Copeia 2009, No. 2, 209–213

The Phylogenetic Position of the Snapping Turtles (Chelydridae) Based on Nucleotide Sequence Data

Christopher H. Chandler[1] and Fredric J. Janzen[1]

We augmented existing genetic datasets with nucleotide sequences from Recombination Activase Gene 1 (*RAG-1*) and an intron in the RNA fingerprint protein 35 (*R35*) for *Macrochelys temminckii* (Chelydridae), and from the nuclear gene Glyceraldehyde-3-phosphate Dehydrogenase (*GAPDH*) for *Macrochelys temminckii*, *Apalone spinifera* (Trionychidae), *Trachemys scripta* (Emydidae), *Sternotherus odoratus* (Kinosternidae), *Staurotypus triporcatus* (Kinosternidae), *Chelydra serpentina* (Chelydridae), *Dermochelys coriacea* (Dermochelyidae), *Platysternon megacephalum* (Platysternidae), and *Chelonia mydas* (Cheloniidae) to address the phylogenetic relationships of the Chelydridae. Our study finds support for a sister group relationship between Chelydridae and either the clades Kinosternoidea (Kinosternidae + Dermatemydidae) or Chelonioidea (Cheloniidae + Dermochelyidae). Further analysis also suggests that the speciation events leading to these clades occurred in rapid succession, within approximately one million years of each other.

THE phylogenetic relationship of snapping turtles (Chelydridae) to other chelonians remains ambiguous despite recent work (Krenz et al., 2005; Parham et al., 2006; Iverson et al., 2007). Early studies, especially using morphological data, first suggested a close affinity with the Big-Headed Turtle (*Platysternon megacephalum*; Platysternidae; e.g., Shaffer et al., 1997), but this hypothesis has largely been refuted by genetic data (Krenz et al., 2005; Parham et al., 2006; Shaffer et al., 2008), suggesting that the similarity between these two groups is the result of convergence. Current research indicates that the closest relative of the Chelydridae may be the Chelonioidea (sea turtles; Cheloniidae + Dermochelyidae; Fig. 1A) or Kinosternoidea (mud and musk turtles and the Mesoamerican River Turtle; Kinosternidae + Dermatemydidae; Krenz et al., 2005; Fig. 1B), or that the Chelydridae are sister to Chelonioidea + Testudinoidea (Emydidae + Platysternidae + Geoemydidae + Testudinidae; Parham et al., 2006; Fig. 1C).

Difficulties in resolving the phylogenetic placement of groups such as the Chelydridae with molecular data may sometimes be caused by long branch attraction or repulsion (Felsenstein, 1978; Siddall, 1998). However, the addition of new taxa to datasets may help by splitting long branches into shorter segments. Therefore, to evaluate these competing hypotheses and clarify the phylogenetic position of the Chelydridae, we sequenced approximately 4000 bp of the nuclear genes *RAG-1* and *R35*, which have proved useful in resolving family-level relationships, from another chelydrid (*Macrochelys temminckii*) to supplement existing datasets. We also sequenced approximately 500 bp of the *GAPDH* gene from exemplars of seven turtle families. Finally, we used a Power Analysis to estimate how much resolution our dataset should provide, to assess whether any apparent polytomies might be meaningful or merely due to insufficient data.

MATERIALS AND METHODS

Taxa and genes.—We restricted our analysis to eight focal species that have already been examined in previous phylogenetic analyses using gene sequence data: *Apalone spinifera* (Trionychidae), *Trachemys scripta* (Emydidae), *Platysternon megacephalum* (Platysternidae), *Sternotherus odoratus* (Kinosternidae), *Staurotypus triporcatus* (Kinosternidae), *Chelydra serpentina* (Chelydridae), *Dermochelys coriacea* (Dermochelyidae), and *Chelonia mydas* (Cheloniidae). We also obtained sequences from the Alligator Snapping Turtle (*Macrochelys temminckii*; Chelydridae), which has not been included in previous molecular phylogenetic studies of turtles.

Sequences from *Apalone spinifera*, *Trachemys scripta, Platysternon megacephalum*, *Sternotherus odoratus*, *Staurotypus triporcatus*, *Chelydra serpentina*, *Dermochelys coriacea*, and *Chelonia mydas* were obtained from previous studies, including portions of Recombination Activase Gene 1 (*RAG-1*) and the RNA fingerprint protein 35 (*R35*) intron, which are both nuclear (Fujita et al., 2004; Krenz et al., 2005). We collected nucleotide sequences from *RAG-1* and the *R35* intron for *M. temminckii* to augment the previously published dataset. We also sequenced approximately 500 bp of intron XI of the nuclear gene Glyceraldehyde-3-phosphate Dehydrogenase (*GAPDH*) for all nine study species, using the same individuals as Krenz et al. (2005).

DNA sequencing.—We extracted total genomic DNA from a *Macrochelys temminckii* tissue sample from Florida, as previously described (Krenz et al., 2005). Using this template, we amplified approximately 3000 bp of Recombination Activase Gene 1 (*RAG-1*), in three separate fragments using the primer pairs RAGF1/RAGR2, RAGF3/RAGR4, and RAGF5, RAGR5 (Krenz et al., 2005), in 25 μL volumes with

[1] Department of Ecology, Evolution, and Organismal Biology, Iowa State University, 253 Bessey Hall, Ames, Iowa 50011; E-mail: (CHC) cholden@iastate.edu; and (FJJ) fjanzen@iastate.edu. Send reprint requests to CHC.
Submitted: 17 September 2007. Accepted: 17 October 2008. Associate Editor: D. Kizirian.
 DOI: 10.1643/CH-07-206

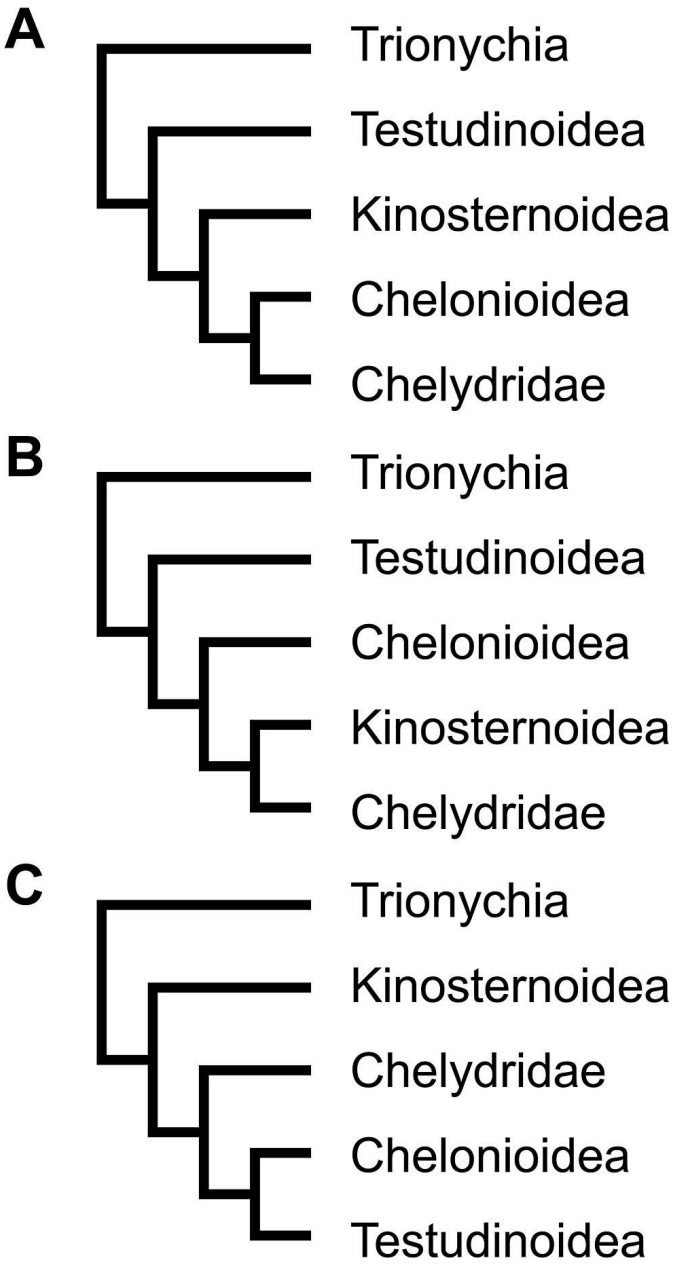

Fig. 1. Recent hypotheses for the phylogenetic position of the snapping turtles (Chelydridae) supported by molecular data from Krenz et al. (2005) and Parham et al. (2006). Trionychia is the clade that includes Trionychidae + Carettochelyidae (Joyce et al., 2004).

10–70 ng of purified genomic DNA, 1X PCR buffer (16 mM $(NH_4)_2SO_4$, 67 mM Tris-Cl, 0.01% Tween-20; Bioline, Taunton, MA), 3.0 mM $MgCl_2$ (Bioline), 1.0 mM dNTPs (Promega, Madison, WI), 0.5 µM forward and reverse primer, and 0.6 u *Taq* Biolase DNA polymerase (Bioline). Thermocycling consisted of an initial denaturation at 94°C for 5 min, then 35 cycles of denaturation at 94°C for 30 sec, annealing at 55°C for 1 min, and extension at 72°C for 90 sec, followed by a final 5-min 72°C extension phase using an Eppendorf Mastercycler Gradient. We amplified approximately 1000 bp of the *R35* intron with primer pair R35Ex1/R35Ex2 (Fujita et al., 2004) and approximately 500 bp of the *GAPDH* gene with primer pair GapdL890/GAPDH950 (Friesen et al., 1997; Spinks and Shaffer, 2005). These amplifications used the same reaction conditions as the *RAG-1* reactions, but with 60°C and 63°C annealing temperatures for their respective primers.

PCR products were purified by electrophoresis in 1.5% TBE agarose gels in the presence of ethidium bromide and DNA fragments of known lengths. Bands of expected lengths were excised from gels and further purified using a Qiagen Gel Extraction Kit. For samples that failed to sequence following initial attempts with gel extracts, we purified PCR products directly using ExoSAP-IT (USB Corporation, Cleveland, OH). We used gel extracts or purified PCR products as template for cycle sequencing reactions using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Mix (PE Applied Biosystems, Foster City, CA). We ran reactions in 10 µL volumes using approximately 25 ng PCR product, 1.3 µL 5X Sequencing Buffer, 2.5 µM primer, and 1.3 µL Terminator Ready Reaction Mix. Cycle sequencing reactions were subjected to 45 cycles of 96°C for 30 seconds, 50°C for 30 seconds, and 60°C for 4 minutes. Cycle sequencing reaction products were purified using CentriSep (Sephadex) spin columns (Princeton Separations, Freehold, NJ). Purified cycle sequencing reactions were submitted to the Iowa State University DNA Sequencing and Synthesis Facility for electrophoresis using an Applied Biosystems 377 Automated Sequencer. Sequences generated in this study were deposited in the GenBank database under accession numbers FJ234440–FJ234450. DNA fragments were assembled into contiguous sequences using BioEdit v7.0.1 (Hall, 1999) and aligned using ClustalW v1.4 (Thompson et al., 1994) through the menu option implemented in the BioEdit interface (Hall, 1999) using the default parameters.

Phylogenetic analysis.—We assessed the suitability of *GAPDH* sequences for phylogenetic analysis using a saturation plot. We conducted Maximum Parsimony (MP) and Maximum Likelihood (ML) analyses using PAUP* v4.0b10 (Swofford, 2003) and Bayesian Analysis (BA) with MrBayes v3.0b4 (Huelsenbeck and Ronquist, 2001). All analyses were performed using rooted trees, with *A. spinifera* as the outgroup. For ML and BA analyses, we selected models of molecular evolution using Modeltest v3.6 (Posada and Crandall, 1998) and MrModelTest v2.2 (Nylander, 2004).

Heuristic searches were performed in MP and ML analyses with ten replicates of random taxon addition, tree bisection-reconnection (TBR) branch swapping, zero-length branches collapsed to yield polytomies, and gaps coded as missing data. In Bayesian analyses, four Markov chains were calculated simultaneously, initiated with random starting trees, and run for 2.0×10^6 generations, with chains sampled every 1000 generations. Stationarity was confirmed graphically, and the first 200 of 2000 trees were discarded (burnin = 200). We estimated statistical support for nodes using 10000 bootstrap replicates for MP and 100 bootstrap replicates for ML analyses, while support for nodes in Bayesian analyses was assessed using posterior probabilities.

Finally, we performed a Power Analysis using the difference of a proportion method of Walsh et al. (1999) to estimate the minimum length of time between successive speciation events that a dataset of this size can resolve. We estimated the average substitution rate by computing the proportion of sites exhibiting nucleotide substitutions between pairs of taxa, dividing by twice the estimated divergence time between the two taxa, and averaging over

all pair-wise comparisons; divergence time estimates were obtained from Near et al. (2005). Then, beginning with an arbitrary desired resolvable divergence time, from the average substitution rate we calculated the proportion of sites expected to undergo nucleotide substitutions under the desired resolvable divergence time. We used this proportion to estimate the effect size index, h, using the formula given in Walsh et al. (1999). Finally, we used the effect size index to calculate the sample size (N, number of base pairs) required to achieve that resolution, with the formula $N = N_{.10}/(100h^2)$, where $N_{.10} = 1237$, the sample size required for an effect size index of $h = 0.10$, with $\alpha = 0.05$ and $\beta = 0.80$ (Walsh et al., 1999). By adjusting the desired resolvable divergence times and repeating this process, we were able to estimate numerically the minimum divergence time resolvable by our dataset.

RESULTS

Saturation plots, showing the pair-wise number of transitions or transversions versus the patristic distance between taxa, demonstrated that *GAPDH* is suitable for phylogenetic analysis at this taxonomic level, since it was not saturated for transitions or transversions (data not shown). Partition homogeneity tests revealed that all loci were combinable, but analyses of *GAPDH* by itself, however, yielded a slightly different tree topology in some cases, whereas *RAG*-1 and *R35* were congruent. Therefore, we present results from analyses of *GAPDH* alone, *RAG-1*+*R35*, and all three genes combined.

Our final alignment consisted of 2800 bp of *RAG-1*, 1092 bp of *R35*, and 881 bp of *GAPDH*, with 683 variable characters, of which 231 were parsimony-informative. Of the 881 bp of *GAPDH*, 168 were variable and 68 were parsimony-informative; these statistics have already been described for *RAG-1* and *R35* (Fujita et al., 2004; Krenz et al., 2005).

For ML analysis, ModelTest selected the TIM+I+G model using the Akaike information criterion (AIC) for the combined analysis, the HKY+G for the *GAPDH* dataset, and the TrN+G model for the *R35*+*RAG-1* dataset (Posada and Crandall, 1998). For Bayesian Analysis, MrModelTest selected the GTR+G model for the combined dataset, the HKY+G model for the *GAPDH* dataset, and the GTR+G model for the *R35*+*RAG-1* dataset, again using AIC (Nylander, 2004). In the few cases where hierarchical likelihood ratio tests and AIC selected different optimal models in ModelTest and MrModelTest, we present only the model selected by AIC, but both models always produced identical tree topology and similar bootstrap support.

All analyses supported a monophyletic Chelydridae. Analysis of *GAPDH* alone grouped the Chelonioidea with the Emydidae and Platysternidae with weak support, with this clade connected to the Kinosternoidea and Chelydridae in a three-way polytomy (Fig. 2). Analysis of *R35* and *RAG-1* suggested a sister relationship between the Kinosternoidea and Chelydridae, with weak (<70% bootstrap) support in MP analyses and moderate support in ML (about 70% bootstrap) and Bayesian (about 0.85 posterior probability) analyses (Fig. 2). Maximum likelihood and Bayesian analyses of the combined dataset yielded the same topology with slightly stronger support, but the MP analysis was inconclusive, resulting in a large polytomy (Fig. 2).

The average nucleotide substitution rates were 0.0020, 0.0003, and 0.0005 substitutions per site per million years

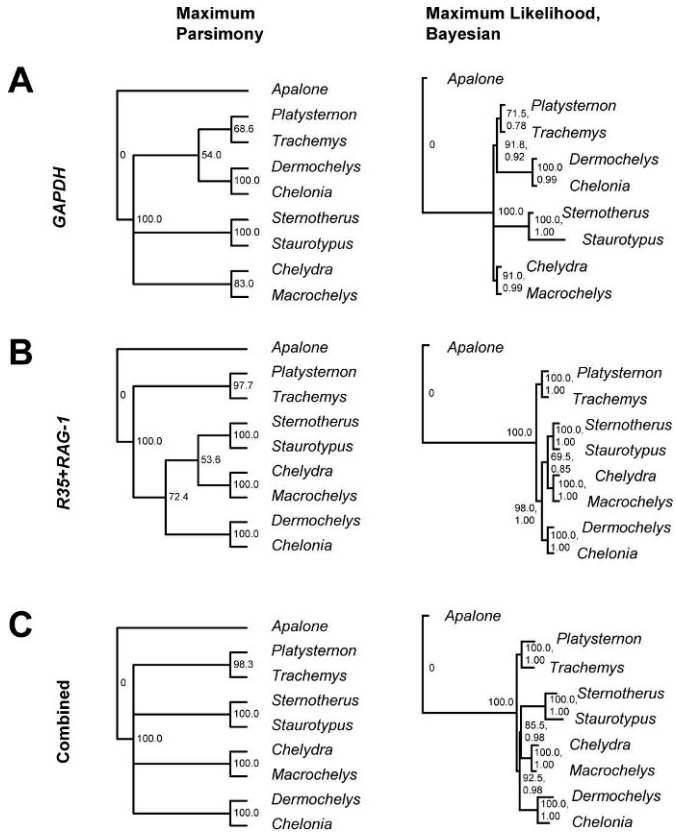

Fig. 2. Phylogenies produced by analysis of (A) 881 bp of *GAPDH*, (B) 2800 bp of *RAG-1* plus 1092 bp of *R35*, and (C) all three genes combined, using Maximum Parsimony (left) and Maximum Likelihood and Bayesian analyses (right). Numbers next to nodes indicate bootstrap support in the Maximum Parsimony trees and Maximum Likelihood trees and posterior probabilities in the Bayesian trees. Maximum Likelihood and Bayesian analyses always produced identical topologies, so their node support values are shown on the same trees. Branch lengths depicted on the right are derived from Maximum Likelihood analyses; phylograms are not all shown on the same scale.

for *GAPDH*, *R35*+*RAG-1*, and all three genes combined, respectively. Power Analysis using these estimates indicated that our *GAPDH*, *R35*+*RAG-1*, and combined datasets could resolve speciation events separated by at least 0.89 million years, 1.45 million years, and 0.64 million years, respectively (given $\alpha = 0.05$ and $\beta = 0.80$).

DISCUSSION

In this study, we supplemented existing datasets with both additional taxa and additional loci to explore the phylogenetic position of chelydrid turtles. We also restricted our analysis to the taxa most relevant to the problem at hand. Despite these new data, we were unable to reject with substantial certainty any of the hypotheses regarding the phylogenetic relationship between the Chelydridae and other turtles (Fig. 2), although the relationships suggested by Parham et al. (2006) were not supported strongly. None of our results implied a close relationship between the Chelonioidea and the Chelydridae, which is suggestive that these two groups are not sister taxa. However, support for key nodes was relatively weak, so additional data are needed before that hypothesis can be ruled out completely.

Rapid divergence events may explain the ambiguous relationships among these groups. Ancient rapid radiations

such as this are especially difficult to resolve, because the large amount of subsequent change that has occurred may obscure an already weak phylogenetic signal (Whitfield and Lockhart, 2007). Additionally, incomplete sorting of ancestral polymorphisms is a problem in cases of rapid speciation. Therefore, sequencing many additional loci may be of little use in solving this problem, because incomplete sorting could cause different unlinked loci to represent different evolutionary histories—that is, only a fraction of the gene trees will reflect the species tree (Whitfield and Lockhart, 2007).

Power Analysis showed that the size of our dataset should provide resolution of at least 0.64 to 1.45 million years. Although the approach we used to obtain this estimate is relatively simplistic compared to other available methods, it tends to be more conservative. For example, taking into account homoplasy by estimating pair-wise genetic distances using a Jukes-Cantor correction or Maximum Likelihood, rather than using the raw proportion of sites with substitutions, would suggest that our dataset could resolve even more rapid successive speciation events (Walsh et al., 1999). Therefore, our inability to completely reject alternative phylogenetic hypotheses is unlikely to result from inadequate data. Considering the large degree of morphological divergence among these groups of turtles, and that the splits between other turtle families are separated by millions of years (Near et al., 2005), the divergence of the Chelonioidea, Kinosternoidea, Chelydridae, and Emydidae + Platysternidae may be explained by near-simultaneous speciation events.

Our results show that support for phylogenetic hypotheses may not always be strengthened by including additional relevant taxa while excluding those outside the unresolved nodes, or by the addition of new loci. While we did not find support for a sister relationship between the Chelydridae and Chelonioidea or Testudinoidea, we were unable to resolve the phylogenetic position of the Chelydridae with great confidence. Our Power Analysis suggests that these higher-level clades of turtles diverged from one another relatively rapidly, within 0.64–1.45 million years, and that this diversification process may be a ''hard'' polytomy that is unlikely to be resolved even by large genetic datasets.

ACKNOWLEDGMENTS

We thank J. Krenz for help with the molecular work, sharing the original dataset, and for assistance with analysis, and S. McGaugh and E. Myers for additional help with molecular work and phylogenetic analysis. The *M. temminckii* sample was provided by P. Moler to D. Weisrock and J. Krenz. The Janzen lab group and H. Shaffer provided helpful constructive criticism on earlier drafts of the manuscript. This research was funded by a Professional Advancement Grant from the Iowa State University Graduate and Professional Student Senate and by NSF DEB-0089680 and DEB-0640932. CHC was supported by an Iowa State University Graduate College Fellowship, an Iowa State University Office of Biotechnology Fellowship, and an assistantship funded by the Iowa State University Center for Integrated Animal Genomics.

LITERATURE CITED

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Systematic Zoology 27:401–410.

Friesen, V. L., B. C. Congdon, H. E. Walsh, and T. P. Birt. 1997. Intron variation in marbled murrelets detected using analyses of single stranded conformational polymorphisms. Molecular Ecology 6:1047–1058.

Fujita, M. K., T. N. Engstrom, D. E. Starkey, and H. B. Shaffer. 2004. Turtle phylogeny: insights from a novel nuclear intron. Molecular Phylogenetics and Evolution 31:1031–1040.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41:95–98.

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

Iverson, J. B., R. M. Brown, T. S. Akre, T. J. Near, M. Le, R. C. Thomson, and D. E. Starkey. 2007. In search of the tree of life for turtles. Chelonian Research Monographs 4:85–106.

Joyce, W. G., J. F. Parham, and J. A. Gauthier. 2004. Developing a protocol for the conversion of rank-based taxon names to phylogenetically defined clade names, as exemplified by turtles. Journal of Paleontology 78: 989–1013.

Krenz, J. G., G. J. P. Naylor, H. B. Shaffer, and F. J. Janzen. 2005. Molecular phylogenetics and evolution of turtles. Molecular Phylogenetics and Evolution 37:178–191.

Near, T. J., P. A. Meylan, and H. B. Shaffer. 2005. Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. The American Naturalist 165:137–146.

Nylander, J. A. A. 2004. MrModelTest. Evolutionary Biology Centre, Uppsala University. http://www.abc.se/~nylander

Parham, J., C. R. Feldman, and J. L. Boore. 2006. The complete mitochondrial genome of the enigmatic bigheaded turtle (*Platysternon*): description of unusual genomic features and the reconciliation of phylogenetic hypotheses based on mitochondrial and nuclear DNA. BMC Evolutionary Biology 6:11.

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

Shaffer, H. B., P. Meylan, and M. L. McKnight. 1997. Tests of turtle phylogeny: molecular, morphological, and paleontological approaches. Systematic Biology 46:235–268.

Shaffer, H. B., D. E. Starkey, and M. K. Fukita. 2008. Molecular insights into the systematics of the Snapping Turtles (Chelydridae), p. 44–49. *In*: Biology of the Snapping Turtle (*Chelydra serpentina*). A. C. Steyermark, M. S. Finkler, and R. J. Brooks (eds.). Johns Hopkins University Press, Baltimore, Maryland.

Siddall, M. E. 1998. Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris Zone. Cladistics 14:209–220.

Spinks, P. Q., and H. B. Shaffer. 2005. Range-wide molecular analysis of the western pond turtle (*Emys marmorata*): cryptic variation, isolation by distance, and their conservation implications. Molecular Ecology 14:2047–2064.

Swofford, D. L. 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachussetts. http://paup.csit.fsu.edu/

Thompson, J. D., D. G. Higgins, and T. J. Gibbons. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research 22:4673–4680.

Walsh, H. E., M. G. Kidd, T. Moum, and V. L. Friesen. 1999. Polytomies and the power of phylogenetic inference. Evolution 53:932–937.

Whitfield, J. B., and P. J. Lockhart. 2007. Deciphering ancient rapid radiations. Trends in Ecology and Evolution 22:258–265.