

Department of Biological Sciences, Southeastern Louisiana University, Hammond, LA, USA

## Proopiomelanocortin (POMC) and testing the phylogenetic position of turtles (Testudines)

ROSEMARY E. BECKER, ROLDÁN A. VALVERDE and BRIAN I. CROTHER

### Abstract

Morphological and molecular studies have inferred multiple hypotheses for the phylogenetic relationships of Testudines. The hypothesis that Testudines are the only extant anapsid amniotes and the sister taxon of diapsid amniotes is corroborated by morphological studies, while the hypothesis that Testudines are diapsid amniotes is corroborated by more recent molecular and morphological studies. In this study, the placement of Testudines is tested using the full length cDNA sequence of the polypeptide hormone precursor proopiomelanocortin (POMC). Because only extant taxa have been used, the hypotheses being tested are limited to the following (1) Testudines as the sister taxon of Archosauria, (2) Testudines included in Archosauria and the sister taxon of Crocodylia, (3) Testudines as the sister taxon of Lepidosauria, (4) Testudines as the sister taxon of Sauria, and (5) Testudines as the sister taxon of a monophyletic Mammalia–Sauria clade. Neither Maximum likelihood, Bayesian, or maximum parsimony analyses are able to falsify the hypothesis of (Archosauria (Lepidosauria, Testudines)) and as such is the preferred inference from the POMC data.

**Key words:** POMC – Testudines – parsimony – likelihood – Bayesian – direct optimization – partitions

### Introduction

The placement of Testudines among amniotes has long been uncertain. Multiple hypotheses of the evolutionary relationships of turtles and other amniotes are supported by morphological and molecular data. Gaffney (1980) inferred that turtles are sister to all amniotes, but the hypothesis that Synapsida is the sister taxon of Testudines and Diapsida is corroborated by the structure of the aortic arch and ear region in amniotes (Gaffney 1980). This latter hypothesis has been corroborated by morphological and molecular studies (see below). Other studies suggest alternative hypotheses about the relationship of turtles and other diapsid amniotes; therefore, Testudines may be the only extant descendent of extinct anapsid amniotes, or they may be diapsid amniotes that have lost their temporal fenestrae.

A series of studies have inferred that Testudines are anapsid amniotes (Gregory 1946; Gauthier et al. 1988a,b; Lee 1993, 1997; Laurin and Reisz 1995; Reisz and Laurin 1991). When considering only extant taxa, these hypotheses show Testudines as the anapsid sister taxon of extant Diapsida (Reisz 1997; Meyer and Zardoya 2003).

A third hypothesis suggests Testudines are anapsid amniotes, and pareiasaurs are the sister taxon of Testudines (Gregory 1946; Lee 1993, 1997). The Paresiasauria, which contains the most recent common ancestor of turtles and pareiasaurs, and all its descendants, was first inferred by Lee (1993), and subsequently corroborated when Lee (1997) recoded and reanalyzed characters described by deBraga and Rieppel (1997).

It has also been suggested that turtles are the sister group to, or actually are, diapsid reptiles (Hedges 1994; Rieppel and deBraga 1996; deBraga and Rieppel 1997; Zardoya and Meyer 1998; Hedges and Poling 1999; Kumazawa and Nishida 1999; Mannen and Li 1999; Rieppel and Reisz 1999; Cao et al. 2000; Meyer and Zardoya 2003; Rest et al. 2003; Hill 2005; Iwabe et al. 2005; Werneburg and Sánchez-Villagra 2009). There are five main hypotheses of turtle–diapsid relationships:

(1) Archosauria (includes dinosaurs, birds, crocodylians) is the sister taxon of Testudines, (2) Testudines is part of Archosauria and sister to Crocodylia (3) extinct Sauropterygia (includes Mesozoic marine lizards like plesiosaurs), which are sister to Lepidosauriformes (includes lepidosaurs and extinct relatives), are the closest relatives of turtles, (4) Testudines are sister to Lepidosauria (rhynchocephalians, lizards), and (5) Testudines are not diapsids but are the sister taxon of Sauria, which is a clade containing Archosauria and Lepidosauria.

In this study we add a new data set [proopiomelanocortin (POMC) sequence] to the question of the placement of Testudines in amniote phylogeny. This data set is not meant to be a magic gene in the sense that it will provide ‘the’ answer, but instead is examined because of its unique multiple components and because high molecular weight polypeptide hormones and polypeptide hormone precursor sequences have been used successfully in phylogenetic analyses of deeper relationships (Dores et al. 1996).

Proopiomelanocortin is a polypeptide hormone precursor, which is a member of the opioid/orphanin gene family along with Proenkephalin (Proenk), Prodorphin (Prodyn), and Proorphanin (Proorph), though Proorph differs significantly from the other polypeptide hormone precursors in the family and does not bind to opioid receptors (Danielson and Dores 1999). These genes are found only in vertebrates, and are hypothesized to be the result of a series of gene duplication events. It has been hypothesized that Proenk duplicated and gave rise to POMC; the second duplication event resulted in Proorph; and the third duplication event produced Prodyn (Danielson and Dores 1999).

Proopiomelanocortin undergoes posttranslational modification to produce multiple hormones and has conserved regions, which become hormones. Variable spacer regions are also present. Having both conserved and variable regions suggest POMC may be a useful phylogenetic tool. In this study, we generated *Trachemys scripta* (Shoepff, 1792) (turtle, Pond Slider) and *Agkistrodon piscivorus* (Lacépède, 1789) (snake, Cottonmouth) POMC cDNA sequences and used 36 additional POMC cDNA sequences from GenBank (Table S1) to investigate the existing hypotheses for the placement of

Corresponding author: Brian I. Crother (bcrother@selu.edu)

Contributing authors: Rosemary E. Becker (rosemary.becker@selu.edu), Roldan Valverde (roldan.valverde@selu.edu)

Testudines among amniotes using maximum parsimony, maximum likelihood (ML), and Bayesian phylogenetic analyses. We partitioned the sequences based on posttranslational processing of POMC and codon position, and direct optimization (DO) using parsimony criterion. Because only extant taxa could be used this limited the testable hypotheses to (1) Testudines as the sister taxon of Archosauria, (2) Testudines included in Archosauria and the sister taxon of Crocodylia, (3) Testudines as the sister taxon of Lepidosauria, (4) Testudines as the sister taxon of Sauria (the clade containing Archosauria and Lepidosauria), and (5) Testudines as the sister taxon to a monophyletic Mammalia–Sauria clade.

## Methods and Materials

### Micro dissection of pituitary glands and RNA extraction

Pituitary glands were micro dissected from *T. scripta* and *A. piscivorus*, placed in RNAlater (Ambion Inc., Austin, TX, USA) and stored at  $-80^{\circ}\text{C}$ . RNA was isolated using TRIzol Reagent (Invitrogen, Carlsbad, CA, USA). Because of the small size of the pituitary, RNA was extracted from three pituitary glands (approximately 25 mg of tissue) at once.

### Initial 300bp fragment of POMC sequence

SuperScript II reverse transcriptase (RT; Invitrogen) was used to produce a cDNA for *T. scripta*. The poly (dT)<sub>17</sub> primer GACTCGA-GTCGGATCCATCGATTTTTTTTTTTTTTTTTT (Dores et al. 1999) was used to prime the DNA polymerization reaction of messenger RNAs. Several degenerate primer sets were designed against the conserved amino acid regions of 'MEHFRWG' and 'MFGGYRK.' These sequences yielded the primers PC2002F (CAYTTYMGITGG-GGIAARC) and PC1001R (CATRAAICCICCRTAICKYTT) which were used in a polymerase chain reaction (PCR) with an initial denaturing step of two minutes at  $94^{\circ}\text{C}$ , followed by 30 cycles at ( $95^{\circ}\text{C} \times 45\text{ s}$ ,  $50.4^{\circ}\text{C} \times 60\text{ s}$ ,  $72^{\circ}\text{C} \times 60\text{ s}$ ), and a 2 min extension at  $72^{\circ}\text{C}$ . Under these conditions, the reaction yielded a single amplicon of about 300bp of the *T. scripta* POMC cDNA sequence.

SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen) was used to produce a cDNA for *A. piscivorus*. The *A. piscivorus* cDNA first strand was made using 8  $\mu\text{l}$  of RNA, the oligo poly (dT)<sub>17</sub> primer and following the Invitrogen protocol. PCR was a 20  $\mu\text{l}$  reaction using GoTaq Green Master Mix (Promega, Madison, WI, USA) and the degenerate primer set PC1001F (ATGGARCAAYTTYMGITGGGG) and PC1001R with an initial four minutes at  $95^{\circ}\text{C}$ , 30 cycles at ( $95^{\circ}\text{C} \times 30\text{ s}$ ,  $50^{\circ}\text{C} \times 30\text{ s}$ ,  $72^{\circ}\text{C} \times 30\text{ s}$ ), and a five minute extension at  $72^{\circ}\text{C}$ . 10  $\mu\text{l}$  of *A. piscivorus* POMC PCR product was run along with 100bp DNA Ladder (Promega) on a 1% agarose gel at 100 V for 25 min showing a 300bp band. The remaining 10  $\mu\text{l}$  of PCR product mixed with 10  $\mu\text{l}$  of water were purified using EdgeBioSystems QuickStep™2 PCR Purification kit (Single Cartridges) (Edge BioSystems, Gaithersburg, MD, USA).

The purified PCR products were cloned into a TOPO-4 vector (Invitrogen) and transformed into TOP 10 cells (Invitrogen). Twenty microlitre of the transformed *E. coli* cells were plated on 50  $\mu\text{g}\ \mu\text{l}^{-1}$  ampicillin agar plates. Three colonies were selected and grown up in 3 ml and then 10 ml 50  $\mu\text{g}\ \mu\text{l}^{-1}$  ampicillin cultures. QIAprep miniprep plasmid DNA purification system (Qiagen, Valencia, CA, USA) was used to purify one of the 10 ml cultures. The QIAprep spin miniprep kit protocol was followed, except the plasmid DNA was eluted with

nuclease free water in the final step. The purified plasmid DNA was digested with EcoRI to identify the *A. piscivorus* POMC PCR insert. The plasmid DNA was subsequently sequenced.

### Obtaining 3' end of POMC sequence

The same cDNA as described above was used to produce the POMC 3' end PCR product. For *T. scripta* the Invitrogen (Carlsbad, CA) GeneRacer protocol was followed using the primers ppPOMCf (GCCCTGTTGCTGGCCCTCCTGCTTCAGA) and the GeneRacer 3' Primer (GCTGTCAACGATACGCTACGTAACG). For *A. piscivorus*, 20  $\mu\text{l}$  PCR was carried out using GoTaq Green Master Mix (Promega) and the degenerate primer PC1001F and primer PC1024R. The subsequent product was transformed, grown, and three plasmid DNA samples were sequenced.

### Obtaining 5' end of POMC sequence

Fresh RNA was extracted from pituitary glands of *T. scripta* and *A. piscivorus* using TRIzol Reagent (Invitrogen) as described above. RaceReady cDNA was made by following the Invitrogen GeneRacer protocol using 7  $\mu\text{l}$  of RNA. The original RNA was dephosphorylated, the mRNA cap structure was removed, and the RNA oligo was ligated onto the RNA. The mRNA was reverse transcribed following the Invitrogen GeneRacer protocol using SuperScript III. The purified *T. scripta* and *A. piscivorus* 5' end PCR products were cloned into TOPO-4 vectors (Invitrogen), transformed, grown, and sequenced.

### Analysis of POMC sequence fragments

Vector NTI version 10.0 (Invitrogen) was used to align the three copies of the *T. scripta* and *A. piscivorus* POMC cDNA 3' and 5' sequence fragments to determine the complete *T. scripta* and *A. piscivorus* POMC cDNA sequences and to generate the predicted amino acid sequences.

### Phylogenetic analysis

#### Sequence alignment

*Agkistrodon piscivorus* and *T. scripta* full length POMC amino acid and nucleotide sequences and 36 other POMC amino acid and nucleotide sequences from GenBank (Table S1) were aligned using BioEdit Sequence Alignment Editor. The amino acid sequences were aligned and this alignment was used as a guide to align the nucleotide sequences. Nucleotides before the start codon and after the stop codon were removed prior to analyses. The first of these was the region in between the CLIP-like region associated with  $\gamma$ -MSH and  $\alpha$ -MSH, and the second was the region in between CLIP and  $\delta$ -MSH (Fig. 1).

#### Outgroups

Because this is basically a study on amniote and tetrapod phylogeny, exemplars from each major amniote and tetrapod were included (see Table S1). A number of outgroup taxa were employed in the analyses, including two cartilaginous fishes, 14 bony fishes (primitive and derived forms), and two lung fishes.

#### Analytical approaches

A number of different methods were used to analyze this data set. Each of them can be recommended for different reasons with the only certainty they have in common is that none are perfect. As such,

POMC-N		ACTH		$\delta$ -MSH	$\beta$ -LPH	
POMC-N	$\gamma$ -MSH	$\alpha$ -MSH	CLIP	$\delta$ -MSH	$\beta$ -MSH	$\beta$ -END

Fig. 1. Four partitions and seven partitions of proopiomelanocortin cDNA based on posttranslational modification events

we inferred phylogenetic hypotheses from parsimony, likelihood and Bayesian approaches to see if there were any consistent patterns in testing the five hypotheses of turtle relationships. We also employed DO because of the nature of the POMC sequences with large gaps and rapidly evolving regions and because we were curious to see how DO performed relative to the more standard methods. The same logic follows for our use of various statistics to assess confidence in the analyses. Because they either suffer from problems (e.g. unrealized assumptions for the bootstrap) or there is simply a lack of understanding on how to statistically interpret them (e.g. decay index), we were interested in seeking common patterns of support or lack thereof to help inform us on the utility of POMC for studies at this phylogenetic level. Finally, we recognize that these analyses are inferring a gene tree and not necessarily a species tree. The interesting question then is, does the gene tree corroborate any of the five hypotheses?

#### Maximum parsimony

PAUP\*4.0b10 (Swofford 2002) was used to conduct an equally weighted parsimony analysis of the entire POMC cDNA sequence alignment across 38 taxa. In this analysis, gaps were treated as missing data, except for three regions (gaps in base pair regions 151–174, 292–420, and 562–768) where insertion/deletion (indel) events presumably occurred and were treated as a fifth character state (Ogden and Rosenberg 2007). First, each nucleotide in the indel was coded as a fifth character. In the second parsimony analysis, the entire indel is coded as a single fifth character. A heuristic search was conducted using 1000 replicates of random stepwise addition. Branch swapping was performed using tree-bisection and reconnection (TBR). Character states were optimized using accelerated transformation. Robustness of the relationships was assessed by bootstraps (2000 replicates), character jackknife (2000 replicates) with 25% character deletion, taxon jackknifing, and decay indices (TreeRot v. 2b; Sorenson and Franzosa 2007).

Because of the results of the second parsimony analysis, which coded each indel as a single character, one additional maximum parsimony (MP) analysis was conducted. The analysis was the same as described above, except it topologically constrained Mammalia to a monophyletic group.

#### Maximum likelihood

ModelTest Server 1.0 running ModelTest 3.8 (Posada and Crandall 1998) was used to determine the model that best fit the data for the entire POMC cDNA sequence alignment using the corrected Akaike Information Criterion as per Posada and Buckley (2004) and Posada and Crandall (2001). All gaps were treated as missing data because general time reversible models do not allow for a fifth character state. Garli0.951-GUI (Zwickl 2006) was used to perform the ML analysis using the GTR+I+ $\Gamma$  model of nucleotide evolution and to conduct 2000 bootstrap replicates.

#### Bayesian inference

MrBayes v3.0b4 (Ronquist and Huelsenbeck 2003) was used to perform a Bayesian inference (BI) of the entire POMC cDNA sequence alignment. In the BI as in the ML analysis, the GTR+I+ $\Gamma$  model of nucleotide evolution was used. In the analysis, the MrBayes v3.0b4 default of four chains, three of which were heated, one of which was cold, and the default uniform priors were used to run 5 000 000 generations saving every 100th tree. Two independent runs of BI were performed to test for convergence of posterior probabilities; if the posterior probabilities from the individual runs were within three units, the topologies were considered convergent. Tracer v1.4 (<http://tree.bio.ed.ac.uk/software/tracer/>) was used to graphically look at burn-in to determine the number of trees to discard. Posterior probabilities from the tree file produced in MrBayes v3.0b4 and filtering the tree file using five constraint trees (the five hypotheses: Testudines sister to Sauria, Archosauria sister to the Testudines–Lepidosauria clade, Lepidosauria sister to the Archosauria–Testudines clade, Testudines sister to a Mammalia–Sauria clade, and Testudines as the sister taxon of Crocodylia) were executed in PAUP\*4.0b10 (Swofford 2002). If fewer than 5% of the trees produced in the Bayesian tree file support a hypothesis, then the hypothesis can be statistically rejected under Bayesian criterion (Weisrock et al. 2006).

#### Partitions

To explore the POMC cDNA sequence data, the sequences were partitioned five ways based on codon position and posttranslational processing forming a total of six partitioning schemes (Fig. 1). These partitioning schemes are as follows (Table S2): the entire POMC cDNA as one partition ( $P_1$ ) as is described above, the entire POMC gene with a partition for each codon position ( $P_C$ ), one partition for each of the four regions of POMC based on posttranslational processing ( $P_4$ ), one partition for each of the codon positions of the four regions of POMC ( $P_{4C}$ ), one partition for seven regions of POMC based on posttranslational processing ( $P_7$ ), and one partition for each of the codon positions of the seven regions of POMC ( $P_{7C}$ ). To determine the best partitioning scheme for this data set, each partitioning scheme was analysed using MrBayes v3.0b4 (Ronquist and Huelsenbeck 2003). Model selection for each partition and the overall analysis were conducted as described above. The 2ln Bayes factor was used to determine the best hypothesis for partitioning the data as described in Kass and Raftery (1995), Nylander et al. (2004), and Brandley et al. (2005). Posterior probabilities and tree filtering for constraint trees were executed as described earlier.

#### Direct optimization

POY version 4.0 Beta Build 2602 (Varón et al. 2008) was used to infer phylogeny using DO (Wheeler 1996). Because the *A. piscivorus* and the *Coturnix japonica* Temminck and Schlegel, 1849 (bird, Japanese Quail) POMC cDNA are partial sequences, the POMC cDNA was divided into two fragments. The first fragment contains nucleotides from the start codon to the beginning of the *C. japonica* partial cDNA sequence; the second fragment contains nucleotides homologous with the first nucleotide of the *C. japonica* partial sequence to the nucleotides in the stop codon. Though the multiple sequence alignment was used to determine the nucleotides homologous with the first nucleotide of the *C. japonica* partial sequence, the POMC cDNA sequences were unaligned when analyzed using DO. For *Heterodontus portusjacksoni* (Meyer, 1793) (Port Jackson shark) 174 nucleotides and for *Chimaera phantasma* Jordan and Schneider, 1900 (cartilaginous fish, Chimaera) 189 nucleotides present only in Chondrichthyes and associated  $\delta$ -MSH and its CLIP-like region, were removed to reduce sequence length variation as per Giribet (2001). A heuristic search was conducted using 500 replicates of random stepwise addition. Branch swapping was performed using TBR. The dynamic characters used in the analysis were then transformed into corresponding static homology characters, which were output into a file read by MacClade4. MacClade4 v.4.06 (Maddison and Maddison 2003) was used to optimize the static homology characters using accelerated transformation and diagnoses the synapomorphies for each monophyletic group.

Because of the results from DO analysis, two additional MP analyses using PAUP\*4.0b10 (Swofford 2002) were conducted. The first of these analyses was the same as described above, except all gaps were treated as missing data. The second analysis also treated all gaps as missing data, and it topologically constrained Mammalia to a monophyletic group.

## Results

### *Trachemys scripta* and *Agkistrodon piscivorus* cDNA Sequences

The full length *T. scripta* POMC cDNA sequence included 1255 bases and the poly A tail. The *A. piscivorus* POMC cDNA sequence included 1000 bases and the poly A tail. The *A. piscivorus* POMC cDNA sequence appears to be incomplete and is missing the first 303 bases.

Our turtle sequence was produced before that of *Chrysemys* (= *Trachemys*) *scripta* (Shoureshi et al. 2007) and differs by 10 internal nucleotides from the sequence published in GeneBank (accession number DQ986316), including two deletions, in addition to a longer sequence on both 5' and 3' ends. We do not know if the differences are because of sequencing errors or intrinsic population differences.

### Sequence alignment

After the alignment of POMC cDNA, it became apparent that there are three regions where gaps are the result of an insertion or deletion event. These regions are from nucleotide position 151–174, from nucleotide 292–420, and from 562 to 768. These regions are all multiples of three and are thought to be indels involving codons.

### Phylogenetic analyses

MP, which treated indels as a fifth character, identified 586 parsimony informative, 197 parsimony uninformative, and 163 constant characters and yielded six most parsimonious trees (mpts) with 3277 steps. In each of the six mpts, there exists a monophyletic amniote group, diagnosed by one unambiguous and 37 ambiguous synapomorphies, containing a monophyletic mammalian group, diagnosed by 22 unambiguous and 41 ambiguous synapomorphies, and a monophyletic saurian (Sauria = archosaurs + lepidosaurs) group including Testudines, diagnosed by 21 ambiguous synapomorphies. A monophyletic amphibian clade is sister to the amniote group. In all six of the mpts, Testudines is sister to Lepidosauria, diagnosed by 25 ambiguous synapomorphies on one of the six mpts and 23 ambiguous synapomorphies on the strict consensus tree, with Archosauria sister to the Testudines–Lepidosauria clade (Fig. 2). The strict consensus tree and the 50% majority rule

consensus tree (not shown) have the same topology. Bootstrap, jackknife, taxon-jackknife, and decay indices are shown in Fig. 3.

Of the 951 base pairs included in the second MP analysis, which coded each entire indel as a single fifth character, 456 are parsimony informative characters, 170 are parsimony uninformative, and 325 are constant characters. The MP analysis resulted in nine mpts with 2891 steps having a CI of 0.400, a RI of 0.578, and a RCI of 0.231. The resulting topology is different than the topology recovered from the first MP analysis in which each gap in the indel is coded as a fifth character state. First, *Monodelphis domestica* (Wagner, 1842) (mammal, grey short-tailed opossum) is not in a clade with the other mammalian taxa, breaking up the monophyly of Mammalia. Instead, *M. domestica* is the sister taxon of the remaining tetrapods. Second, Aves is the sister taxon of the mammalian clade, excluding *M. domestica*, which is supported by 19 ambiguous synapomorphies on one of the nine mpts and 18 ambiguous synapomorphies on the 50% majority rule consensus tree, therefore disrupting the monophyly of Archosauria and the monophyly of the clade containing Sauria and Testudines, which was a result of the first MP analysis. In all nine of the mpts, Testudines and Lepidosauria are sister taxa. This clade is diagnosed by two unambiguous and 24 ambiguous synapomorphies on one of the mpts and by two unambiguous and 22 ambiguous synapomorphies on the

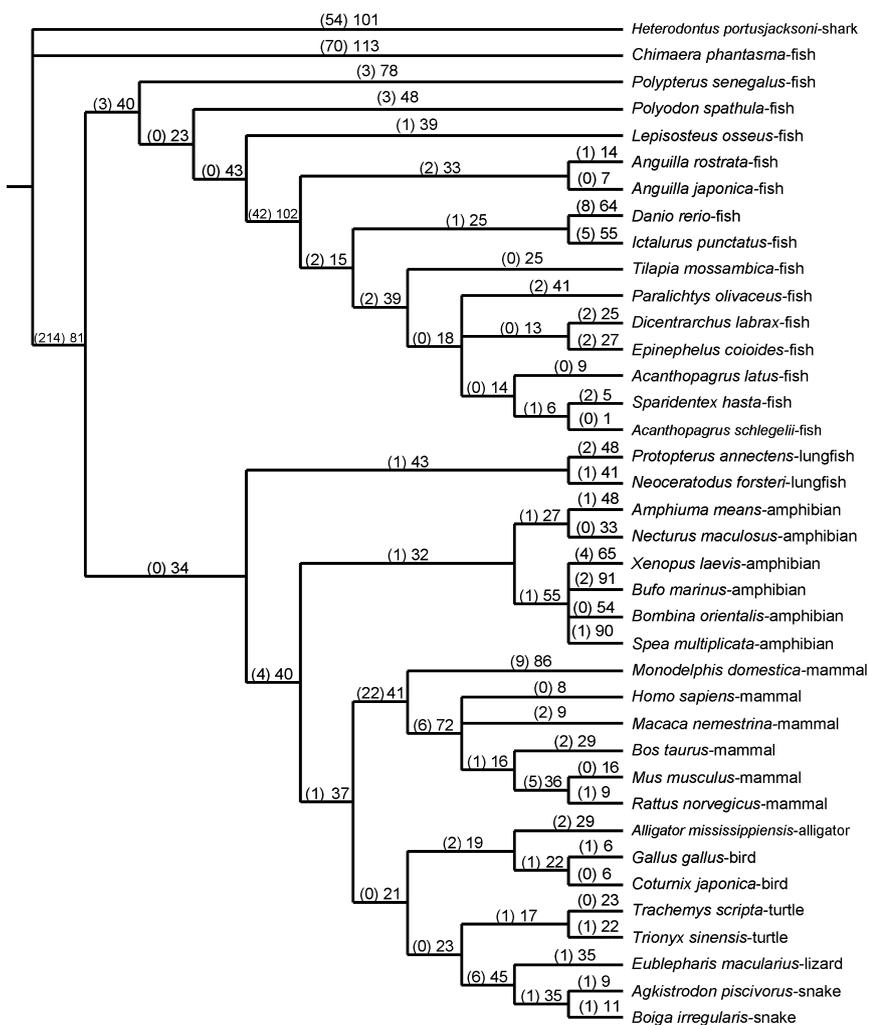


Fig. 2. Strict consensus cladogram of six most parsimonious trees with 3277 steps showing unambiguous (in parentheses) and ambiguous synapomorphies. With regard to the position of turtles as sister to lepidosaurs, this topology is the same as in the parsimony analyses where gaps are treated as missing

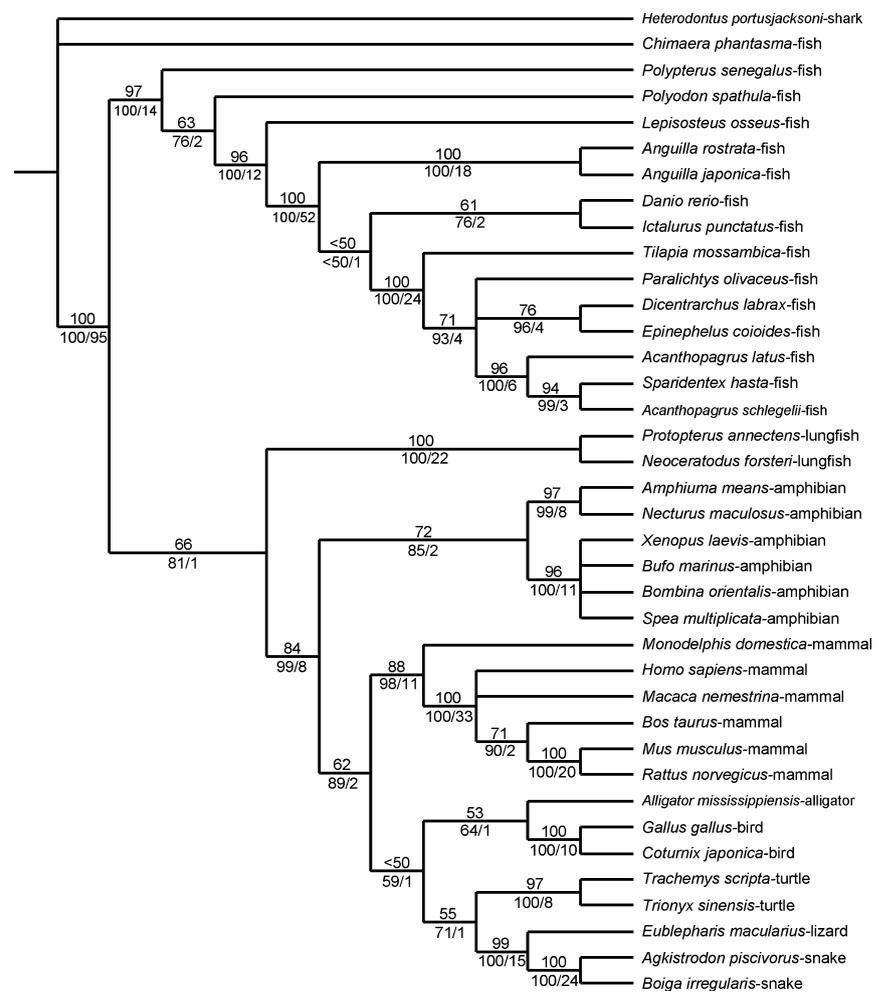


Fig. 3. Strict consensus tree of six most parsimonious trees having 3277 steps showing bootstrap proportions above the branches. The 25% deletion character jackknife values (left of slash) and decay indices (right of slash) are below the branches

50% majority rule consensus tree (not shown). The strict consensus tree does not recover a monophyletic amphibian clade and does not resolve the relationships of *M. domestica* (grey short-tailed opossum), Caudata, Anura, and the remaining amniotes. The 50% majority rule consensus tree recovers a monophyletic amphibian clade and resolves these relationships (not shown).

Constraining the monophyly of Mammalia for a MP analysis in which each of the three entire indels are coded as a single fifth character results in six mpts with 2896 steps having a CI of 0.400, a RI of 0.577, and a RCI of 0.230. This topology is five steps longer than when Mammalia is not constrained to a monophyletic group. These six mpts are identical to the results from the initial MP analysis (Fig. 2), where each gap in the indel is coded as a fifth character state. There is a monophyletic amniote group, containing a monophyletic Mammalia and a monophyletic Sauria, which includes Testudines. A monophyletic amphibian clade is sister to the amniotes. In all six of the mpts, Testudines is sister to Lepidosauria, with Archosauria as the sister taxon (not shown).

Maximum likelihood analysis resulted in a single tree with less resolution than the six mpts. The ML tree infers amniotes and amphibians as sister taxa. Like the mpts, the amniote clade contains a monophyletic Mammalia and a Sauria-Testudines clade. Unlike MP, the ML analysis is unable to resolve the relationships among the Testudines, Lepidosauria, and Archosauria monophyletic groups (not shown, but see Fig. 4).

The initial ( $P_1$ ) Bayesian analysis, like both the MP and ML analyses, the Bayesian tree has a monophyletic amniote group whose sister taxon is a monophyletic Amphibia. Within the amniotes, there are the same two monophyletic groups: a Mammalia and Sauria-Testudines. In the Bayesian analysis, as in ML, the relationships of Testudines, Lepidosauria, and Archosauria are unresolved (Fig. 4). After using constraint trees, two of the five hypotheses for the placement of Testudines among amniotes can be statistically rejected based on Bayesian criterion (Weisrock et al. 2006). The hypothesis of Testudines as the sister taxon of Sauria was shown in 45.0% of the trees. The second hypothesis, which places Archosauria as the sister taxon of the Testudines-Lepidosauria clade, was shown in 34.6% of the trees. The third hypothesis, which places Lepidosauria as the sister taxon of the Archosauria-Testudines clade, was shown in 16.8% of the trees, and the fourth hypothesis, which places Testudines as the sister taxon of the Mammalia-Sauria clade, is rejected, because it was only shown in 0.2% of the trees (Table S3). The fifth hypothesis for the placement of Testudines as the sister taxon of Crocodylia was rejected because for all of the BIs the posterior probability describing the Crocodylia-Aves clade was 100, which excludes the possibility that any trees would support a Testudines-Crocodylia clade.

Further Bayesian analyses of POMC cDNA was based on partitioning the POMC cDNA sequence data in five ways based on codon position and posttranslational processing. Based on the 2Ln Bayes factor interpretation (Table S6), the

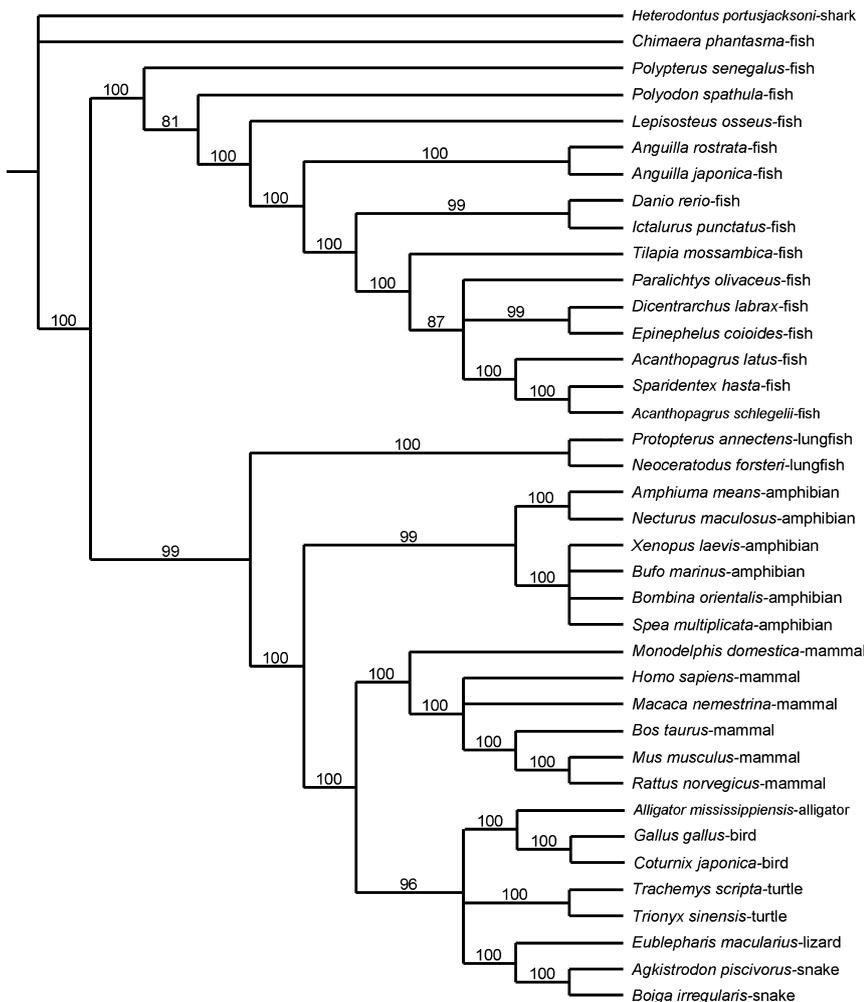


Fig. 4. Bayesian consensus tree and phylogram of the  $P_7$  partitioning scheme showing posterior probabilities. With regard to the position of turtles, this topology is the same as those inferred for maximum likelihood and the Bayesian partitioned analyses  $P_1$  and  $P_c$ . See text for partition details

$P_{7C}$  partitioning scheme is the model which best fits the POMC cDNA sequence data (Table S5). Similar to MP, ML, and  $P_1$  Bayesian analyses, the  $P_{7C}$  Bayesian majority rule tree has a monophyletic amniote group and its sister taxon is a monophyletic amphibian clade. Within the amniote group, there is a mammalian clade and a clade containing Sauria and Testudines. Differing from the prior three analyses, in the  $P_{7C}$  BI, the Testudines clade is the sister taxon of Sauria, which has a posterior probability of 0.88 (Fig. 5). The constraint tree filter rejected the hypothesis placing Lepidosauria as the sister taxon of the Archosauria–Testudines clade and the hypothesis placing Testudines as the sister taxon of a Mammalia–Sauria clade, because fewer than 5% of the trees in the tree file from the BI support these relationships (Weisrock et al. 2006). However, the hypothesis that Testudines is the sister taxon to Sauria, which was present in 88.3% of the trees produced by the BI (Table S3), and the hypothesis that Archosauria is the sister taxon of the Testudines–Lepidosauria clade, which is present in 9.4% of the trees produced by the BI (Table S3), could not be rejected.

The trees produced by BI of the  $P_c$  and  $P_7$  partitioning schemes have nearly identical topologies as the initial  $P_1$  BI, which has no partitions, except the tree produced by the  $P_7$  partitioning scheme had no resolution for the four anurans included in the analysis. All three of these trees have similar posterior probabilities (only  $P_7$  results shown, Fig. 4). The trees produced by the BI of the  $P_4$  and the  $P_{4C}$  partitioning

schemes have similar topologies to the  $P_{7C}$  partitioning scheme, except there are differences among all three topologies in the relationships of the operational taxonomic units (OTUs) in the mammalian clade. Also, the trees produced by the BI of the  $P_4$  and the  $P_{4C}$  partitioning schemes have much lower posterior probabilities describing the Sauria clade, 0.58 and 0.60 respectively (not shown but see Fig. 5). After filtering the remaining partitioning schemes using the alternative hypotheses as had been done with the  $P_1$  and the  $P_{7C}$  partitioning schemes, only the hypothesis placing Testudines as the sister taxon of a Mammalia–Sauria clade can be rejected (Table S3).

Direct optimization resulted in a single mpt with 2992 steps (Fig. 6). The mpt produced by DO has a different topology than the trees produced by MP, ML, and BI. *M. domestica* is not in a clade with the other mammalian taxa, instead, *M. domestica* is the sister taxon of the remaining tetrapods. Aves is the sister taxon of the mammalian clade, excluding *M. domestica*, which is supported by four unambiguous and 35 ambiguous synapomorphies, disrupting the monophyly of Archosauria and the clade containing Sauria and Testudines, which was inferred by MP, ML, and BI. Instead, DO inferred Testudines as the sister to (Crocodylia (Aves, Mammalia)).

Reanalysis of the multiple sequence alignment with MP with gaps treated as missing data resulted in nine mpts of 2887 steps with 453 parsimony informative characters, 170 variable but parsimony uninformative characters, and 323 constant characters. The 50% majority rule consensus tree, like the tree from

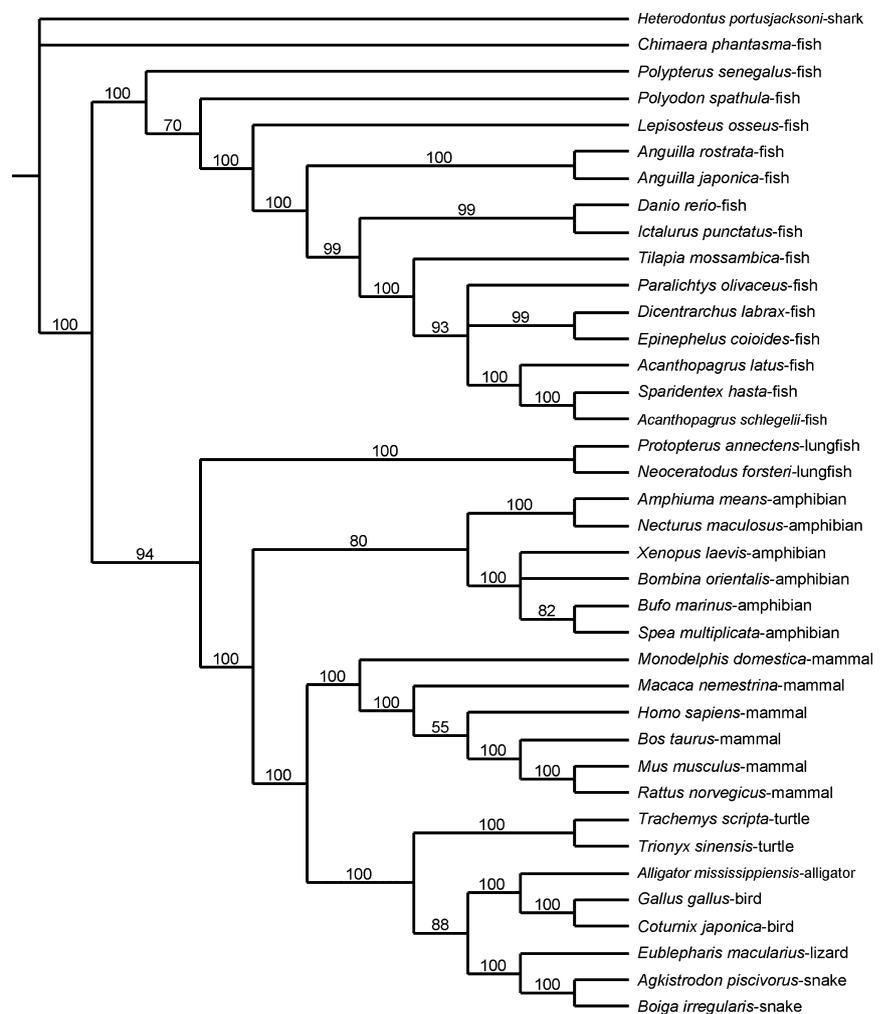


Fig. 5. Bayesian consensus tree and phylogram of the  $P_{7C}$  partitioning scheme showing posterior probabilities. With regard to the position of turtles, this topology is the same as those inferred in Bayesian partitioned analyses  $P_4$  and  $P_{4c}$ . See text for partition details

DO does not have a monophyletic Mammalia (excludes *M. domestica*), and that clade is the sister taxon of Aves, which is supported by 18 ambiguous synapomorphies on the 50% majority rule consensus tree and 19 ambiguous synapomorphies of one of the nine mpts (not shown). The reanalysis also inferred Lepidosauria–Testudines sister to Mammalia–Aves–Crocodilia, but DO did not result in Lepidosauria–Testudines. MP treating all gaps as missing data and constraining the monophyly of Mammalia, results in six mpts with 2893 steps (an additional six steps). The six mpts have an identical topology to the six mpts that resulted from the initial MP analysis, and therefore the strict consensus tree and 50% majority rule consensus tree also have identical topologies. There is a monophyletic amniote group, containing a monophyletic mammalian group and a monophyletic saurian group, which includes Testudines. A monophyletic amphibian clade is sister to the amniote group. In all six of the mpts, Testudines is sister to Lepidosauria, with the Archosauria as the sister taxon (not shown but see Figs 2 and 3).

## Discussion

### Analyses and hypotheses

Phylogenetic analysis of POMC cDNA sequences was conducted to investigate the relationship of Testudines among extant amniotes. Using only extant amniotes limits the possible

hypotheses (1) Testudines as the sister taxon of Archosauria, (2) Testudines included in Archosauria and the sister taxon of Crocodilia, (3) Testudines as the sister taxon of Lepidosauria, (4) Testudines as the sister taxon of Sauria, and (5) Testudines as the sister taxon of a monophyletic Mammalia–Sauria clade. After performing MP, ML, and BI analyses, only the MP analyses provided resolution to Archosauria, Lepidosauria, and Testudines evolutionary relationships. As shown in Figs 2 and 3, the MP analysis, in which each gap in indel is coded as a fifth character state (Ogden and Rosenberg 2007), fails to reject the hypothesis of (Testudines, Lepidosauria), with Archosauria as the sister taxon.

When an indel is longer than one amino acid or three nucleotides for protein coding sequences, as is the case in this multiple sequence alignment, and gaps are coded as fifth character states, it suggests that each gap position in the indel is independent from the next gap (Eernisse and Kluge 1993). If adjacent gaps are actually the result of a single indel, the effect the indel has on the analysis will be inflated if each position is treated as a fifth character (Simmons and Ochoterena 2000), as was the case in the first MP analysis. Thus, a second MP analysis, in which the entire indel is coded as a single fifth character, was also conducted. This MP analysis also fails to reject the hypothesis (Testudines, Lepidosauria).

A bootstrap proportion of 55% and a jackknife value of 71% described the Testudines–Lepidosauria clade (Fig. 3).

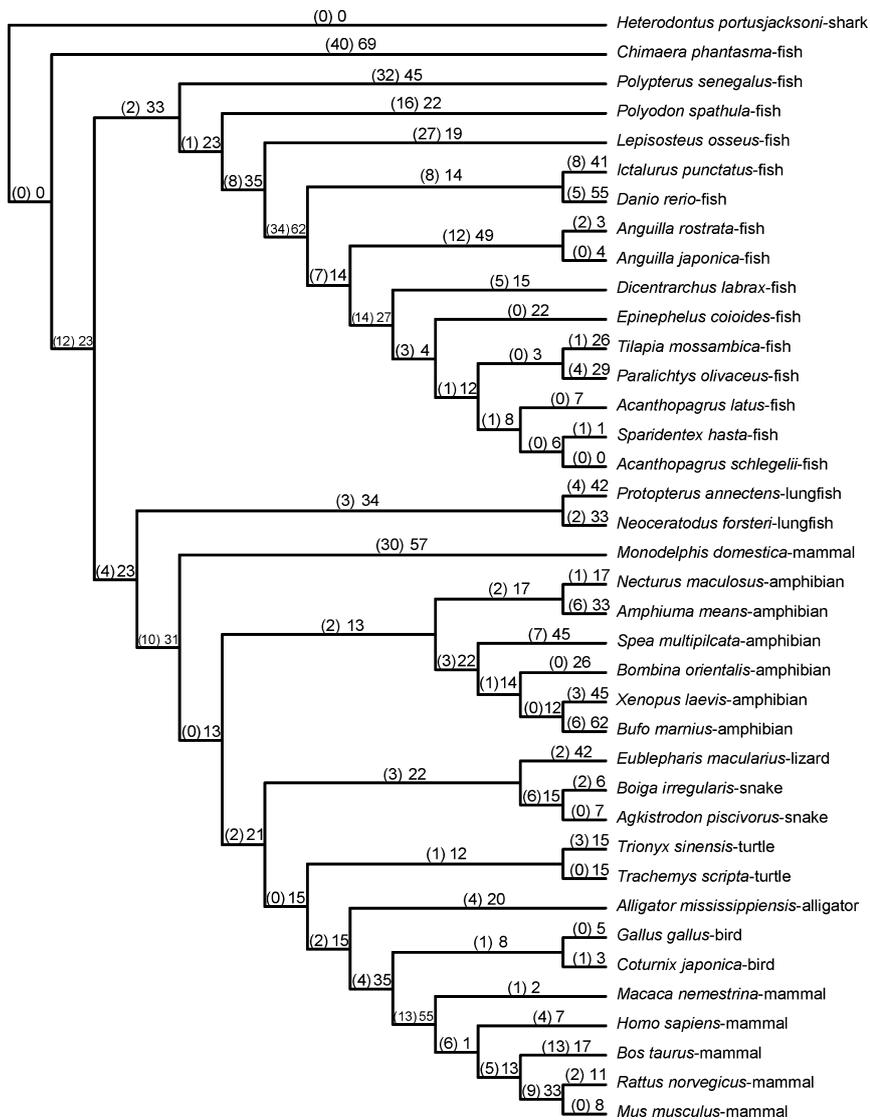


Fig. 6. Direct optimization single most parsimonious tree with 2992 steps showing unambiguous synapomorphies in parentheses and ambiguous synapomorphies diagnosing each clade using accelerated transformation. Note the non-monophyly of Mammalia and Archosauria

Hillis and Bull (1993) suggested that bootstrap values are not useful indicators of repeatability and that a second data set may not produce the same results as the first data set with high bootstrap proportions. Most bootstrap proportions over 50% are considered highly conservative, and the bootstrap proportion is actually lower than the probability that the described clade is the true clade if there are equal rates of change among taxa, <20% of the changes are internodal, and the tree topology is symmetrical (Hillis and Bull 1993). The bootstrap proportion and the jackknife value are just above 50% for the Testudines–Lepidosauria clade. Because the criteria for the Hillis and Bull (1993) interpretation of bootstrap proportions are not met, these values suggest that compared to the whole data set, relatively few characters support this clade (23 of 586 in the strict consensus tree).

Taxon jackknife analysis investigates the usefulness of phylogenetic trees by testing the robustness of data sets. If a data set is robust, removing one taxon at a time and then reanalyzing the data set should not change the tree topology, because the evolutionary relationships depicted in the hypothesis should be present in multiple taxa in the analysis (Lanyon 1985; Siddall 1995). In this study, all of the monophyletic groups shown in the strict consensus tree of the six mpts are

consistent with the removal of at least 80% of the taxa. The Testudines–Archosauria clade is present in 83% of the trees produced by the taxon jackknife analysis (Fig. 3). This clade is not present with the removal of the following OTUs: *A. piscivorus*, *T. scripta*, *Gallus gallus* (Linnaeus, 1758) (bird, Red Junglefowl), *Bos taurus* (Linnaeus, 1758) (mammal, domestic cattle), *Spea multiplicata* (Cope, 1863) (amphibian, Mexican Spadefoot), and *Neoceratodus forsteri* (Krefft, 1870) (fish, Australian Lungfish). Three of these six OTUs are in the saurian clade, and therefore directly involved in the hypotheses being investigated. The *A. piscivorus* POMC cDNA is a partial sequence, and PAUP\*4.0b10 deals with missing data by assigning the character state to the taxon with missing data that would be most parsimonious with the given tree (Swofford 2002). This method of handling missing data would reinforce the sister relationship between *A. piscivorus* and *Boiga irregularis* (Merrem, 1802) (snake, Brown Treesnake), and perhaps relationships of these and other closely related taxa. The *T. scripta* sequence is one of two turtle sequences used in this analysis, because its removal affects the topology of the mpts, there is phylogenetic information in this sequence that is not present in any other sequences. The *G. gallus* POMC cDNA sequence is the only complete bird sequence included in the

analysis. The first one-third of the quail, *C. japonica*, POMC cDNA sequence is missing, which contains approximately 60% of the parsimony informative characters. This region contains POMC-N, of which only a small portion actually codes for  $\gamma$ -MSH, making this region more variable than other regions of the POMC gene. Removing the *G. gallus* sequence removes approximately 60% of the parsimony informative characters for Aves, a critical group for investigating the placement of Testudines among amniotes. The inconsistencies of phylogenetic information present in *A. piscivorus*, *T. scripta*, and *G. gallus* POMC cDNA sequences cause changes in the topology of the most parsimonious trees when these OTUs are removed from the parsimony analysis; including more sequences and completing existing sequences could help to resolve these inconsistencies.

Bremer support or decay indices describe the number of additional steps needed to break-up a particular monophyletic group. Bremer (1988) suggested that when investigating phylogenetic relationships using parsimony, trees other than the most parsimonious trees should also be investigated. If trees that include more steps than the most parsimonious trees also contain the monophyletic groups in the most parsimonious tree, it is a good indication that the monophyletic groups are true groups (Bremer 1988). The decay index for the Testudines–Lepidosauria clade is one (Fig. 3), and is only found in the mpts, and therefore is not considered strongly supported.

Although there is low statistical confidence in the Testudines–Lepidosauria hypothesis produced by MP, there is no resolution among Archosauria, Lepidosauria, and Testudines in the ML or the initial BI analyses (Fig. 4). It is possible that the lack of resolution in the ML and BI trees is because of the fact that gaps in the alignment cannot be treated as a fifth character state in general time reversible models of evolution. Instead, the gaps had to be treated as missing data, which results in a loss of data when it is apparent in the alignment that an insertion or deletion event occurred (Giribet and Wheeler 1999; Simmons and Ochoterena 2000; Simmons et al. 2001). Another possibility for the lack of resolution in the ML and BI analyses is the need for a model of evolution that better fits the data. Perhaps using a single model of evolution for the entire POMC cDNA sequence may not have been the best choice. The POMC gene is a protein coding sequence that codes for a variety of hormones. The various regions of the gene coding for different hormones and the different codon positions may be on different evolutionary paths and therefore should be partitioned and analyzed using different models of nucleotide evolution to prevent a compromise in the estimation of nuisance parameters (Yang 1996; Huelsenbeck and Rannala 2004; Brandley et al. 2005).

Five partitioning schemes were established to determine if partitioning the POMC sequence based on posttranslational modification and codon position could produce a model that better fit the POMC cDNA data than the model chosen for the entire sequence. The  $P_{7C}$  partitioning scheme was the only partitioning scheme able to reject one of the hypotheses for the placement of Testudines among amniotes (Table S3). The  $P_{7C}$  Bayesian analysis does not reject the hypothesis of (Testudines, Sauria), which has a posterior probability of 88 (Fig. 5). This result is in conflict with (Testudines, Lepidosauria) which was supported by MP. However, it did not statistically reject (Testudines, Lepidosauria) and there are concerns with the validity of using the  $P_{7C}$  partitioning scheme as a model of evolution.

Bayes factors work by minimizing the number of partitions, and therefore minimizing the random error in estimating parameters that results from fewer nucleotides existing in each partition (Brandley et al. 2005). However, it is difficult to determine whether Bayes factors can actually provide an adequate balance between the complexity or number of partitions in a model and estimation of error (Nylander et al. 2004). The  $P_{7C}$  partitioning scheme sectioned the POMC cDNA into seven regions based on posttranslational modification of the various hormones produced by POMC, and each partition was further divided into the three codon positions producing 21 partitions. When using the  $P_{7C}$  partitioning scheme, all of the partitions had fewer than 100 nucleotides. In fact, 18 of the 21 partitions had fewer than 50 nucleotides, and 9 had fewer than 30 nucleotides, with the three smallest partitions containing only 18 nucleotides (Table S6). With so few nucleotides representing each partition and estimating the nuisance parameters for each partition, it seems likely that random error will influence the phylogeny. While little is known about how small a partition can be before random error plays a role in the phylogenetic analysis (Brandley et al. 2005), it has been shown that decreasing the number of characters in an analysis is correlated with decreasing phylogenetic accuracy (Wiens and Servedio 1998; Hillis et al. 2003), which raises concern with the  $P_{7C}$  partitioning scheme.

Both Brandley et al. (2005) and Nylander et al. (2004) questioned if the  $2Ln$  Bayes factor interpretations (Table S7) are strong enough to apply in a phylogenetic context. When Nylander et al. (2004) used Bayes factors to select the model of nucleotide evolution, the model selected by the Bayes factors resulted in a tree that was between approximately 600–900  $\ln$  likelihood units better than all the other models suggested. However, in this study, the selected model, partitioning scheme  $P_{7C}$ , was  $<100$   $\ln$  likelihood units greater than the initial unpartitioned  $P_1$  model, and was  $<10$   $\ln$  likelihood units greater than the  $P_7$  partitioning scheme model (Table S4). A  $2Ln$  Bayes factor of 9.034 supports the  $P_{7C}$  partitioning scheme over the  $P_7$  partitioning scheme (Table S5). According to the interpretation (Table S7), this is strong support to use the  $P_{7C}$  partitioning scheme, but as stated above, this only results in an increase of fewer than 10  $\ln$  likelihood units going from a harmonic mean  $\ln$  likelihood of  $-12443.415$  to  $-12453.815$ , which is likely not significant when looking at the overall likelihood. It is reasonable to instead use the  $P_7$  partitioning scheme, which was unable to resolve the evolutionary relationships of saurian amniotes and did not reject any hypotheses for the placement of Testudines among amniotes (Table S3), to minimize the number of partitions and subsequently minimize the random error.

Castoe and Parkinson (2006) found that as models of nucleotide substitution became more complex, the posterior probabilities supporting clades tended to be higher and in some instances changed the topology of the Bayesian consensus tree. Though Castoe and Parkinson (2006) credited increased posterior probabilities to more complex models providing more accurate measures of nodal support; they also suggested that more complex models could over-inflate posterior probabilities. The  $P_4$ ,  $P_{4C}$ , and the  $P_{7C}$  partitioning schemes all support the Archosauria–Lepidosauria clade with posterior probabilities of 58, 60, and 88, respectively (Fig. 5). The most complex partitioning scheme has by far the highest posterior probability. Complex partitioning schemes can cause overparameterization, especially when parameters cannot be

meaningfully estimated from the data, i.e. there are a small number of nucleotides estimating the parameters for each partition (Castoe et al. 2004). Determining the model of evolution for each partition separately may also result in overparameterization or overfitting, because it is difficult to know when two parameters should be shared across partitions (Nylander et al. 2004). It seems likely that the  $P_{7C}$  partitioning scheme is overparameterized, which causes the precision of the model to be poor, predictions from the model to be imprecise, and unnecessary sampling variance because of estimating extra parameters, which can lead to compromised phylogenetic accuracy (Burnham and Anderson 1998; Cunningham et al. 1998; Lemmon and Moriarty 2004). In this case, the high posterior probability describing the Sauria clade is likely inflated as a result of overparameterization.

Another issue with the  $P_{7C}$  partitioning scheme arises with the partitions based on the posttranslational modification of POMC. These partitions are not strictly the nucleotides that code for the amino acids which compose the hormones. The partitions are approximations based on the location of paired basic amino acids, which are the sites where endopeptidases cleave the polypeptide precursor sequence. The hormones are then further modified and the paired basic amino acids are removed (Dores 1990). Because the partitions are approximations, they may not accurately represent sequence fragments having different evolutionary paths, therefore making the partitions arbitrary.

### Direct optimization

In DO, indels and base substitutions are seen as transformation events from ancestral to descendant nucleotide sequences and are used to directly build a phylogeny given an optimality criterion (Wheeler 1996). This differs from the process of producing a multiple sequence alignment followed by phylogeny building, because sequences with length variation in multiple sequence alignments will have gaps, which are coded either as missing data or as a fifth nucleotide. When gaps are coded as a fifth nucleotide, they have the potential to become a character diagnosing a clade; while in DO, indels are transformation events resulting in a sequence that can diagnose a clade, consistent with the ideographic character concept of Grant and Kluge (2004; Kluge 2007).

The most parsimonious tree resulting from the DO analysis (Fig. 6) differs from the mpts of MP analysis (Figs 2 and 3). The difference in treatment of indels in DO using parsimony criterion and in MP analysis of multiple sequence alignment may have caused the differences in topology of the mpts.

Also, the POMC cDNA sequences were modified before DO analysis to reduce sequence length variation as per Giribet (2001). First, the sequences were divided into two sections. The *A. piscivorus* and *C. japonica* are partial cDNA sequences, and were not present in the first section of unaligned sequences. Second, 174 and 189 nucleotides were removed from *H. portusjacksoni* and *C. phantasma* respectively. These nucleotides code for the  $\delta$ -MSH and CLIP-like regions, which are only present in Chondrichthyes, and caused major length variations. These modifications of the input sequence data may have caused the differences in the mpt from DO and MP analysis.

A major difference between the topologies of the mpts resulting from DO and MP analyses is that the mammalian clade is not monophyletic in the DO analysis (Figs 2, 3 and 6). The sequence alignment used for the MP analysis contains an indel not present in the DO alignment and critical to the monophyly of mammals. After conducting MP analysis coding all gaps as missing data, the mammalian clade is disrupted, as is the saurian clade (not shown), which is a similar topology to the DO results (Fig. 6). If Mammalia is constrained to a monophyletic group during MP analysis coding all gaps as missing data, the saurian clade is restored with Lepidosauria and Testudines as sister taxa (Figs 2 and 3). Though in the MP analysis, regardless of treatment of gaps, Lepidosauria and Testudines are sister taxa. If there was a monophyletic mammalian clade resulting from DO, which seems reasonable given the numerous studies failing to reject the monophyly of Mammalia (Meyer and Zardoya 2003 and references within), it seems likely there could also be a monophyletic saurian group.

### Conclusions

Neither ML nor BI of full POMC cDNA sequence resulted in resolution of the evolutionary relationships of Archosauria, Lepidosauria, and Testudines (Fig. 7). The  $P_{7C}$  (and  $P_4$  and  $P_{4c}$ ) partitioned BI analysis failed to falsify the hypothesis that Testudines is the sister taxon of Sauria, but this partitioning

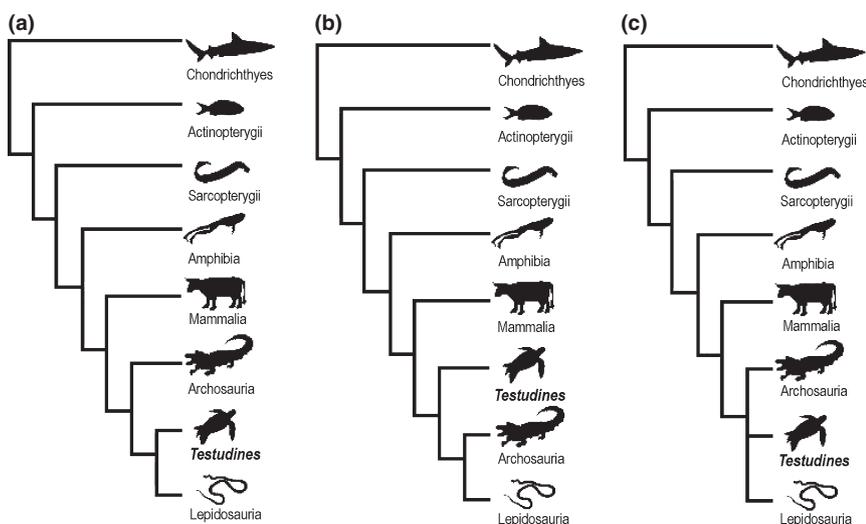


Fig. 7. The summary figure showing which analysis supports which turtle hypothesis. (a) is supported by MP analyses; (b) is supported by Bayesian analysis of partitioning schemes  $P_4$ ,  $P_{4C}$ , and  $P_{7C}$ ; (c) is supported by Bayesian analysis of partitioning schemes  $P_1$ ,  $P_C$ , and  $P_7$  and maximum likelihood (ML) analysis. (a) is the only hypothesis not rejected by the Bayesian, MP or ML analyses

scheme and its selection using Bayes factors appears to be flawed. The  $P_{7C}$  partitioned Bayesian analysis also could not statistically reject the hypothesis that Archosauria is the sister taxon of a Lepidosauria–Testudines clade, even though the inference with highest posterior probabilities showed (Testudines (Lepidosauria, Archosauria)). All the BI analyses did statistically reject the hypothesis that Testudines is the sister taxon of a Mammalia–Sauria clade. The MP analyses consistently inferred (Archosauria (Lepidosauria, Testudines)). Bootstrap, jackknife, and decay index statistics for the most parsimonious hypothesis of (Testudines, Lepidosauria) showed that relatively few characters diagnose the sister relationship, which will break down with one additional evolutionary step. The taxon jackknife analysis reported inconsistencies in the data from *A. piscivorus*, *T. scripta*, and *G. gallus* suggesting more OTUs and complete cDNA sequences need to be used. Nevertheless, with none of the BI, MP or ML analyses rejecting Archosauria as the sister taxon of a Testudines–Lepidosauria clade, it is probably the most robust hypothesis of the evolutionary relationships for taxa included in the analysis of POMC cDNA sequences. Our hypothesis corroborates previous work that also failed to falsify the hypothesis (Testudines, Lepidosauria) (Hedges 1994; Hill 2005).

### Acknowledgments

K. Piller kindly commented on the ms. and provided laboratory advice. Wulf Schleip and Wolfgang Wüster generously translated the Zusammenfassung. This work was partially funded by LEQSF (2006-09)-RD-A-26 to Roldán Valverde.

### Zusammenfassung

*Proopiomelanocortin (POMC) und die Untersuchung der Phylogenetischen Position der Schildkröten (Testudines)*

Morphologische und molekulargenetische Untersuchungen haben zu verschiedenen Hypothesen über die Verwandtschaftsverhältnisse der Testudines geführt. Die Hypothese, dass Testudines die einzigen noch lebenden anapsiden Amnioten und das Schwestertaxon der diapsiden Amnioten sind, wird durch morphologische Untersuchungen bekräftigt, während neuere morphologische und molekulargenetische Studien eher die Hypothese stützen, dass Testudines diapside Amnioten sind. In dieser Studie untersuchen wir die Stellung der Testudines mittels vollständiger cDNA-Sequenzen des polypeptiden Prohormons Proopiomelanocortin (POMC). Da nur lebende Taxa verwendet werden, können nur folgende Hypothesen getestet werden: 1) Testudines ist das Schwestertaxon der Archosauria, 2) Testudines sind in Archosauria enthalten und somit das Schwestertaxon der Crocodylia, 3) Testudines sind das Schwestertaxon der Lepidosauria, 4) Testudines sind das Schwestertaxon der Sauria und 5) Testudines sind das Schwestertaxon einer monophyletischen Mammalia–Sauria Gruppe. Weder Maximum-Likelihood, Bayesian oder Maximale Parsimony Analysen waren in der Lage, die Hypothese (Archosauria (Lepidosauria, Testudines)) zu falsifizieren, sodass diese die bevorzugte Folgerung aus den POMC Daten darstellt.

### References

deBraga M, Rieppel O (1997) Reptile phylogeny and the interrelationships of turtles. *Zool J Linn Soc* **120**:281–354.  
 Brandley MC, Schmitz A, Reeder TW (2005) Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Syst Biol* **54**:373–390.  
 Bremer K (1988) The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* **42**:795–803.

Burnham KP, Anderson DR (1998) Model selection and Multimodel Inference: A Practical Information-theoretic Approach. Springer-Verlag, New York.  
 Cao Y, Sorenson MD, Kumazawa Y, Mindell DP, Hasegawa M (2000) Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. *Gene* **259**:139–148.  
 Castoe TA, Parkinson CL (2006) Bayesian mixed models and the phylogeny of pitvipers (Viperidae: Serpentes). *Mol Phylogenet Evol* **39**:91–110.  
 Castoe TA, Doan TM, Parkinson CL (2004) Data partitions and complex models in Bayesian analysis: the phylogeny of gymnophthalmid lizards. *Syst Biol* **53**:448–469.  
 Cunningham CW, Zuh H, Hillis DM (1998) Best-fit maximum likelihood models for phylogenetic inference: empirical tests with known phylogenies. *Evolution* **52**:978–987.  
 Danielson PB, Dores RM (1999) Molecular evolution of the opioid/orphanin gene family. *Gen Comp Endocrinol* **113**:169–186.  
 Dores RM (1990) The proopiomelanocortin family. *Prog Clin Biol Res* **342**:22–27.  
 Dores RM, Rubin DA, Quinn TW (1996) Is it possible to construct phylogenetic trees using polypeptide hormone sequences? *Gen Comp Endocrinol* **103**:1–12.  
 Dores RM, Sollars C, Danielson P, Lee J, Alrubanjan J, Joss JMP (1999) Cloning of a POMC cDNA from the Pituitary of the Australian Lungfish, *Neoceratodus forsteri*: Analyzing Trends in the Organization of the Prohormone Precursor, Proopiomelanocortin. *Gen Comp Endocrinol* **116**:433–444.  
 Eernisse DJ, Kluge AG (1993) Taxonomic congruence versus total evidence, and Amniote phylogeny inferred from fossils, molecules, and morphology. *Mol Biol Evol* **10**:1170–1195.  
 Gaffney ES (1980) Phylogenetic relationships of the major groups of amniotes. In: Panchen AL (ed.), *The Terrestrial Environment and The Origin of Land Vertebrates*. Academic Press, London, pp 593–610.  
 Gauthier J, Kluge AG, Rowe T (1988a) The early evolution of the amniota. In: Benton MJ (ed.), *The Phylogeny and Classification of the Tetrapods*. Clarendon Press, Oxford, pp 103–153.  
 Gauthier J, Kluge AG, Rowe T (1988b) Amniote phylogeny and the importance of fossils. *Cladistics* **4**:105–209.  
 Giribet G (2001) Exploring the behavior of POY, a program for direct optimization of molecular data. *Cladistics* **17**:S60–S70.  
 Giribet G, Wheeler WC (1999) On gaps. *Mol Phylogenet Evol* **13**:132–143.  
 Grant T, Kluge AG (2004) Transformation series as an ideographic character concept. *Cladistics* **20**:23–31.  
 Gregory WK (1946) Pareiasaurs versus placodonts as near ancestors to turtles. *Bull Am Mus Nat Hist* **86**:275–326.  
 Hedges SB (1994) Molecular evidence for the origin of birds. *Proc Natl Acad Sci USA* **91**:2621–2624.  
 Hedges SB, Poling LL (1999) A molecular phylogeny of reptiles. *Science* **283**:998–1001.  
 Hill RV (2005) Integration of morphological data sets for phylogenetic analysis of Amniota: the importance of integumentary characters and increased taxonomic sampling. *Syst Biol* **54**:530–547.  
 Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* **42**:182–192.  
 Hillis DM, Pollack DD, McGuire JA, Zwickl DJ (2003) Is sparse taxon sampling a problem for phylogenetic inference? *Syst Biol* **52**:124–126.  
 Huelsenbeck JP, Rannala B (2004) Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Syst Biol* **53**:904–913.  
 Iwabe N, Hara Y, Kumazawa Y, Shibamoto K, Saito Y, Miyata T, Katoh K (2005) Sister group relationship of turtles to the Bird-Crocodylian clade revealed by nuclear DNA-coded proteins. *Mol Biol Evol* **22**:810–813.  
 Kass RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc* **90**:773–795.  
 Kluge AG (2007) Completing the neo-Darwinian synthesis with an event criterion. *Cladistics* **23**:613–633.  
 Kumazawa Y, Nishida M (1999) Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical

- evidence for archosaurian affinity of turtles. *Mol Biol Evol* **16**:784–792.
- Lanyon SM (1985) Detecting internal inconsistencies in distance data. *Syst Zool* **34**:397–403.
- Laurin M, Resiz RR (1995) A reevaluation of early amniote phylogeny. *Zool J Linn Soc* **113**:165–223.
- Lee MSY (1993) The origin of the turtle body plan: bridging a famous morphological gap. *Science* **261**:1716–1720.
- Lee MSY (1997) Pareiasaur phylogeny and the origin of turtles. *Zool J Linn Soc* **120**:197–280.
- Lemmon AR, Moriarty EC (2004) The importance of proper model assumption in Bayesian phylogenetics. *Syst Biol* **53**:265–277.
- Maddison WP, Maddison DR (2003) *MacClade4* v. 4.06. Sinauer Associates, Sunderland.
- Mannen H, Li SS-L (1999) Molecular evidence for a clade of turtles. *Mol Phylogenet Evol* **13**:144–148.
- Meyer A, Zardoya R (2003) Recent advances in the (molecular) phylogeny of vertebrates. *Ann Rev Ecol Evol Syst* **34**:311–338.
- Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL (2004) Bayesian phylogenetic analysis of combined data. *Syst Biol* **53**:47–67.
- Ogden TH, Rosenberg MS (2007) How should gaps be treated in parsimony? A comparison of approaches using simulation. *Mol Phylogenet Evol* **42**:817–826.
- Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst Biol* **53**:793–808.
- Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- Posada D, Crandall KA (2001) Selecting the best-fit model of nucleotide substitution. *Syst Biol* **50**:580–601.
- Reisz RR (1997) The origin and early evolutionary history of amniotes. *Trends Ecol Evol* **12**:218–222.
- Reisz RR, Laurin M (1991) *Owenetta* and the origin of turtles. *Nature* **349**:324–326.
- Rest JS, Ast JC, Austin CC, Waddell PJ, Tibbetts EA, Hay JM, Mindell DP (2003) Molecular systematic of primary reptilian lineages and the tuatara mitochondrial genome. *Mol Phylogenet Evol* **29**:289–297.
- Rieppel O, deBraga M (1996) Turtles as diapsid reptiles. *Nature* **384**:453–455.
- Rieppel O, Reisz RR (1999) The origin and early evolution of turtles. *Annu Rev Ecol Syst* **30**:1–22.
- Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574.
- Shoureshi P, Baron A, Szynskie L, Dores RM (2007) Analyzing the evolution of beta-endorphin post-translational processing events: studies on reptiles. *Gen Comp Endocrinol* **153**:148–154.
- Siddall ME (1995) Another monophyly index: revisiting the jackknife. *Cladistics* **11**:33–56.
- Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. *Syst Biol* **49**:369–381.
- Simmons MP, Ochoterena H, Carr TG (2001) Incorporation, relative homoplasy, and effect of gap characters in sequence-based phylogenetic analyses. *Syst Biol* **50**:454–462.
- Sorenson MD, Franzosa E (2007) *TreeRot*, Version 3. Boston University, Boston.
- Swofford DL (2002) *PAUP\**. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4. Sinauer Associates, Sunderland.
- Varón A, Vinh LS, Bomash I, Wheeler WC (2008) POY 4.0 Beta 2602. American Museum of Natural History, New York. <http://research.amnh.org/scicomp/projects/poy.php>.
- Weisrock DW, Shaffer HB, Storz BL, Storz SR, Voss SR (2006) Multiple nuclear gene sequences identify phylogenetic species boundaries in the rapidly radiating clade of Mexican ambystomatid salamanders. *Mol Ecol* **15**:2489–2503.
- Werneburg I, Sánchez-Villagra MR (2009) Timing of organogenesis support basal position of turtles in the amniote tree of life. *BMC Evol Biol* **9**:82.
- Wheeler WC (1996) Optimization alignment: the end of multiple sequence alignment in phylogenetics? *Cladistics* **12**:1–9.
- Wiens JJ, Servedio MR (1998) Phylogenetic analysis and intraspecific variation: performance of parsimony, likelihood, and distance methods. *Syst Biol* **47**:228–253.
- Yang Z (1996) Maximum likelihood models for combined analyses of multiple sequence data. *J Mol Evol* **42**:587–596.
- Zardoya R, Meyer A (1998) Complete mitochondrial genome suggests diapsid affinities of turtles. *Proc Natl Acad Sci USA* **95**:14226–14231.
- Zwickl DJ (2006) Genetic Algorithm Approaches for The Phylogenetic Analysis of Large Biological Sequence Datasets Under The Maximum Likelihood Criterion. Ph.D. dissertation, The University of Texas, Austin.

### Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1.** POMC cDNA and Amino Acid Sequences from GenBank.

**Table S2.** Partitioning schemes used for Bayesian analysis and the model of nucleotide evolution selected for each partition.

**Table S3.** Percentage of trees supporting alternative hypotheses for the placement of Testudines among amniotes produced by the Bayesian analysis of each partitioning scheme.

**Table S4.** The harmonic mean of the Ln likelihood of the 45,000 trees produced and sampled for each Bayesian partitioning scheme of POMC cDNA, and the Ln Bayes factors calculated by Tracer v1.4 for each partitioning scheme of POMC cDNA.

**Table S5.** 2Ln Bayes factors for calculated for each Bayesian partitioning scheme of POMC cDNA.

**Table S6.** The number of characters, variable characters, and parsimony informative characters present in each partition of the six partitioning schemes used during Bayesian analysis of the Proopiomelanocortin (POMC) gene.

**Table S7.** Interpretations of 2Ln Bayes factors as modified from Kass and Raftery (1995) by Nylander et al. (2004).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.