Evolution, 59(8), 2005, pp. 1768–1782

FOSSIL CALIBRATIONS AND MOLECULAR DIVERGENCE TIME ESTIMATES IN CENTRARCHID FISHES (TELEOSTEI: CENTRARCHIDAE)

THOMAS J. NEAR,[1,2] DANIEL I. BOLNICK,[3,4,5] AND PETER C. WAINWRIGHT[3,6]
[1]*Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996-1610*
[2]*E-mail: tnear@utk.edu*
[3]*Center for Population Biology, University of California, One Shields Avenue, Davis, California 95616*
[4]*Section of Integrative Biology, University of Texas at Austin, 1 University Station C0930, Austin, Texas 78712-0253*
[5]*E-mail: danbolnick@mail.utexas.edu*
[6]*E-mail: pcwainwright@ucdavis.edu*

Abstract.—Molecular clock methods allow biologists to estimate divergence times, which in turn play an important role in comparative studies of many evolutionary processes. It is well known that molecular age estimates can be biased by heterogeneity in rates of molecular evolution, but less attention has been paid to the issue of potentially erroneous fossil calibrations. In this study we estimate the timing of diversification in Centrarchidae, an endemic major lineage of the diverse North American freshwater fish fauna, through a new approach to fossil calibration and molecular evolutionary model selection. Given a completely resolved multi-gene molecular phylogeny and a set of multiple fossil-inferred age estimates, we tested for potentially erroneous fossil calibrations using a recently developed fossil cross-validation. We also used fossil information to guide the selection of the optimal molecular evolutionary model with a new fossil jackknife method in a fossil-based model cross-validation. The centrarchid phylogeny resulted from a mixed-model Bayesian strategy that included 14 separate data partitions sampled from three mtDNA and four nuclear genes. Ten of the 31 interspecific nodes in the centrarchid phylogeny were assigned a minimal age estimate from the centrarchid fossil record. Our analyses identified four fossil dates that were inconsistent with the other fossils, and we removed them from the molecular dating analysis. Using fossil-based model cross-validation to determine the optimal smoothing value in penalized likelihood analysis, and six mutually consistent fossil calibrations, the age of the most recent common ancestor of Centrarchidae was 33.59 million years ago (mya). Penalized likelihood analyses of individual data partitions all converged on a very similar age estimate for this node, indicating that rate heterogeneity among data partitions is not confounding our analyses. These results place the origin of the centrarchid radiation at a time of major faunal turnover as the fossil record indicates that the most diverse lineages of the North American freshwater fish fauna originated at the Eocene-Oligocene boundary, approximately 34 mya. This time coincided with major global climate change from warm to cool temperatures and a signature of elevated lineage extinction and origination in the fossil record across the tree of life. Our analyses demonstrate the utility of fossil cross-validation to critically assess individual fossil calibration points, providing the ability to discriminate between consistent and inconsistent fossil age estimates that are used for calibrating molecular phylogenies.

Key words.—Divergence times, fossil calibration, *Lepomis*, *Micropterus*, molecular clock, penalized likelihood.

Received January 18, 2005. Accepted June 1, 2005.

Integration of fossil information with molecular phylogenies offers potential for efforts using molecular dating methods to reconstruct the timing of diversification of organismal lineages. Enthusiasm for these methods has been tempered by the fact that the initial promise of the ''molecular clock,'' where molecular evolutionary rates are assumed to be uniform across organismal lineages (Zuckerkandl and Pauling 1962, 1965), has been confounded by the observation of persistent and widespread heterogeneity in molecular evolutionary rates among and within lineages (Britten 1986). The effects of molecular rate heterogeneity, both among lineages and data partitions, on molecular divergence time estimates is an area of active research. There are now a number of compensatory strategies including: (1) pruning lineages exhibiting deviation from uniform molecular evolutionary rates (Takezaki et al. 1995), (2) using multiple models of molecular evolution with differing molecular evolutionary rates on different branches of phylogenetic trees (Yoder and Yang 2000), (3) modeling the evolution of the rate of evolution using Bayesian methods (Thorne et al. 1998; Huelsenbeck et al. 2000), and (4) using nonparametric and semiparametric models of molecular rate evolution (Sanderson 1997, 2002, 2003).

Substantial progress has been made in accounting for rate heterogeneity in molecular divergence time estimation; however, there has been far less attention devoted to issues involved with using fossils to calibrate molecular phylogenies. Fossil calibrations are a key step in converting relative molecular divergence times to absolute age; therefore, erroneous fossil dates will thwart molecular divergence time estimates. Several sources of error contribute to inaccurate fossil dates, with the most common arising from the fact that the fossil record is incomplete and will consistently produce underestimates of lineage ages (Marshall 1990a; Springer 1995). Other sources of error include taxonomic misidentification of fossils and their erroneous placement onto phylogenetic trees (Benton and Ayala 2003) and erroneous estimates of the geologic ages of fossil-bearing rock formations (Conroy and van Tuinen 2003). Error will also result if fossil-inferred minimal age estimates are applied to crown groups in phylogenetic trees instead of the appropriate stem lineages (Doyle and Donoghue 1993; Magallón and Sanderson 2001).

Several studies have attempted to assess the consistency among multiple fossil calibration points, by determining the similarity of dates inferred from molecular data using different fossil calibrations (Norman and Ashley 2000; Renner et al. 2001; Soltis et al. 2002; Near et al. 2003a, 2005; van Tuinen and Dyke 2003; Near and Sanderson 2004). These efforts have produced mixed results, as some studies have

demonstrated fairly high consistency among all calibration points (Renner et al. 2001; Near et al. 2003a, 2005; Near and Sanderson 2004), and others have identified particular fossil calibrations that appear inconsistent (Soltis et al. 2002; van Tuinen and Dyke 2003). Recently, methods have been developed that attempt to assess the consistency of fossil calibration points in molecular divergence time estimations using a cross-validation between fossil and molecular age estimates (Near and Sanderson 2004; Near et al. 2005). These methods have proven useful in identifying inconsistent fossil calibration points in datasets for angiosperms, mammals, and turtles. Once inconsistent fossils are removed from molecular dating analyses, investigators may have greater confidence in the results, since the consistent fossils offer maximal agreement between fossil and molecular estimated divergence times (Near et al. 2005). In addition, Near and Sanderson (2004) have introduced a novel method of fossil-based model cross-validation, which uses fossil information to select the optimal model of molecular evolution for penalized likelihood analysis.

Molecular Divergence Time Estimates and the Evolution of the North American Freshwater Fish Fauna

The North American fish fauna is diverse (Briggs 1986; Lundberg et al. 2000), and is comprised of lineages that span ray-finned fish diversity. In addition, there are several endemic clades, including some that are currently restricted to North America but occur as fossils on several other continents (Wiley 1976; Grande and Bemis 1991, 1998). Previous attempts to reconstruct the timing and origin of the diverse North American freshwater fish fauna have used earliest occurrences of taxa in the fossil record, as well as the historical geomorphology of continental connections and fragmentation (Patterson 1981; Cavender 1986). The consensus from these studies is that the most diverse lineages of North American freshwater fishes originated at a time near the Eocene-Oligocene transition (approximately 34 million years ago; mya), where there is a signature of both origin and diversification of new lineages, and extinction of Paleogene lineages in the fossil record (Cavender 1986; Wilson and Williams 1992).

There are numerous phylogenetic hypotheses for North American freshwater fishes that include both fossil and extant taxa; these provide a rich opportunity to reconstruct and compare the timing of speciation among a set of diverse lineages (Lundberg 1975; Wiley 1976; Grande and Bemis 1991; Lundberg 1992; Smith 1992a,b; Wilson 1992; Wilson and Williams 1992; Stearley and Smith 1993). However, using fossil information to calibrate molecular phylogenies has only recently been applied to very limited components of the North American freshwater fish fauna (Dowling et al. 2002; Smith et al. 2002; Near et al. 2003a; Near and Benard 2004).

In this study we apply fossil cross-validation methods to a phylogeny of Centrarchidae, a clade of freshwater fishes with all species, both fossil and extant, endemic to North America (Etnier and Starnes 1993). Centrarchids are an excellent model system for investigating the origin of the diverse North American freshwater fish fauna and exploring issues of fossil calibration in molecular dating. We have generated a robust phylogenetic hypothesis for all 32 extant centrarchid species derived from three mtDNA gene regions and four nuclear loci. Centrarchids are well represented in the fossil record of Cenozoic freshwater deposits in North America (Smith 1981; Cavender 1986), thus providing many potential calibration points. A calibrated centrarchid molecular phylogeny may also provide critical external calibrations for molecular phylogenies of important and diverse lineages of North American freshwater fishes that have a poor fossil record and essentially no calibration points (Near and Benard 2004). In addition, a calibrated molecular phylogeny of Centrarchidae can provide valuable data to the study of functional morphology (Wainwright and Lauder 1992) and the evolution of reproductive isolation (Bolnick and Near 2005).

A total of 10 fossil calibration points is applied to the centrarchid molecular phylogeny, and fossil-cross validation is used to evaluate individual fossil calibration points. Using these methods as a guide to identify a group of consistent fossil calibrations, we then use a new jackknife method to determine which fossil-inferred date should be treated as a fixed minimum age estimate in a fossil-based model cross-validation, which is used to select the optimal smoothing parameter value for penalized likelihood analysis. We present a chronogram that depicts our hypothesis of the timing of the origin and diversification of centrarchid fishes, and use this chronogram to discuss the evolution of the diverse North American freshwater fish fauna, with regard to the fossil record and paleoclimate of North America.

Materials and Methods

Collection and Alignment of DNA Sequences

We collected sequence data from mitochondrial and nuclear genes from all 32 centrarchid and four outgroup species using specimens and methods outlined in Near et al. (2004), except we added DNA sequences of four additional genes, and an additional specimen of *Lepomis microlophus* collected from the Neeches River, Jasper County, Texas. The mtDNA and nuclear gene sequences for this specimen of *L. microlophus* were not ready when we performed our earlier analyses of centrarchid phylogeny (Near et al. 2004). A total of five gene regions were sampled from the mitochondrial genome: the protein-coding NADH subunit 2 (ND2); the large subunit ribosomal RNA (16S rRNA); and three transfer RNA (tRNA) genes: methionine (Met), tryptophan (Trp), and alanine (Ala). We also sequenced four nuclear genes, S7 ribosomal protein intron 1 (S7), calmodulin intron 4 (CaM), *Tmo4C4*, and rhodopsin (RH). GenBank accession numbers for all sequences used are given in the Appendix (available online only at http://dx.doi.org/10.1554/05-030.1.s1). Primers for polymerase chain reaction (PCR) amplification of the RH locus were Rho-1 5′-GTC CAT ATG AAT ACC CTC AGT ACT ACC-3′, and Rho-2 5-′TCT TTC CGC AGC ACA ACG TGG-3′. Complete double-stranded sequences were obtained from PCR products using four sequencing primers; RHint2F (TAC TAC CTW GTC ARC CCW GCA GC), RHint3F (GCA ARC CCA TYA GCA ACT TCC G), RHint2R (GTG GTG ATC ATG CAG TGG CGG A), and RHint3R (CTC RGG RAT GTA MCG RGA CCA GCC). Conditions for PCR and primer sequences of all the other genes sampled in this study are available in other sources

(Streelman and Karl 1997; Chow and Hazama 1998; Chow and Takeyama 2000; Near et al. 2003a,b).

Protein coding genes, tRNAs, and introns were aligned with ClustalW (Thompson et al. 1997). The 16S rRNA sequences were aligned based on secondary structure, designating nucleotides as either paired and unpaired. Existing models of 16S rRNA secondary structure were used in conjunction with ClustalX to produce a final alignment (Gutell and Fox 1988; Gutell et al. 1993; De Rijk et al. 2000). Following Near et al. (2004), two species each from Percidae and Nototheniidae were used as outgroup taxa in all analyses (online Appendix).

Model Selection and Phylogenetic Analyses

Phylogenetic hypotheses of centrarchid relationships were generated with a combined data strategy using a partitioned mixed-model Bayesian (Ronquist and Huelsenbeck 2003) analysis (pMM Bayesian) with posterior probabilities estimated using metropolis-coupled Markov chain Monte Carlo (MC3; Larget and Simon 1999; Huelsenbeck et al. 2001). Fourteen data partitions were identified: three codons from each of the three protein coding genes (ND2, *Tmo4C4*, and RH), paired and unpaired nucleotide sites in the 16S rRNA, and single partitions for each of the two introns (S7 and CaM) and the pooled tRNA genes. The optimal maximum likelihood (ML) model of sequence evolution for each data partition was assessed with hierarchical likelihood ratio tests (LRT) using the computer program Modeltest 3.0 (Posada and Crandall 1998). The different models of sequence evolution that were selected for each particular data partition were assigned in the computer program MrBayes 3.0 (Ronquist and Huelsenbeck 2003) with the APPLYTO command, and model parameter values were estimated for each data partition using UNLINK commands. The models used in the pMM Bayesian analyses differed by no more than three parameters. Each model had a substitution matrix (one, two, or six substitution rates) corresponding to JC or F81, HKY85 or K80, and GTR models of sequence evolution, among-site rate variation (equal versus gamma distributed rates), and whether or not the presence of invariant sites was modeled (Swofford et al. 1996). MrBayes 3.0 was run with 4×10^6 generations to ensure convergence of the MC3 algorithm in the estimation of tree topology and branch lengths.

In addition to the Bayesian analyses, phylogenetic relationships of centrarchids were estimated from the concatenated DNA dataset with maximum parsimony. A heuristic tree search with 100 random addition sequence replicates was performed using the computer program PAUP* 4.0 (Swofford 2000). Relative clade support in maximum parsimony optimality criteria was assessed with a bootstrap analysis using PAUP* with 2000 pseudoreplicates.

Heterogeneity of nucleotide substitution rates among centrarchid lineages for the concatenated seven-gene dataset was assessed using a likelihood ratio test comparing rate variable (nonclock) and rate constant (molecular clock) models of sequence evolution. The significance of the likelihood ratio test was determined by comparing to a chi-square distribution with $s - 2$ degrees of freedom, where s equals the number of taxa in the phylogenetic tree (Felsenstein 1981).

The Centrarchid Fossil Record, Calibration Points, and Absolute Age Estimates

Centrarchid fishes have a rich fossil record that extends from the Eocene to the Pleistocene; however, the oldest fossils that are clearly assignable to extant lineages are from the Miocene (Uyeno and Miller 1963; Smith 1981; Cavender 1986). Our examination of the literature resulted in 10 potential fossil calibration points that serve as minimal age estimates, with the absolute age estimates of these fossils extending from 16.0 mya to 2.4 mya (Table 1). In all cases, these calibrations were the oldest fossil occurrences of a particular clade in the geological record. Fossils were placed in the centrarchid phylogeny using shared apomorphies among extant and fossil taxa, and the dating of crown versus stem group nodes by the fossil calibrations followed strategies outlined in Doyle and Donoghue (1993) and Magallón and Sanderson (2001). Absolute age estimation of individual fossils was greatly facilitated by the fact that most centrarchid fossils occur in well-studied formations that also bear mammal fossils (Woodburne 2004a). The estimated absolute ages of these fossil-bearing formations were taken from the literature, and are the result of one, or combination of, biostratigraphic, radioisotope, and magnetostratigraphic data (Table 1).

An earlier study of molecular dating in centrarchids used two fossil calibration points, and a cross-validation analysis indicated that the two calibrations resulted in consistent molecular age estimates (Near et al. 2003a). However, recent morphological analysis of the *Micropterus* fossil used in Near et al. (2003a) indicates that this fossil was incorrectly identified and is not *Micropterus* or any other centrarchid (G. R. Smith, pers. comm.). Therefore, the Toledo Bend fossil, initially identified as *Micropterus*, was not included as a calibration point in the analyses reported here.

Molecular estimates of divergence times often result in much older dates than inferred from the fossil record (Foote et al. 1999; Smith and Peterson 2002). The reason for this discrepancy may be due to the fact that the fossil record is incomplete and first appearances of clades are inadequate indicators of minimal ages. We attempted to assess the completeness of the fossil record for five of the 10 calibration points using gap analysis, which is a statistical method for placing confidence intervals on the lower bound of fossil ranges (Strauss and Sadler 1989; Marshall 1990a,b; Springer 1995). If the fossil horizons are randomly distributed between the earliest fossil and the last occurrence (in this case all lineages are extant), then the confidence interval for the unseen gaps at the lower bound of the fossil record is estimated from the formula given in Strauss and Sadler (1989):

$$P = 1 - (1 + \alpha)^{-(n-1)} \quad (1)$$

where P is the confidence level (e.g., 0.95), α is the confidence interval expressed as a fraction of the observed stratigraphic range, and n is the number of fossil horizons plus one for the extant taxa. The confidence interval is determined by solving for α (Marshall 1990a). Following Marshall (1990a,b), we applied the collective fossil record of particular species and lineages to estimate the lower bounds with a confidence level at 0.95. Lower bound estimates were cal-

TABLE 1. Centrarchid fossil calibrations, estimated geologic ages of fossils, and estimated lower bound of age estimate for selected nodes. mya, millions of years ago.

	Fossil taxon	Age (mya)	Gap analysis age at node (mya)[1]	Location	Reference	Source for age
C1	*Micropterus* spp.	16.0	23.2	Lower Snake Creek local fauna, Sioux Co., Nebraska	Matthew 1924	Tedford et al. 1987
C2	*Archoplites clarki*	15.5	18.7	Clarkia Lake Beds, Latah Co., Idaho	Smith and Miller 1985	Golenberg et al. 1990
C3	*Lepomis cf. microlophus*	13.5	NA	Lower Valentine Formation, Brown Co., Nebraska	Smith 1962	Tedford et al. 1987
C4	*Micropterus* spp.	12.0	NA	Wakeeney local fauna, Trego Co., Kansas	Wilson 1968	Wilson 1968; Tedford et al. 1987
C5	*Pomoxis* sp.	12.0	NA	Wakeeney local fauna, Trego Co., Kansas	Wilson 1968	Wilson 1968; Tedford et al. 1987
C6	*L. kansasensis*	6.6	29.5	Rhino Hill Quarry, Logan Co., Nebraska	Hibbard 1936	Passey et al. 2002
C7	*L. cyanellus*	3.9	5.3	Rexroad 3 local fauna, Meade Co., Kansas	Smith 1962	Repenning 1987
C8	*Ambloplites rupestris*	3.9	NA	Rexroad 3 local fauna, Meade Co., Kansas	Smith 1962	Repenning 1987
C9	*L. humilis*	3.4	7.2	Sand Draw local fauna, Brown Co., Nebraska	Smith and Lundberg 1972	Repenning 1987
C10	*L. megalotis*	2.4	NA	Rita Blanca Lake Deposits, Hartley Co., Texas	Koster 1969	Lindsay et al. 1975; Repenning 1987

[1] 95% lower bound age estimate.

culated only for lineages that were represented by at least three fossils in the geologic record.

Estimation of Divergence Times

Mean branch lengths from all post-burn-in trees resulting from the pMM Bayesian analysis were used in divergence time estimation, and these were generated in MrBayes using the SUMT command. Since we were able to reject a molecular clock model, all divergence times were estimated with the penalized likelihood method (Sanderson 2002), using the computer program r8s version 1.7 (Sanderson 2003; Near and Sanderson 2004). In analyses performed for fossil cross-validation, the optimal smoothing parameter to apply in penalized likelihood was determined using sequence-based cross-validation as outlined by Sanderson (2002). Subsequent penalized likelihood analyses using the set of consistent fossil age estimates used the fossil-based model cross-validation to determine the optimal smoothing parameter value (Near and Sanderson 2004). In all penalized likelihood analyses, fossil calibrations were either fixed in r8s using the FIXAGE command, or treated as minimal age constraints using the CONSTRAIN MIN_AGE command. These analyses are not meant to endorse penalized likelihood over other commonly used strategies such as Bayesian methods to estimate divergence times from molecular phylogenies (Thorne et al. 1998); however, penalized likelihood appears to be the most appropriate method to assess the veracity of individual calibration points (Near and Sanderson 2004; Near et al. 2005).

Uncertainty in the divergence time estimates was explored using two methods. First, error contributed from data sampling was estimated using a nonparametric bootstrap procedure outlined in Baldwin and Sanderson (1998). One thousand bootstrap replicates were generated using the computer program CodonBootstrap version 2.1 (Bollback 2001), with the pMM Bayesian tree and PAUP* commands inserted after every replicate. Upon execution of the bootstrap data file in PAUP*, branch lengths were calculated on the pMM Bayesian tree at each replicate and the trees with branch lengths were imported into a Nexus formatted tree file. Confidence intervals were estimated by calculating the central 95% distribution of divergence time estimates from the 1000 bootstrap replicates at a given node using the PROFILE command in r8s (Sanderson and Doyle 2001). We note that this error estimate does not account for the possibility of suboptimal solutions in penalized likelihood. This can have the effect of producing confidence intervals on the divergence times that are too narrow.

In addition to data sampling error, divergence time estimates may be affected by uncertainty in the phylogenetic reconstruction. Using a pMM Bayesian analysis allows us to estimate both the mean and 95% credibility interval of substitution numbers for individual branches. This credibility interval can be calculated by taking the central 95% distribution of estimates from the 30,000 post-burn-in Bayesian trees. However, these branch lengths are measured in substitutions, not in absolute time (millions of years). To get a credibility interval for actual divergence times, we imported the 30,000 post-burn-in trees into r8s, and used the PROFILE command to transform each tree into a chronogram. The resulting distribution of lengths for each node across the 30,000 chronograms represents the Bayesian 95% credibility interval for branch lengths, translated into time, and serves as an estimate of phylogenetic uncertainty. Ideally, a single estimate of uncertainty should be used that combines error in

sampling, Bayesian estimation, and likelihood estimation, but there is currently no method for acquiring such a compound confidence interval. All divergence time estimates were calculated in millions of years ago and were placed into the context of geologic ages following revisions to the Cenozoic geochronology (Berggren et al. 1995).

Assessing Consistency of Fossil Calibration Points: Fossil Cross-Validation

We explored the consistency of the 10 individual centrarchid calibration points using fossil cross-validation (Near et al. 2005). Fossil cross-validation measures the agreement between a single fossil calibration point and all other fossil calibration points included in the analysis in order to identify fossil calibrations that generate inconsistent molecular age estimates, as compared to the known fossil absolute ages.

Given a phylogenetic tree with multiple nodes dated with fossil information, we fixed the age of a single node using the fossil calibration and calculated the difference between the molecular age estimate and the absolute fossil age estimates for all other fossil-dated nodes in the centrarchid phylogeny. When the fossil age at node χ is used as the single fixed calibration point, and multiple nodes in the centrarchid phylogeny are dated with fossil information, Near et al. (2005) defined

$$D_i = (\mathrm{MA}_i - \mathrm{FA}_i) \tag{2}$$

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 as the single fixed minimum age estimate in a penalized likelihood analysis.

Values of D_i were used in a three-step procedure to identify and remove inconsistent fossil calibrations from the centrarchid molecular dating analysis (Near et al. 2005). First, for each fixed fossil calibration, we calculated SS_χ, which is the sum of the squared D_i values:

$$\mathrm{SS}_\chi = \sum_{i \neq \chi} D_i^2. \tag{3}$$

Each fossil calibration was then ranked based on the magnitude of its SS_χ score, and we identified the fossil calibration with the highest SS_χ score to be the most inconsistent and that with the lowest the least inconsistent relative to all other fossil calibration points. Second, we calculated the average squared deviation of D_i values for all 10 fossil calibrations in the analysis (s):

$$s = \frac{\sum_{\chi=1}^{n} \sum_{i \neq \chi} D_i^2}{n(n-1)} \tag{4}$$

where n is equal to the number of fossil calibrated nodes. To determine the impact of removing individual fossil calibrations on s, we removed all D_i values involving the fossil with the greatest SS_χ value and recalculated s based on the D_i values involving the remaining fossil calibrations (Near et al. 2005). This process of eliminating the fossil calibration with the highest SS_χ value was continued until only two fossil calibrations remained, and these two were the ones with the lowest and second lowest SS_χ values. If all fossil calibration points represent equally accurate minimal absolute age estimates for the nodes at which they are phylogenetically placed, then we expect s to change very little as fossil calibrations are removed from the analysis. However, removal of fossils that are inaccurate relative to other calibration points should produce an appreciable decrease in s (Near et al. 2005). We visually inspected the behavior of s as fossils were removed by plotting this value as the fossil calibrations were removed. The significance of removal of fossils on s was determined by comparing the variance of SS_χ values across n nodes to the variance of SS_χ across the $n - 1$ nodes remaining after removal of the ith calibration point. The SS_χ values were assumed to be independent samples, since each one is calculated from different fixed fossil calibration points. We used a one-tailed F-test based on $n - 1$ degrees of freedom numerator (dfn) and the difference in the number of comparisons and dfn as the degrees of freedom denominator, treating the s values as random samples taken from a normal population (Zar 1984). The F-test was to test the hypothesis that there is no difference in the variance of s before and after removing fossil calibrations. We expect a decrease in the variance of s as inconsistent fossils are removed, because fossils that are consistent with one another will have an overall lower variance in s. Since the F-test assumes that the data are drawn from a normally distributed population, and the SS_χ values prior to the removal of any fossil calibrations were not normally distributed (Kolmogorov-Smirnov test, $\mathrm{df} = 90$, $P < 0.001$), we log-transformed the SS_χ values, and normality could not be rejected (Kolmogorov-Smirnov test, $\mathrm{df} = 90$, $P = 0.117$). We performed the F-tests on these log-transformed values, allowing us to assess the significance in the change of variance before and after removing a particular fossil calibration. When a very inaccurate fossil is removed from the analysis, we expect a significant reduction in the variance in the squared differences between molecular and fossil age estimates (D_i) across all fossil calibrated nodes in the centrarchid phylogeny.

Determination of Optimal Smoothing Parameter: Fossil-Based Model Cross-Validation

Previous uses of penalized likelihood have relied on a cross-validation method to determine the optimal level of rate-smoothing, using the fit between the estimated number of substitutions on a given branch and the ability of the model to predict the number of substitutions when the branch is removed from the analysis (Sanderson 2002). In this study we use a different and novel strategy, fossil-based model cross-validation, which uses the consistent fossil ages to determine the optimal smoothing parameter value (Near and Sanderson 2004). The method uses a set of minimum and/or maximum age constraints and a single fixed minimal age. Sequentially, each individual node that is dated with a fossil constraint was relaxed and all ages and parameters were re-estimated, resulting in a new molecular age estimate for the node with the relaxed constraint. If this new molecular age was younger than the minimum, or older than the maximum fossil constraint, then the estimate was considered to violate the constraint and the node was given a score equal to the absolute value of the difference in age between the new mo-

lecular estimated age and the fossil constraint. If the new estimated age did not violate the minimum and/or maximum constraint, it was given a score equal to zero (Near and Sanderson 2004). Scores for each node with a fossil constraint were summed across the tree to obtain an overall cross-validation score, which in turn was calculated for a range of smoothing parameter values. The smoothing value that resulted in the lowest cross-validation score was selected for the given molecular dataset (Near and Sanderson 2004).

To objectively identify the single fossil that would be treated as a fixed minimum age estimate among the set of consistent fossils in the fossil-based model cross-validation analysis, we further scrutinized the consistent fossils with a jackknife analysis. To determine which fossil was the most consistent, relative to all other consistent calibrations, all but one calibration (i) were fixed as minimum age estimates in r8s, and the percent deviation ($D\%_i$) between the molecular and fossil age estimates at the ith calibration point was calculated as

$$D\%_i = \frac{|D_i|}{\mathrm{FA}_i}(100). \quad (5)$$

The calibration point with the lowest $D\%_i$ value was used as a fixed calibration point in fossil-based model cross-validation, and the other five calibrations were treated as minimum age constraints.

Results

DNA Sequences, Models, and Phylogenetic Inferences

Alignment of the nucleotide sites from the seven gene regions sequenced for this study resulted in a phylogenetic dataset comprising a total of 5553 aligned base pairs. Ten different models of sequence evolution were selected for the 14 designated data partitions, and only three models were shared among multiple data partitions (Table 2). The phylogenetic tree resulting from the pMM Bayesian analysis is similar to phylogenies produced from maximum parsimony and pMM Bayesian analyses of a dataset containing three of the seven gene regions used in this study (Near et al. 2004); however, our new phylogeny boasted a greater number of resolved nodes in both pMM Bayesian and maximum parsimony analysis. When compared to Near et al. (2004), our new phylogeny had three additional interspecific nodes supported with significant Bayesian posterior probabilities, and two additional nodes that scored high bootstrap pseudoreplicate values in maximum parsimony analysis (Fig. 1).

Despite the fact that the seven-gene dataset resulted in greater phylogenetic resolution and overall node support relative to other DNA studies of centrarchid phylogeny (Roe et al. 2002; Near et al. 2004; Harris et al. 2005), there are few topological differences between our new phylogeny and these earlier efforts. In fact, many of the precladistic hypotheses of centrarchid phylogenetic relationships were quite compatible with a mtDNA and nuclear gene phylogeny that is very similar to our new phylogeny (Near et al. 2004). This is an important point, as it indicates that the phylogeny of Centrarchidae used to examine fossil calibration of molecular phylogenies is not a radical departure from traditional views of centrarchid relationships (Bailey 1938; Branson and Moore 1962).

There were four major clades of centrarchids resolved in both the pMM Bayesian tree (Fig. 1) and the maximum parsimony trees (not shown). First, the monotypic lineage *Acantharchus pomotis* was the sister taxon of all other centrarchids. The second clade was a well-supported lineage containing the genera *Centrarchus*, *Enneacanthus*, *Pomoxis*, *Archoplites*, and *Ambloplites*. The two remaining clades (*Lepomis* and *Micropterus*) were sister lineages, and this relationship was supported with significant Bayesian posterior probabilities and

Table 2. 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[1]
Mitochondrial genes				
ND2 all sites	GTR	6	yes	gamma distributed
ND2 1st codon	K80	2	no	gamma distributed
ND2 2nd Codon	HKY85	2	yes	gamma distributed
ND2 3rd codon	GTR	6	yes	gamma distributed
tRNA	K80	2	yes	gamma distributed
16S rRNA all sites	GTR	6	yes	gamma distributed
16S rRNA unpaired sites	GTR	6	no	gamma distributed
16S rRNA paired sites	HKY85	6	yes	gamma distributed
Nuclear genes				
S7 ribosomal protein intron 1	HKY85	2	no	gamma distributed
Calmodulin intron 1	HKY85	2	no	gamma distributed
Tmo-4C4 all sites	K81	6	no	gamma distributed
Tmo-4C4 1st codon	F81	1	no	gamma distributed
Tmo-4C4 2nd codon	F81	1	no	equal
Tmo-4C4 3rd codon	K80	2	no	gamma distributed
Rhodopsin all sites	HKY85	2	yes	gamma distributed
Rhodopsin 1st codon	JC	1	yes	gamma distributed
Rhodopsin 2nd codon	F81	1	yes	gamma distributed
Rhodopsin 3rd codon	HKY85	2	no	gamma distributed

[1] Among-site rate variation.

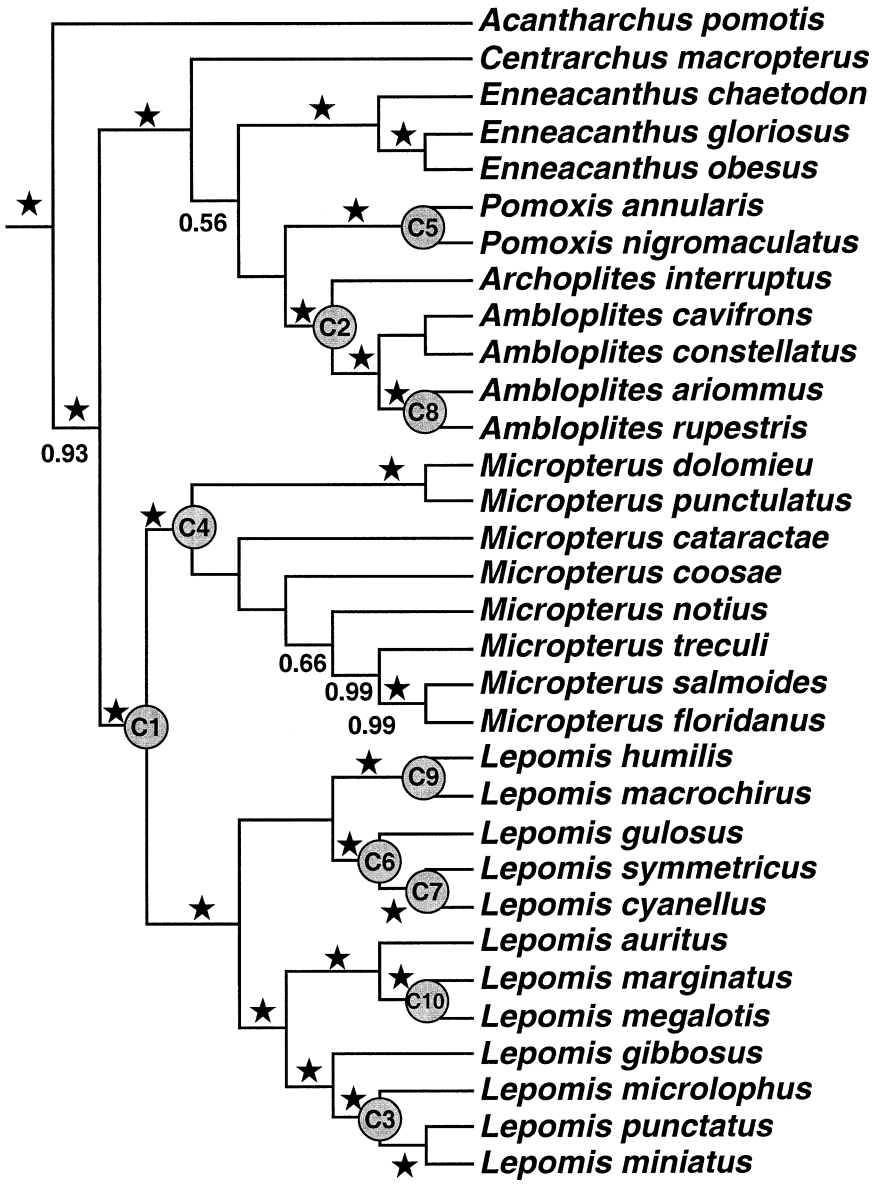

FIG. 1. A new phylogenetic tree of all 32 centrarchid species resulting from the partitioned mixed-model Bayesian analysis of the seven-gene dataset. Bayesian posterior probabilities for nodes less than 1.00 are indicated. Nodes that scored 80% or greater in maximum parsimony bootstrap analysis are marked with a star. Circled ''C'' numbers designate nodes dated with fossil information (Table 1).

a high (>80%) maximum parsimony bootstrap pseudoreplicate score (Fig. 1). Also worth noting was the fact that all polytypic genera of Centrarchidae (*Enneacanthus*, *Pomoxis*, *Ambloplites*, *Micropterus*, and *Lepomis*) were monophyletic in all phylogenetic analyses (Fig. 1).

Centrarchid Fossils and Calibration Points

Examination of the literature documenting the fossil record of North American freshwater fishes yielded 10 potential calibration points for the centrarchid molecular phylogeny (Table 1). Based on morphological identification of the fossils relative to extant species we were able to place these 10 fossils onto the centrarchid phylogeny (C1–C10, Fig. 1). The only exception to this strategy of fossil placement was the assignment of calibration point 4 (C4), which followed the recommendation of G. R. Smith (pers. obs.). In all instances, fossil calibrations provided minimum age estimates for the stem group relative to placement on the phylogeny (Doyle and Donoghue 1993; Magallón and Sanderson 2001). Fossil calibration points were relatively well dispersed on the centrarchid phylogeny, with no parts of the tree lacking fossil calibrations (Fig. 1). The ages of the 10 fossil calibrations ranged from 16.0 to 2.4 mya (Table 1).

Using equation (1), we performed gap analysis (Strauss and Sadler 1989; Marshall 1990a,b; Springer 1995) to estimate the 95% lower bound on the age estimate for five of the 10 nodes dated with fossil information (Table 1). We also estimated that the lower bound for the most recent common ancestor of all *Lepomis* species was 15.7 mya (Fig. 1), despite the fact that there was no specific fossil that could be assigned

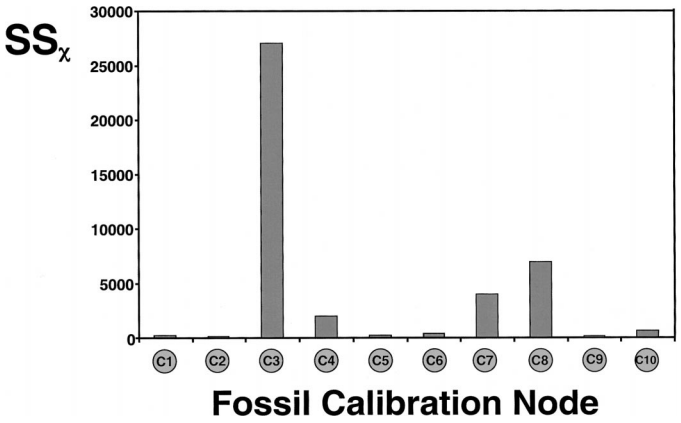

FIG. 2. Histogram of the disagreement between molecular and fossil age estimates, as measured by the sum of the squared differences between molecular and fossil age estimates (SS_χ) for a given fossil calibration when it was used as the single calibration point.

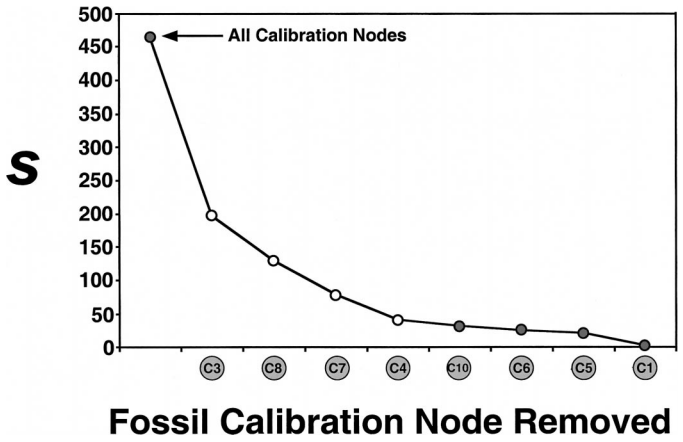

FIG. 3. Plot illustrating the effect of removing fossil calibration points on the overall disagreement between molecular and fossil age estimates (*s*). Open points indicate that the removal of that fossil calibration resulted in a significant reduction in the variance of the log-transformed differences between molecular and fossil-based estimates of node age. Fossil calibrations 2 and 9 are not on the figure, because they were the last two calibrations remaining after the removal of calibration 1.

to this node. The estimated 95% lower bound ages for calibrations C6 and C9 exceeded twice the value of the absolute minimum fossil age (Table 1), and three and five fossils, respectively, in the geological record represented these two nodes. The numbers of fossils for the three other nodes ranged from nine to 17, and the disparity between the absolute and 95% lower bound age estimates were considerably less than those calculated for C6 and C9 (Table 1).

Fossil Cross-validation

Cross-validation analysis indicated that using any one of several calibration points as a fixed minimum age in penalized likelihood analysis resulted in large deviations between fossil and molecular age estimates for all other relaxed fossil dated nodes (Fig. 2). We note that it may be helpful to determine the directionality in deviation between the fossil and molecular age estimates; however, none of the calibration points exhibited an appreciable negative mean deviation seen in other datasets (Near et al. 2005). The sequence of fossil removal from the cross-validation analysis was determined by ranking the calibration points based on the magnitude of its SS_χ score (Fig. 2). In Figure 3 we illustrate the effect of removing these fossils on the magnitude of *s*. Removal of the first four fossil calibrations resulted in a 91% decrease in *s*, and removal of the next four calibration points had virtually no effect on the magnitude of *s* (Fig. 3). Sequential one-tailed *F*-tests resulted in significant differences between the overall variance of *s* after the removal of C3, C8, C7, and C4; however, there were no significant *F*-test results after the subsequent removal of each remaining fossil calibration point (Fig. 3).

Based on the results of the fossil cross-validation and *F*-tests, we identified calibrations points C3, C8, C7, and C4 as inconsistent and removed them from the molecular dating analysis. To further explore the relative agreement among calibration points designated as inconsistent versus consistent, we plotted the *s* score and standard error values within each of these two categories of calibration points (Fig. 4). The consistent fossils identified in the cross-validation analysis had a significantly lower *s* score (Fig. 4) than the set of inconsistent fossil calibrations.

We explored a possible effect of proximity of fossil calibrations on the centrarchid phylogeny by plotting the percent deviation between molecular and fossil ages (D_i/FA_i) versus the number of nodes separating two calibrations on the phylogeny (Fig. 5). This plot shows that fossil calibrations that have a node distance of one exhibit very similar deviations as calibrations that are separated by 11 nodes on the phylogeny (Fig. 5). Based on the lack of relationship between deviation of molecular and fossil age estimates with node distance between fossil calibrations, and the fact that the final six selected fossils are well spaced on the phylogeny (Fig. 1), it does not appear that the proximity of fossils on the phylogeny is biasing the results of the fossil cross-validation analysis.

The mean percent deviation ($D\%_i$) of individual consistent fossil calibrations in the jackknife analyses ranged from 5.6% to 54.1% (Fig. 6). The calibration C2, *Archoplites clarki* (Table 1) had the lowest $D\%_i$ score; this calibration was fixed for the fossil-based model cross-validation, and subsequent penalized likelihood analyses.

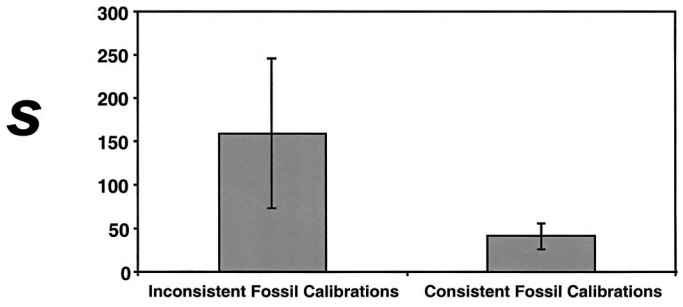

FIG. 4. Histogram of the overall disagreement between molecular and fossil ages (*s*) for inconsistent fossils and consistent fossils. The error bars are standard error.

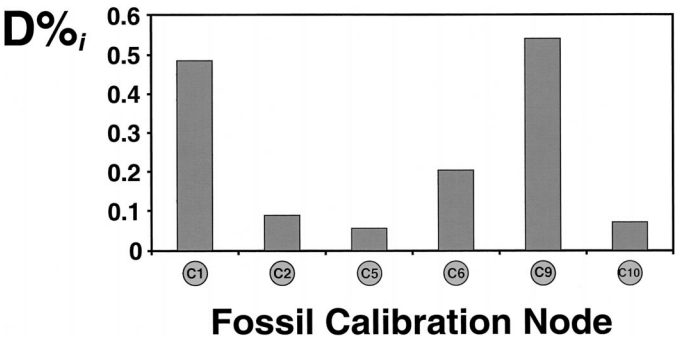

FIG. 5. Histogram of jackknife analysis of percent deviation between molecular and fossil age estimates ($D\%_i$) for consistent fossil calibrations.

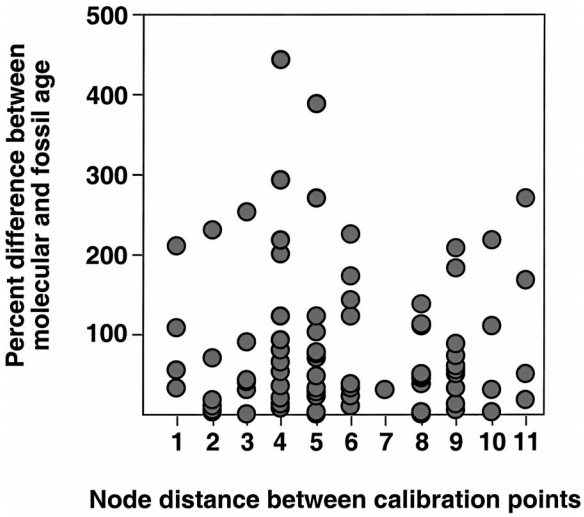

FIG. 6. Plot of percent deviation between fossil and molecular age estimates versus node distance between the two fossil calibrations on the phylogeny (Fig. 1).

Estimation of Divergence Times in Centrarchidae

We calibrated the centrarchid molecular phylogeny using a single fixed absolute minimum age (calibration C2; Table 1) and five minimum age constraints (calibrations C1, C5, C6, C9, and C10; Table 1). Fossil-based model cross-validation of the smoothing parameter for penalized likelihood analysis identified 0.10 as the optimal smoothing parameter value (Sanderson 2002; Near and Sanderson 2004). The results of the penalized likelihood analysis are presented in Table 3 and Figure 7. The confidence intervals estimated from the bootstrap and Bayesian methods were very similar for most nodes, with exceptions observed at the oldest of the nodes in the centrarchid chronogram (Fig. 7; Table 3). Nodes I (C8), N (C4), Y (C7), and AD (C4) were calibration points

TABLE 3. Estimated ages, bootstrap estimated confidence intervals, Bayesian confidence intervals, and fossil ages of inconsistent and minimal age constraint calibration nodes in the centrarchid chronogram (Fig. 7). Fixed and minimal age constraint calibrations are highlighted in bold. mya, millions of years ago.

Node	Age estimate (mya)	Bootstrap estimated confidence interval (mya)	Estimated from Bayesian credibility intervals (mya)	Difference between fossil and molecular age estimate (mya)
A	33.59	±3.58	±5.77	no fossil
B	28.94	±2.43	±3.28	no fossil
C	22.83	±2.80	±2.79	no fossil
D	21.60	±1.62	±2.35	no fossil
E	19.18	±4.23	±1.47	no fossil
C2	**15.50**	**fixed**	**fixed**	**fixed**
G	8.84	±1.20	±1.27	no fossil
H	7.34	±1.35	±1.35	no fossil
I	2.93	±0.63	±0.76	−1.04
C5	**12.00**	**±1.02**	**±1.56**	**0.00 (constrained)**
K	11.32	±1.59	±1.96	no fossil
L	4.03	±0.92	±1.00	no fossil
C1	**24.81**	**±2.43**	**±2.98**	**8.81 (constrained)**
N	8.40	±1.29	±1.53	−3.60
O	6.45	±1.33	±1.33	no fossil
P	5.93	±1.37	±1.35	no fossil
Q	5.10	±1.22	±1.23	no fossil
R	4.11	±1.06	±1.04	no fossil
S	2.84	±0.82	±0.80	no fossil
T	1.67	±0.45	±0.49	no fossil
U	14.64	±1.31	±1.61	no fossil
V	13.11	±1.29	±1.88	no fossil
C9	**5.17**	**±0.82**	**±0.96**	**1.84 (constrained)**
C6	**6.60**	**±0.41**	**±0.82**	**0.00 (constrained)**
Y	2.59	±0.43	±0.55	−1.31
Z	9.81	±1.06	±1.35	no fossil
AA	4.19	±0.57	±0.59	no fossil
C10	**2.72**	**±0.38**	**±0.82**	**0.32 (constrained)**
AC	7.84	±1.00	±1.25	no fossil
AD	3.46	±0.51	±0.32	10.04
AE	1.67	±0.41	±0.45	no fossil

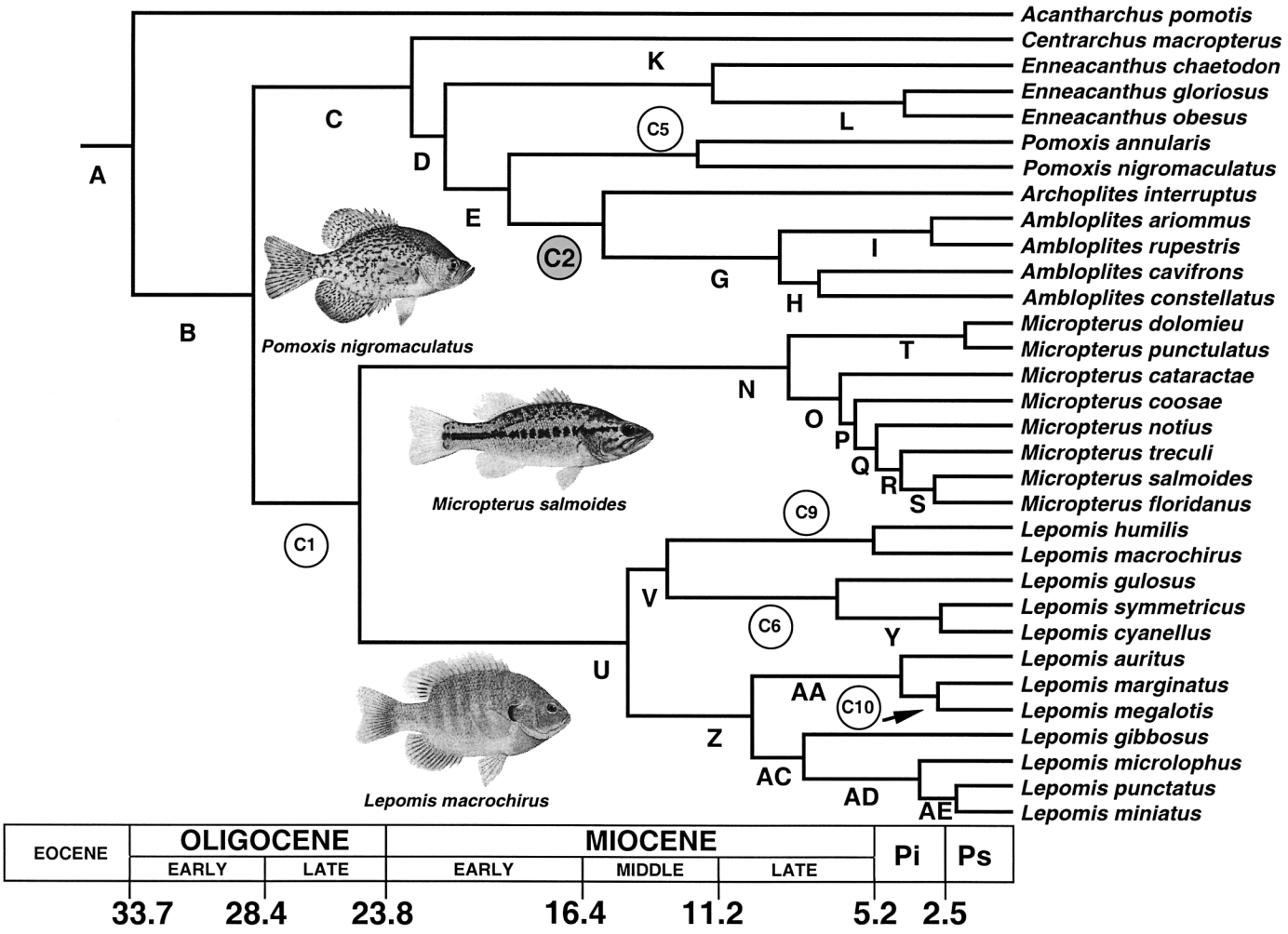

FIG. 7. Time-calibrated phylogeny (chronogram) of Centrarchidae based on molecular dating using a single calibration point as a fixed age (shaded circle), and five calibration points as minimal age constraints (open circles). The chronogram is calibrated against the geological time scale (Berggren et al. 1995). Exact age estimates for all nodes are in Table 3. Pi, Pliocene; Ps, Pleistocene. Fishes redrawn from Forbes and Richardson (1920).

identified as inconsistent and removed from the penalized likelihood analysis, and in all four cases the fossil ages of the inconsistent nodes fell outside the confidence intervals of the penalized likelihood age estimates (Table 3).

The age of the most recent common ancestor of Centrarchidae was 33.59 mya ± 3.58 (bootstrap estimate), ±5.77 (estimate from Bayesian credibility intervals). When separate data partitions were analyzed in penalized likelihood analyses, all partitions resulted in similar age estimates of the most recent common ancestor of Centrarchidae (Table 4). Of the three major polytypic centrarchid clades, the lineage comprising *Centrarchus*, *Enneacanthus*, *Pomoxis*, *Archoplites*, and *Ambloplites* was the oldest at 22.83 ± 2.80, ± 2.79. The crown group node age of *Lepomis* was 14.64 mya ± 1.31, ± 1.61, and the age of the most recent common ancestor of *Micropterus* was 8.40 mya ± 1.29, ± 1.53 (Fig. 7; Table 3). The youngest sister species pairs were *M. dolomieu* and *M. punctulatus* (1.67 mya ± 0.45, ± 0.49), and *L. punctatus* and *L. miniatus* (1.67 mya ± 0.41, ± 0.45). The oldest sister species pair was *Pomoxis annularis* and *P. nigromaculatus* (12.0 mya); however, the age at this node was constrained by a fossil calibration (C5). When this calibration was relaxed in the jackknife analysis the penalized likelihood estimate of this node's age (11.05 mya) was very similar to the fossil age. The oldest sister species pair whose most recent common ancestral node was not a calibration point was *Ambloplites constellatus* and *A. cavifrons* (7.34 mya ± 1.35, ± 1.35).

TABLE 4. The age of the most recent common ancestor of Centrarchidae resulting from penalized likelihood analysis of separate nucleotide data partitions. Ages given in millions of years ago.

Partition	Age
ND2 1st codon	39.17
ND2 2nd codon	31.47
ND2 3rd codon	36.02
tRNA	32.42
16S unpaired	38.18
S7	35.91
CaM	33.60
RH all positions	39.20
Tmo4C4 all positions	33.72

Discussion

Fossil Calibration and Molecular Dating in Centrarchidae

Fossil information is informative about divergence times but may be subject to substantial error (Lee 1999; Smith and Peterson 2002). Recent trends in molecular dating methods have emphasized the inclusion of multiple fossil calibration points with the goals of more precisely gauging the degree of molecular evolutionary rate heterogeneity and approaching greater accuracy in molecular age estimates (Springer et al. 2003; Yoder and Yang 2004). However, application of the fossil-cross validation method reveals that not all fossil calibrations are equally accurate and some fossils can introduce significant error to molecular dating analyses (Figs. 2 and 3). These conclusions are borne out in our analyses of the 10 fossils available to calibrate the centrarchid molecular phylogeny. We show that penalized likelihood analyses using individual fossils as calibration points result in different levels of agreement between molecular and fossil age estimates for all other nodes (Fig 2).

Perhaps the most interesting result in the application of the fossil cross-validation method to centrarchid fossils was the ability to reduce the overall disagreement between molecular and fossil age estimates as the apparently erroneous fossil calibrations are removed from the analysis (Fig. 3). It is important to point out that the fossil cross-validation method is not identifying different internally consistent sets of fossils. The amount of error among the four fossil calibrations identified as inconsistent was significantly greater than the error observed among the six fossil calibrations designated as consistent (Fig. 4). When all 10 fossil calibration points are included in the penalized likelihood analysis, regardless of their performance in the fossil cross-validation analysis, the estimated age of the most recent common ancestor of Centrarchidae (node A, Fig. 7) is 23.3 mya. This result is more than 10 million years less than the age estimated from penalized likelihood analysis using only the consistent fossil calibration points (Table 3, Fig. 7), and is grossly inconsistent with the ages of the earliest documented centrarchid fossils (Cavender 1986).

When closely examined, the four inconsistent fossils reveal two interesting patterns that deserve further scrutiny with regard to fossil calibrations in molecular dating analyses. First, the molecular age estimates for three of these four calibration nodes are younger than the fossil age estimate (Table 3), and two of these fossils are relatively young dates at apical nodes in the centrarchid phylogeny (Fig. 1, Table 1). In addition to problems of estimating ages at these apical nodes due to phenomena such as ancestral polymorphism (Edwards and Beerli 2000), one possible reason for the pattern that the fossils at these nodes are older than the molecular age estimates would be an undersampling of species in the molecular phylogeny. Even though we have sampled all 32 recognized centrarchid species, at least 25% of all centrarchid species are hypothesized to be polytypic, and currently recognized species may comprise undescribed cryptic species complexes (Mayden et al. 1992). The possibility of unrecognized cryptic centrarchid species has also been revealed with analyses of intraspecific mtDNA variation in three *Lepomis* species, where the age of the intraspecific most recent common ancestor ranged between 2.6 and 4.3 mya (Bermingham and Avise 1986). If there are unrecognized cryptic species along branches calibrated by fossils C7 and C8, we could be assigning fossils to nodes that are too young, since the older nodes involving possible cryptic lineages were not sampled in our phylogenetic analyses. Second, there is the possibility of lineage extinction resulting in an erroneous assignment of a fossil calibration point to a node in the molecular phylogeny. The fossil calibration C4 may be an example (Fig. 1, Table 1). Based on morphological examination (G. R. Smith, pers. obs.), this fossil shares characters with the sister species pair *Micropterus dolomieu* and *M. punctulatus* (node T, Fig. 7). Because fossils date the stem group of a lineage, this fossil date was assigned to node N in Figure 7, which is the most recent common ancestor node of all *Micropterus* species. The length in absolute age between nodes N and T is 6.73 mya, and this is among the longest internode lengths in the centrarchid chronogram (Fig. 7). If there are missing nodes between nodes C1 and N that represent extinct centrarchid lineages, and characters used to identify fossil C4 to the stem lineage of node T are actually symplesiomorphic in *Micropterus,* then the actual assignment of the fossil C4 should be at some point between nodes C1 and N. Assigning C4 to node N forces the age to be substantially older than the ages estimated for this node using other consistent fossils (Table 3, Fig. 7)

There are several ways that a fossil calibration can introduce error in molecular divergence time estimates and these include underestimation of the fossil age due to incompleteness of the fossil record, assignment of the fossil to the wrong node in a phylogeny, and error in dating fossil-bearing geological formations. One benefit of the fossil cross-validation approach used here is that it represents a method to identify fossils that are inconsistent within a broader sample of calibration points, or at least it identifies fossils that should elicit further scrutiny (Near et al. 2005). Our results in Centrarchidae are encouraging and reflect similar patterns of fossil calibration error as revealed by fossil cross-validation in various disparate lineages such as mammals, monocot angiosperms, and turtles (Near et al. 2005; Near and Sanderson 2004).

Centrarchid Chronograms and the Evolution of the North American Freshwater Fish Fauna

The origin of the North American freshwater fish fauna has previously been investigated from the perspective of the fossil record, and there is a marked period of extinction as well as first appearance of lineages at the Eocene-Oligocene transition (Cavender 1986; Wilson and Williams 1992). In our analyses, penalized likelihood analysis results in an age of 33.59 mya $\pm$ 3.58, $\pm$ 5.77, for the most recent common ancestor of living Centrarchidae (Fig. 7, Table 3), and this is similar to the age of the first appearance of centrarchids in the fossil record (Cavender 1986). Unfortunately, these oldest centrarchid fossils have never been formally described taxonomically, thus preventing them from contributing calibration points for the centrarchid molecular phylogeny. If any of these undescribed centrarchid fossils are older than 35 million years, and if phylogenetically they are assigned

as stem lineages below the centrarchid crown group, then the age of centrarchids may actually be older than the estimate from the molecular based penalized likelihood analysis.

With regard to higher relationships of teleost fishes, Centrarchidae is thought to represent a basal lineage of Perciformes (Gosline 1966; Nelson 1994), and with more than 9000 species and 145 taxonomic families, Perciformes is the largest taxonomic order of vertebrates (Nelson 1994). Perciformes is noteworthy not only for its rich species diversity, but also for a marked sudden appearance in the fossil record near the Cretaceous-Tertiary Boundary approximately 65 mya (Patterson 1993). Since the oldest perciform fossils date to 65 mya, the molecular age estimate of 33.59 mya for the most recent common ancestor indicates that Centrarchidae is a relatively young perciform clade. This result either means that Centrarchidae are unlikely to be a basal perciform lineage or it indicates that the group existed for some time before the common ancestor of the extant species.

Previous molecular age estimates of centrarchids are similar to the results from our penalized likelihood analysis (Fig. 7, Table 3). For example, Bermingham and Avise (1986) estimated intraspecific coalescent ages for three *Lepomis* species (*L. microlophus*, *L. gulosus*, and *L. macrochirus*) and these ages are all younger than the age of the node relating these species and their sister species in the chronogram (Fig. 7). A previous analysis using a Langley-Fitch method to estimate divergence times in *Micropterus* with two fossil calibration points and two mtDNA protein coding genes estimated that the age of most recent common ancestor of *Micropterus* was 11.17 mya ± 0.63 (Near et al. 2003a). The age of this node in our analyses was 8.40 ± 1.29, ± 1.53, and the near overlap in the error estimates between these two ages, as well as similar age estimates for the internal nodes in the *Micropterus* phylogeny indicate a close similarity in the results of these two analyses, despite using different fossil calibrations, analytical methods, and molecular datasets. In addition, when using the six consistent calibration points, we observed agreement between fossil and molecular ages for nodes without specific fossils to calibrate the molecular phylogeny. For example, there was close agreement between the molecular age estimate (14.64 mya ± 1.31, ± 1.61) and the fossil gap analysis lower bound age (15.7 mya) for the most recent common ancestor of *Lepomis* (node U; Tables 2 and 4, Fig. 7).

The timing of diversification in Centrarchidae, and the appearance in the fossil record at, or near, this time in most of the diverse extant clades of North American freshwater fishes, indicates that the Eocene-Oligocene transition was a critical time in the origin of this diverse fauna. The early portion of the Eocene was the warmest part of the early Cenozoic (Woodburne 2004b), and at this time the interior of North America was dominated by warm temperate, subtropical, and tropical forests (Graham 1999). The Late Eocene was characterized by a dramatic cooling trend that culminated in the formation of major Antarctic glaciations during the transition between the Eocene and the Early Oligocene (Zachos et al. 2001; Woodburne 2004b). The Eocene-Oligocene climatic transition impacted both terrestrial and marine habitats, as evidenced by significant extinction and taxonomic turnover in such disparate groups of organisms as benthic foraminifera (Thomas 1992), echinoderms (McKinney et al. 1992), cetaceans (Fordyce 1992), and terrestrial plants (Graham 1999). It is compelling that the pattern of extinction and lineage origination seen in the North American freshwater fish fossil record (Cavender 1986; Wilson and Williams 1992), and the molecular date for the most recent common ancestor of Centrarchidae, seems to have occurred at a time of global climate change and strong signatures of extinction across the tree of life. This pattern supports the idea that the origin of the most diverse lineages of North American freshwater fishes may have resulted from the significant climate changes that have been invoked to explain patterns of extinction and lineage origination of other organismal lineages during the Eocene-Oligocene transition (Berggren and Prothero 1992).

The strategies and methods used in this study to estimate divergence times of Centrarchidae provide a unique opportunity to calibrate molecular phylogenies in other North American freshwater fish clades that have rich fossil records (i.e., Cyprinidae, Catostomidae, and Ictaluridae). Once robust hypotheses regarding absolute ages for several of these clades are available, we can attempt to reconstruct the timing of origination and diversification of an entire continental freshwater fish fauna. This phylogenetic and temporal data could provide invaluable information to investigate, at an unprecedented scale, the evolution and historical structuring of the complex communities of fishes occupying freshwater habitats in North America (Mayden 1987; Webb et al. 2002).

Acknowledgments

We are deeply grateful to H. B. Shaffer and M. J. Sanderson with whom the methods of fossil cross-validation and fossil-based model cross-validation presented in this paper were developed. Also, the vast majority of the DNA sequencing work was carried out in H. B. Shaffer's laboratory. G. R. Smith graciously shared valuable information about centrarchid fossils. G. B. Pauly, M. K. Fujita, and C. D. Hulsey provided laboratory assistance. We received excellent comments on the manuscript from J. A. Fordyce, P. R. Hollingsworth, C. D. Hulsey, and A. M. Simons. TJN was supported with a postdoctoral fellowship from the Center for Population Biology at the University of California, Davis (UCD), and the H. H. Ross and P. W. Smith Memorial Funds from the Illinois Natural History Survey provided initial research funds. DIB was supported by a National Science Foundation (NSF) graduate research fellowship, scholarships from UCD and the ARCS foundation, and postdoctoral support from NSF DEB 0089716 to M. Turelli and J. Gillespie. Additional funding was provided by NSF IBN-0076436 to PCW.

Literature Cited

Bailey, R. M. 1938. A systematic revision of the centrarchid fishes with a discussion of their distribution, variations, and probable interrelationships. University of Michigan, Ann Arbor.

Baldwin, B. G., and M. J. Sanderson. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Natl. Acad. Sci. USA 95:9402–9406.

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

Berggren, W. A., and D. R. Prothero. 1992. Eocene-Oligocene climatic and biotic evolution: an overview. Pp. 1–28 *in* D. R. Proth-

ero and W. A. Berggren, eds. Eocene-Oligocene climatic and biotic evolution. Princeton Univ. Press, Princeton, NJ.
Berggren, W. A., D. V. Kent, I. C. C. Swisher, and M.-P. Aubry. 1995. A revised Cenozoic geochronology and chronostratigraphy. Pp. 129–212 *in* W. A. Berggren, D. V. Kent, M.-P. Aubry and J. Hardenbol, eds. Geochronology, time scales and global stratigraphic correlation. Society for Economic Paleontology and Mineralogy, Tulsa, OK.
Bermingham, E., and J. C. Avise. 1986. Molecular zoogeography of freshwater fishes in the southeastern United States. Genetics 113:939–965.
Bollback, J. P. 2001. CodonBootstrap, version 2.1. Available via http://brahms.ucsd.edu/jon_bollback.html.
Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution 59: 1754–1767.
Branson, B. A., and G. A. Moore. 1962. The lateralis components of the acoustico-lateralis system in the sunfish family Centrarchidae. Copeia 1962:1–108.
Briggs, J. C. 1986. Introduction to the zoogeography of North American fishes. Pp. 1–16 *in* C. H. Hocutt and E. O. Wiley, eds. The zoogeography of North American freshwater fishes. John Wiley & Sons, New York.
Britten, R. 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science 231:1393–1398.
Cavender, T. M. 1986. Review of the fossil history of North American freshwater fishes. Pp. 699–724 *in* C. H. Hocutt and E. O. Wiley, eds. The zoogeography of North American freshwater fishes. John Wiley & Sons, New York.
Chow, S., and K. Hazama. 1998. Universal PCR primers for S7 ribosomal protein genes in fish. Mol. Ecol. 7:1255–1256.
Chow, S., and H. Takeyama. 2000. Nuclear and mitochondrial DNA analyses reveal four genetically separated breeding units of the swordfish. J. Fish. Biol. 56:1087–1098.
Conroy, C. J., and M. van Tuinen. 2003. Extracting time from phylogenies: positive interplay between fossil and genetic data. J. Mammal. 84:444–455.
De Rijk, P., J. Wuyts, Y. V. d. Peer, T. Winkelmans, and R. D. Wachter. 2000. The European large subunit ribosomal RNA database. Nucleic Acids Res. 28:177–178.
Dowling, T. E., C. A. Tibbets, W. L. Minckley, and G. R. Smith. 2002. Evolutionary relationships of the plagopterins (Teleostei: Cyprinidae) from cytochrome *b* sequences. Copeia 2002:665–678.
Doyle, J. A., and M. J. Donoghue. 1993. Phylogenies and angiosperm diversification. Paleobiology 19:141–167.
Edwards, S. V., and P. Beerli. 2000. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54:1839–1854.
Etnier, D. A., and W. C. Starnes. 1993. The fishes of Tennessee. University of Tennessee Press, Knoxville.
Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368–376.
Foote, M., J. P. Hunter, C. M. Janis, and J. J. Sepkoski. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. Science 283:1310–1314.
Forbes, S. A., and R. E. Richardson. 1920. The fishes of Illinois. Illinois Natural History Survey, Springfield, IL.
Fordyce, R. E. 1992. Cetacean evolution and Eocene/Oligocene environments. Pp. 368–381 *in* D. R. Prothero and W. A. Berggren, eds. Eocene-Oligocene climatic and biotic evolution. Princeton Univ. Press, Princeton, NJ.
Golenberg, E. M., D. E. Giannasi, M. T. Clegg, C. J. Smiley, M. Durbin, D. Henderson, and G. Zurawski. 1990. Chloroplast DNA sequence from a Miocene *Magnolia* species. Nature 344: 656–658.
Gosline, W. A. 1966. The limits of the fish family Serranidae, with notes on other lower percoids. Proc. Calif. Acad. Sci. 33:91–112.
Graham, A. 1999. Late Cretaceous and Cenozoic history of North American vegetation. Oxford Univ. Press, Oxford, U.K.
Grande, L., and W. E. Bemis. 1991. Osteology and phylogenetic relationships of fossil and recent paddlefishes (Polyodontidae) with comments on the interrelationships of Acipenseriformes. J. Vertebr. Paleontol. 11(Mem. 1):1–121.
———. 1998. A comprehensive phylogenetic study of amiid fishes (Amiidae) based on comparative skeletal anatomy: an empirical search for interconnected patterns of natural history. J. Vertebr. Paleontol. 18(Mem. 4):1–690.
Gutell, R. R., and G. E. Fox. 1988. A compilation of large subunit RNA sequences presented in structural format. Nucleic Acids Res. 16:r175–r225.
Gutell, R. R., M. W. Gray, and M. N. Schnare. 1993. A compilation of large subunit (23S and 23S-like) ribosomal RNA structures. Nucleic Acids Res. 21:3055–3074.
Harris, P. M., K. J. Roe, and R. L. Mayden. 2005. A mitochondrial DNA perspective on the molecular systematics of the sunfish genus *Lepomis* (Actinopterygii: Centrarchidae). Copeia 2005: 340–346.
Hibbard, C. W. 1936. Two new sunfish of the family Centrarchidae from middle Pliocene Kansas. Univ. Kans. Sci. Bull. 24: 177–185.
Huelsenbeck, J. P., B. Larget, and D. Swofford. 2000. A compound Poisson process for relaxing the molecular clock. Genetics 154: 1879–1892.
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.
Koster, W. J. 1969. Fishes of the Rita Blanca Lake deposits. Pp. 135–139 *in* R. Y. Anderson and D. W. Kirkland, eds. Paleoecology of an Early Pleistocene lake on the High Plains of Texas. Geological Society of America, Boulder, CO.
Larget, B., and D. L. Simon. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16:750–759.
Lee, M. S. Y. 1999. Molecular clock calibrations and metazoan divergence dates. J. Mol. Evol. 49:385–391.
Lindsay, E. H., N. M. Johnson, and N. D. Opdyke. 1975. Preliminary correlation of North American land mammal ages and geomagnetic chronology. Pap. Paleontol. Mus. Univ. Mich. 12:111–119.
Lundberg, J. G. 1975. Fossil catfishes of North America. Pap. Paleontol. Mus. Univ. Mich. 11:1–51.
———. 1992. The phylogeny of ictalurid catfishes: a synthesis of recent work. Pp. 392–420 *in* R. L. Mayden, ed. Systematics, historical ecology, and North American freshwater fishes. Stanford Univ. Press, Stanford, CA.
Lundberg, J. G., M. Kottelat, G. R. Smith, M. L. J. Stiassny, and A. C. Gill. 2000. So many fishes, so little time: an overview of recent ichthyological discovery in continental waters. Ann. MO. Bot. Gard. 87:26–62.
Magallón, S., and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762–1780.
Marshall, C. R. 1990a. The fossil record and estimating divergence times between lineages: maximum divergence times and the importance of reliable phylogenies. J. Mol. Evol. 30:400–408.
———. 1990b. Confidence intervals on stratigraphic ranges. Paleobiology 16:1–10.
Matthew, W. D. 1924. Third contribution to the Snake Creek fauna. Bull. Am. Mus. Nat. Hist. 50:59–210.
Mayden, R. L. 1987. Historical ecology and North American highland fishes: a research program in community ecology. Pp. 210–222 *in* W. J. Matthews and D. C. Heins, eds. Community and evolutionary ecology of North American stream fishes. Univ. of Oklahoma Press, Norman, OK.
Mayden, R. L., B. M. Burr, L. M. Page, and R. R. Miller. 1992. The native freshwater fishes of North America. Pp. 827–863 *in* R. L. Mayden, ed. Systematics, historical ecology, and North American freshwater fishes. Stanford Univ. Press, Stanford, CA.
McKinney, M. L., K. J. McNamara, B. D. Carter, and S. K. Donovan. 1992. Evolution of paleogene echinoids: a global and regional view. Pp. 349–367 *in* D. R. Prothero and W. A. Berggren, eds. Eocene-Oligocene climatic and biotic evolution. Princeton Univ. Press, Princeton, NJ.
Near, T. J., and M. F. Benard. 2004. Rapid allopatric speciation in logperch darters (Percidae: *Percina*). Evolution 58:2798–2808.
Near, T. J., and M. J. Sanderson. 2004. Assessing the quality of

molecular divergence time estimates by fossil calibrations and fossil-based model selection. Philos. Trans. R. Soc. B 359: 1477–1483.
Near, T. J., D. I. Bolnick, and P. C. Wainwright. 2004. Investigating phylogenetic relationships of sunfishes and black basses (Actinopterygii: Centrarchidae) using DNA sequences from mitochondrial and nuclear genes. Mol. Phylogenet. Evol. 32: 344–357.
Near, T. J., T. W. Kassler, J. B. Koppelman, C. B. Dillman, and D. P. Philipp. 2003a. Speciation in North American black basses, *Micropterus* (Actinopterygii: Centrarchidae). Evolution 57: 1610–1621.
Near, T. J., J. J. Pesavento, and C.-H. C. Cheng. 2003b. Mitochondrial DNA, morphology and the phylogenetic relationships of Antarctic icefishes (Notothenioidei: Channichthyidae). Mol. Phylogenet. Evol. 28:87–98.
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. Am. Nat. 165:137–146.
Nelson, J. S. 1994. Fishes of the world. Wiley, New York.
Norman, J. E., and M. V. Ashley. 2000. Phylogenetics of perissodactyla and tests of the molecular clock. J. Mol. Evol. 50: 11–21.
Passey, B. H., T. E. Cerling, M. E. Perkins, M. R. Voorhies, J. M. Harris, and S. T. Tucker. 2002. Environmental change in the Great Plains: an isotopic record from fossil horses. J. Geol. 110: 123–140.
Patterson, C. 1981. The development of the North American fish fauna: a problem of historical biogeography. Pp. 265–281 *in* P. L. Forey, ed. The evolving biosphere. Cambridge Univ. Press, Cambridge, U.K.
———. 1993. An overview of the early fossil record of acanthomorphs. Bull. Mar. Sci. 52:29–59.
Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818.
Renner, S. S., G. Clausing, and K. Meyer. 2001. Historical biogeography of Melastomataceae: the roles of tertiary migration and long-distance dispersal. Am. J. Bot. 88:1290–1300.
Repenning, C. A. 1987. Biochronology of the microtine rodents of the United States. Pp. 236–268 *in* M. O. Woodburne, ed. Cenozoic mammals of North America. Univ. of California Press, Berkeley, CA.
Roe, K. J., P. M. Harris, and R. L. Mayden. 2002. Phylogenetic relationships of the genera of North American sunfishes and basses (Percoidei: Centrarchidae) as evidenced by the mitochondrial cytochrome *b* gene. Copeia 2002:897–905.
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. Mol. Biol. Evol. 14:1218–1231.
———. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 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. Am. J. Bot. 88:1499–1516.
Smith, A. B., and K. J. Peterson. 2002. Dating the time of origin of major clades. Annu. Rev. Earth Planet. Sci. 30:65–88.
Smith, C. L. 1962. Some Pliocene fishes from Kansas, Oklahoma, and Nebraska. Copeia 1962:505–520.
Smith, G. R. 1981. Late Cenozoic freshwater fishes of North America. Annu. Rev. Ecol. Syst. 12:163–193.
———. 1992a. Introgression in fishes: significance for paleontology, cladistics, and evolutionary rates. Syst. Biol. 41:41–57.
———. 1992b. Phylogeny and biogeography of the Catostomidae, freshwater fishes of North America and Asia. Pp. 778–826 *in* R. L. Mayden, ed. Systematics, historical ecology, and North American freshwater fishes. Stanford Univ. Press, Stanford, CA.
Smith, G. R., and J. G. Lundberg. 1972. The Sand Draw fish fauna. Pp. 40–54 *in* M. F. Skinner and C. W. Hibbard, eds. Pleistocene preglacial and glacial rocks and faunas of North Central Nebraska. American Museum of Natural History, New York.
Smith, G. R., and R. R. Miller. 1985. Taxonomy of fishes from Miocene Clarkia Lake beds, Idaho. Pp. 75–83 *in* C. J. Smiley, ed. Late Cenozoic history of the Pacific Northwest. American Association for the Advancement of Science, San Francisco, CA.
Smith, G. R., T. E. Dowling, K. W. Gobalet, T. Lugaski, D. K. Shiozawa, and R. P. Evans. 2002. Biogeography and timing of evolutionary events among Great Basin fishes. Pp. 175–234 *in* R. Hershler, D. B. Madsen and D. R. Currey, eds. Great Basin aquatic systems history. Smithsonian Inst. Press, Washington, DC.
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. Proc. Natl. Acad. Sci. USA 99:4430–4435.
Springer, M. S. 1995. Molecular clocks and the incompleteness of the fossil record. J. Mol. Evol. 41:531–538.
Springer, M. S., W. J. Murphy, E. Eizirik, and S. J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc. Natl. Acad. Sci. USA 100:1056–1061.
Stearley, R. F., and G. R. Smith. 1993. Phylogeny of the Pacific trout and salmon (*Oncorhynchus*) and genera of the family Salmonidae. Trans. Am. Fish. Soc. 122:1–33.
Strauss, D., and P. M. Sadler. 1989. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. Math. Geol. 21:411–427.
Streelman, J. T., and S. A. Karl. 1997. Reconstructing labroid evolution with single-copy nuclear DNA. Proc. R. Soc. Lond. B 264:1011–1020.
Swofford, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.
Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic inference. Pp. 407–514 *in* D. M. Hillis, C. Moritz, and B. K. Mable, eds. Molecular systematics. Sinauer, Sunderland, MA.
Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. Mol. Biol. Evol. 12: 823–833.
Tedford, R. H., M. F. Skinner, R. W. Fields, J. M. Rensberger, D. P. Whistler, T. Galusha, B. E. Taylor, J. R. Macdonald, and S. D. Webb. 1987. Faunal succession and the biochronology of the Arikareean through Hemphillian interval (Late Oligocene through earliest Pliocene epochs) in North America. Pp. 153–210 *in* M. O. Woodburne, ed. Cenozoic mammals of North America. Univ. of California Press, Berkeley.
Thomas, E. 1992. Middle Eocene-Late Oligocene bathyal benthic foraminifera (Weddell Sea): faunal changes and implications for ocean circulation. Pp. 245–271 *in* D. R. Prothero and W. A. Berggren, eds. Eocene-Oligocene climatic and biotic evolution. Princeton Univ. Press, Princeton, NJ.
Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 22:4673–4680.
Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15:1647–1657.
Uyeno, T., and R. R. Miller. 1963. Summary of Late Cenozoic freshwater fish records for North America. Occ. Pap. Mus. Zool. Univ. Mich. 631:1–34.
van Tuinen, M., and G. J. Dyke. 2003. Calibration of galliform molecular clocks using multiple fossils and genetic partitions. Mol. Phylogenet. Evol. 30:74–86.
Wainwright, P. C., and G. V. Lauder. 1992. The evolution of feeding biology in sunfishes (Centrarchidae). Pp. 472–491 in R. L. Mayden, ed. Systematics, historical ecology, and North American freshwater fishes. Stanford Univ. Press, Stanford, CA.
Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue.

2002. Phylogenies and community ecology. Annu. Rev. Ecol. Syst. 33:475–505.

Wiley, E. O. 1976. The phylogeny and biogeography of fossil and recent gars (Actinopterygii: Lepisosteidae). Univ. Kans. Mus. Nat. Hist. Misc. Publ. 64:1–111.

Wilson, M. V. H. 1992. Importance for phylogeny of single and multiple stem-group fossil species with examples from freshwater fishes. Syst. Biol. 41:462–470.

Wilson, M. V. H., and R. R. G. Williams. 1992. Phylogenetic, biogeographic, and ecological significance of early fossil records of North American freshwater teleostean fishes. Pp. 224–244 *in* R. L. Mayden, ed. Systematics, historical ecology, and North American freshwater fishes. Stanford Univ. Press, Stanford, CA.

Wilson, R. L. 1968. Systematics and faunal analysis of a Lower Pliocene vertebrate assemblage from Trego County, Kans. Contrib. Mus. Paleontol. Univ. Mich. 22:75–126.

Woodburne, M. O. 2004a. Late Cretaceous and Cenozoic mammals of North America. Columbia Univ. Press, New York.

———. 2004b. Global events and the North American mammalian biochronology. Pp. 315–343 *in* M. O. Woodburne, ed. Late Cretaceous and Cenozoic mammals of North America. Columbia Univ. Press, New York.

Yoder, A. D., and Z. Yang. 2000. Estimation of primate speciation dates using local molecular clocks. Mol. Biol. Evol. 17: 1081–1090.

———. 2004. Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context. Mol. Ecol. 13:757–773.

Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292:686–693.

Zar, J. H. 1984. Biostatistical analysis. Prentice-Hall, Englewood Cliffs, NJ.

Zuckerkandl, E., and L. Pauling. 1962. Molecular disease, evolution, and genic heterogeneity. Pp. 189–225 *in* M. Kasha and B. Pullman, eds. Horizons in biochemistry. Academic Press, New York.

———. 1965. Evolutionary divergence and convergence in proteins. Pp. 97–166 *in* V. Bryson and H. J. Vogel, eds. Evolving genes and proteins. Academic Press, New York.

Corresponding Editor: A. Yoder