Sci. USA 92:10217–10221.

ATCHLEY, W. R., AND W. M. FITCH. 1997. A natural classification of the basic helix-loop-helix class of transcription factors. Proc. Natl. Acad. Sci. USA 92:10217–10221.

BRAUN, M. J., P. L. DEININGER, AND J. W. CASEY. 1985. Nucleotide sequence of a transduced *myc* gene from a defective feline leukemia provirus. J. Virol. 55:177–183.

BROCHU, C. A. 1997. Fossils, morphology, divergence timing, and the phylogenetic relationships of *Gavialis*. Syst. Biol. 46:479–522.

BROCHU, C. A. 1999a. Phylogenetics, taxonomy, and historical biogeography of Alligatoroidea. J. Vertebr. Paleontol. 19S:9–100.

BROCHU, C. A. 1999b. Taxon sampling and reverse successive weighting. Syst. Biol. 48:808–813.

BROCHU, C. A. 2001. Crocodylian snouts in space and time: Phylogenetic approaches toward adaptive radiation. Am. Zool. 41:564–585.

BROCHU, C. A. 2003. Phylogenetic approaches toward crocodylian history. Annu. Rev. Earth Planet. Sci. 31:357–397.

BROCHU, C. A., AND L. D. DENSMORE III. 2001. Crocodile phylogenetics: A summary of current progress. Pages 3–8 *in* Crocodilian biology and evolution (G. C. Grigg, F. Seebacher, and C. E. Franklin, eds.). Surrey Beatty and Sons, Chipping Norton, New South Wales, Australia.

BROCHU, C. A., M. L. BOUARÉ, F. SISSOKO, E. M. ROBERTS, AND M. A. O'LEARY. 2002. A dyrosaurid crocodyliform braincase from Mali. J. Paleont. 76:1060–1071.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

BUSCALIONI, A. D., F. ORTEGA, D. B. WEISHAMPEL, AND C. M. JIANU. 2001. A revision of the crocodyliform *Allodaposuchus precedens* from the Upper Cretaceous of the Hateg Basin, Romania: Its relevance in the phylogeny of Eusuchia. J. Vertebr. Paleontol. 21:74–86.

CLARK, J. M. 1994. Patterns of evolution in Mesozoic Crocodyliformes. Pages 84–97 *in* In the shadow of the dinosaurs (N. C. Fraser and H.-D. Suess, eds.). Cambridge Univ. Press, New York.

COLLUM, R. G., D. F. CLAYTON, AND F. W. ALT. 1991. Structure and expression of canary *myc* family genes. Mol. and Cell. Biol. 11:1770–1776.

DENSMORE, L. D., III 1983. Biochemical and immunological systematics of the order Crocodilia. Evol. Biol. 15:397–465.

DENSMORE, L. D., III AND H. C. DESSAUER. 1984. Low levels of protein divergence detected between *Gavialis* and *Tomistoma*: Evidence for crocodilian monophyly? Comp. Biochem. and Physiol. 77B:715–720.

DENSMORE., L. D., III AND P. S. WHITE. 1991. The systematics and evolution of the Crocodilia as suggested by restriction endonuclease analysis of mitochondrial and ribosomal DNA. Copeia 1991:602–615.

ERICSON, P. G. P., U. S. JOHANSSON, AND T. J. PARSONS. 2000. Major divisions in oscines revealed by insertions in the nuclear gene c-*myc*: A novel gene in avian phylogenetics. Auk 117:1069–1078.

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

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791.

FRISCHAUF, A.-M., H. LEHRACH, A. POUSTKA, AND N. MURRAY. 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. 170:827–842.

GALTIER, N., AND M. GOUY. 1998. Inferring pattern and process: Maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol. Biol. Evol. 15:871–879.

GATESY, J., AND G. D. AMATO. 1992. Sequence similarity of 12S ribosomal segment of mitochondrial DNAs of gharial and false gharial. Copeia 1992:241–243.

GATESY, J., G. AMATO, M. NORELL, R. DESALLE, AND C. HAYASHI. 2003. Combined support for wholesale taxic atavism in gavialine crocodylians. Syst. Biol. 52:403–422.

GATESY, J., R. DESALLE, AND W. WHEELER. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. Mol. Phylogenet. Evol. 2:152–157.

GRAUR, D., W. A. HIDE, AND W.-H. LI. 1991. Is the guinea-pig a rodent? Nature 351:649–652.

GRAYBEAL, A. 1994. Evaluating the phylogenetic utility of genes: A search for genes informative about deep divergences among vertebrates. Syst. Biol. 43:174–193.

HADDRATH, O., AND A. J. BAKER. 2001. Complete mitochondrial DNA genome sequences of extinct birds: Ratite phylogenetics and the vicariance biogeography hypothesis. Proc. R. Soc. Lond. B 268:939–945.

HASEGAWA, M., H. KISHINO, AND T. YANO. 1985. Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

HASS, C. A., M. A. HOFFMAN, L. D. DENSMORE III, AND L. R. MAXSON. 1992. Crocodilian evolution: Insights from immunological data. Mol. Phylogenet. Evol. 1:193–201.

HEDGES, S. B., AND L. L. POLING. 1999. A molecular phylogeny of reptiles. Science 283:998–1001.

HUELSENBECK, J. P., AND F. RONQUIST. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

HUELSENBECK, J. P., F. RONQUIST, R. NIELSEN, AND J. P. BOLLBACK. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.

IRESTEDT, M., U. S. JOHANSSON, T. J. PARSONS, AND P. G. P. ERICSON. 2001. Phylogeny of major lineages of suboscines (Passeriformes) analysed by nuclear DNA sequence data. J. Avian Biol. 32:15–25.

JOHANSSON, U. S., T. J. PARSONS, M. IRESTEDT, AND P. G. P. ERICSON. 2001. Clades within the "higher land birds", evaluated by nuclear DNA sequences. J. Zool. Syst. Evol. Res. 39:37–51.

KISHINO, H., AND M. HASEGAWA. 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.

LOCKHART, R. J., M. A. STEEL, M. D. HENDY, AND D. PENNY. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11:605–612.

MARCU, K. B., S. A. BOSSONE, AND A. J. PATEL. 1992. Myc function and regulation. Annu. Rev. Biochem. 61:809–860.

MCCRACKEN, K. G., J. HARSHMAN, D. A. MCCLELLAN, AND A. D. AFTON. 1999. Data set incongruence and correlated character evolution: An example of functional convergence in the hind-limbs of stifftail diving ducks. Syst. Biol. 48:683–714.

MIYAMOTO, M. M., C. A. PORTER, AND M. GOODMAN. 2000. C-*myc* gene sequences and the phylogeny of bats and other eutherian mammals. Syst. Biol. 49:501–514.

MOHAMMAD-ALI, K., M. ELADERI, AND F. GALIBERT. 1995. Gorilla and orangutan c-*myc* nucleotide sequences: Inferences on hominid phylogeny. J. Mol. Evol. 41:262–276.

NAYLOR, G. J. P., AND D. C. ADAMS. 2001. Are the fossil data really at odds with the molecular data? Morphological evidence for Cetartiodactyla phylogeny reexamined. Syst. Biol. 50:444–453.

NORELL, M. A. 1989. The higher level relationships of the extant Crocodylia. J. Herpetol. 23:325–335.

NORELL, M. A., J. M. CLARK, AND J. H. HUTCHISON. 1994. The Late Cretaceous alligatoroid *Brachychampsa montana* (Crocodylia): New material and putative relationships. Am. Mus. Novit. 3116:1–26.

POE, S. 1996. Data set incongruence and the phylogeny of crocodilians. Syst. Biol. 45:393–414.

PRENDERGAST, G. C. 1997. Myc structure and function. Pages 1–28 *in* Oncogenes as transcriptional regulators. Volume 1. Retroviral oncogenes (M. Yaniv and J. Ghysdael, eds.). Birkhäuser Verlag, Basel.

PRYCHITKO, T. M., AND W. S. MOORE. 1997. The utility of DNA sequences of an intron from the b-fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). Mol. Phylogenet. Evol. 8:193–204.

RIEPPEL, O. 1994. Osteology of *Simosaurus gaillardoti* and the relationships of stem-group Sauropterygia. Fieldiana Geol. N.S. 28:1–85.

RIEPPEL, O., AND R. R. REISZ. 1999. The origin and early evolution of turtles. Annu. Rev. Ecol. Syst. 30:1–22.

RYAN, K. M., AND G. D. BIRNIE. 1996. *Myc* oncogenes: The enigmatic family. Biochem. J. 314:713–721.

SALISBURY, S. W., AND P. M. A. WILLIS. 1996. A new crocodylian from the Early Eocene of southeastern Queensland and a preliminary investigation of the phylogenetic relationships of crocodyloids. Alcheringa 20:179–226.

SAMBROOK, J., E. F. FRITSCH, AND T. MANIATIS. 1989. Molecular cloning: A laboratory manual, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SEBER, G. A. F. 1973. The estimation of animal abundance and related parameters. Griffin, London.

SHEDLOCK, A. M., M. C. MILINKOVITCH, AND N. OKADA. 2000. SINE evolution, missing data, and the origin of whales. Syst. Biol. 49:808–817.

SHEDLOCK, A. M., AND N. OKADA. 2000. SINE insertions: Powerful tools for molecular systematics. BioEssays 22:148–160.

SWOFFORD, D. L. 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pages 295–333 *in* Phylogenetic analysis of DNA sequence data (M. M. Miyamoto and J. Cracraft, eds.). Oxford Univ Press, New York.

SWOFFORD, D. L. 1998. PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Sunderland, Massachusetts.

TAMURA, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition–transversion and G+C content biases. Mol. Biol. Evol. 9:678–687.

TRUEMAN, J. W. H. 1998. Reverse successive weighting. Syst. Biol. 47:733–737.

VAN DIJK, M. A. M., E. PARADIS, S. F. CATZEFLIS, AND W. W. DE JONG. 1999. The virtues of gaps: Xenarthran (edentate) monophyly supported by a unique deletion in a A-crystallin. Syst. Biol. 48: 94–106.

WATSON, D. K., E. P. REDDY, P. H. DUESBERG, AND T. S. PAPAS. 1983. Nucleotide sequence analysis of the chicken c-*myc* gene reveals homologous and unique coding regions by comparison with the transforming gene of avian myelocytomatosis virus MC29, Δ*gag-myc*. Proc. Natl. Acad. Sci. USA 80:2146–2150.

WEBER, J. L., AND C. WONG. 1993. Mutation of human short tandem repeats. Hum. Mol. Genet. 2:1123–1128.

WHITE, P. S., AND L. D. DENSMORE III. 2001. DNA sequence alignments and data analysis methods: Their effect on the recovery of crocodylian relationships. Pages 29–37 *in* Crocodilian biology and evolution (G. C. Grigg, F. Seebacher, and C. E. Franklin, eds.). Surrey Beatty & Sons, Chipping Norton, New South Wales, Australia.

First submitted 12 August 2002; reviews returned 10 November 2002; final acceptance 3 February 2003
Associate Editor: Allan Baker