**Systematic Biology Advance Access published May 28, 2009**


DOI:10.1093/sysbio/syp005

# Conflicting Mitochondrial and Nuclear Phylogenies for the Widely Disjunct *Emys* (Testudines: Emydidae) Species Complex, and What They Tell Us about Biogeography and Hybridization

PHILLIP Q. SPINKS* AND H. BRADLEY SHAFFER

*Department of Evolution and Ecology and Center for Population Biology, University of California, Davis, CA 95616, USA; *Correspondence to be sent to: Department of Evolution and Ecology, University of California, Davis, CA 95616, USA; E-mail: pqspinks@ucdavis.edu.*

*Abstract.*—Understanding the mechanisms by which widely disjunct members of a clade came to occupy their current distribution is one of the fundamental challenges of biogeography. Here, we used data from 7 nuclear and 1 mitochondrial gene to examine the phylogenetic and biogeographic history of *Emys*, a clade of turtles that is broadly disjunct in western and eastern North America and Europe. We found strong disagreement between mitochondrial and nuclear gene trees, with mitochondrial DNA supporting the monophyly of the North American taxa (*marmorata* + *blandingii*) to the exclusion of the European *orbicularis*, and nuclear genes supporting the monophyly of (*blandingii* + *orbicularis*) to the exclusion of *marmorata*. We used fossil-calibrated molecular chronograms, in combination with supporting evidence from the fossil record and paleoclimatology, to identify a potential example of ancient hybridization and mitochondrial gene capture 12 million years ago, which explains this discrepancy. Based on the weight of evidence, we argue that the invasion of Eurasia by *Emys orbicularis* occurred about 16 Ma via a trans-Beringian land bridge. The case of *Emys* emphasizes how single-gene trees can be strongly affected by population processes, including hybridization, and that the effects of these processes can persist through long periods of evolutionary history. Given the chaotic state of the current taxonomy of these turtles, our work also emphasizes the care that should be used in implementing taxonomic changes based on 1 or a few gene trees and the importance of taking a conservative approach in renaming or splitting higher taxa based on apparent nonmonophyly. [*Actinemys*; *Clemmys*; *Emydoidea*; *Emys*; *HNF-1α*; nuclear phylogeny; *R35; RELN*.]

Over the last 3 decades, the majority of vertebrate phylogenetic and phylogeographic analyses have been based almost exclusively on mitochondrial DNA (mtDNA). The reasons for this reliance reflect both the biology of the molecule (mtDNA is nonrecombining and has a rapid coalescent time; Avise et al. 1987; Moritz et al. 1987) and the more pragmatic concerns that nuclear DNA (nuDNA) markers have been available for only a handful of model organisms. Although the general utility of mtDNA for phylogenetic and phylogeographic analyses is well established, a growing body of evidence suggests that inferences based on any individual gene, and mtDNA in particular, should be tempered with caution (Funk and Omland 2003; Chan and Levin 2005). Incongruence between mtDNA- and nuDNA-based phylogenetic reconstructions is commonly observed (Ferris et al. 1983; Shaw 2002; Leaché and McGuire 2006; Robertson et al. 2006; Peters et al. 2007; Spinks and Shaffer 2007; Good et al. 2008) and has been attributed to mechanisms including low mutation rate of the nuDNA sequences, incomplete lineage sorting, introgression/hybridization, and natural selection (Avise et al. 1987; Rand 2001; Sanderson and Shaffer 2002; Funk and Omland 2003; Ballard and Whitlock 2004; Ballard and Rand 2005). Ultimately, the mitochondrial genome is a single, large genetic locus that can provide but a single perspective on the evolutionary history of a group (Zhang and Hewitt 2003; Ballard and Whitlock 2004). Thus, mtDNA alone is often inadequate for phylogeographic/phylogenetic analyses, especially in the face of complex evolutionary scenarios including introgression, hybridization, and/or selection (Chan and Levin 2005).

When mitochondrial–nuclear incongruence has been discussed in the literature, it is generally in the context of recent species radiations because it is here where the long waiting time for multigene monophyly has been emphasized (Hudson and Coyne 2002). For example, a recent survey (Funk and Omland 2003) reported paraphyly or polyphyly in 23% of 2319 species-level mtDNA analyses, and this is probably a conservative estimate of how often inferences based on mtDNA are at odds with species history. A number of authors have reported an excess of replacement:silent substitutions in mtDNA, providing strong evidence of selection on mitochondria or mitochondrial cellular interactions (reviewed in Gerber et al. 2001). Discriminating selection from alternative forces shaping nucleotide composition is difficult (Gerber et al. 2001), but in many instances, mtDNA is probably not selectively neutral (Ballard and Kreitman 1995; Gerber et al. 2001; Rand 2001; Ballard and Whitlock 2004; Ballard and Rand 2005).

Although the potential effects of selection and lineage sorting on relatively recent (<1 Ma) phylogeographic/phylogenetic reconstruction seem clear, there has been far less focus on mitochondrial–nuclear incongruence for more ancient phylogeny problems. In principle, if lineage sorting is always complete prior to speciation and if no insidious problems of introgression persist, all gene trees should eventually come to reflect species trees and any single gene should reflect phylogeny accurately. However, the time required for all gene trees to reflect any given species tree depends on both within-species historical demography (Pamilo and Nei 1988; Degnan and Rosenberg 2006) and the species birth–death process (Nee et al. 1994). In practice,

hybridization and incomplete lineage sorting in the distant past can persist into phylogeny reconstruction in the present, particularly if extinction has not eliminated problematic lineages. Discriminating between lineage sorting, ancient hybridization, and more recent hybridization as causes of gene tree conflict can be particularly challenging and requires simultaneous analysis of mitochondrial and multiple nuclear markers (Peters et al. 2007; Zhang and Sota 2007; Good et al. 2008).

Here, we focus on the *Emys* species complex (Testudines: Emydidae), a group of 4 species of freshwater aquatic turtles that has been the subject of several phylogenetic analyses and taxonomic revisions over the last decade (Fig. 1). We were originally drawn to this system for biogeographic, morphological, and phylogenetic reasons, and our pursuit of these basic questions revealed a deeper issue in gene tree/species tree conflict. Biogeographically, these turtles have a bizarre distribution, with widely allopatric taxa in western and eastern North America and Europe (Fig. 2). In Europe, *Emys orbicularis* and the recently recognized *Emys trinacris* are sympatric (Fritz et al. 2005, 2006). However, they are allopatric with respect to the North American species *Emys marmorata* and *Emys blandingii*, which are themselves separated by a distributional gap of at least 1500 km (Fig. 2). Although it has been hypothesized that *orbicularis*-like animals may have crossed the trans-Beringian land bridge into Europe sometime in the early Miocene ~20 Ma (Fritz 1998), relatively little solid fossil evidence is available to help shed light on this fascinating biogeographic distribution (Holman and Parmalee 2005). On the morphological side, plastral kinesis is often viewed as an important taxonomic feature in turtles; yet, it is variable within *Emys*: *E. marmorata* has an akinetic plastron, whereas *blandingii*, *orbicularis*, and *trinacris* all have hinged plastra. A deeper understanding of the evolution of both the distribution and the morphology of the group would benefit tremendously from a strong phylogeny, and even more so from a well-calibrated chronogram. Unfortunately, a well-resolved phylogeny for the *Emys* complex has been remarkably difficult to resolve (Bickham et al. 1996; Feldman and Parham 2002; Stephens and Wiens 2003).

In this study, we used deep within-taxon sampling and multiple unlinked genetic markers to reconstruct the phylogeny of the *Emys* species complex. Previous analyses have been based on extremely sparse sampling of taxa/species and limited character sampling. Our approach was to sample multiple individuals per species from across their ranges and multiple nuclear markers per individual (Supplementary Appendix, http://www.oxfordjournals.org/our journals/sysbio/). In total, we collected sequence data for 70 individuals in the genus *Emys* representing all 4 species, plus 13 outgroups (which included all but 1 Emydinae species) for 1 mitochondrial and up to 7 nuclear loci. We estimated divergence times among members of the *Emys* species complex to explore the timing and likely dispersal route of *orbicularis* to Europe and to provide mechanistic insights into the nuclear/mitochondrial gene tree discordance that we discovered. For the remainder of this paper, we follow Feldman and Parham (2002) in using the name *Emys* to refer to the apparent clade consisting of *blandingii, marmorata, orbicularis*, and *trinacris*. For a recent discussion of the taxonomic confusion surrounding these species, see Turtle Taxonomy Working Group (2007a, 2007b).

### *Brief Review of Emys Biology and Phylogenetic History*

*Emys* are medium-sized turtles that range from highly aquatic (*marmorata, orbicularis*) to semiterrestrial (*blandingii*). They tend to be habitat generalists, occurring in lakes, ponds, and slow-moving rivers across their broad ranges (Fig. 2; Ernst and Barbour 1989; Ernst et al. 1994). *Emys orbicularis* is the only Old World member of the diverse family Emydidae and is widely distributed across Europe and western Asia. *Emys marmorata* is distributed along the North American Pacific Coast from Baja California Norte, Mexico to Washington, USA. *Emys blandingii* is patchily distributed across northeastern North America from Nebraska eastward to Nova Scotia and from Ohio to Ontario, Canada (Conant and Collins 1991). Recent phylogeographic work based largely on mtDNA variation has demonstrated a number of distinct phylogroups in both *E. orbicularis* (Lenk et al. 1999) and *E. marmorata* (Spinks and Shaffer 2005) but little variation within *E. blandingii* (Mockford et al. 2005, 2006). Based on both molecular and morphological differentiation, *E. trinacris* was recently removed from *E. orbicularis*; it is narrowly distributed in Sicily and adjacent mainland Italy (Fritz et al. 2005, 2006).

Previously, the western species *marmorata* was allied with 3 eastern North American emydine turtles under the genus *Clemmys*. As molecular data were brought to bear (mostly mtDNA and some allozyme), phylogenetic analyses of *Clemmys* indicated that the 4 contained species (*guttata*, *insculpta*, *marmorata*, and *muhlenbergii*) did not comprise a monophyletic group. Rather, *Clemmys marmorata* from western North America appeared to be closely related to *E. orbicularis* (from Europe) and *Emys* (= *Emydoidea*) *blandingii* (eastern North America) (Bickham et al. 1996; Burke et al. 1996; Holman and Fritz 2001; Feldman and Parham 2002). As a consequence, the taxonomy of the subfamily Emydinae has been modified to reflect these newer phylogenetic hypotheses. This taxonomy is still in flux, reflecting both phylogenetic uncertainty (Fig. 1; see also Turtle Taxonomy Working Group 2007a) and differing nomenclatural approaches to the problem (summarized in Turtle Taxonomy Working Group 2007b). Currently, the phylogenetic relationships among *Emys* and an adequate explanation for the biogeographic distribution of this group remain obscure.

## Materials and Methods

### *Sample Collection*

Our sampling strategy was based on the phylogeographic analyses of Lenk et al. (1999) and Spinks and

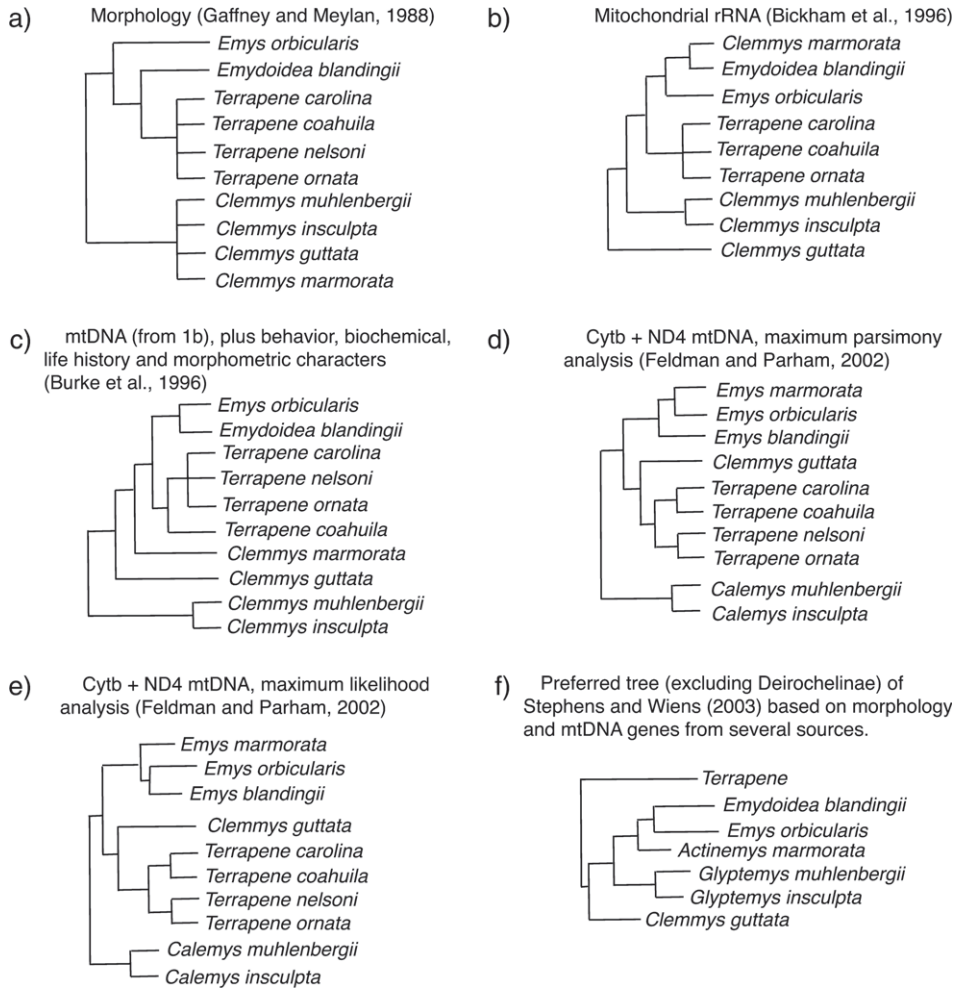


FIGURE 1. Previous phylogenetic hypotheses for the Emydinae showing incongruent topologies and inconsistent taxonomy. Phylogeny 1a was based on morphological analyses of 25 characters (Gaffney and Meylan 1988), whereas Phylogeny 1b was based on 16S mtDNA (Bickham et al. 1996). Phylogeny 1c was based on the 16S mtDNA data from Bickham et al. (1996) plus 30 morphological characters and a total of 8 behavioral, allozyme, chromosomal, and life history characters (Burke et al. 1996). Phylogenies 1d and 1e were based on concatenated mitochondrial *cytb* and ND4 sequences (Feldman and Parham 2002), and Phylogeny1f was based on 116 osteological and 109 external morphology characters plus *cytb*, control region, ND4, and 16S mtDNA sequences from various authors (Stephens and Wiens 2003).

Shaffer (2005) for *orbicularis* and *marmorata*, respectively. We sampled 1–3 individuals per population to span the geographic range of each species while also including at least 2 individuals of each major mtDNA clade. Our samples include 35 *orbicularis* and 21 *marmorata* samples, plus 2 *E. trinacris* from its restricted range (Fritz et al. 2005, 2006). Our *E. blandingii* sampling was less thorough and included 12 samples (10 localities) that encompass the range of the species from Nebraska and Michigan eastward to New York and Nova Scotia and north to Ontario, Canada (Supplementary Appendix). Tissue samples included blood and tail snips from live animals and muscle/liver from frozen and ethanol-preserved tissue specimens. Our outgroup sampling consisted of 13 taxa, including 8 emydid taxa (6 additional emydine species plus 2 deirochelyines), 2 geoemydids, a testudinid, a platysternid, and a cheloniid (Supplementary Appendix), providing a broad sample of closely and more distantly related cryptodiran turtles (Krenz et al. 2005; Parham et al. 2006).

### *DNA Extraction, PCR Amplification, and Sequencing*

Eskandar Pouyani (University of Heidelberg, Germany) generously provided us with DNA extracts for most *orbicularis* specimens. For the remaining taxa, DNA was extracted from muscle or blood using a salt extraction protocol (Sambrook and Russell 2001). Our sequence data included some sequences from GenBank, but most data were generated for this analysis. Our data consisted of fragments of 1 mitochondrial and up to 7 nuclear loci, generally from the same individual (Supplementary Appendix). All PCR products were amplified using 15–20 μL volume Taq-mediated PCRs with cycling conditions and primer sequences that can be found in the reference(s) following each locus listed


FIGURE 2. Map showing distribution of the *Emys* species complex across the northern hemisphere. This figure appears in color in the online version of *Systematic Biology*.

below. Our mtDNA sequence data consisted of nearly complete *cytb* sequences (~1070 bp; Spinks et al. 2004), whereas our nuclear markers included Intron 1 of the fingerprint protein 35 (*R35*, ~1000 bp; Fujita et al. 2004), Intron 2 of the hepatocyte nuclear factor 1α (*HNF-1α*, ~800 bp; Primmer et al. 2002), and Intron 61 of the reelin gene (*RELN*, ~1150 bp; Spinks and Shaffer 2007). *R35* has proven to be a phylogenetically informative marker for turtles (Engstrom et al. 2004; Fujita et al. 2004; Spinks et al. 2004; Krenz et al. 2005), and in a recent analysis of Asian box turtles (genus *Cuora*), *HNF-1α* was more variable and *RELN* was only slightly less variable than *R35* (Spinks and Shaffer 2007), suggesting that these markers might also be informative. We also included sequence data from the nuclear recombination activase gene 1 (*RAG-1*, ~790 bp; Krenz et al. 2005), Intron 5 of the transforming growth factor beta-2 gene (*TGFB2*, ~950 bp; Primmer et al., 2002), and 2 anonymous nuclear loci: TB29 and TB73 (~590 and ~660 bp, respectively; Shaffer and Thomson 2007; Thomson and Shaffer 2008). All PCR products were sequenced in both directions on ABI 3730 automated sequencers at the Univeristy of California, Davis, Division of Biological Sciences sequencing facility (http://dnaseq.ucdavis.edu/). Patterns from the sequencing chromatograms indicated that some individuals were heterozygous for length polymorphisms at all loci, except *RAG-1* (Bhangale et al. 2005). By sequencing in both directions, we were able to obtain sequence data from most of the introns for the putative length-polymorphic individuals.

### *Phylogenetic Analyses*

Sequences were aligned, checked for nucleotide ambiguities, and the *cytb* fragments were translated using MacClade 4.06 (Maddison and Maddison 2003) to check for the presence of nuclear–mitochondrial pseudogenes (numts). MacClade 4.06 was also used to identify identical sequences that were excluded from phylogenetic analyses. We performed phylogenetic analyses under maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference on individual markers and concatenated nuclear markers. Individual markers were coded in 2 different ways, depending on how heterozygotes were considered. Either individuals were considered to be the terminal OTU (in which case heterozygous positions were coded as ambiguous) or haplotypes were considered to be the terminal OTU (in which case an individual's haplotypes must be derived from its composite nucleotides). We used the Phase2.1.1 software (Stephens et al. 2001; Stephens and Donnelly 2003) to reconstruct probable pairs of haplotypes for each *Emys* sequence. Phase2.1.1 uses a Bayesian approach to reconstruct haplotypes from population

genotypic data. We used the default settings, except that we ran 5 iterations/locus and accepted haplotype reconstructions with Bayesian posterior probability (BPP) of $\geq$0.95 only; reconstructed haplotypes with $\leq$0.94 BPP were coded as missing data. In cases where a sequence contained a single polymorphic site, phase was determined manually. We ran all individual gene tree analyses on both individual genotype and haplotype data to confirm that they returned similar gene trees and to check for possible shared haplotypes between species that might provide evidence of recent hybridization and introgression. However, when data were concatenated, we only used genotypic data because there is no biologically meaningful way to concatenate haplotypes across individuals.

The MP and ML analyses were performed using PAUP* 4.0b10 (Swofford 2002) with 10 random stepwise heuristic searches and tree bisection-reconnection branch swapping (MP analyses) or subtree pruning-regrafting branch swapping (ML). We bootstrapped each MP and ML data set with 100 pseudoreplicates/analysis (Felsenstein 1985), but each ML bootstrap analysis was limited to 100 h (1 h/pseudoreplicate) of computation time. For ML and Bayesian analyses, models of molecular evolution for parameter estimation were selected using the DT-Modsel software package (Minin et al. 2003), and Modeltest 3.06 (Posada and Crandall 1998) was used to report parameter values (estimated with PAUP*) for use in ML analyses. For Bayesian analyses, we assessed several partitioning strategies for various combinations of codons/loci (see below) and selected the optimal partitioning strategy based on Bayes factors. Bayesian and ML topologies were compared using the tree filter option in PAUP* to determine if the ML tree topologies were contained within the 95% credible set of the BPP distribution of trees from the corresponding Bayesian analyses. Due to computational constraints, all analyses of haplotype data were conducted with RaxML (Stamatakis et al. 2008) and MrBayes through the CIPRES Web portal (www.phylo.org) to carry out ML bootstrap and Bayesian analyses.

### *Topological, Rate Heterogeneity, and Neutrality Tests*

In order to compare and contrast the phylogenetic signal from mtDNA versus nuDNA, we performed phylogenetic analyses on mtDNA and individual and concatenated nuclear loci (Shaffer et al. 1991, 1997; Chippindale and Wiens 1994; Gadagkar et al. 2005). We did not concatenate our mtDNA and nuDNA data for analyses because the mtDNA versus nuDNA topologies were incongruent (see below). Given the uncertain relationships among the *Emys* species complex, we used Bayesian tests of monophyly to further assess relationships recovered from single-locus analyses. The posterior probability is the probability that a tree is correct conditioned on the model and data (Huelsenbeck and Rannala 2004), thus the Bayesian framework offers a relatively simple way to evaluate monophyly (Linnen and Farrell 2007). For each nuclear locus, we constructed 3 constraint trees (using MacClade v4.06) for each of the 3 rooted possible relationships among *Emys* species (*trinacris* was lumped with *orbicularis* because these species were generally not reciprocally monophyletic). These constraints were imported into PAUP* and, with all outgroup taxa deleted, used to filter the corresponding postburn-in Bayesian trees from each locus. If less than 1.25% of trees ($\alpha = 0.05$, but after Bonferroni correction $\alpha = 0.05/4$) were retained, then the null hypothesis of a sister group relationship was rejected (Miller et al. 2002; Buschbom and Barker 2006; Linnen and Farrell 2007).

Nucleotide substitution rate heterogeneity at each locus and for concatenated data sets was tested among lineages using the likelihood-ratio test (Felsenstein 1988). To carry out these tests, we used PAUP* to compute $-\ln L$ scores for each data partition with and without a molecular clock enforced. The likelihood-ratio test statistic was then compared with a $\chi^2$ distribution with $N - 2$ degrees of freedom, where $N$ was equal to the number of taxa (Felsenstein 1988).

Under neutral conditions, the ratio of intraspecific nucleotide polymorphism to interspecific nucleotide sequence divergence should be similar across loci. Deviations from neutrality are often associated with recent strong selection on 1 or more genes (Kimura 1983), and such deviations may indicate that recent, gene-specific selection has occurred within species. Therefore, we used multilocus Hudson, Kreitman, and Aguadé (HKA) tests (Hudson et al. 1987) to assess the possibility that recent hybridization among species of *Emys*, combined with strong, posthybridization selection leading to gene capture, might explain the discordant gene genealogies observed within the *Emys* species complex. We used the HKA software (Hey 2004) to carry out multilocus HKA tests. For these tests, we used PAUP* to calculate uncorrected "P" pairwise sequence distances, and polymorphic sites were determined visually in MacClade 4.06. We performed a single 8-locus HKA test for each pairwise species comparison, including ((*blandingii*+*marmorata*), *orbicularis/trinacris*),((*blandingii*+ *orbicularis/trinacris*), *marmorata*), and ((*orbicularis/trinacris* + *marmorata*), *blandingii*). Significance was evaluated using 10 000 simulations.

### *Bayesian Phylogenies and Bayesian Estimates of Species Trees*

Partitioned-model Bayesian analyses were performed using MrBayes v3.1.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). To determine optimal partitioning strategies, we followed the approach of Nylander et al. (2004) and performed Bayesian analyses on all combinations of "natural" partitions, including 5 combinations for the *cytb* data, as well as 5 combinations for the nuDNA data. For example, we analyzed *cytb* as a single partition, by codon position, and by all possible combinations of codon positions (Table 1). Similarly, the nuDNA data were analyzed as a single partition, by locus, and by combinations of *RAG-1* codon positions

TABLE 1. Models of molecular evolution for parameter estimation, partitioning strategy, and Bayes factor comparisons relative to the preferred model for the 83-taxon cytochrome *b* (*cytb*) data set

| | | | Bayes factors (harmonic means) | | | |
|---|---|---|---|---|---|---|
| | Partitioning strategy | No. of partitions | Run No. 1 | Run No. 2 | Total | Bayes factor comparisons |
| *cytb* A | *cytb* (all codons combined) | 1 | −6267.61 | −6260.56 | −6266.91 | Preferred |
| *cytb* B | *cytb* (1st and 2nd) + 3rd | 2 | −6320.67 | −6314.43 | −6319.98 | −106.14 |
| *cytb* C | *cytb* (1st and 3rd) + 2nd | 2 | −6536.23 | −6541.4 | −6540.71 | −547.6 |
| *cytb* D | *cytb* (2nd and 3rd) + 1st | 2 | −6628.96 | −6629.68 | −6629.39 | −724.96 |
| *cytb* E | *cytb* 1st + 2nd + 3rd | 3 | −6288.1 | −6291.64 | −6290.97 | −48.12 |

Note: These data were generated using the empirical branch length prior. Bayes factor comparisons favored Model A (shaded), where the data were allocated to a single partition. Models are shown in ascending order of parameters such that Model "A" has the least and "E" has the most parameters. Values for Bayes factor comparisons are twice the difference in the −log of the Bayes factor from the more compared with the less parameter-rich model. Models for each partition included *cytb* 1st positions = TVMef+G, *cytb* 2nd positions = TrN+I, *cytb* 3rd positions = TrN+G, *cytb* 1st and 2nd positions = HKY+I+G, *cytb* 1st and 3rd and *cytb* all positions = K81uf+I+G, *cytb* 2nd and 3rd positions = TIM+I+G.

and introns (Tables 2 and 3). Some of our nuclear loci exhibited nucleotide rate heterogeneity, and Marshall et al. (2006) have shown that the use of partition models in MrBayes can lead to overly long trees if among-partition rate variation (APRV) is not accommodated. Therefore, we invoked the *prset ratepr = variable* option in MrBayes v3.1.1 in order to accommodate APRV in our partitioned model analyses. In addition, we compared tree lengths from ML analyses to the 95% highest posterior density (HPD) of tree lengths from the BPP distribution to determine if ML versus Bayes tree lengths were similar. The Tracer v1.4 software (Rambaut and Drummond 2007) was used to generate the 95% HPD estimates.

Bayesian analyses of all partitions and sets of partitions were performed with 2 replicates and 4 chains for $5 \times 10^6$ generations. Chains were sampled every $10^3$ generations, and stationarity was determined as the point at which the potential scale reduction factor equaled 1, and when the −log-likelihood (−ln *L*) scores plotted against generation time reached a stationary value. The first 25% of samples were discarded as burn-in. We then compared Bayes factors (harmonic means) from the various partitioning strategies and selected the optimal partitioning strategy using the guidelines set out by Kass and Raftery (1995).

The phylogenetic analyses described above were based on single markers or concatenated nuclear markers. However, it is clear that gene trees can be incongruent with the species tree (Kellog et al. 1996; Degnan and Rosenberg 2006; Pollard et al. 2006; Edwards et al. 2007; Liu and Pearl 2007). Thus, we used the Bayesian estimation of species trees software (BEST v1.6; Liu and Pearl 2007) to simultaneously estimate gene trees from multiple nuclear loci and the overall species tree in a coalescent framework. Two-step BEST analyses were performed on both the 3-gene, 83-taxon, and 7-gene, 26-taxon data sets with the same models/partitions as for the Bayesian phylogenetic analyses described above. For these analyses, *trinacris* was considered as a separate, labeled taxon from *orbicularis*.

### *Divergence Time Estimates*

We used the program r8s v1.70 (Sanderson 2003) to estimate divergence times. Because some of our data violated the assumption of a molecular clock (see below), we used the penalized likelihood method (Sanderson 2002) implemented in r8s. We used ML trees generated from the mtDNA (83 taxon, *cytb*) and concatenated nuDNA (26 taxon, 7 loci) sequence data sets. Intraspecific branch lengths were relatively short throughout these trees, so we pruned most terminal taxa

TABLE 2. Models of molecular evolution for parameter estimation, partitioning strategy, and Bayes factor comparisons relative to the preferred model for the 83-taxon 3-locus data set

| | | | Bayes factors (harmonic means) | | | |
|---|---|---|---|---|---|---|
| | Partitioning strategy | No. of partitions | Run No. 1 | Run No. 2 | Total | Bayes factor comparisons |
| 3A | All concatenated | 1 | −7916.55 | −7916.38 | −7916.47 | 35.44 |
| 3B | (*HNF-1α* and *RELN*) + R35 | 2 | −7906.98 | −7903.73 | −7906.32 | 15.14 |
| 3C | (*HNF-1α* and *R35*) + *RELN* | 2 | −7913.09 | −7915.04 | −7914.48 | 31.46 |
| 3D | (*RELN* and R35) + *HNF-1α* | 2 | −7894.01 | −7899.44 | −7898.75 | Preferred |
| 3E | *HNF-1α* + *RELN* + *R35* | 3 | −7900.76 | −7903.90 | −7903.25 | −9 |

Note: These data were generated using the empirical branch length prior. Bayes factor comparisons favored Model D (shaded), where the data were allocated to 2 partitions: *RELN* and *R35* combined plus *HNF-1α*. Models are shown in ascending order of parameters such that Model "A" has the least and "E" has the most parameters. Values for Bayes factor comparisons are twice the difference in the −log of the Bayes factor from the more compared with the less parameter-rich model. Models for each partition were *HNF-1α* = K80+G, *RELN*, *R35*, (*HNF-1α* and *R35*), and 3 loci concatenated = HKY+G, (*HNF-1α* and *RELN*) = K81uf+G.

TABLE 3. Models of molecular evolution for parameter estimation, partitioning strategy, and Bayes factor comparisons relative to the preferred model for the 26-taxon nuDNA data set

| | | | Bayes factors (harmonic means) | | | |
|---|---|---|---|---|---|---|
| | Partitioning strategy | No. of partitions | Run No. 1 | Run No. 2 | Total | Bayes factor comparisons |
| 7A | All concatenated | 1 | −14182.76 | −14183.28 | −14183.05 | 69.72 |
| 7B | RAG + introns concatenated | 2 | −14167.43 | −14172.16 | −14171.47 | 46.56 |
| 7C | RAG by codon + introns concatenated | 4 | −14277.07 | −14277.31 | −14277.2 | 258.02 |
| 7D | RAG + 6 introns | 7 | −14146.41 | −14148.8 | −14148.19 | Preferred |
| 7E | RAG by codon + 6 introns | 9 | −14151.67 | −14150.99 | −14151.39 | −6.4 |

Note: Bayes factor comparisons favored Model 7D (shaded). Models are shown in ascending order of parameters such that the "A" models have the least and the "E" models have the most parameters. Values for Bayes factor comparisons are twice the difference in the −log of the Bayes factor from the more compared with the less parameter-rich model. Models for each partition were *HNF-1α* and RAG 1st and 3rd positions = K80+G, *RELN* and RAG 1st positions = HKY, *R35*, TB29, *TGFB2*, RAG (all positions), and RAG 1st and 2nd positions = HKY+G, RAG 2nd positions = F81, RAG 3rd positions = K80, RAG 2nd and 3rd positions = K80+I, all 6 introns concatenated = K81uf+G, all 7 loci concatenated = K81uf+G.

prior to analysis in r8s and used the truncated Newton algorithm with a smoothing parameter selected using cross-validation (Sanderson 2002). We used the fossil calibrations and estimated divergence times reported by Near et al. (2005, Fig. 6). In addition, we determined the central 95% confidence interval for each node in the respective mtDNA and concatenated nuDNA trees as described in the r8s manual (http://loco.biosci.arizona.edu/r8s/r8s1.7.manual.pdf). First, we used the program seqboot (Felsenstein 1993) to generate 100 bootstrap pseudoreplicates of each data set. These replicate data sets were imported into PAUP* and used in conjunction with the optimal models of sequence evolution and ML trees to obtain 100 sets of branch lengths for the respective mtDNA and nuDNA topologies. These trees were imported into r8s, and the confidence intervals were generated using the PROFILE command.

## RESULTS

### *mtDNA Phylogeny*

We generated *cytb* sequence data for 55 of 83 individuals (GenBank Accession Nos. EU787021–EU787075). The remaining sequences were downloaded from GenBank (Supplementary Appendix). The *cytb* data matrix was nearly complete, with 2% missing data (TreeBase matrix accession number M4306). Our mtDNA data set consisted of up to 1070 bp for 83 individuals. Thirty-eight sequences were redundant and excluded from phylogenetic analysis, leaving 45 unique sequences (32 ingroup and 13 outgroup taxa) for analysis. Among-lineage rate heterogeneity was not significant ($P = 0.15$). Figure 3 shows the *cytb* ML tree with BPPs and bootstrap support values (ML and MP) as indicated. Bayes factor comparisons favored a 2-partition model (Model B, Table 1). And, although our *cytb* ML tree was contained within the 95% credible set of trees from the Bayesian posterior distribution, the tree length of the ML tree (1.72) was not contained within the 95% HPD of tree lengths (2.097–2.772) from the Bayesian analysis. Incorrect branch length priors can lead to overly long Bayesian tree length estimates (Yang and Rannala 2005). Therefore, we used an empirical Bayesian approach and used an equation derived by Brown et al. (2009) to estimate an empirical branch length prior to calculated as the natural log of 0.5 divided by the average expected ML branch length. Using this new branch length prior, we repeated all Bayesian analyses, and in these subsequent analyses, Bayes factor comparisons favored the single-partition model (Model A, Table 1). However, the ML tree length was still outside the 95% HPD of tree lengths from the Bayesian analysis (1.97–2.55). Tree length estimates from analyses using MrBayes in conjunction with partitioned models can infer overly long branches, probably with no effect on topology (Marshall et al. 2006). We compared the original majority rule consensus tree of the posterior distribution of trees from our original Bayesian analysis with the majority rule consensus tree of the posterior distribution of trees generated using the empirical branch length prior. These consensus trees were topologically identical, and BPP support values varied only slightly between the 2 analyses (not shown). Based on these results, all 4 species (*blandingii, marmorata, orbicularis,* and *trinacris*) were recovered as monophyletic. *Emys blandingii* and *marmorata* were recovered as sister taxa with strong support from the Bayesian analyses (BPP = 0.96, bootstrap support values: ML = 68, MP = 62). Like Lenk et al. (1999) and Fritz et al. (2005), we recovered *E. trinacris* as the sister group to *orbicularis*, again with strong support from the Bayesian analysis only (BPP = 0.97, ML = 54, MP = 64).

### *83 Taxon, Nuclear Phylogenies*

For our nuDNA sequence data set, most of the sequences were generated for this analysis. Twenty-six were downloaded from GenBank but were previously collected from the same individual turtles by us (Supplementary Appendix). Our nuclear intron data included up to 2872 bp for 83 turtles, including 768 bp of *HNF-1α* (EU787076–EU787158), 1126 bp of *RELN* (EU787292–EU787372), and 978 bp of *R35* (EU787159–EU787218). As with the mtDNA data set, this matrix had very low levels of missing data (about 3%) (TreeBase matrix

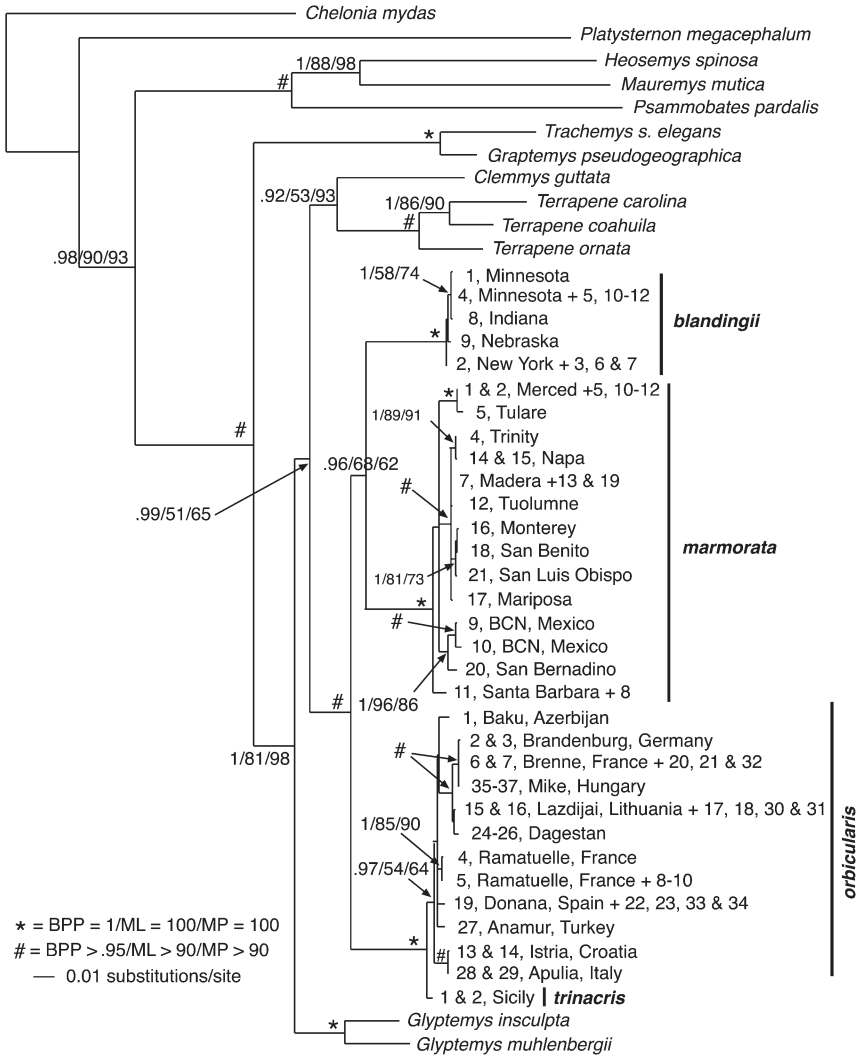


FIGURE 3. Maximum-likelihood phylogeny of the 83-taxon mitochondrial *cytb* data set (1070 bp). Estimated ML model parameters conform to the K81uf+I+G model of sequence evolution. $-\ln L = 6583.958$, rate matrix: A–C = 1, A–G = 8.6807, A–T = 0.3082, C–G = 0.3082, C–T = 0.6807, G–T = 1. Base frequencies: A = 0.32, C = 0.33, G = 0.10, and T = 0.25. Proportion of invariant sites ($I$) = 0.4627, and $\gamma$-shape parameter = 1.1628. Terminals are collection localities (Supplementary Appendix), followed by US states for *Emys blandingii*, California counties for *Emys marmorata* (except for the Mexican samples), and city/region and country for *Emys orbicularis* and *Emys trinacris*. Sample numbers to the right of terminals indicate those samples that had identical haplotypes and were excluded from phylogenetic analyses. For example, *E. blandingii* Sample 5 (Wisconsin) and Samples 10–12 (Michigan) were identical to Sample 4 (Minnesota). "#" indicates nodes with support values of 100 from all analytical methods. "*" indicates nodes with BPPs ≥0.95 and MP/ML bootstrap values ≥90. Numerical values indicate other BPP/ML/MP support values.

accession number M4306). Concatenated sequences from 17 individuals were redundant and excluded from analysis, leaving 66 terminals for analysis. The likelihood-ratio tests showed significant among-lineage rate heterogeneity for *R35* and *RELN* ($P \leq 0.003$), but not for *HNF-1α* ($P = 0.18$).

Phylogenetic analyses of each gene separately revealed that monophyletic Emydidae and Deirochelyinae were the only consistently well-supported higher groups (Supplementary Figs. S1–S3). At the species level, *blandingii* was recovered as monophyletic at all 3 loci, but with strong support from *RELN* only. The monophyly of *marmorata* and *orbicularis* was strongly supported by 2 genes, and these results were generally upheld regardless of whether inferred haplotypes or sequences were analyzed (Supplementary Figs. S1–S3). In addition, the haplotype analyses revealed that *blandingii* was monophyletic at *HNF-1α*, *R35*, and *RELN* (with strong support at *R35* and *RELN*), whereas *marmorata* was monophyletic at *HNF-1α* and *RELN* (but with no support), and *orbicularis/trinacris* was strongly supported as monophyletic at *R35* and *RELN* (Supplementary Figs. S1–S3). Gene trees from *HNF-1α* and *R35* (Supplementary Figs. S1 and S2, respectively) both suggest that *orbicularis* and *blandingii* form a clade (but with weak support), rather than the *blandingii–marmorata* sister group suggested by the mtDNA; the *RELN* tree was equivocal on this point (Supplementary Fig. S3). To further explore relationships among *Emys* species, we used SH tests and Bayesian tests

of monophyly. We tested 4 hypotheses, including 1) *blandingii* and *marmorata* as sister taxa, 2) *blandingii* sister to (*orbicularis/trinacris*), 3) *marmorata* sister to (*orbicularis/trinacris*), and 4) *E. marmorata*, *G. muhlenbergii*, *G. insculpta*, and *C. guttata* as monophyletic (the latter hypothesis is the "old" *Clemmys* clade depicted in Fig. 1*a*). No hypothesis could be rejected based on the SH tests or Bayesian tests of monophyly for *RELN*. However, all hypotheses except (*blandingii* + (*orbicularis/trinacris*)) were rejected based on Bayesian tests of monophyly for *HNF-1α* and *R35* (Table 4).

Given the low levels of phylogenetic resolution for any 1 nuclear gene tree, we concatenated these data (nonhaplotype) to explore any common phylogenetic signal that might emerge from these data. Figure 4 shows the ML tree from the concatenated data with the optimum $-\ln L$ score and with support values as indicated. Bayes factor comparisons favored a 2-partition strategy: *RELN* and *R35* combined, plus *HNF-1α* (Model 3D, Table 2). The ML tree was contained within the 95% credible set of trees from the Bayesian posterior distribution, but the ML tree length (0.238) was far outside the Bayesian 95% HPD of tree lengths (10.69–14.98). Again, we repeated all Bayesian analyses using an empirical branch length prior. Subsequent Bayes factor comparisons favored the same partitioning strategy (Model 3D, Table 2), and with this empirical branch length prior, the Bayesian HPD of tree lengths (0.238–0.232) included our ML tree length estimate. Once again, we note that the topology of the majority rule consensus tree of the posterior distribution of trees from our original Bayesian analysis was identical to the majority rule consensus tree of the posterior distribution of trees generated using the empirical branch length prior, and BPP support values varied only slightly between the 2 analyses (not shown).

The 3-gene concatenated nuDNA tree was reasonably well resolved and provides strong support for many results that were found or suggested in the individual nuclear gene analyses. Emydidae, Emydinae, and Deirochelyinae were all strongly supported, as were *marmorata*, *blandingii*, and *orbicularis/trinacris*. In contrast to the mtDNA-based results, but consistent with several of the individual gene–nuclear gene analyses, the tree from the 83-taxon concatenated nuDNA data set recovered (*blandingii* + (*orbicularis/trinacris*)) with strong support (BPP $\geq$ 0.95, ML and MP $\geq$ 90), and *orbicularis* was paraphyletic with respect to *trinacris*. The *Emys* clade was recovered as monophyletic, but with low statistical support. When we applied tests of monophyly to the 4 sister-group hypotheses for species of *Emys* (see above), once again all hypotheses except (*blandingii* + (*orbicularis/trinacris*)) were rejected under the Bayesian test of monophyly (Table 4).

### *Nuclear 26-Taxon Phylogeny*

To gain additional nuclear perspectives on the phylogeny of this group, we bolstered our nuDNA locus sampling by constructing a 26-individual data set containing a minimum of 2 individuals per *Emys* species plus the 13 outgroups. We added up to 788 bp of *RAG-1* (EU787244–EU787265) plus additional GenBank sequences (Supplementary Appendix), 590 bp of TB29 (EU787266–EU787291), 663 bp of TB73 (EU787373–EU787398), and 952 bp of *TGFB2* (EU787219–EU787243),

TABLE 4. Bayesian tests of monophyly and SH tests

| $H_0$: Monophyletic clade constraints | 83-taxon data set: Bayesian topology tests (% of trees retained) *HNF-1α* | *R35* | *RELN* | Concatenated nuDNA | SH tests: Concatenated nuDNA |
|---|---|---|---|---|---|
| *Clemmys* sensu Gaffney and Meylan (1988) | 0.04 | 0.11 | 0 | 0 | Best |
| *blandingii* sister to *marmorata* | 0.03 | 0 | 13.7 | 0 | $P = 0.268$ |
| *blandingii* sister to *orbicularis* | 33.5 | 5.4 | 14.7 | 100 | $P = 0.382$ |
| *marmorata* sister to *orbicularis* | 0.03 | 0 | 14.4 | 0 | $P = 0.268$ |

| | 26-taxon data set: Bayesian topology tests (% of trees retained) *HNF-1α* | *R35* | *RELN* | *RAG-1* | TB29 | TB73 | *TGFB2* | Concatenated nuDNA | SH tests: Concatenated nuDNA |
|---|---|---|---|---|---|---|---|---|---|
| *Clemmys* sensu Gaffney and Meylan (1988) | 0.11 | 0.36 | 0 | 0 | 0 | 0 | 0 | 0.04 | $P = 0.000$ |
| *blandingii* sister to *marmorata* | 0.31 | 0 | 29.5 | 0.05 | 6.1 | 2 | 1.5 | 0 | $P = 0.386$ |
| *blandingii* sister to *orbicularis* | 32 | 29 | 38.2 | 88 | 6.9 | 95 | 1.4 | 78 | Best |
| *marmorata* sister to *orbicularis* | 0.25 | 0 | 30.4 | 0.15 | 6.6 | 1 | 1.4 | 0 | $P = 0.386$ |

Notes: Three constraint trees were constructed to test for monophyletic sister group relationships among *Emys* species and to test the monophyly of *Clemmys* sensu Gaffney and Meylan (1988). *Emys trinacris* was lumped with *orbicularis* for these tests because these 2 species were not reciprocally monophyletic (see text). Values in cells are the number of trees retained by the corresponding filter. The number of post-burn-in trees for each test was 7500, and significance level was 0.0125 ($\alpha = 0.05/4$ after Bonferroni correction). Thus, the null hypothesis of a monophyletic *Clemmys* or monophyletic sister group relationships were rejected if fewer than 94 trees were retained. For the SH tests, all 4 topologies were specified a priori.

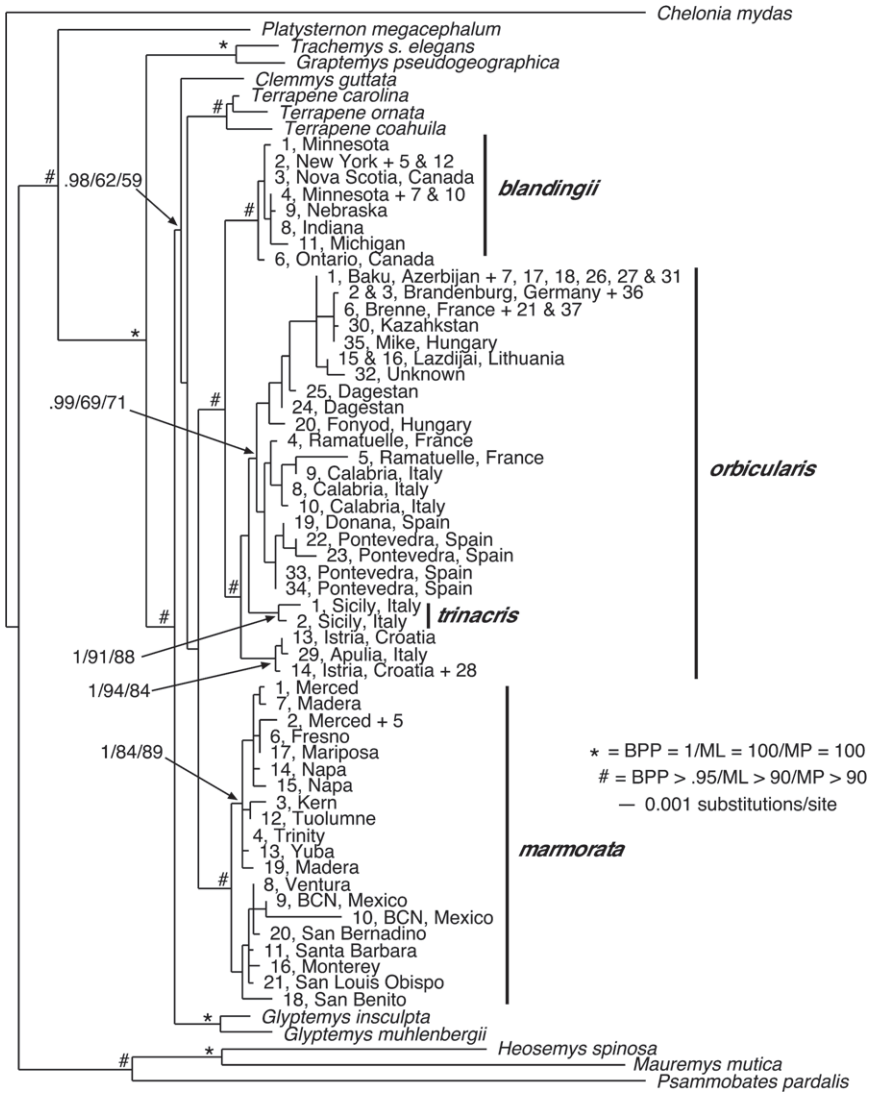


FIGURE 4. Maximum-likelihood phylogeny based on the 83-taxon concatenated nuDNA data set (2872 bp). Nuclear loci included *HNF-1α*, *RELN*, and *R35*. Estimated model parameters conform to the HKY+G model of sequence evolution. $-\ln L = 7931.2518$, Ti/tv ratio = 1.5788. Base frequencies: A = 0.29, C = 0.19, G = 0.21, and T = 0.31. Proportion of invariant sites ($I$) = 0, and $\gamma$-shape parameter = 0.4551. Terminal labels and node symbols as in Figure 3. Dashed branches are not drawn to scale.

for a 7-gene total of up to 5838 bp of nuDNA sequence data. This matrix was relatively complete, with 5% missing data (TreeBase matrix accession number M4307), but there was a slight discrepancy in the number of base pairs for *RELN* between the 26- and 83-taxon data sets (1104 bp vs. 1126 bp, respectively) due to a 22 bp insertion within *E. orbicularis* No. 1. Five individual loci (*RAG-1, R35, RELN*, TB73, and *TGFB2*) and the concatenated nuDNA data set showed significant among-lineage rate heterogeneity ($P \leq 0.025$), but *HNF-1α* and TB29 did not ($P \geq 0.052$). We analyzed all 7 nuclear loci as a concatenated data set (Fig. 5*a*) and as individual loci (Fig. 5*b–h*). Figure 5*a* shows the optimal ML tree. Bayes factor comparisons favored the 7-partition model (Model 7D, Table 3), but in the initial Bayesian analyses of these partitions, the 2 runs did not converge. However, after changing the temperature parameter to 0.1, subsequent runs converged. Our ML topology was contained within the 95% credible set of the posterior distribution of trees from the Bayesian analysis, and the ML tree length (0.188) was contained within the 95% HPD of tree lengths from the Bayesian analysis (0.185–0.211). Based on the concatenated data, Emydidae, Emydinae, and Deirochelyinae were monophyletic with strong support (BPP $\geq$ 0.95, ML and MP $\geq$ 90), as were the emydine genera *Terrapene, Glyptemys,* and *Emys* (the first 2 with strong support; Fig. 5*a*). We recovered a virtual polytomy among most emydine genera, with no intergeneric relationships receiving strong statistical support. However, the monophyly of *blandingii* + *orbicularis*/*trinacris* was again strongly supported (BPP $\geq$ 0.95, ML and MP $\geq$ 90). Based on SH tests, no sister group relationships among *Emys* could be rejected ($P \geq 0.386$; Table 4), but the monophyly of *Clemmys* sensu Gaffney and Meylan (1988) was rejected ($P \leqslant 0.01$). Results of the Bayesian tests of monophyly for *HNF-1α*, *R35*, and

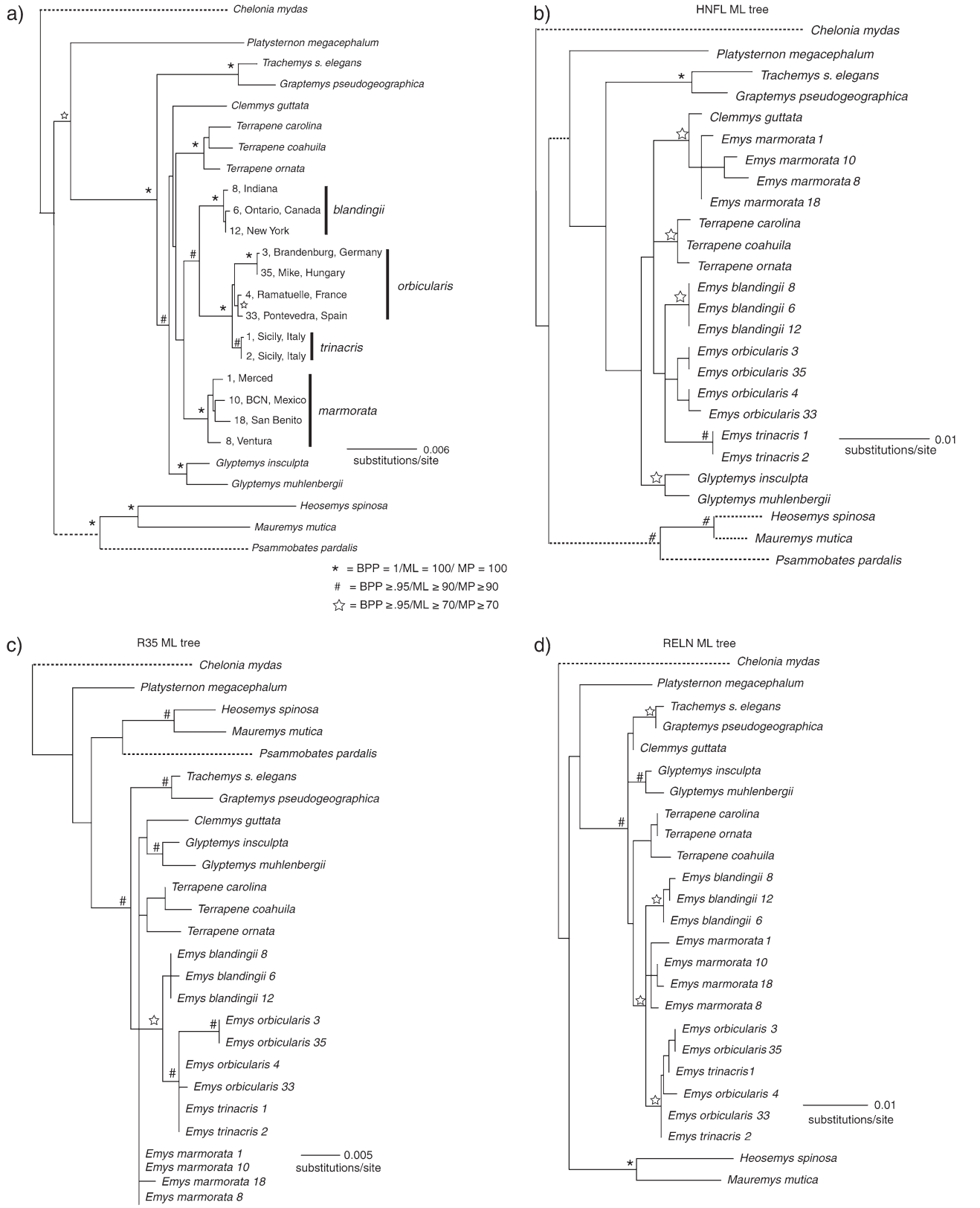


FIGURE 5. (Continued)

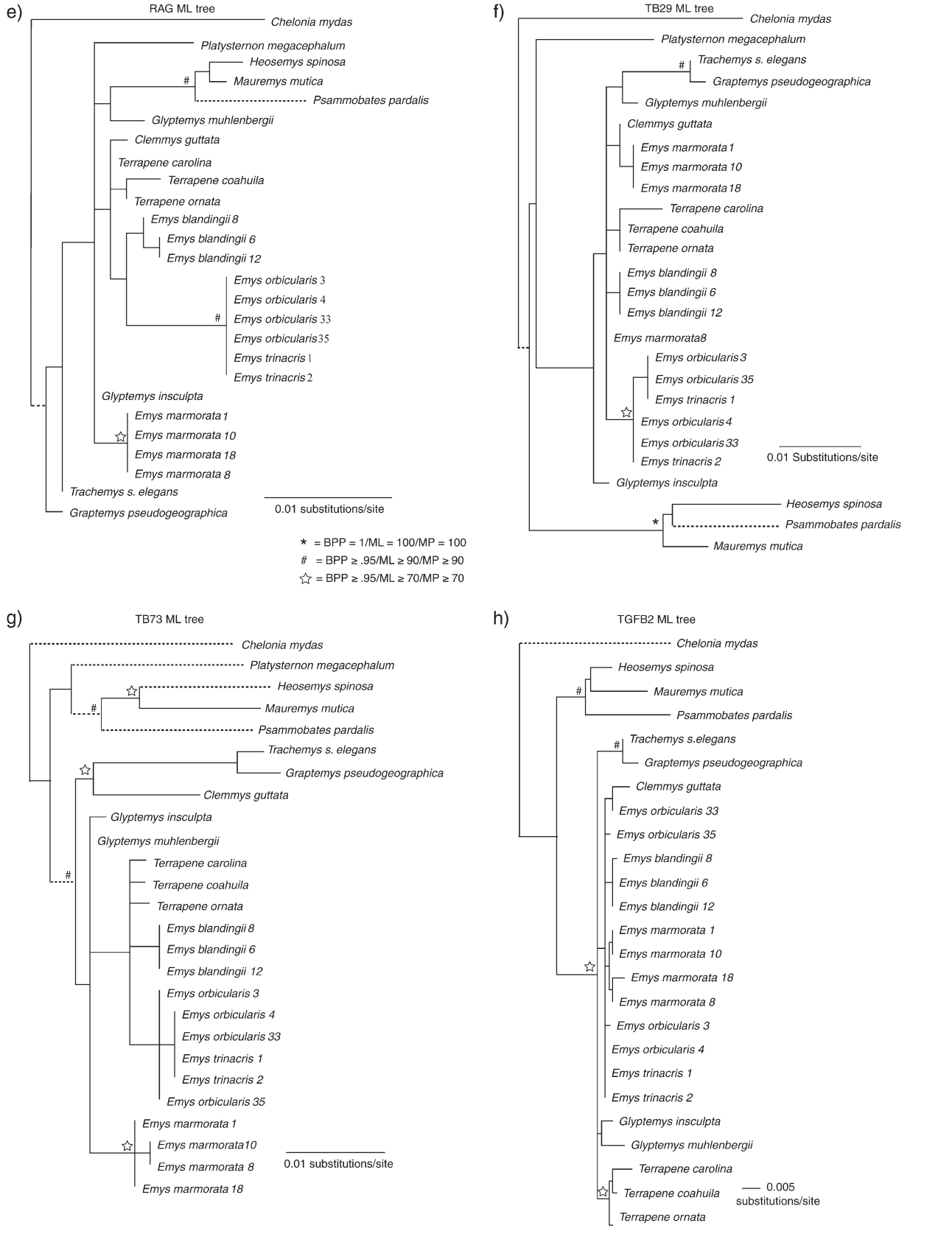


FIGURE 5. (Continued)

*RELN* were the same as for the 83-taxon data set, but the results were varied for the remaining 4 loci (Table 4). For example, *Clemmys* sensu Gaffney and Meylan (1988) was rejected at all loci and with the concatenated data. The evidence on interspecific *Emys* relationships was less clear than that for the 83-taxon data set, but points in the same direction. The hypothesis of *blandingii* + *orbicularis/trinacris* was not strongly rejected at any locus and was the favored hypothesis for all loci except *TGFB2* (Table 4). Again we note that *marmorata* was never recovered as sister to *blandingii* for any individual gene, whereas the *blandingii* + *orbicularis/trinacris* clade was recovered from analyses of *HNF-1α*, *R35*, and *RAG*, suggesting a consistent conflict with the mtDNA results.

Results from the BEST analyses indicated strong support for the monophyly of Emydidae only. A large percentage (0.93) of trees from the posterior tree distribution from the 7-gene data set recovered *blandingii* + *orbicularis/trinacris*, although the percentage of trees from the 3-gene data set was much less (Fig. 6). In addition, results from the HKA tests indicated no significant departure from neutrality ($P \geq 0.31$; Table 5), suggesting that there is no evidence of recent, strong selective sweeps on any nuclear or mitochondrial markers.

### *Estimated Divergence Times*

We show 2 time-calibrated phylogenies (chronograms) in Figure 7. Our estimated divergence times derived from mtDNA and nuDNA sequences varied somewhat among data sets, but were roughly congruent for most nodes. For example, the estimated divergence time for the Emydinae was about 29 Ma based on the nuDNA and 23 Ma based on mtDNA (Nodes Emydinae 2 and Emydinae 1; Fig. 7). Among the box turtle (*Terrapene*) species, divergence times (Nodes B1 and B2) were also similar (nuDNA ∼12 Ma and mtDNA ∼11 Ma), but they were less similar for the inferred split between *Glyptemys insculpta* and *G. muhlenbergii* (nuDNA = 20 Ma, Node E2 and mtDNA = 13 Ma, Node E1).

Divergence time estimates for the *Emys* species complex were more problematic due to the variable phylogenetic position of *marmorata* with respect to *orbicularis/trinacris* and *blandingii*. Estimated divergence times between *marmorata* and its sister group vary widely depending on the data set used. For nuDNA, the split between *marmorata* and the remaining *Emys* (Node C2) dates to about 23 Ma, whereas mtDNA estimates date the split between *marmorata* and *blandingii* (Node C1) at around 12 Ma. However, we see a very different situation for *blandingii*. The *blandingii* to (*orbicularis/trinacris*) divergence based on mtDNA and nuDNA was very similar at ∼17 Ma (Nodes D1 and D2; Table 6).

## Discussion

### *Phylogeny of the* Emys *Complex*

Among closely related taxa, it is often the case that mtDNA and nuDNA data sets are incongruent and sometimes strongly so, and our mtDNA phylogeny differed not only from all previous phylogenetic analyses, but also from our nuDNA phylogenies (compare Figs. 1, and 3–5). The mtDNA-based phylogeny was well resolved, with most interspecific relationships receiving strong support from at least 1 phylogenetic method. However, the *Emys* species complex, and within it the *marmorata* + *blandingii* clade, received strong Bayesian support only (BPP = 0.99) (Fig. 3). The ML and Bayesian topologies generated from the 83- and 26-taxon concatenated nuDNA data sets were congruent, recovering the *blandingii* + *orbicularis/trinacris* clade with strong support, although the monophyly of the *Emys* species complex was not strongly supported (Figs. 4–5). In addition, results of the BEST analysis, although not conclusive, indicated some support for the monophyly of *blandingii* + *orbicularis/trinacris* (Fig. 6). Thus, the most striking discrepancy among our analyses, and between our analyses and previous analyses shown in Figure 1, was the inconsistent position of *marmorata* with respect to *blandingii* and *orbicularis/trinacris*.

All species have a singular evolutionary history, although individual gene trees may trace different components of that overall history. We can be fairly certain that *marmorata* occupies incongruent mitochondrial versus nuclear phylogenetic positions, but whether this incongruence is a reflection of the actual species history of the *Emys* complex or is an artifact of the analyses is less clear. The incongruence between data partitions in the placement of *marmorata* could be due to lineage sorting or hybridization/introgression, and discriminating between these processes can be difficult (Holder et al. 2001; Buckley et al. 2006; Peters et al. 2007; Good et al. 2008), especially in light of the extremely short interspecific branch lengths of our nuDNA phylogeny (Degnan and Rosenberg 2006). Results of the HKA tests indicate no significant deviation from neutrality, and there were no shared haplotypes among *Emys* species (except among *orbicularis* and *trinacris*), suggesting that both nuclear and mitochondrial loci appear to have sorted to monophyly. Thus, there is no

FIGURE 5 (previous page). *a*) Maximum-likelihood tree based on the 26-taxon, 7-locus nuDNA data set (5838 bp). Nuclear loci included *HNF-1α*, *RAG*, *RELN*, *R35*, TB29, TB73, and *TGFB2*. Estimated model parameters conform to the K81uf+G model of sequence evolution. −ln $L$ = 14136.934, rate matrix: A–C = 1, A–G = 2.6547, A–T = 0.6783, C–G = 0.0.6783, C–T = 2.6547, G–T = 1. Base frequencies: A = 0.30, C = 0.20, G = 0.21, and T = 0.29. Proportion of invariant sites ($I$) = 0, and γ-shape parameter = 0.5054. Node symbols as in Figure 3. Dashed branches are not drawn to scale. *b*) *HNF-1α* ML tree, selected model = K80+G. *c*) *R35* ML tree, selected model = HKY+G. *d*) *RELN* ML tree selected model = HKY. *e*) *RAG* ML tree, selected model = HKY+G. *f*) TB29 ML tree, selected model = HKY+G. *g*) TB73 ML tree, selected model = HKY+I. *h*) *TGFB2* ML tree selected model = HKY+G.

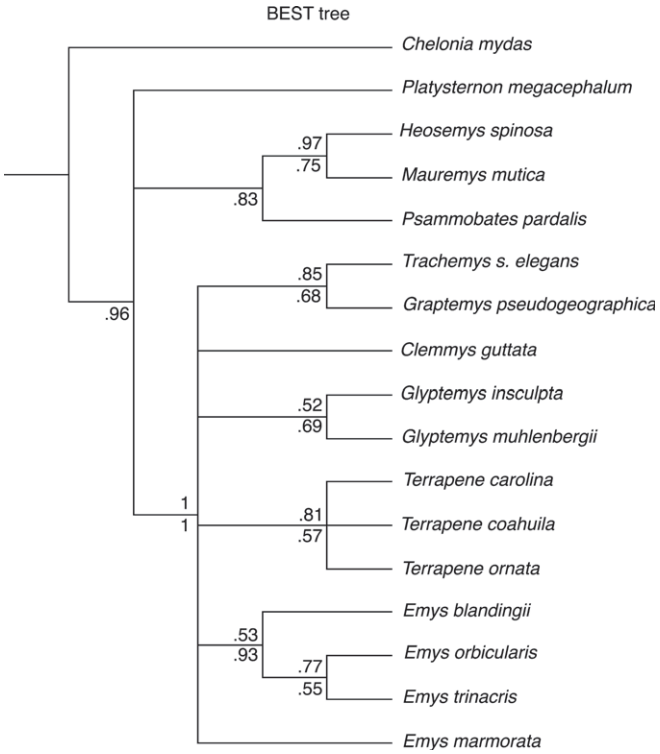


FIGURE 6. The majority rule consensus of the posterior distribution of trees from the BEST analyses of both the 3 and 7-gene data sets. Above and below branches are majority rule consensus tree values from the 3- and 7-gene data sets, respectively.

evidence for recent gene flow among *blandingii*, *marmorata*, and *orbicularis/trinacris*, and no evidence of recent selective sweeps within species. Under the assumption that the mtDNA- and nuDNA-based phylogenies accurately reflect these gene tree histories, and in view of our estimated divergence times, we hypothesize that relatively ancient introgression and mitochondrial capture in *Emys* is the best explanation for all available data.

### *Estimated Divergence Times and Among-Tree Disagreement*

Estimating divergence times from molecular data is a difficult problem (Doyle and Donoghue 1993; Graur and Martin 2004). Perhaps the most contentious point is dealing with the error/precision of calibration points and the resulting divergence time estimates based on these calibrations (Graur and Martin 2004; Hedges and Kumar 2004; Near et al. 2005). We follow most researchers in interpreting our divergence time estimates as minimum times useful for understanding the history of a group, and especially for developing biogeographic hypotheses. To this end, we are most concerned here with understanding the differences in divergence times estimated from mtDNA versus nuDNA. Because both sets of divergence time estimates were generated using the same fossil calibration points, any error or lack of precision of the calibration points should not influence these differences. In addition, we used fossils that were deemed to be consistent dating points based on the cross-validation procedure proposed by Near et al. (2005), which should eliminate some of the most extreme outlier fossil dates.

Given our taxon and data sampling for *Emys*, the incongruence between the mitochondrial and the nuclear gene trees within this group is probably an accurate representation of these gene tree histories. Although lineage sorting could explain such well-supported incongruence, we favor the interpretation that the incongruence seen here is due to ancient reticulation between *blandingii* and *marmorata* such that *marmorata* captured *blandingii* mitochondrial haplotypes, probably sometime in the late Miocene/early Pliocene (see below). Our reasoning is based on the following assumptions and observations: 1) both for the genus *Emys* and for the species *marmorata*, *blandingii*, and *orbicularis/trinacris*, the mitochondrial gene and 5/7 nuclear loci appear to have sorted to monophyly; 2) no nuclear locus (0/7) recovered *marmorata* as sister to *blandingii*, whereas 3/7 loci recovered *blandingii* as sister to *orbicularis/trinacris*, 3) Bayesian tests of monophyly rejected the hypothesis that *blandingii* and *marmorata* are sister taxa (and failed to reject the hypothesis that *blandingii* and *orbicularis/trinacris* were sister taxa) at all nuclear loci except *TGFB2*; 4) the BEST analyses recovered *blandingii* + *orbicularis/trinacris* with reasonable support; and 5) under the model of hybridization and mitochondrial gene capture of *blandingii* mtDNA by *marmorata*, the divergence between *blandingii* and *orbicularis* should

TABLE 5. HKA tests for pairwise species comparisons among *Emys* species

| | | Polymorphic sites | | | | Interspecific divergence | |
|---|---|---|---|---|---|---|---|
| | | *blandingii* | | *marmorata* | | | |
| Pairwise comparison | Locus | Observed | Expected | Observed | Expected | Observed | Expected |
| *blandingii* vs. *marmorata* | *Cytb* | 5 | 10.04 | 46 | 46.76 | 6.2 | 8.16 |
| Sum of deviations = 15.3313 | *RELN* | 5 | 3.48 | 16 | 15.97 | 0.58 | 2.8 |
| Degrees of freedom = 14 | *HNF-1α* | 0 | 3.7 | 22 | 17.24 | 1.24 | 3.01 |
| $P = 0.36$ | *R35* | 6 | 3.51 | 15 | 15.84 | 0.48 | 2.8 |
| | *RAG* | 2 | 0.53 | 1 | 2.52 | 0.7 | 0.86 |
| | TB29 | 1 | 0.49 | 2 | 2.34 | 0.44 | 0.8 |
| | TB73 | 3 | 0.86 | 2 | 4.07 | 0.99 | 1.4 |
| | *TGFB2* | 2 | 1.4 | 8 | 7.26 | 0.49 | 2.4 |
| | | *blandingii* | | *orbicularis* | | | |
| *blandingii* vs. *orbicularis* | *Cytb* | 5 | 8.92 | 42 | 44.87 | 7.2 | 6.86 |
| Sum of deviations = 4 15.9769 | *RELN* | 5 | 4.98 | 28 | 25.81 | 0.78 | 3.91 |
| Degrees of freedom = 14 | *HNF-1α* | 0 | 2.08 | 14 | 11.13 | 0.49 | 1.68 |
| $P = 0.31$ | R35 | 6 | 4.95 | 27 | 25.65 | 0.57 | 3.89 |
| | RAG | 2 | 0.36 | 0 | 2.1 | 0.9 | 0.58 |
| | TB29 | 1 | 0.32 | 1 | 1.85 | 0.56 | 0.51 |
| | TB73 | 3 | 1.33 | 7 | 7.73 | 0.69 | 2.13 |
| | *TGFB2* | 2 | 1.07 | 6 | 5.86 | 0.2 | 1.65 |
| | | *marmorata* | | *orbicularis* | | | |
| *marmorata* vs. *orbicularis* | *Cytb* | 46 | 46.52 | 42 | 50.24 | 6.68 | 12.41 |
| Sum of deviations = 8.35148 | *RELN* | 16 | 19.22 | 28 | 21.71 | 0.66 | 5.28 |
| Degrees of freedom = 14 | *HNF-1α* | 22 | 15.87 | 14 | 18.24 | 1.22 | 4.4 |
| $P = 0.86$ | R35 | 15 | 18.27 | 27 | 20.98 | 0.84 | 5.07 |
| | RAG | 1 | 0.84 | 0 | 1.03 | 1.2 | 0.46 |
| | TB29 | 2 | 1.4 | 1 | 1.72 | 0.65 | 0.76 |
| | TB73 | 2 | 3.88 | 7 | 4.76 | 1.14 | 2.11 |
| | *TGFB2* | 8 | 5.99 | 6 | 6.32 | 0.42 | 2.97 |

Note: All tests indicated no significant deviation from neutrality. For these tests, *trinacris* was lumped with *orbicularis*.

be similar for both gene partitions, but the divergence between *blandingii* and *marmorata* should be very different, reflecting the timing of gene capture rather than the timing of species splitting. Our combined results provide strong, independent lines of nuclear evidence that *blandingii* is most likely the true sister species to *orbicularis*/*trinacris* and that the mtDNA-based relationship of *marmorata* and *blandingii* as sister taxa is an artifact.

Under the hypothesis that *marmorata* captured *blandingii*'s mtDNA during an ancient hybridization event, the *blandingii–orbicularis*/*trinacris* split (Nodes D1 and D2; Fig. 7) would not have been affected by *blandingii–marmorata* reticulation. Estimated mtDNA versus nuDNA divergence times at Nodes D1 and D2 are virtually identical (~17 million years each), and each falls well within 1 standard deviation of the other (Table 6). However, the divergence time estimates for *marmorata* from its sister group (Nodes C1 and C2) should have been affected by this reticulation event. The estimated ages of these 2 nodes differ by almost a factor of 2 (12 vs. 23 million years for mtDNA and nuDNA, respectively) and are more than 2 standard deviations different from each other (Table 6). We interpret this concordance between independent nuclear and mitochondrial gene trees for the *blandingii–orbicularis* divergence, but the much younger and phylogenetically different placement of *marmorata* in the mtDNA tree compared with the nuclear tree, as strong evidence that hybridization and subsequent capture of *blandingii* mtDNA by *marmorata* occurred about 12 Ma. Lineage sorting is an alternative explanation for the mtDNA tree, but this hypothesis does not predict identical date estimates for one, but not another, divergence time.

### *Historical Biogeography*

The temporal framework afforded by our estimated divergence times potentially sheds new light on the problematic evolutionary and biogeographic history of the *Emys* species complex. Based on our chronograms (Fig. 7), the divergence among ancestral *Emys* lineages occurred about 23 Ma (based on Node C2), with subsequent divergence of the ancestral *blandingii* and *orbicularis* lineages around 17 Ma (Nodes D1 and D2), followed by reticulation of *marmorata* with *blandingii* around 12 Ma (Node C1). Fossil and paleogeologic data provide additional relevant evidence. Fossil material assignable to the *Emys* species complex is fairly extensive, with at least 1 fossil taxon assigned to each of the 3 main *Emys* lineages. However, given the historical and current confusion over the taxonomy and interpretation of morphological characters within the group, the phylogenetic placement of fossil taxa assignable to the *Emys* species complex remains tentative. For example, Holman (1995) described *Emydoidea hutchisoni* from the medial Miocene of Nebraska, but later (Holman 2002) referred most of the material that formed the basis for the original description and redescription of *Emydoidea*

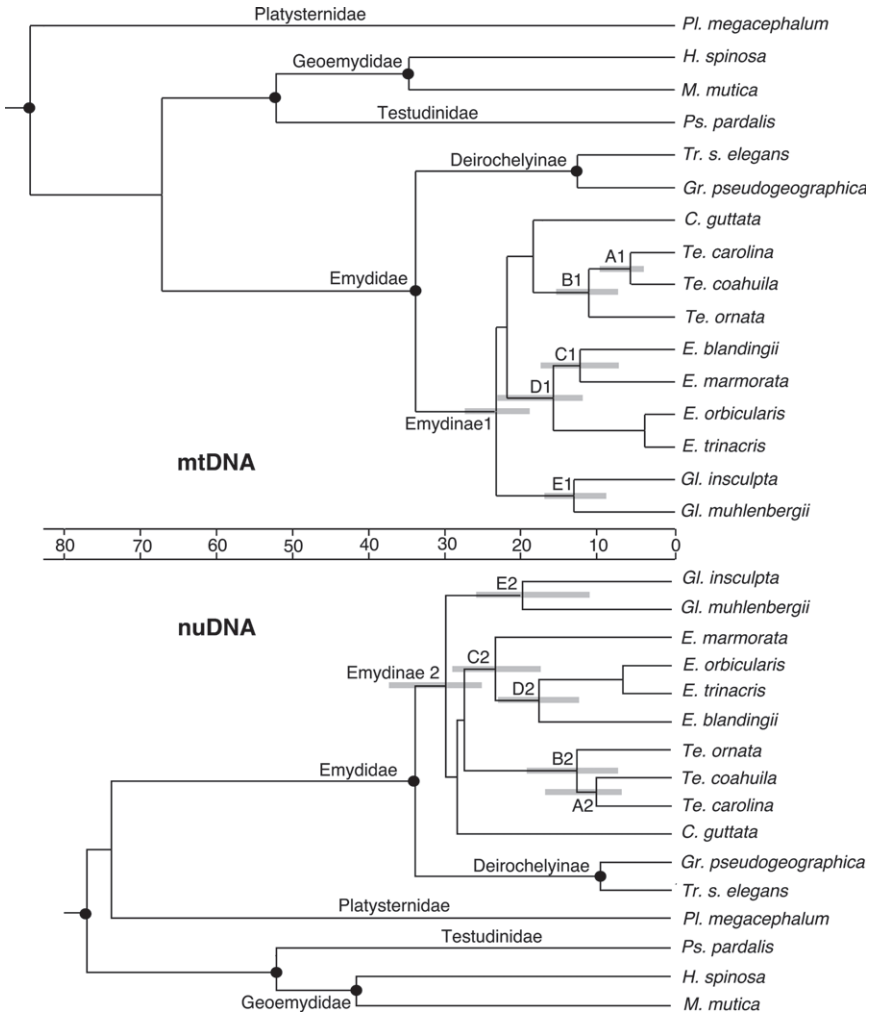


FIGURE 7. Time-calibrated phylogenies (chronograms) and estimated divergence times for the Emydinae (Table 6). The mitochondrial chronogram was generated from the 83-taxon *cytb* ML tree, whereas the nuclear chronogram was generated from the 26-taxon, 7-locus ML tree. Central scale is in millions of years. Fixed and constrained calibration are shown as darkened circles and include 1) the most recent common ancestor (MRCA) of Emydidae, Geoemydidae, Testudinidae, and Platysternidae constrained at 70 ± 6.37 Ma; 2) the MRCA of Geoemydidae + Testudinidae fixed at 52 Ma; 3) Geoemydidae: the MRCA of *Mauremys mutica* and *Heosemys spinosa*, constrained at 37.27 ± 5.78 Ma, Emydidae: the MRCA of *Trachemys s. elegans* and *Emys marmorata*, fixed at 34 Ma, and Deirochelyinae: the MRCA of *Trachemys s. elegans* and *Graptemys pseudogeographica*, constrained at 15.36 ± 3.16 Ma (Near et al. 2005, and references therein). Gray bars indicate error bars for divergence time estimates generated here.

*hutchisoni* to the genus *Chrysemys*, a distantly related extant deirochelyine genus.

What does seem clear from the fossil record is that the *Emys* complex was diversified, and present in both Europe and North America by about 15 Ma, consistent with our molecular estimate of 17 Ma for the split of European *orbicularis* from its sister group. Hutchison (1981) reported on a fragmentary *Emydoidea* hyoplastra that was "generalized enough" to form an ancestral morphotype of *blandingii* + *orbicularis* from the medial Barstovian (14–13 Ma) of Nebraska, and fossil material clearly assignable to extant *blandingii* is dated to ∼5–4 Ma (Holman 1995). The oldest fossil *Emys* in Europe were, interestingly enough, originally described as *Emydoidea*, including *Emydoidea taraschuki* from the middle to late Miocene of Ukraine (Chkhikvadze 1980, as cited in Danilov 2005) and *Emydoidea sukhanovi* from the late Miocene of eastern Europe (Chkhikvadze 1983, as cited in Fritz 1998). Fritz (1998) relegated these extinct taxa to the *E. orbicularis/trinacris* lineage, with the earliest date at 16–11 Ma based on *E. taraschuki* material. Fossils assignable to the *marmorata* lineage are much younger and include the extinct *Clemmys owyheensis* known from the Hemphillian Pliocene (∼5 Ma) of Oregon and Idaho (Brattstrom and Sturn 1959; Zug 1969), as well as Pliocene to recent Pleistocene *marmorata* fossils from California, Oregon, and Washington (Brattstrom and Sturn 1959; Gustafson 1978).

TABLE 6. Mean, SD, and 95% confidence interval of estimated divergence times for the chronogram in Figure 7

| Node | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|
| | | Cytochrome *b* | | |
| A1 | 6.3 | 1.1 | 4.1 | 9.8 |
| B1 | 10.9 | 1.67 | 7.6 | 15.5 |
| C1 | 12 | 1.94 | 7.8 | 17.6 |
| D1 | 16.6 | 2.1 | 12 | 23.2 |
| E1 | 13 | 1.57 | 9 | 17 |
| Emydinae 1 | 23 | 1.97 | 19 | 27.7 |
| | | Concatenated nuclear loci | | |
| A2 | 10.3 | 1.92 | 6.9 | 17.3 |
| B2 | 12.4 | 2.21 | 7.7 | 19.4 |
| C2 | 23.1 | 2.47 | 17.7 | 29.1 |
| D2 | 17.2 | 2.4 | 12.5 | 23.3 |
| E2 | 19.6 | 2.9 | 11.3 | 26 |
| Emydinae 2 | 29.4 | 2.01 | 25.2 | 37.7 |

Note: Time-calibrated dates were generated using fixed and constrained node dates from Near et al. (2005). SD = standard deviation.

### *Post Hoc Biogeographic Hypothesis*

Our phylogenetic results and estimated divergence times suggest 2 important dates in the history of *Emys* diversification and provide a temporal context for the diversification of Emydinae in North America. Two major historical warming episodes might have had a profound impact on the diversity and distribution of emydine turtles. Our divergence time estimates place the root diversification among the major emydine lineages somewhere around 29–23 Ma (Fig. 7 and Table 6), corresponding to the late Oligocene warming period ~27–24 Ma (Zachos et al. 2001). The extremely short branch lengths of the nuDNA phylogeny (Figs. 4–5) may indicate a rapid pulse of cladogenesis within Emydinae during this relatively warm period. This pattern of expansion and rapid diversification has been identified in a number of diverse taxa (Webb 1977; Sage 2001; Steiner et al. 2005), lending additional support to this interpretation.

The immigration of *Emys* into Europe ~17 Ma presumably occurred via the early to mid-Miocene trans-Beringian land bridge that was exposed ~24–14 Ma (Sanmartin et al. 2001; Burbrink and Lawson 2007). Alternate land routes that could have provided a dispersal route for *Emys* from the New to the Old World include the Thulean and DeGreer land bridges. However, these routes were open much earlier (~65–39 Ma), and the climate during this period was probably too cold for northerly turtle migrations (Burbrink and Lawson 2007). Finally, the putative *blandingii–marmorata* reticulation probably occurred sometime after the nuDNA *blandingii–orbicularis* split (~17 Ma) and prior to the mtDNA *blandingii–marmorata* split (~12 Ma). This period (17–12 Ma) closely coincides with the mid-Miocene climatic optimum (~18–14 Ma) when relatively warm and moist climatic conditions prevailed in North America (Zachos et al. 2001). Although hard fossil evidence is lacking, these would have been optimal conditions for *marmorata* and *blandingii* to expand their ranges, reestablish contact, and hybridize. As climates cooled and dried from the mid-Miocene to the Pleistocene, their ranges would have contracted to their current allopatric state.

Our hypothesis is complex, but it is consistent with the fossil record, molecular divergence time estimates, and the paleoclimatic record. In favor of the scenario of mitochondrial introgression, we note that instances of mitochondrial introgression/fixation among allopatrically distributed taxa are becoming more frequently observed in a variety of taxa (Bernatchez et al. 1995; Wilson and Bernatchez 1998; Peters et al. 2007; Robertson et al. 2006; Melo-Ferreira et al. 2005, 2007; Good et al. 2008).

## CONCLUSIONS: A TAXONOMIC PLEA

From a phylogenetic perspective, incongruent data partitions are generally viewed as a problem requiring resolution. However, if one is able to confidently reconstruct species relationships, conflicting data partitions can provide unique insights into ancient evolutionary events and processes. Even relatively ancient processes can be illuminated by the simultaneous analysis of mitochondrial and multiple nuclear sequences in a temporal framework. Our understanding of *Emys* biology, paleontology, and taxonomy should all be viewed in light of the probable reticulate history of this clade revealed by our analysis.

Our results also echo those from a growing body of evidence indicating the need to approach taxonomic revisions cautiously. The original allocation proposed by Holbrook (1842) of *blandingii* to the same genus as *orbicularis* now appears to accurately reflect phylogenetic history, whereas subsequent decisions to remove *blandingii* from *Emys* have obscured the close relationship of these broadly disjunct species (Parham and Feldman 2002; Spinks et al. 2004; Turtle Taxonomy Working Group 2007a). The evidence for *Emys* monophyly inclusive of *marmorata* is less clear because *marmorata* is probably carrying *blandingii* mitotypes, and we found no consistent or strong statistical support from analyses of nuclear loci for its phylogenetic placement. Given their apparently close phylogenetic affinities however, we feel the continued recognition of 3 genera (*Emydoidea* for *blandingii*, *Actinemys* for *marmorata*, and *Emys* for *orbicularis/trinacris*) obscures, rather than illuminates, the phylogenetic relationships and fascinating biogeographic history of these turtles. Perhaps most importantly, this case study indicates the importance of moving forward cautiously with new taxonomies and waiting until a full set of strongly supportive data are in hand before cluttering the taxonomic literature with new names and combinations (Parham et al. 2006).

## SUPPLEMENTARY MATERIAL

Supplementary appendices can be found at http://www.oxfordjournals.org/our journals/sysbio/.

## FUNDING

This work was supported by the National Science Foundation (DEB 0213155, DEB 0516475, DEB 0507916, DEB 0817042), and the University of California, Davis, Agricultural Experiment Station.

## ACKNOWLEDGMENTS

For tissue samples, we thank Dan Holland, Eskandar Pouyani (University of Heidelberg, Germany), Cesar Ayres, (Universitario A Xunqueira), Carla Cicero (Museum of Vertebrate Zoology, Berkeley), Raymond Farrell, Robert Zappalorti (Herpetological Consultants, Inc.), Charles Innis (New England Aquarium), John Iverson (Earlham College), Tibor Kovács and Susan McGaugh (Iowa State University), Anne Meylan (Florida Fish and Wildlife Commission), Peter Meylan (Eckerd College), Steve Mockford (Dalhousie University), Tamás Molnár (Kapsovár University), Matt Osentoski, and Chris Tabaka (Binder Park Zoo). J. Brown and A. Lemmon shared their manuscript (currently in review) that allowed us to calculate empirical branch length priors for our Bayesian analyses. For helpful discussion on the manuscript, we thank Bob Thomson, Justen Whittall, Dave Starkey, and members of the Shaffer Lab. This work benefited greatly from valuable comments provided by Jack Sullivan, Elizabeth Jockusch, and 2 anonymous reviewers.

## REFERENCES

Avise J.C., Arnold J., Ball R.M., Bermingham E., Lamb T., Nigel J.E., Reeb C.A., Saunders N.C. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Ann. Rev. Ecol. Syst. 18:489–522.

Ballard J.W.O., Kreitman M. 1995. Is mitochondrial DNA a strictly neutral marker? Trends. Evol. Ecol. 12:485–488.

Ballard J.W.O., Rand D.M. 2005. The population biology of mitochondrial DNA and its phylogenetic implications. Ann. Rev. Ecol. Evol. Syst. 36:621–642.

Ballard J.W.O., Whitlock M.C. 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13:729–744.

Bernatchez L., Glémet H., Wilson C.A., Danzmann R.G. 1995. Introgression and fixation of arctic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). Can. J. Fish. Aquat. Sci. 52:179–185.

Bhangale T.R., Rieder M.J., Livingston R.J., Nickerson D.A. 2005. Comprehensive identification and characterization of diallelic insertion-deletion polymorphisms in 330 human candidate genes. Hum. Mol. Gen. 14:59–69.

Bickham J.W., Lamb T., Minx P., Patton J.C. 1996. Molecular systematics of the genus *Clemmys* and the intergeneric relationships of emydid turtles. Herpetologica. 52:89–97.

Brattstrom B.H., Sturn A. 1959. A new species of fossil turtle from the Pliocene of Oregon, with notes on other fossil *Clemmys* from western North America. Bull. So. Cal. Acad. Sci. 58:65–71.

Brown J.M., Hedtke S.M., Lemmon A.R., Lemmon E.M. 2009. When trees grow too long: investigating the causes of highly inaccurate bayesian branch-length estimates. Syst. Biol. In press.

Buckley T.R., Cordeiro M., Marshall D.C., Simon C. 2006. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada* Dugdale). Syst. Biol. 55: 411–425.

Burbrink F.T., Lawson R. 2007. How and when did Old World ratsnakes disperse into the New World? Mol. Phylogenet. Evo. 43: 173–189.

Burke R.L., Leuteritz T.E., Wolf A.J. 1996. Phylogenetic relationships of emydine turtles. Herpetologica. 52:572–584.

Buschbom J., Barker D. 2006. Evolutionary history of vegetative reproduction in *Porpidia* s.l. (lichen-forming Ascomycota). Syst. Biol. 55:471–484.

Chan K.M.A., Levin S.A. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution. 59:720–729.

Chippindale P.T., Wiens J.J. 1994. Weighting, partitioning, and combining characters in phylogenetic analysis. Syst. Biol. 43: 278–287.

Conant R., Collins J.T. 1991. A field guide to reptiles and amphibians of eastern and central North America. Boston: Houghton Mifflin.

Danilov I.G. 2005. Die fossilen Schildkroten Europas. In: Fritz U., editor. Handbuch der Reptilien und Amphibien Europas. Schildkroten II (Cheloniidae, Dermochelyidae, Fossile Schildkroten). Wiebelsheim (Germany): Aula p. 329–441.

Degnan J.H., Rosenberg N.A. 2006. Discordance of species trees with their most likely gene trees. PLOS. Genet. 2:762–768.

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

Edwards S.V., Liu L., Pearl D.K. 2007. High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA. 104:5936–5941.

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

Ernst C.H., Barbour R.W. 1989. Turtles of the world. Washington (DC): Smithsonian Institution Press.

Ernst C.H., Lovich J.E., Barbour R.W. 1994. Turtles of the United States and Canada. Washington, D.C.: Smithsonian Institution Press.

Feldman C.R., Parham J.F. 2002. Molecular phylogenetics of emydine turtles: taxonomic revision and the evolution of shell kinesis. Mol. Phylogenet. Evol. 22:388–398.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 39:783–791.

Felsenstein J. 1988. Phylogenies from molecular sequences: inference and reliability. Ann. Rev. Genet. 22:521–565.

Felsenstein J. 1993. PHYLIP: phylogenetic inference package. Version 3.5d. Seattle (WA): University of Washington.

Ferris S.D., Sage R.D., Huang C., Nielsen J.T., Ritte U., Wilson A.C. 1983. Flow of mitochondrial DNA across a species boundary. Proc. Natl. Acad. Sci. USA. 80:2290–2294.

Fritz U. 1998. Introduction to zoogeography and subspecific differentiation in *Emys orbicularis* (Linnaeus, 1758). Proceedings of the *Emys* symposium; 1996 Oct. 4-6; Dresden, Germany. Rheinbach (Germany): Deutsche Gesellschaft fr Herpetologie und Terrarienkunde. p. 1–28.

Fritz U., d'Angelo S., Pennisi M.G., Lo Valvo M. 2006. Variation of Sicilian pond turtles, *Emys trinacris*—what makes a species cryptic? Amphibia-Reptilia. 27:513–529.

Fritz U., Fattizzo T., Guicking D., Tripepi S., Pennisi M.G., Lenk P., Roger U., Wink M. 2005. A new cryptic species of pond turtle from southern Italy, the hottest spot in the range of the genus *Emys* (Reptilia, Testudnies, Emydidae). Zool. Scripta. 34:351–371.

Fujita M.F., Engstrom T.N., Starkey D.E., Shaffer H.B. 2004. Turtle phylogeny: insights from a novel nuclear intron. Mol. Phylogenet. Evol. 31:1031–1040.

Funk D.J., Omland K.E. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Ann. Rev. Ecol. Evol. Syst. 34:397–423.

Gadagkar S.R., Rosenberg M.S, Kumar S. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Exp. Zool. 304B:64–74.

Gaffney E.S., Meylan P.A. 1988. A phylogeny of turtles. In: Benton M.J., editor. Phylogeny and classification of the tetrapods. Oxford: Clarendon Press.p. 157–219.

Gerber A.S., Loggins R., Kumar S., Dowling T.E. 2001. Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? Ann. Rev. Genet. 35:539–566.

Good J.M., Hird S., Reid N., Demboski J., Steppan S.J., Martin-Nims T.R., Sullivan J. 2008. Ancient hybridization and mitochondrial capture between two species of chipmunks. Mol. Ecol. 17: 1313–1327.

Graur D., Martin W. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends. Genet. 20:80–86.

Gustafson E.P. 1978. The vertebrate faunas of the Pliocene Ringold Formation, south-central Washington. Bull. Mus. Nat. Hist. Oregon. 23:1–62.

Hedges S.B., Kumar S. 2004. Precision of molecular time estimates. Trends. Genet. 20:242–247.

Hey J. 2004. HKA (a computer program for tests of natural selection). Available from: URL http://lifesci.rutgers.edu/heylab/~heylab/heylabsoftware.htm#HKA. Newark (NJ): Rutgers University.

Holbrook J.E. 1842. North American herpetology. Facsimile Reprints in Herpetology [reprinted in 1976 by Society for the Study of Amphibians and Reptiles]. Philadelphia (PA): J. Dobson.

Holder M.T., Anderson J.A., Holloway A.K. 2001. Difficulties in detecting hybridization. Syst. Biol. 50:978–982.

Holman J.A. 1995. A new species of *Emydoidea* (Reptilia: Testudines) from the late Barstovian (Medial Miocene) of Cherry County, Nebraska. J. Herpetol. 29:548–553.

Holman J.A. 2002. The status of *Emydoidea hutchisoni* Holman, 1995 (Testudines:Emydidae). Mich. Academician. 2:393–394.

Holman J.A., Fritz U. 2001. A new emydine species from the Medial Miocene (Barstovian) of Nebraska, USA with a new generic arrangement for the species of *Clemmys* sunsu McDowell (1964) (Reptilia: Testudines: Emydidae). Zool. Abh. 51:321–344.

Holman J.A., Parmlee D. 2005. Noteworthy turtle remains from the Late Miocene (Late Hemphillian) of northeastern Nebraska. Tex. J. Sci. 57:307–310.

Hudson R.R., Kreitman M., Aquadé M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116: 153–159.

Hudson R.R., Coyne J.A. 2002. Mathematical consequences of the genealogical species concept. Evolution. 56:1557–1565.

Huelsenbeck J.P., Rannala B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53:904–913.

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

Hutchison J. 1981. *Emydoidea* (Emydinae, Testudines) from the Barstovian (Miocene) of Nebraska. Paleobios. 37:1–6.

Kass R.E., Raftery A.E. 1995. Bayes factors. J. Am. Stat. Assoc. 90: 773–795.

Kellogg E.A., Appels R., Mason-Gamer R.J. 1996. When genes tell different stories: the diploid genera of Triticeae (Gramineae). Syst. Bot. 21:321–347.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

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

Leaché A.D., McGuire J.A. 2006. Phylogenetic relationships of horned lizards (*Phrynosoma*) based on nuclear and mitochondrial data: evidence for a misleading mitochondrial gene tree. Mol. Phylogenet. Evol. 39:628–644.

Lenk P., Fritz U., Joger U., Wink M. 1999. Mitochondrial phylogeography of the European pond turtle, *Emys orbicularis* (Linnaeus 1758). Mol. Ecol. 8:1911–1922.

Linnen C.R., Farrell B.D. 2007. Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera:Diprionidae) sawflies. Evolution. 61:1417–1438.

Liu L., Pearl D.K. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of species phylogeny using estimated gene tree distributions. Syst. Biol. 56:504–514.

Maddison W.P., Maddison D.R. 2003. MacClade. Sunderland (MA): Sinauer Associates.

Marshall D.C., Simon C., Buckley T.R. 2006. Accurate branch length estimation in partitioned Bayesian analyses requires accommodation of among-partition rate variation and attention to branch length priors. Syst. Biol. 55:993–1003.

Melo-Ferreira J., Boursot P., Randi E., Kryukov A., Suchentrunk F., Ferrand N., Alves P.C. 2007. The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula. Mol. Ecol. 16: 605–618.

Melo-Ferreira J., Boursot P., Suchentrunk F., Ferrand N., Alves P.C. 2005. Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in Northern Iberia. Mol. Ecol. 14:2459–2464.

Miller R.E., Buckley T.R., Manos P.S. 2002. An examination of the monophyly of morning glory taxa using Bayesian phylogenetic inference. Syst. Biol. 51:740–753.

Minin V., Abdo Z., Joyce P., Sullivan J. 2003. Performance-based selection of likelihood models for phylogeny estimation. Syst. Biol. 52:674–683.

Mockford S.W., Herman T.B., Snyder M., Wright J.M. 2006. Conservation genetics of Blanding's turtle (*Emydoidea blandingii*) and its application in the identification of evolutionarily significant units. Conserv. Gen. 8:209–219.

Mockford S.W., McEachern L., Herman T.B., Snyder M., Wright J.M. 2005. Population genetic structure of a disjunct population of Blanding's turtle (*Emydoidea blandingii*) in Nova Scotia. Biol. Conserv. 123:373–380.

Moritz C., Dowling T.E., Brown W.M. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann. Rev. Ecol. Syst. 18:269–292.

Near T.J., Meylan P.A., Shaffer H.B. 2005. Assessing concordance of fossil calibration points in molecular clock studies: An example using turtles. Am. Nat. 165:137–146.

Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Phil. Trans. R. Soc. Lond. B, 344:305–311.

Nylander J.A.A., Ronquist F., Huelsenbeck J.P., Nieves-Aldrey J.L. 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53:47–67.

Pamilo P., Nei M. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Parham J.F., Feldman C.R. 2002. Generic revisions of emydine turtles. Turtle Tortoise Newslett. 6:28–30.

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

Parham J.F., Türkozan O., Stuart B.L., Arakelyan M., Shafei S., Papenfuss T.J. 2006. Genetic evidence for premature taxonomic inflation in Middle Eastern tortoises. Proc. Cal. Acad. Sci. 57:955–964.

Peters J.L., Zhuravlev Y., Fefelov I., Logie A., Omland K.E. 2007. Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). Evolution 61:1992–2006.

Pollard D., Iyer V.N., Moses A.M., Eisen M.B. 2006. Whole genome phylogeny of the *Drosophila melanogaster* species subgroup: widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. PLoS. Genet. 2(10):e173.

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

Primmer C.R., Borge T., Lindell J., Sætre G.P. 2002. Single nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. Mol. Ecol. 11:603–612.

Rambaut A., Drummond A.J. 2007. Tracer version 1.4. Available from: URL http://beast.bio.ed.ac.uk/Tracer.

Rand D.M. 2001. The units of selection on mitochondrial DNA. Ann. Rev. Ecol. Syst. 32:415–448.

Robertson A.V., Ramsden C., Niedzwiecki J., Fu J., Bogart J.P. 2006. An unexpected recent ancestor of unisexual *Ambystoma*. Mol. Ecol. 15:3339–3351.

Ronquist F., Huelsenbeck J.P. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.

Sage R.F. 2001. Environmental and evolutionary preconditions for the origin and diversification of the $C_4$ photosynthetic syndrome. Plant. Biol. 3:202–213.

Sambrook J., Russell D.W. 2001. Molecular cloning: a laboratory manual. 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Sanderson M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

Sanderson M.J. 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., Shaffer H.B. 2002. Troubleshooting molecular phylogenetic analyses. Ann. Rev. Ecol. Syst. 33:49–72.

Sanmartin I., Enghoff H., Ronquist F. 2001. Patterns of animal dispersal, vicariance and diversification in the holarctic. Bio. J. Linn. Soc. 73:345–390.

Shaffer H.B., Clark J.M, Kraus F. 1991. When molecules and morphology clash: a phylogenetic analysis of the North American ambystomatid salamanders (Caudata: Ambystomatidae). Syst. Biol. 40:284–303.

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

Shaffer H.B., Thomson R.C. 2007. Using single nucleotide polymorphisms (SNPs) to delimit species in recent radiations. Syst. Biol. 56: 896–906.

Shaw K. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proc. Natl. Acad. Sci. USA. 25:16122–16127.

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

Spinks P.Q., Shaffer H.B. 2007. Conservation phylogenetics of the Asian box turtles (Geoemydidae, *Cuora*): mitochondrial introgression, numts, and inferences from multiple nuclear loci. Conserv. Genet. 8:641–657.

Spinks P.Q., Shaffer H.B., Iverson J.B., McCord W.P. 2004. Phylogenetic hypotheses for the turtle family Geoemydidae. Mol. Phylogenet. Evol. 32:164–182.

Stamatakis A., Hoover P., Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web-servers. Syst. Biol. 75: 758–771.

Steiner S., Tilak M., Douzery E.J.P., Catzeflis F.M. 2005. New DNA data from a transthyretin nuclear intron suggests an Oligocene to Miocene diversification of living South America opossums (Marsupialia: Didelphidae). Mol. Phylogenet. Evol. 35:363–379.

Stephens M., Donnelly P. 2003. A comparison of bayesian methods for haplotype reconstruction. Am. J. Hum. Genet. 73:1162–1169.

Stephens M., Smith N.J., Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

Stephens P.R., Wiens J.J. 2003. Ecological diversification and phylogeny of emydid turtles. Biol. J. Linn. Soc. 79:577–610.

Swofford D.L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Sunderland, MA; Sinauer Associates.

Thomson R.C., Shedlock A.M., Edwards S.V., Shaffer H.B. 2008. Developing markers for multilocus phylogenetics in non-model organisms: A test case with turtles. Mol. Phylogenet. Evol. 49:514–525.

Turtle Taxonomy Working Group. 2007a. An annotated list of modern turtle terminal taxa, with comments on areas of taxonomic instability and recent change. In: Shaffer H.B., FitzSimmons N.N., Georges A., Rhodin A.G.J., editors. Defining turtle diversity: Proceedings of a workshop on genetics, ethics, and taxonomy of freshwater turtles and tortoises. Chelonian Research Monographs No. 4. Lunenburg (MA): Chelonian Research Foundation. p. 173–199.

Turtle Taxonomy Working Group. 2007b. Turtle taxonomy: methodology, recommendations, and guidelines. In: Shaffer H.B., FitzSimmons N.N., Georges A., Rhodin A.G.J., editors. Defining turtle diversity: Proceedings of a workshop on genetics, ethics, and taxonomy of freshwater turtles and tortoises. Chelonian Research Monographs No. 4. Lunenburg (MA): Chelonian Research Foundation. p. 73–84.

Webb S.D. 1977. A history of savanna vertebrates in the New World. Part I: North America. Ann. Rev. Ecol. Syst. 8:355–380.

Wilson C.C., Bernatchez L. 1998. The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). Mol. Ecol. Notes. 7:127–132.

Yang Z., Rannala, B. 2005. Branch-Length Prior Influences Bayesian Posterior Probability of Phylogeny. Syst. Biol. 54:455–470.

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

Zhang A., Sota T. 2007. Nuclear gene sequences resolve species phylogeny and mitochondrial introgression in *Leptocarabus* beetles showing trans-species polymorphisms. Mol. Phylogenet. Evol. 45:534–546.

Zhang D.X., Hewitt G.M. 2003. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Mol. Ecol. 12:563–584.

Zug G.R. 1969. Fossil Chelonians, *Chrysemys* and *Clemmys,* from the Upper Pliocene of Idaho. Great. Basin. Nat. 29:82–87.

*First submitted 2 November 2007; reviews returned 7 February 2008; final acceptance 21 November 2008*
*Associate Editor: Elizabeth Jockusch*