Assessing Concordance of Fossil Calibration Points in Molecular Clock Studies: An Example Using Turtles

Thomas J. Near,[1,*] Peter A. Meylan,[2,†] and H. Bradley Shaffer[3,‡]

1. Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996-1610;
2. Natural Sciences, Eckerd College, St. Petersburg, Florida 33711;
3. Section of Evolution and Ecology and Center for Population Biology, University of California, Davis, California 95616

Submitted July 2, 2004; Accepted November 8, 2004;
Electronically published December 29, 2004

Online enhancement: appendix.

abstract: Although still controversial, estimation of divergence times using molecular data has emerged as a powerful tool to examine the tempo and mode of evolutionary change. Two primary obstacles in improving the accuracy of molecular dating are heterogeneity in DNA substitution rates and accuracy of the fossil record as calibration points. Recent methodological advances have provided powerful methods that estimate relative divergence times in the face of heterogeneity of nucleotide substitution rates among lineages. However, relatively little attention has focused on the accuracy of fossil calibration points that allow one to translate relative divergence times into absolute time. We present a new cross-validation method that identifies inconsistent fossils when multiple fossil calibrations are available for a clade and apply our method to a molecular phylogeny of living turtles with fossil calibration times for 17 of the 22 internal nodes in the tree. Our cross-validation procedure identified seven inconsistent fossils. Using the consistent fossils as calibration points, we found that despite their overall antiquity as a lineage, the most species-rich clades of turtles diversified well within the Cenozoic. Many of the truly ancient lineages of turtles are currently represented by a few, often endangered species that deserve high priority as conservation targets.

Keywords: penalized likelihood, divergence time, Testudines, cross-validation, molecular clock, fossil record.

* E-mail: tnear@utk.edu.

† E-mail: meylanpa@eckerd.edu.

‡ E-mail: hbshaffer@ucdavis.edu.

Am. Nat. 2005. Vol. 165, pp. 137–146. 0003-0147/2005/16502-40514$15.00.

The molecular clock hypothesis proposes that genes and gene products evolve as a random Poisson process, and the degree of genetic divergence between any two species is indicative of the time since common ancestry (Zuckerkandl and Pauling 1965). The molecular clock hypothesis has been the primary basis for methods of molecular dating that attempt to relate the degree of genetic divergence among species or lineages to their time of evolutionary divergence or the age of their most recent common ancestor. Recent work on molecular dating has emphasized the considerable error that may be present in fossil calibration and in reconstructing patterns of DNA sequence evolution (Graur and Martin 2004), and strategies have been proposed that minimize the impact of these errors. It is now clear that heterogeneity in DNA substitution rates among lineages is virtually ubiquitous (Britten 1986; Bromham and Penny 2003). In the past several years, many methods have been developed that take into account molecular rate heterogeneity among lineages, allowing molecular dating methods to be applied to these previously intractable groups (Sanderson 1998; Arbogast et al. 2002). These methods range from those that identify and remove nonclocklike subsets of the data, such as linearized tree and quartet methods (Takezaki et al. 1995; Cooper and Penny 1997), to methods that include all of the data and attempt to estimate local rates for branches and lineages in the phylogeny using nonparametric, semiparametric, and Bayesian methodologies (Sanderson 1997, 2002; Thorne et al. 1998).

Transforming relative into absolute divergence times requires an additional step of fossil calibration, allowing one to compare the rates of evolutionary processes across disparate clades and species (Sanderson 1998; Magallon and Sanderson 2001). Calibration is usually accomplished by reference to a fossil of known age that can be confidently assigned to a node in a molecular phylogeny; this then serves as a calibration point that is treated as a fixed, minimal age estimate for the lineage (Marshall 1990*b*; Sanderson 1998). However, the use of fossils for dating nodes is subject to error, and there are several reasons why a given fossil may incorrectly date a node, leading to in-

congruence between age estimates derived from fossils and molecular dating methods (Smith and Peterson 2002; Benton and Ayala 2003; Bromham and Penny 2003). The incompleteness of the fossil record will consistently lead to an underestimation of the age of any given lineage (Marshall 1990*b*), and the magnitude of this error bias will depend on the difference between the fossil age and the actual lineage age (Springer 1995). Severe but directionally random errors can also result when fossils are placed erroneously into a phylogenetic tree (Lee 1999; Benton and Ayala 2003), when the phylogeny itself is in error, or when the geological age estimates of fossil-bearing rocks are in error (Conroy and van Tuinen 2003). Finally, error will occur if minimal age estimates are applied to the crown group that a fossil subtends rather than the appropriate stem lineage in a phylogenetic tree (Doyle and Donoghue 1993; Magallon and Sanderson 2001).

Proposed solutions to account for error in the fossil record include methods to provide confidence intervals for lower bound age estimates based on the density of a particular taxon or lineage in the fossil record (Strauss and Sadler 1989; Marshall 1990*a*, 1990*b*) and strategies for identifying phylogenetically misplaced or otherwise "rogue" fossils based on the consistency between the branching sequence in phylogenetic trees (clade rank) and relative stratigraphic position of fossils. The rationale for these latter methods stems from the observation that if fossils are perfect estimates of lineage ages and they are assigned correctly to nodes in the phylogeny, then there should be a significant correlation between the order of branching events in the phylogenetic tree and their respective fossil ages (Norell and Novacek 1992; Huelsenbeck 1994; Benton 1995).

The potential errors introduced by erroneous fossil information has resulted in criticism of molecular dating studies that rely on a single fossil calibration point (Alroy 1999; Lee 1999; Conroy and van Tuinen 2003; Graur and Martin 2004), and several authors have recently advocated the use of multiple fossil calibration points (Smith and Peterson 2002; Soltis et al. 2002; Conroy and van Tuinen 2003). By using multiple calibration points, a variance in the age estimate for a given node in the phylogeny, using different fossil dates, can potentially be used to calculate confidence intervals around molecular age estimates (Smith and Peterson 2002). While this approach has merit, it may also be the case that some fossils are so inaccurate that one would be better off eliminating them rather than including them in a multifossil calibration of a molecular phylogeny. However, no general strategy has been developed that allows one to identify calibration points that form an internally consistent, and therefore reliable, age estimate across a tree versus those that are inconsistent and potentially erroneous for any of the reasons discussed above (Lee 1999). This is a particularly insidious problem since correctly classified fossils are virtually always underestimates of lineage ages. Thus, for the presumably most common sources of error, the distribution of that error will not be normal around a node. Instead, the more erroneous fossil dates will have the effect of biasing the analysis farther from the true date in the direction of artificially young divergence times (Bromham et al. 2000).

In this study, we present a new method that allows one to identify misleading fossils when multiple calibration points are available for a given clade. We then apply our new method by calibrating a multifossil molecular clock in turtles. Many features of turtles make them an attractive system to examine issues of fossil calibration in molecular divergence time estimation. Turtles exhibit a limited taxonomic diversity that facilitates the estimation of phylogenetic relationships for all major extant lineages, they have a rich fossil record dating to the Triassic, and most fossil taxa can be readily assigned to nodes in the turtle phylogeny using morphological synapomorphies (Gaffney and Meylan 1988; Shaffer et al. 1997). With the availability of multiple fossil calibration points, we introduce a cross-validation method that measures the consistency between fossil and molecular age estimates using each of the fossil calibration points separately. We identify fossils that are inconsistent with regard to the degree of deviation between fossil and molecular age estimates, and we assess the effect of removing the inconsistent fossils on the overall differences between molecular and fossil age estimates at all other fossil-calibrated nodes. Through these analyses, we demonstrate that our newly developed methods provide an objective and general approach to identify fossils that may lead to erroneous molecular age estimates. Furthermore, we show that for trees with a dense collection of potential calibration points, one may be better off using fewer but more accurate fossils than all of the available data.

Material and Methods

Molecular Data and Phylogeny Inference

A phylogeny of 23 species representing all of the major lineages of living turtles was inferred from a data set containing 892 base pairs (bp) of the mitochondrial encoded cytochrome *b* (cyt*b*) gene (Shaffer et al. 1997), 2,790 bp of the nuclear encoded protein coding recombinase activating gene 1 (RAG-1), and 1,009 bp of the nuclear encoded R35 intron (Engstrom et al. 2004; Fujita et al. 2004; Spinks et al. 2004). Species sampled and GenBank accession numbers can be found in the appendix in the online edition of the *American Naturalist*. Details regarding the phylogenetic performance of RAG-1 will be presented else-

Table 1: Turtle fossils and dates used as calibration points

Node	Date (mya)	Fossil taxon	Reference
1	210.0	*Proterochersis*	Gaffney 1986, 1990
2	110.0	*Sandownia harrisi*	Meylan et al. 2000
3	110.0	*Araripemys barretoi*	Meylan and Gaffney 1991
4	110.0	*Cearachelys placidoi*	Gaffney et al. 2001
5	110.0	*Santanachelys gaffneyi*	Hirayama 1998
6	100.0	*Aspideretes maortuensis*	Yeh 1965
7	65.0	*Hoplochelys*	Hutchison 1981
8	52.0	*Hadrianus majusculus*	Hutchison 1980
9	50.0	*"Ocadia" crassa*	Lapparent de Broin 2001
10	50.0	*Baltemys*	Hutchison 1991
11	71.0	*Yaminuechelus gasparinii*	de la Fuente et al. 2001
12	34.0	*Chrysemys antiqua*	Hutchison 1996
13	90.0	Lindholmemydidae	Sukanov 2000
14	18.0	*Pelusios rusingae*	Williams 1954
15	15.0	*Chelodina* and *Elseya*	Gaffney et al. 1989
16	11.6	*Chelus*	Wood 1976
17	5.0	*Trachemys inflata*	Jackson 1988

Note: mya = millions of years ago.

where (J. G. Krenz, G. J. P. Naylor, H. B. Shaffer, and F. J. Jansen, unpublished manuscript). We identified seven separate data partitions from the three codon positions from cyt*b* and RAG-1 and a single partition for the R35 intron. The optimal model of sequence evolution for each data partition was assessed with hierarchical likelihood ratio tests using the computer program Modeltest 3.0 (Posada and Crandall 1998). We conducted a phylogenetic analysis of the combined three gene data set using a partitioned mixed-model Bayesian analysis, with posterior probabilities estimated using Markov chain Monte Carlo (Larget and Simon 1999; Huelsenbeck et al. 2001). The optimal models of sequence evolution determined for each data partition were used in the computer program MrBayes 3.0 (Ronquist and Huelsenbeck 2003) with the APPLYTO command and appropriate model parameter values estimated for each data partition using UNLINK commands. We ran four chains simultaneously in each analysis and repeated the analysis four separate times. Each MrBayes search was run with 4×10^6 generations, providing convergence in estimations of the tree topology, branch lengths, and parameter values for the models of DNA sequence evolution and posterior probability of node support. Branch lengths for the estimation of divergence times were estimated as the mean branch lengths of all post-burn-in trees resulting from the partitioned mixed-model Bayesian analysis, and these were generated using the SUMT command in MrBayes. To confirm that our estimates of phylogenetic relationships were robust, we also ran a parsimony analysis on the full molecular data set in PAUP* (Swofford 2000) using a heuristic tree search with 100 random sequence addition replicates and assessing nodal support with traditional bootstrap scores based on 2,000 pseudoreplicates.

Turtle Fossils and Absolute Age Estimates

We were able to assign a minimal absolute age estimate to 17 of the 22 internal nodes in the turtle phylogeny (fig. A1 in the online edition of the *American Naturalist*). Placement of fossils on the inferred molecular phylogeny was guided by cladistic analysis of 115 morphological characters presented in earlier studies of turtle phylogenetic relationships (Shaffer et al. 1997). Minimal age estimates were assigned to stem groups, which are the most inclusive group of taxa that contains all extant and extinct members of the clade (table 1). Stem group age estimates were applied to the most recent common ancestor node between the crown group associated with the stem group and the extant sister group of that crown group (Doyle and Donoghue 1993; Magallon and Sanderson 2001; Near et al. 2003).

Tests of Rate Heterogeneity and Estimation of Divergence Times

We tested for nucleotide substitution rate heterogeneity among lineages for each of the three sampled gene regions and the concatenated data set using a likelihood ratio test comparing rate-variable and rate-constant models of sequence evolution (Huelsenbeck and Crandall 1997; Near et al. 2003). The likelihood ratio test statistic was compared to a χ^2 distribution with $N - 2$ degrees of freedom, where N equals the number of taxa included in the analysis (Huelsenbeck and Crandall 1997).

Subsequent to the discovery of significant rate heterogeneity, we estimated divergence times using the penalized likelihood method (Sanderson 2002), as implemented in the computer program r8s (Sanderson 2003). For each iteration of divergence time estimation, cross-validation determined the optimal smoothing parameter as outlined in Sanderson (2002) and implemented in r8s. The effect of data sampling on divergence time estimates was assessed with bootstrap resampling of the nucleotide data. One hundred bootstrap pseudoreplicates were generated using the seqboot program in the Phylip software package (Felsenstein 1993), replicates were imported into PAUP*, and branch lengths were calculated on the Bayesian tree using the optimal model of sequence evolution. Confidence intervals of divergence time estimates were determined by calculating the central 95% distribution of bootstrap-generated divergence time estimates at a given node using the PROFILE command in r8s (Sanderson and Doyle 2001).

Assessing Consistency of Divergence Time Estimates from Individual Fossil Absolute Age Estimates

We developed a novel cross-validation method to measure the agreement, or consistency, between any one fossil calibration point and other available fossil calibrations. Our goal was to identify the set of fossils that all yield relatively consistent calibrations for a tree, as well as those that yield discordant age calibrations that may result in inaccurate absolute age estimates.

Given a tree with multiple fossils in which each provides an independent calibration point, we first calculated, for each potential calibration fossil, the difference between the molecular and observed fossil age estimate for all other fossil-dated nodes on the tree. When the fossil age at node χ is used as a single calibration point, and fossil age estimates are available for n nodes in the phylogeny, then we define $D_i = (\mathrm{MA}_i - \mathrm{FA}_i)$, where FA_i is the fossil age estimate and MA_i is the molecular age estimate for node i using the fossil calibration at node χ. We then calculate $\overline{D}_\chi$ as

$$\overline{D}_\chi = \frac{\sum_{i \neq \chi} D_i}{n - 1},$$

the average D_i for all available nodes based on the fossil calibration at node χ. In a given iteration of the cross-validation analysis, the fossil age for a single node (χ) was used as the calibration point in penalized likelihood analysis, and the $\overline{D}_\chi$ and its standard error were calculated from the remaining 16 fossil-dated calibration point nodes. A plot of $\overline{D}_\chi$ (fig. 1, *top*) provides a visual assessment of the performance of each fossil, although the interdependence of each $\overline{D}_\chi$ with all other values (because each fossil and its associated error contribute to all other values of D) limits any statistical analyses of these values.

We used a three-step procedure to sequentially identify and remove inconsistent fossils from the analysis. First, for each fossil calibration, we calculated SS, which is simply the sum of the squared differences between the molecular (MA) and fossil (FA) age estimates at all other fossil-dated nodes (fig. 1, *bottom*):

$$\mathrm{SS}_\chi = \sum_{i \neq \chi} D_i^2.$$

We then ranked each calibration point based on the magnitude of its SS score and identified the fossil with the greatest SS score to be the most inconsistent with the other fossils in the analysis. Second, we calculated an average squared deviation of D_i values for all 17 fossil calibrations in the analysis (s):

$$s = \frac{\sum_{\chi=1}^{n} \sum_{i \neq \chi} D_i^2}{n(n-1)},$$

where n is equal to the total number of observations of D_i. To determine the overall impact of eliminating fossil calibration points from our analysis, we removed the fossil with the greatest SS and recalculated s based on the remaining 16 fossils. We continued this process until only the two fossil calibration points with the lowest and second lowest SS values remained. We expect s to incrementally decrease by a small constant amount as fossils are removed if all calibration points are approximately equally accurate. However, extreme outliers that provide very inaccurate calibrations with respect to other fossils should show an appreciable drop in s when removed from the analysis, and we visually assessed this effect in figure 2. Finally, in order to determine the effect of fossil calibration removal on the variance of s, we compared the variance of s across n nodes with the variance of s for all $n - 1$ nodes remaining after removal of the ith calibration point. The significance of change in variance before and after fossil calibration removal was determined using a one-tailed F-test based on $n - 1$ degrees of freedom. The expectation is that the removal of a very inaccurate fossil calibration point will result in a significantly reduced variance in the differences between molecular and fossil age estimates across all dated nodes in the phylogeny. Removal of these inconsistent fossils should result in a consistent set of fossils that accurately calibrates a molecular phylogeny.

Four of the oldest fossils used as calibration points in our molecular clock analyses are the same age (110 million

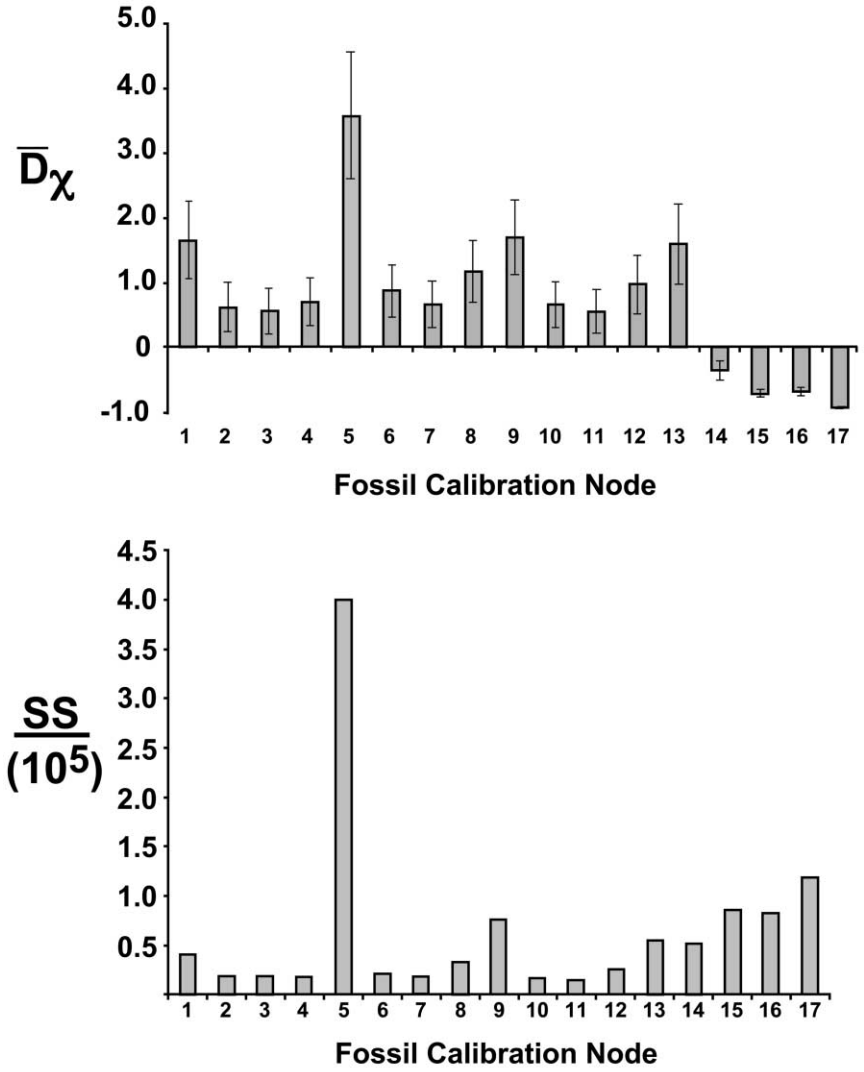

Figure 1: *Top*, Histogram of the mean deviation ($\overline{D}$) between molecular and fossil age estimates for all nodes using a single fossil-dated node as a calibration point. *Bottom*, Histogram of the SS values for a given fossil calibration node when it was used as the single calibration point.

years ago [mya]), and three of these were recovered from the Santana Formation in Brazil (nodes 3–5; table 1). A potential issue in using these fossil dates as calibration points is that there are two sets of phylogenetically nested nodes that are dated with fossils of the same age (nodes 3 and 4, and nodes 2 and 5; fig. A1). In both cases, one of these two redundant calibrations is presumably erroneous, unless the intervals between the diversification events at nodes 3 and 4 and nodes 2 and 5 were tightly spaced in evolutionary time. We thus viewed these as potentially erroneous calibration points before the analysis began.

Results and Discussion

Phylogenetic Analyses and Tests of Rate Heterogeneity

Likelihood ratio tests identified four different models of DNA substitution for the seven data partitions (table A1 in the online edition of the *American Naturalist*). Our partitioned mixed-model Bayesian analysis of the same 23 turtle species used in previous analyses (Shaffer et al. 1997) resulted in a well-resolved phylogeny, and the majority of nodes were supported with significant (≥0.95) Bayesian posterior probabilities (fig. A1). Maximum parsimony analysis of the concatenated data set resulted in a tree

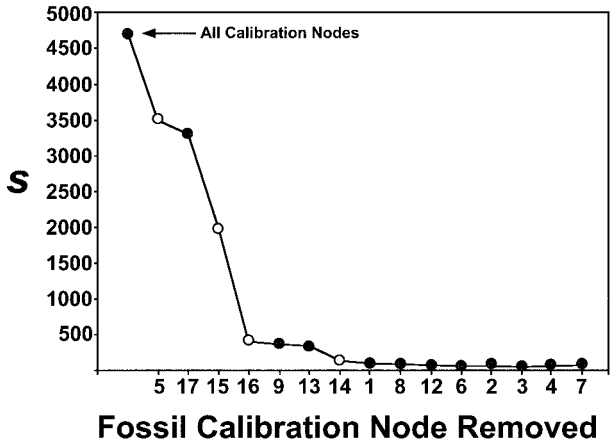

Figure 2: Plot illustrating the effect of removing fossil calibration points on *s*. Open points indicate that the removal of that fossil calibration resulted in a significant reduction in the variance of *s*, based on a one-tailed *F*-test.

topology very similar to that presented in figure A1; the only exception was that *Chelydra* was the sister taxon of the clade containing *Chelonia* and *Dermochelys*. In addition, maximum parsimony bootstrap support was very high for most nodes, with only five nodes recovered in <95% of the bootstrap pseudoreplicates (fig. A1). The tree topology is very similar to previous hypotheses of turtle relationships based on parsimony analyses of mtDNA sequences and morphology (Shaffer et al. 1997), with one important difference. Previous analyses identified the genera *Chelydra* and *Platysternon* as sister taxa, based primarily on the weight of morphological evidence. However, in the current tree, these lineages appear to be distantly related (fig. A1), further calling into question the validity of this hypothesized monophyletic group. Each of the three gene regions, as well as the concatenated nucleotide data set, shows significant nucleotide substitution rate heterogeneity among lineages (table A2 in the online edition of the *American Naturalist*).

Consistency of Divergence Time Estimates from Individual Fossil Absolute Age Estimates

Cross-validation analysis of individual fossil calibration points revealed a large average deviation between fossil and molecular age estimates for several fossil-dated nodes that were used to calibrate the molecular phylogeny (fig. 1, *top*). Nodes 14–17 were the most recent fossil ages (table 1), and these nodes produced molecular age estimates for other nodes that were consistently younger than their fossil age estimates, yielding negative $\overline{D}$ values (fig. 1, *top*). This pattern of much younger molecular age estimates, relative to fossil age estimates at other nodes, indicates that the error for fossil nodes 14–17 cannot be explained by lack of preservation, since they result in molecular ages that are far younger than the fossil dates. In contrast, among the fossil calibrations resulting in positive $\overline{D}$ values, only node 5 (dated with one of the three Santana Formation fossils) was substantially larger than all others, approaching a mean deviation of 400% with standard error bars that did not overlap those for any other node (fig. 1, *top*).

The ranking of the fossil calibrations using the SS values (fig. 1, *bottom*) determined the sequence that the fossil calibrations were removed from the analysis. In figure 2, we show both the order of removal of nodes (from left to right) and the effect of removing these calibrations on the magnitude of *s*. The removal of the first four fossil calibration points resulted in a 92% decrease in *s* from 470 to 40 (fig. 2). There was an additional decrease in *s* from 40 to 13 when the next three fossil calibration points were removed, and subsequent removal of the remaining seven fossil calibrations had essentially no impact (13–7.5) on the magnitude of *s* (fig. 2). Based on figure 2 alone, it would appear that removing either the first four or first seven fossils results in a large overall decrease in *s*.

To further explore the contribution of each fossil to a decrease in the variance of *s*, we conducted sequential one-tailed *F*-tests for the removal of each fossil calibration in figure 2. As expected, several of the first seven fossils with the largest SS (nodes 5, 15, and 16) resulted in a significant decrease in *s* (fig. 2). However, the sequential removal of nodes 17, 9, and 13 based on their rank order of SS value did not individually lead to a significant decrease in *s*. This can be seen visually in figure 2, where the removal of fossil 17 leads to a slight but insignificant decrease in *s*; the same is true for the removal of fossils 9 and 13. Our interpretation of these results hinges on the relationship between SS (which was used to determine the order of removal of fossils) and *s*. In using SS to evaluate the consistency of individual fossils, we relied on a measure of the absolute difference (in millions of years) between a molecular age estimate and the corresponding real fossil age for node i based on fossil χ. We favor this measure because it is based on absolute age estimates and their deviations. However, fossils with large proportional errors in the true age of the node that they date will have large SS values; if these are relatively young fossils, then they will often have correspondingly small effects on *s*. Other measures of deviation (e.g., $[(\mathrm{MA} - \mathrm{FA})/\mathrm{FA}]^2$) will change this relationship between SS and *s* (data not shown) in ways that may be preferable for some clades. We view the comparison of different measures of deviance between molecular and fossil age estimates as an important area for future research.

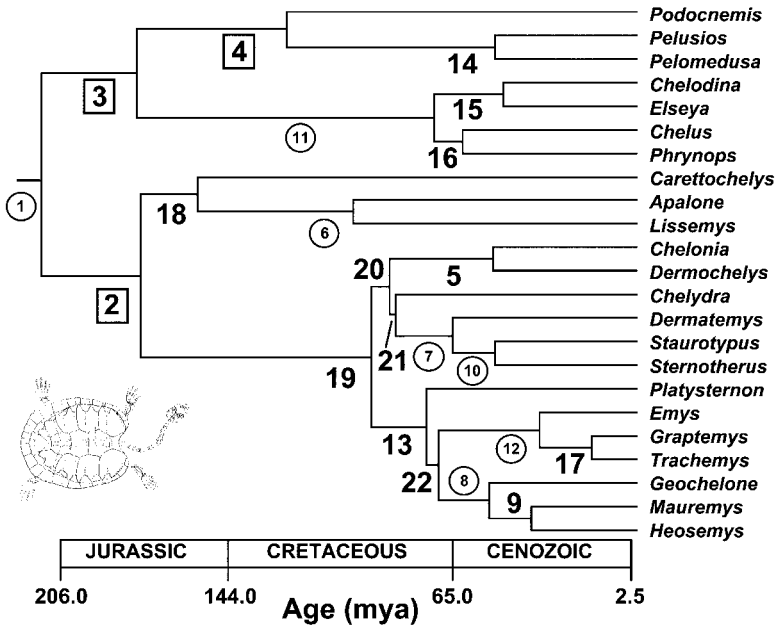

Figure 3: Time-calibrated phylogeny (chronogram) of turtles based on molecular dating estimates of divergence times using a set of seven consistent fossil dates as fixed calibration points (*numbers in circles*) and three more questionable fossils as minimal age constraints (*numbers in squares*). Exact age estimates for all numbered nodes are in table 2.

Based on these analyses, we designated the fossil calibrations with the seven highest SS values as inconsistent (nodes 5, 17, 15, 16, 9, 13, and 14) and removed them from the molecular dating analysis. Viewed another way, the 10 remaining fossil calibration points, individually or in combination, never resulted in a significant drop in *s*, suggesting that they were relatively consistent with each other. Of these 10, three are dated at 110 mya (nodes 2, 3, and 4) and almost certainly cannot all date to a simultaneous time of origin, even though they do not stand out as inconsistent based on figure 2. For example, the date of 110 mya for nodes 3 and 4 is redundant because node 4 is phylogenetically nested within node 3 (fig. A1) and both of these fossils are from the same formation. Because of these potential problems, we took a conservative approach and treated these three fossil calibrations as minimal age constraints. This has the effect in penalized likelihood analysis of constraining the age of the node to be at least 110 mya but permits the estimated age in penalized likelihood analysis to be older than this date.

Thus, we ended up with seven clearly inconsistent fossils that were removed from the analysis, three simultaneously aged fossils that we treated as minimal age constraints, and seven fossil dates that were deemed consistent and used as fixed calibration points. This strategy allows us to use fixed calibration points at both the root node of the turtle phylogeny (node 1) and at more apical nodes as well as minimal age constraints at intermediate nodes in the turtle tree (fig. 3).

A Timescale for Diversification of Major Turtle Lineages

Based on the cross-validation analyses and *F*-tests, we used the set of seven consistent fossils as fixed ages, and three additional minimum-age estimate fossils, for the penalized likelihood analysis to estimate divergence times among the major lineages of turtles. Estimated divergence times and their 95% confidence intervals are provided in table 2. As expected, there was a large difference between the fossil and molecular age for all of the calibration point nodes identified as inconsistent (table 2, last column), and the 95% confidence intervals of all but one of the molecular divergence time estimates did not overlap with the fossil age estimates for the seven inconsistent calibration point nodes (table 2). In addition, the estimated divergence times for two of the three minimal age constraint fossil calibrations dated at 110 mya were substantially higher than the fossil age (nodes 2 and 3; table 2). Node 4 was treated as a minimal age constraint because its fossil calibration was redundant with the older node 3; however, the molecular age estimate for node 4 is much closer than node 3 to the fossil estimated at 110 mya for both of these nodes (table 2), suggesting that it is reasonably accurate.

We show a time-calibrated phylogeny, or chronogram,

Table 2: Estimated divergence times and bootstrap estimates of standard error

Node and status of fossil calibration	Age estimate	Bootstrap estimate of CI	Difference between fossil and molecular age estimate
1. Fixed	210.00	Fixed	NA
2. Minimal age	174.87	±11.28	64.87
3. Minimal age	176.62	±8.37	66.62
4. Minimal age	123.82	±10.62	13.82
5. Inconsistent fossil	50.24	±6.55	−59.76
6. Fixed	100.00	Fixed	NA
7. Fixed	65.00	Fixed	NA
8. Fixed	52.00	Fixed	NA
9. Inconsistent fossil	37.27	±5.78	−12.73
10. Fixed	50.00	Fixed	NA
11. Fixed	71.00	Fixed	NA
12. Fixed	34.00	Fixed	NA
13. Inconsistent fossil	73.99	±18.77	−16.01
14. Inconsistent fossil	49.83	±7.13	31.83
15. Inconsistent fossil	46.74	±5.49	31.74
16. Inconsistent fossil	60.92	±4.84	49.32
17. Inconsistent fossil	15.36	±3.16	10.36
18. No fossil	154.98	±10.78	NA
19. No fossil	93.68	±6.23	NA
20. No fossil	87.03	±8.19	NA
21. No fossil	84.91	±5.64	NA
22. No fossil	69.98	±6.37	NA

Note: Bootstrap estimates of the 95% confidence interval (CI) of age estimates in the turtle chronogram (fig. 3) and differences between molecular and fossil age estimates for inconsistent calibration points (ages in millions of years). NA is not applicable.

in figure 3 based on the seven fixed and three minimal age-constrained calibration points in our analysis. A consistent theme in this chronogram is the antiquity of most of the living families of turtles, with six of nine family-level stem divergences predating the Cretaceous-Cenozoic boundary (fig. 3). The exceptions to this trend include the relatively diverse Kinosternidae (represented here by *Staurotypus* and *Sternotherus*: 23 living species), Testudinidae (*Geochelone*: 60 living species), Geoemydidae (*Mauremys* and *Heosemys*: 67 living species), Chelidae (*Chelodina*, *Elseya*, *Chelus*, and *Phrynops*: 56 living species), and the subfamily Deirochelyinae of Emydidae (*Graptemys* and *Trachemys*: 37 living species). All of these relatively species-rich groups diversified within the Cenozoic, and largely within the last 50 million years (fig. 3). Together, these four families and one subfamily (of the 14 generally recognized families of living turtles and tortoises) account for approximately 243 of 319, or about 76%, of all the living species of turtles on earth. Thus, while the oft-cited notion of turtles as an "ancient" group is true, the most species-rich families have all diversified relatively recently, within the last 50 million years.

The Bottom Line: How Do Inconsistent Fossils Affect Molecular Divergence Time Estimates?

Including all 17 fossils in the molecular clock analysis, regardless of their performance in the cross-validation analyses, provides additional insights into the role of unreliable fossils in date estimates. The inclusion of all fossils had little effect on the divergence time estimates in four of the five nodes in the turtle phylogeny not dated with the fossil record (nodes 19–22). However, the divergence time for the remaining undated node (node 18) changed dramatically when the inconsistent nodes were included. It appears that the effect of including unreliable fossils for node 18 constrained its age to be between 100 mya (node 6; see table 1) and 110 mya (node 2). However, treating the *Sandownia harrisi* fossil at node 2 as a minimal age constraint allowed a much older age estimate of 154.98 mya for node 18. In contrast, the slight effect on divergence time estimates for nodes 19–22 when the inconsistent fossils are included appears to result from their close phylogenetic proximity to a series of reliable fossils (nodes 7, 8, 10, and 12) and their relatively great phylogenetic distance to the closest inconsistent fossil. The key point is

that the magnitude of the impact that unreliable or inconsistent fossils will have on divergence time estimates will depend on tree topology and inferred branch lengths in the vicinity of those unreliable or inconsistent fossils.

Conclusions

A recent critique of molecular clock estimates of divergence times (Graur and Martin 2004) highlighted the potential errors inherent in studies based on single calibration points that are assumed to be error-free estimates of divergence times. These authors echo the concerns of others (Alroy 1999; Lee 1999; Conroy and van Tuinen 2003) that both realistic estimates of error for fossil dates and multiple calibration points are necessary next steps in the refinement of molecular dating analyses. Here, we take this caution one step further in recognizing that not all calibration points are equally reliable; due to uncertainties in the fossil record, our inability to date some strata precisely, and phylogenetic uncertainty, some fossil calibration points will almost certainly be "rogue" outliers that inflate rather than add precision to our molecular estimates of divergence times. Our proposed solution is to seek the set of fossils that provides molecular age estimates that closely correspond to fossil age estimates using a novel cross-validation strategy. Although this involves the often unpopular approach of discarding some (fossil) data in favor of others, it results in a set of calibration points that is internally consistent across a clade with much lower variance of time estimates than would be otherwise possible.

Using the set of seven consistent fixed fossil calibration points and three additional fossil-calibrated minimal age constraints, we found that most families and higher taxa of living turtles predate the Cretaceous/Cenozoic boundary, but more than three-quarters of the living species of turtles fall into five families that originated well within the Cenozoic. Thus, while the major living turtle lineages are indeed ancient, most of this antiquity consists of deeply diverged but species-poor lineages with few living representatives. These species-poor clades comprise the most ancient components of the evolutionary history of turtles. Given that many of them are also under extreme pressure from exploitation for food, medicine, and other human activities (Klemens 2000), our analyses emphasize that they deserve particular attention as conservation targets in the future.

Acknowledgments

D. I. Bolnick, B. M. Fitzpatrick, J. A. Fordyce, A. E. Georges, J. H. Gillespie, and M. J. Sanderson provided valuable discussion of molecular clock methods and fossil calibrations. D. I. Bolnick, C. D. Hulsey, T. H. Oakley, and G. R. Smith read earlier versions of the manuscript. M. Fujita and J. Krenz provided unpublished sequence data, and J. Iverson and A. Rhodin provided access to their unpublished lists of turtle species diversity. T.J.N. was supported by a Center for Population Biology Postdoctoral Fellowship at the University of California, Davis. The National Science Foundation under grants 9727161/0213153, CalFed, and the UC-Davis agricultural station awarded to H.B.S. provided research support.

Literature Cited

Alroy, J. 1999. The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. Systematic Biology 48:107–118.

Arbogast, B. S., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. Annual Review of Ecology and Systematics 33:707–740.

Benton, M. J. 1995. Testing the time axis of phylogenies. Philosophical Transactions of the Royal Society of London B 349:5–10.

Benton, M. J., and F. J. Ayala. 2003. Dating the tree of life. Science 300:1698–1700.

Britten, R. 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398.

Bromham, L., and D. Penny. 2003. The modern molecular clock. Nature Reviews Genetics 4:216–224.

Bromham, L., D. Penny, A. Rambaut, and M. D. Hendy. 2000. The power of relative rates tests depends on the data. Journal of Molecular Evolution 50:296–301.

Conroy, C. J., and M. van Tuinen. 2003. Extracting time from phylogenies: positive interplay between fossil and genetic data. Journal of Mammalogy 84:444–455.

Cooper, A., and D. Penny. 1997. Mass survival of birds across the Cretaceous-Tertiary boundary: molecular evidence. Science 275: 1109–1113.

de la Fuente, M., F. d. Lapparent de Broin, and T. Manera de Bianco. 2001. The oldest and first nearly complete skeleton of a chelid, of the *Hydromedusa* sub-group (Chelidae, Pleurodira), from the Upper Cretaceous of Patagonia. Bulletin de la Societe Geologique de France 172:237–244.

Doyle, J. A., and M. J. Donoghue. 1993. Phylogenies and angiosperm diversification. Paleobiology 19:141–167.

Engstrom, T. N., H. B. Shaffer, and W. P. McCord. 2004. Multiple data sets, high homoplasy and the phylogeny of softshell turtles (Testudines: Trionychidae). Systematic Biology 53:693–710.

Felsenstein, J. 1993. PHYLIP: phylogenetic inference package. Version 3.5c. University of Washington, Seattle.

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.

Gaffney, E. S. 1986. Triassic and early Jurassic turtles. Pages 183–187 *in* K. Padian, ed. The beginnings of the age of dinosaurs. Cambridge University Press, Cambridge.

———. 1990. The comparative osteology of the Triassic turtle *Proganochelys*. Bulletin of the American Museum of Natural History 194:1–263.

Gaffney, E. S., and P. A. Meylan. 1988. A phylogeny of turtles. Pages 157–219 *in* M. J. Benton, ed. The phylogeny and classification of tetrapods. Clarendon, Oxford.

Gaffney, E. S., M. Archer, and A. White. 1989. Chelid turtles from the

Miocene freshwater limestones of Riversleigh Station Northwestern Queensland, Australia. American Museum Novitiates 2959:1–10.
Gaffney, E. S., D. D. A. Campos, and R. Hirayama. 2001. *Cearachelys*, a new side-necked turtle (Pelomeousoides: Bothremydidae) from the early Cretaceous of Brazil. American Museum Novitiates 3319:1–20.
Graur, D., and W. Martin. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends in Genetics 20:80–86.
Hirayama, R. 1998. Oldest known sea turtle. Nature 392:705–708.
Huelsenbeck, J. P. 1994. Comparing the stratigraphic record to estimates of phylogeny. Paleobiology 20:470–483.
Huelsenbeck, J. P., and K. A. Crandall. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annual Review of Ecology and Systematics 28:437–466.
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.
Hutchison, J. H. 1980. Turtle stratigraphy of the Willwood Formation, Wyoming: preliminary results. Papers on Paleontology the Museum of Paleontology University of Michigan 24:115–118.
———. 1981. Homology of the plastral scales of the Kinosternidae and related turtles. Herpetologica 37:73–85.
———. 1991. Early Kinosternidae (Reptilia: Testudines) and their phylogenetic significance. Journal of Vertebrate Paleontology 11:145–167.
———. 1996. Testudines. Pages 337–353 *in* D. R. Protrhero and R. J. Emry, eds. The terrestrial Eocene-Oligocene transition in North America. Cambridge University Press, Cambridge.
Jackson, D. R. 1988. A reexamination of fossil turtles of the genus *Trachemys* (Testudines: Emydidae). Herpetologica 44:317–325.
Klemens, M. W. 2000. Turtle conservation. Smithsonian Institution, Washington, DC.
Lapparent de Broin, F. d. 2001. The European turtle fauna from the Triassic to the Present. Dumerilia 4:155–217.
Larget, B., and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Molecular Biology and Evolution 16:750–759.
Lee, M. S. Y. 1999. Molecular clock calibrations and metazoan divergence dates. Journal of Molecular Evolution 49:385–391.
Magallon, S., and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762–1780.
Marshall, C. R. 1990*a*. Confidence intervals on stratigraphic ranges. Paleobiology 16:1–10.
———. 1990*b*. The fossil record and estimating divergence times between lineages: maximum divergence times and the importance of reliable phylogenies. Journal of Molecular Evolution 30:400–408.
Meylan, P. A., and E. S. Gaffney. 1991. *Araripemys* Price, 1973. Pages 326–334 *in* J. G. Maisey, ed. Santana fossils: an illustrated atlas. TFH, Neptune, NJ.
Meylan, P. A., R. T. J. Moody, C. A. Walker, and S. D. Chapman. 2000. *Sandownia harrisi*, a highly derived trionychoid turtle (Testudines: Cryptodira) from the Early Cretaceous of the Isle of Wight, England. Journal of Vertebrate Paleontology 20:522–532.
Near, T. J., T. W. Kassler, J. B. Koppelman, C. B. Dillman, and D. P. Philipp. 2003. Speciation in North American black basses, *Micropterus* (Actinopterygii: Centrarchidae). Evolution 57:1610–1621.
Norell, M. A., and M. J. Novacek. 1992. The fossil record and evolution: comparing cladistic and paleontological evidence for vertebrate history. Science 255:1690–1693.
Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.
Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.
Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Molecular Biology and Evolution 14:1218–1231.
———. 1998. Estimating rate and time in molecular phylogenies: beyond the molecular clock? Pages 242–264 *in* D. E. Soltis, P. S. Soltis, and J. J. Doyle, eds. Molecular systematics of plants. 2. DNA sequencing. Kluwer, Amsterdam.
———. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 19:101–109.
———. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19:301–302.
Sanderson, M. J., and J. A. Doyle. 2001. Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. American Journal of Botany 88:1499–1516.
Shaffer, H. B., P. Meylan, and M. L. McKnight. 1997. Tests of turtle phylogeny: molecular, morphological, and paleontological approaches. Systematic Biology 46:235–268.
Smith, A. B., and K. J. Peterson. 2002. Dating the time of origin of major clades. Annual Review of Earth and Planetary Sciences 30:65–88.
Soltis, P. S., D. E. Soltis, V. Savolainen, P. R. Crane, and T. G. Barraclough. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. Proceedings of the National Academy of Sciences of the USA 99:4430–4435.
Spinks, P. Q., H. B. Shaffer, J. B. Iverson, and W. P. McCord. 2004. Phylogenetic hypotheses for the turtle family Geoemydidae. Molecular Phylogenetics and Evolution 32:164–182.
Springer, M. S. 1995. Molecular clocks and the incompleteness of the fossil record. Journal of Molecular Evolution 41:531–538.
Strauss, D., and P. M. Sadler. 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Mathematical Geology 21:411–427.
Sukanov, V. B. 2000. Mesozoic turtles of Middle and Central Asia. Pages 309–367 *in* M. J. Benton, M. A. Shishkin, D. M. Unwin, and E. N. Kurochkin, eds. The age of dinosaurs in Russia and Mongolia. Cambridge University Press, Cambridge.
Swofford, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b4. Sinauer, Sunderland, MA.
Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. Molecular Biology and Evolution 12:823–833.
Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution 15:1647–1657.
Williams, E. E. 1954. A new Miocene species of *Pelusios* and the evolution of that genus. Breviora 25:1–7.
Wood, R. C. 1976. Description of fossil *Chelus*. Breviora 435:1–26.
Yeh, H. K. 1965. New materials of fossil turtles of Inner Mongolia. Vertebrata PalAsiatica 9:47–69.
Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. Pages 97–166 *in* V. Bryson and H. J. Vogel, eds. Evolving genes and proteins. Academic Press, New York.

Associate Editor: Axel Meyer

Appendix from T. J. Near et al., "Assessing Concordance of Fossil Calibration Points in Molecular Clock Studies: An Example Using Turtles"

(Am. Nat., vol. 165, no. 2, p. 137)

Species Sampled and GenBank Accession Numbers

Family classification (in bold), species names (both scientific and common), and GenBank accession numbers for cytochrome *b*, R35 intron, and RAG-1. **Carettochelyidae**, *Carettochelys insculpta*, pig-nosed turtle, U81355, AY259571, AY687904; **Chelidae**, *Chelodina longicollis*, common snake-necked turtle, U81356, AY339636, AY687921; *Chelus fimbriatus*, matamata, U81343, AY339640, AY687918; *Elseya latisternum*, serrated snapping turtle, U81354, AY339643, AY687920; *Phrynops gibbus*, Gibba turtle, U81348, AY742455, AY687919; **Cheloniidae**, *Chelonia mydas*, green turtle, U81352, AY339635, AY687907; **Chelydridae**, *Chelydra serpentina*, American snapping turtle, U81357, AY742461, AY687906; **Dermatemydidae**, *Dermatemys mawii*, Central American river turtle, U81364, AY339638, AY687910; **Dermochelyidae**, *Dermochelys coriacea*, leatherback sea turtle, U81363, AY742460, AY687908; **Emydidae**, *Emys marmorata*, western pond turtle, U81344, AY339631, AY687917; *Graptemys psuedogeographica*, false map turtle, U81345, AY742457, AY687916; *Trachemys scripta*, red-eared slider, U81351, AY742458, AY687915; **Geoemydidae**, *Heosemys spinosa*, spiny turtle, U81362, AY434652, AY687913; *Mauremys reevesii*, Chinese three-keeled pond turtle, U81358, AY434567, AY687914; **Kinosternidae**, *Sternotherus odoratus*, common musk turtle, U81350, AY742463, AY687911; *Staurotypus triporcatus*, Mexican giant musk turtle, U81349, AY339633, AY687909; **Pelomedusidae**, *Pelomedusa subrufa*, African helmeted turtle, U81346, AY339639, AY687922; *Pelusios williamsi*, William's mud turtle, U81347, AY339629, AY687923; **Platysternidae**, *Platysternon megacephalum*, big-headed turtle, U81361, AY742462, AY687905; **Podocnemidae**, *Podocnemis expansa*, South American river turtle, U81360, AY742456, AY687924; **Testudinidae**, *Geochelone pardalis*, leopard tortoise, U81354, AY742459, AY687912; **Trionychidae**, *Apalone spinifera*, spiny softshell turtle, U81342, AY259582, AY687901; *Lissemys punctata*, Indian flapshell turtle, U81359, AY259593, AY687902.

Table A1

Summary of models of DNA substitution selected for data partitions using maximum likelihood ratio tests

Data partition	DNA substitution model	No. substitution types	Invariant sites?	Substitution rates[a]
Mitochondrial genes:				
Cytochrome *b* first codon	GTR	6	Yes	Gamma distributed
Cytochrome *b* second codon	GTR	6	Yes	Gamma distributed
Cytochrome *b* third codon	GTR	6	No	Gamma distributed
Nuclear genes:				
RAG-1 first codon	GTR	6	No	Gamma distributed
RAG-1 second codon	HKY85	2	Yes	Gamma distributed
RAG-1 third codon	HKY85	2	No	Gamma distributed
R35 intron	HKY85	2	No	Gamma distributed

[a] Among-site rate variation.

Table A2
Likelihood ratio tests of nucleotide substitution rate heterogeneity among lineages based on a χ^2 distribution with 21 df

Gene	ln likelihood (molecular clock enforced)	ln likelihood (molecular clock not enforced)	Λ	P
Cytochrome *b*	−9,924.89	−9,887.58	74.62	≪.001
RAG-1	−10,669.57	−10,581.23	176.68	≪.001
R35	−5,729.90	−5,626.93	205.94	≪.001
Combined	−27,600.62	−27,442.65	315.94	≪.001

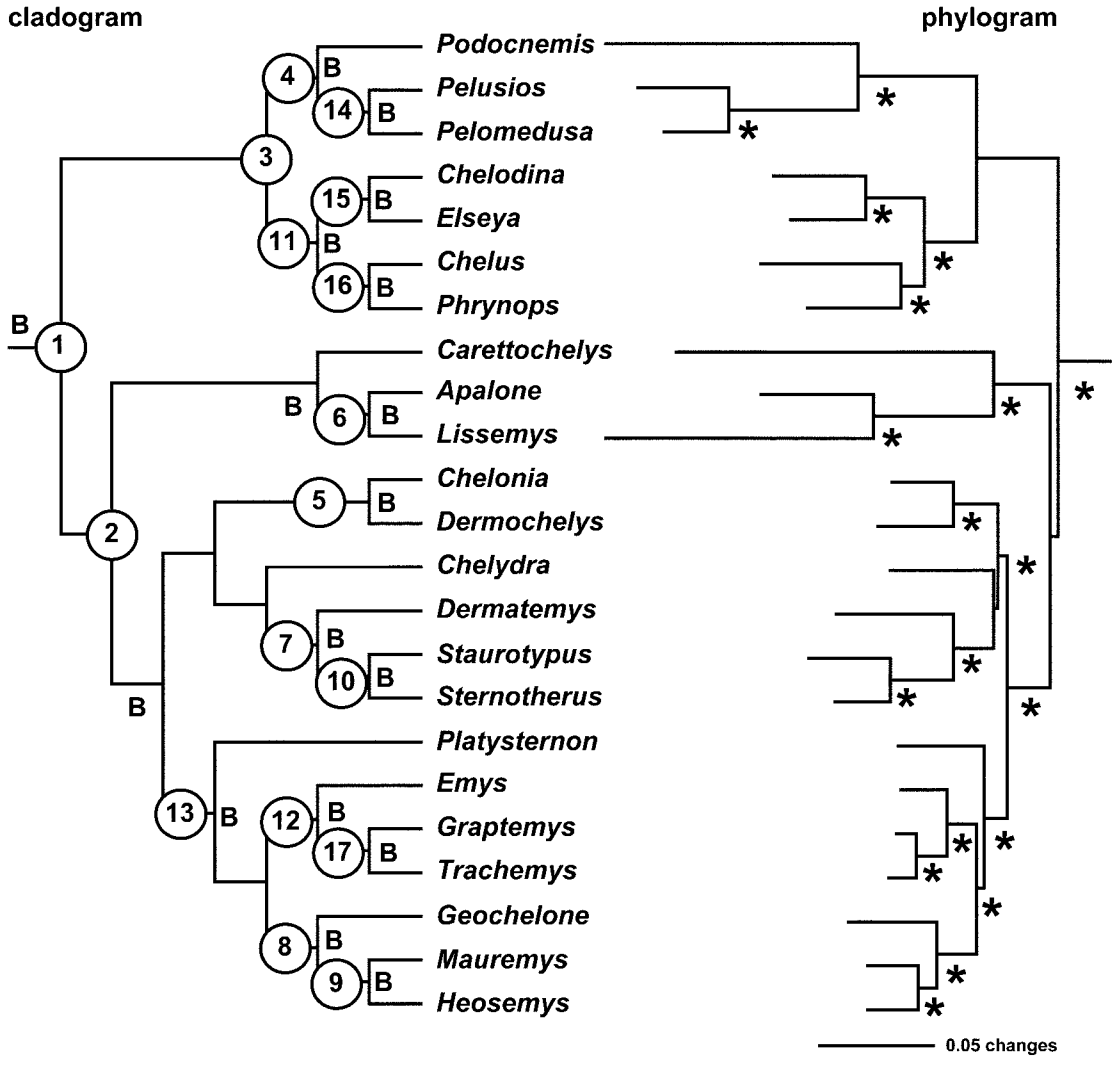

Figure A1: New phylogenetic tree for 23 turtle species representing all major extant lineages (see Shaffer et al. 1997). The tree was inferred from a Bayesian maximum likelihood analysis of a combined data set that included the mitochondrial-encoded cytochrome *b* gene and two nuclear genes, an intron from the R35 gene and an exon from RAG-1. Circled numbers at nodes on the cladogram (*left*) identify all fossil-dated calibration points (table 1). The phylogram depicting Bayesian estimated branch lengths is on the right. Nodes marked with an asterisk on the phylogram were supported with Bayesian posterior probabilities ≥0.95, and those marked with a *B* on the cladogram were supported in ≥95% of the maximum parsimony bootstrap pseudoreplicates.