Phylogeny and temporal diversification of the New World pond turtles (Emydidae)

Phillip Q. Spinks\textsuperscript{a,}\textdagger, Robert C. Thomson\textsuperscript{b}, Evan McCartney-Melstad\textsuperscript{a}, H. Bradley Shaffer\textsuperscript{a}

\textsuperscript{a}Department of Ecology and Evolutionary Biology and La Kretz Center for California Conservation Science, University of California, Los Angeles, CA 90095, USA
\textsuperscript{b}Department of Biology, University of Hawai'i at Mānoa, Honolulu, HI 96822, USA

\textbf{A R T I C L E I N F O}

\textbf{Article history:}
Received 18 November 2015
Revised 3 June 2016
Accepted 7 July 2016
Available online 9 July 2016

\textbf{Keywords:}
Species tree
Chronogram
Divergence time
Fossil constraints
Emydinae
Deirochelyinae

\textbf{A B S T R A C T}

We present a comprehensive multigene phylogeny and time tree for the turtle family Emydidae. Our phylogenetic analysis, based on 30 nuclear and four mitochondrial genes (23,330 total base pairs) sequenced for two individuals for each of the currently recognized species of the subfamily Emydinae and two species from each of the more species-rich Deirochelyinae genera, yielded a well-supported tree that provides an evolutionary framework for this well-studied clade and a basis for a stable taxonomy. We calibrated an emydid time tree using three well-vetted fossils, modeled uncertainty in fossil ages to reflect their accuracy in node dating, and extracted stem/crown ages of a number of key diversification events. We date the age of crown emydids at a relatively young 44 Ma, and the crown age of both contained subfamilies at roughly 30 Ma. One deirochelyine clade, which includes the genera Graptemys, Malaclemys, Pseudemys, and Trachemys and contains 11% of all turtle species, dates to 21 Ma just prior to the mid-Miocene climatic optimum, suggesting a potential causal link between warm, moist conditions and rapid species accumulation of these highly aquatic turtles. Both nuclear DNA data alone and in combination with mitochondrial DNA support the monophyly of an inclusive genus \textit{Emys} containing the old world species \textit{orbicularis} and \textit{trinacris} and the New World \textit{blandingii}, \textit{marmorata} and \textit{pallida}. Given that all members of this group were originally aligned in the genus \textit{Emys} and that the age of the clade is roughly equal to other emydine genera, we strongly support a classification that places these five species in a single genus rather than the alternative three-genus scheme (\textit{Emys} (\textit{orbicularis}, \textit{trinacris}), \textit{Actinemys} (\textit{marmorata}, \textit{pallida}), \textit{Emydoidea} (\textit{blandingii})). The phylogeny and resulting time tree presented here provides a comprehensive foundation for future comparative analyses of the Emydidae that will shed light on the historical ecology and conservation prioritization of this diverse chelonian clade.

© 2016 Elsevier Inc. All rights reserved.

\textbf{1. Introduction}

The increasing ease with which molecular sequence data can be collected is enabling the tree of life to be assembled at an ever-increasing rate with ever-larger data sets. However, while some branches of the tree of life are recovered with high confidence, others remain poorly resolved. It is these parts of the tree that continue to be challenging and often require multiple lines of evidence and large amounts of data coupled with diverse analytical approaches. One such challenging case is the New World pond turtles (family Emydidae). Emydids are semi or fully aquatic turtles, with 53 currently recognized species (Spinks et al., 2014; Turtle Taxonomy Working Group, 2014) that comprise about 16% of global turtle species richness. The family has been the focus of some of the most intensive long-term field studies of vertebrate population biology (Gibbons and Avery, 1990), aging (Congdon et al., 2003) and community ecology (Lindeman, 2000; Stephens and Wiens, 2009) conducted. It contains the first turtle to have its genome fully sequenced (the painted turtle \textit{Chrysemys picta}, Shaffer et al., 2013), the most widely farmed and invasive reptile (the red-eared slider \textit{Trachemys scripta elegans}; Kraus, 2009), and important models for the study of anoxia and mechanisms of sex determination (Bull and Vogt, 1979; Janzen, 1994; Johlin and Moreland, 1933; Ultsch and Jackson, 1982). The family is broadly distributed in North America north of Mexico, but also contains a few taxa that extend across the Greater Antilles, Mexico, Central and South America (Trachemys, Parham et al., 2013, 2015) and Europe (\textit{Emys orbicularis}/\textit{trinacris}, Rogn, 2009). Roughly two thirds of the 53 species fall into one of the IUCN endangerment categories (Turtle

\textsuperscript{\dagger} Corresponding author at: Department of Ecology and Evolutionary Biology, University of California, 621 Charles E. Young Dr. South, Los Angeles, CA 90095-1606, USA.

\textit{E-mail address: pqspinks@ucla.edu} (P.Q. Spinks).
Taxonomy Working Group, 2014), making them an important conserva-
tion target at a global scale.

Although it has been the subject of several recent morphologi-
cal and molecular analyses, many aspects of emydid evolutionary
history remain poorly resolved (Feldman and Parham, 2002; 
Guillon et al., 2012; Parham et al., 2013; Spinks et al., 2013; 
Spinks and Shaffer, 2009; Stephens and Wiens, 2003; Wiens et al., 2010). Given the importance of a strong phylogeny for com-
parative inference and taxonomic stability, resolving the emydid
phylogeny is a critical component of their continued importance
in evolutionary and ecological studies, and conservation.

Emydidae has universally been divided into two reciprocally
monophyletic lineages generally recognized as the subfamilies
Deirochelyinae and Emydinae (Gaffney and Meylan, 1988). 
Deirochelyinae includes six geographically widespread, polytypic
and morphologically variable genera that encompasses most of
the species richness in the family (about 42 recognized species).
Generic boundaries have been stable for the last several decades,
although their interrelationships, numbers of contained species,
and interspecific relationships remain elusive and often con-
tentious. For example, the 14 currently recognized species of
map turtles (genus Graptemys) are often morphologically distinct,
but exhibit very shallow levels of genetic divergence (Ennen et al.,
2010; Lamb et al., 1994) that has stymied most efforts at phylo-
genetic reconstruction. The taxonomy and species composition
of the two other species-rich deirochelyine genera, Trachemys and
Pseudemys (the sliders with 16 species, and river cooters with eight
species, respectively) are also unsettled, potentially reflecting
recent hybridization and introgression that may have resulted in
unintentional taxonomic inflation (Jackson et al., 2012; Parham
et al., 2006a, 2013; Spinks et al., 2013). In contrast, species composi-
tion and delineation within Emydinae is modest (about 11 spe-
cies are generally recognized) and relatively uncontroversial with
the sole exception of the genus Terrapene (Fritz and Hasas, 2014; 
Martin et al., 2013). Intergeneric relationships among the Emydi-
nae, however, continue to thwart resolution with available DNA
data, leading to considerable disagreement on the resulting classi-
fication (Angielczyk et al., 2011; Bickham et al., 1996; Burke et al.,
1996; Feldman and Parham, 2002; Holman and Fritz, 2001; Spinks and
Shaffer, 2009).

One of the greatest stumbling blocks to the phylogenetic resolu-
tion of Emydidae has been the discordance between different data
sets. Phylogenies generated from mitochondrial DNA (mtDNA),
nuclear DNA (nuDNA) plus insertion/deletion characters, and mor-
phology alone or in combination are often discordant (Spinks et al.,
2009; Stephens and Wiens, 2003; Wiens et al., 2010), and mul-
tilocus phylogenies generated from more than one exemplar/specie-
tend to be statistically well supported but different from those
generated using single exemplar sampling (Spinks et al., 2009; 
Wiens et al., 2010). Among closely related groups the individuals
selected for analyses can have a dramatic impact on phylogenetic
topology (Shaw and Small, 2005; Spinks et al., 2013). For example,
contrary to initial expectations, when Spinks et al. (2013) generated
phylogenies from 10 concatenated nuclear loci sampled from >3
individuals/species for the genus Pseudemys (86 individuals in total),
they recovered a poorly resolved tree with no support for the mono-
phly of most species or their interrelationships. However, using
randomly chosen single exemplars drawn from this data set for each
species lineage, they found strong support for most relationships but
little consistency among trees when different individuals were
included. These results indicate that the phylogeny of Pseudemys is
unstable and suggest that other emydid lineages may be similarly
influenced by the individuals selected to represent each putative
species (Spinks et al., 2013).

Phylogenetic discord such as that seen across Emydidae is com-
mon across the tree of life and can be due to confounding biological
processes including incomplete lineage sorting or horizontal gene
transfer (Maddison, 1997; Sang and Zhong, 2000; Kubatko, 2009),
or methodological issues such as model misspecification (Posada
and Buckley, 2004; Tamura, 1994; Yang et al., 1995) or data align-
ment errors (Thorne and Kishino, 1992; Ogden and Rosenberg,
2006). Given their importance in comparative ecology, evolution
and conservation biology, we assembled three data sets to further
resolve the phylogeny and estimate divergence times among the
Emydidae. We focused on intergeneric relationships of the family
and species-level analyses among members of the subfamily Emy-
dinae. Our molecular data consists of four mitochondrial genes
(mtDNA) and 30 nuclear loci generated for 42 taxa (41 ingroup,
one outgroup). We performed Bayesian phylogenetic analyses,
generated species trees, and estimated divergence times among
genera and many emydid species. Our results demonstrate that
most intergeneric relationships within the Emydidae are now com-
ing into focus, as are species-level relationships in the historically
problematic Emydinae. We also provide and discuss divergence
time estimates for the origin of all major emydid lineages based
on a fossil-calibrated time tree for the family.

2. Materials and methods

2.1. Taxon and marker sampling

Taxon sampling consisted of 42 individuals subsampled from all
emydid genera (41 individuals). These 41 samples included two
samples/species except for Glyptemys muhlenbergii (one individ-
ual), and a single Platysternon megacephalum as the outgroup taxon
to the Emydidae (Parham et al., 2006b). Our species sampling of
Graptemys, Pseudemys, and Trachemys is not comprehensive
because these groups are characterized by extremely low levels of
intraspecific genetic divergence and relatively poorly-
delimited species boundaries (Lamb et al., 1994; Thomson unpub-
lished; Spinks et al., 2013; Parham et al., 2013, 2015), and will
require extensive taxon and data sampling for complete resolution.

As a consequence, we cannot make accurate inferences about the
crown age of these genera, although we can still infer stem diver-
gence times between them. DNA was extracted from blood or soft

tissue samples using a salt extraction protocol (Sambrook and
Russell, 1989); see Spinks et al. (2014) for primers and PCR condi-
tions. All PCR products were sequenced by Beckman Coulter Geno-

comics (http://www.beckmangenomics.com/). Chromatograms
indicating heterozygous length polymorphisms (Bhangale et al.,
2005) were encountered for several individuals and we used the
Indelignent v.1.2 software (Dmitriev and Rakitov, 2008, http://

tap.inhs.uiuc.edu/dmitriev/indel.asp) to reconstruct nucleotide
sequences from chromatograms disrupted by heterozygous length
polymorphisms.

2.2. Phylogenetic and divergence time analyses

Alignments were carried out using the MAFFT software (Katoh
et al., 2002) implemented in Geneious v5.1 (Drummond et al.,
2011). Coding regions were translated using Geneious v5.1 to
check for pseudogenes; none were found. We generated phyloge-
genies for the mtDNA and nuDNA data sets individually (four and
30 partitions, respectively) and combined (34 partitions) and esti-
mated divergence times for the combined data set under a Baye-
sian framework using BEAST v.1.8.2 (Drummond and Rambaut,
2007). For these analyses, the data were partitioned by mtDNA
gene (4 partitions) and nuclear locus (30 partitions) and combined
mtDNA + nuDNA (34 partitions). We used an independent HKY
model of sequence evolution for each partition. For divergence
time analyses, we employed the uncorrelated log-normal clock
model (UCLN) with the Yule stochastic branching process prior, and default priors for the remaining parameters except a uniform prior for the UCLN clock mean. We ran the analysis for 50,000,000 generations sampling every 5000 generations and discarding the first 25% of samples as burnin. Log files were examined for satisfactory mixing of the MCMC chains and to determine effective sample sizes (ESS) of >200 using Tracer v.1.6 (Rambaut and Drummond, 2007). We replicated each analysis five times, removed burnin and combined tree files by hand, and generated maximum clade credibility trees using TreeAnnotator v1.8.2. Most of the theses analyses were carried out through the CIPRES Web portal (Miller et al., 2010). In addition, we also generated Bayesian phylogenies for each locus individually using MrBayes v3.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003). The single-locus Bayesian analyses consisted of two independent runs comprising four incrementally heated chains that ran for 50,000,000 generations. We sampled the posterior distribution every 5000 generations, and checked for stationarity by ensuring that the average standard deviation of split frequencies between independent runs approached 0 and the potential scale reduction factor equaled 1. The MCMC samples from all Bayesian analyses were examined in Tracer v1.6 to ensure that all chains were sampling from the same target distribution and that the 25% burnin period was adequate.

To estimate divergence times, we employed three uniform fossil calibration priors defined and justified by Joyce et al. (2013), “Chrysemys” antiqua is the oldest fossil that definitively falls within Emydidae and was used to generate a minimum age estimate for both Emydidae and Emysternia (Emydidae + P. megacephalum, Crawford et al., 2015). The maximum age for Emydidae is less easily defined (Joyce et al., 2013), and we follow Joyce et al. (2013) in using the lindholmemydid taxon Pseudochrysemys gobiensis to place a maximum stem age on the group. Finally, we follow Joyce et al. (2013) in using the fossil taxon “Pseudemys idahoensis” (Gilmore, 1933) as a minimum constraint on the clade consisting of Graptemys + Malaclemys + Trachemys. For detailed discussion and justification of these fossil calibration points, see Joyce et al. (2013). The xml file used for this analysis is available from Dryad (doi: 10.5061/dryad.j47h4).

2.3. Species tree reconstruction

We used the multispecies coalescent model implemented in *BEAST* v1.8.2 (Drummond et al., 2012; Heled and Drummond, 2010) to estimate species trees. We analyzed the data as fully partitioned including the mtDNA data as a single partition (31 partitions total) and included the tree generated from Bayesian analyses of the concatenated nuDNA as the starting tree. We performed initial analyses using the species tree Yule process for the species tree prior and piecewise linear and constant root for the population size model. We iterated through various combinations of molecular substitution and clock models to determine an appropriate model under which the MCMC would mix adequately. Analyses were run for up to 400,000,000 generations sampling every 40,000 generations. We used Tracer (Rambaut and Drummond, 2007) to assess the ESS of the posterior samples. Minimally, the first 25% of samples were discarded as burnin.

Finally, we used the pseudo-likelihood method implemented in MP-EST v. 1.4 (Liu et al., 2010) to estimate a species phylogeny. We used the seqboot module of the Phylip software package (Felsenstein, 2005) to generate 100 bootstrap replicate nucleotide sequence alignments for the mitochondrial and all nuclear loci, and then performed Bayesian phylogenetic analyses on each pseudoreplicate alignment using MrBayes, and the models/settings described previously. This resulted in a total of 3100 analyses that were carried out on a local dual Xeon E5-2630v3 server. We performed analyses on all 100 pseudoreplicate sets of gene trees using MP-EST, and then generated a 50% majority rule consensus tree from the 100 MP-EST trees using the Phyutility software (Smith and Dunn, 2008).

3. Results

3.1. mtDNA phylogeny

Our mtDNA data set was composed of 42 individuals (41 emydids, and one Platysternon outgroup) and up to 2984 base pairs (bp) generated from gene fragments of COI, CYTB, DLOOP, and ND4 (Appendix S1). The matrix contained no missing sequences and ~0.4% missing data (Dryad # doi: 10.5061/dryad.j47h4), and all new sequences were submitted to GenBank (Appendix S1). The maximum clade credibility tree from the Bayesian analysis was well supported across all but four nodes that had Bayesian posterior probabilities (BPP) < 0.95 (Fig. 1). Relationships among emydid genera are consistent with previous mitochondrial analyses (Feldman and Parham, 2002; Spinks and Shaffer, 2009; Spinks et al., 2009; Wiens et al., 2010), but the phylogenetic placement of deirochelyine taxa differed from previous mtDNA analyses of the Emydidae (Fig. 1). In particular, our mtDNA tree recovered Graptemys as paraphyletic with respect to Malaclemys + Trachemys with strong support (Fig. 1). This novel result is at odds with previous analyses (e.g., Spinks et al., 2009; Stephens and Wiens, 2003; Wiens et al., 2010), and with the generally accepted concept of a monophyletic Graptemys. Given the incongruence between mtDNA and nuDNA phylogenies for emydid, the non-monophyly of Graptemys, and the inconsistent placement of Chrysemys and Deirochelys across trees and studies, we consider phylogenies generated from mtDNA only to be generally unreliable phylogenetic hypotheses for the Emydidae.

3.2. Single-locus nuDNA gene trees

Our nuclear loci ranged in size from 496 to 1038 bp (X = 678 bp). All sequences generated here were submitted to GenBank (Appendix S1). Gene trees from individual loci varied greatly in their level of phylogenetic resolution and associated support values, but are mostly consistent with previous analyses (Table S1, see Figs. S1–S8, Supplementary Information for gene trees). For example, 10/30 gene trees recovered the Deirochelyinae–Emydinae split. Across all 30 genes, the number of genera recovered as monophyletic with strong support varied from no genera supported (TB69, ZFP36L) to seven (HMGB2, Spin, Table S1). In addition, the number of strongly supported nodes with BPP ≥ 0.95 ranging from a low of one (ZFP36L) to 24 nodes supported (VIM) (Table S1, Figs. S1–S8).

3.3. Concatenated nuDNA phylogeny

Our nuDNA data set consisted of up to 21,047 bp of aligned sequence data. However we deleted numerous phylogenetically uninformative gaps prior to analyses, decreasing the alignment to 20,346 bp. This matrix contained five missing sequences and 1.7% missing data (Dryad doi: 10.5061/dryad.j47h4). The phylogeny recovered from analyses of these data was completely resolved and well supported at all but two nodes (BPP = 1.0 for 37/40 nodes, Fig. 2). Relationships among emydid and deirochelyine taxa recovered here are incongruent with the nuDNA results of Wiens et al. (2010) that was based on six nuclear loci and single exemplar taxon sampling. In addition, relationships among deirochelyine taxa recovered in our previous analyses based on seven nuclear loci, but multiple samples/species (Spinks et al., 2009) are
somewhat incongruent with those reported here, although those previous analyses were not well supported statistically. For example, Deirochelys is well supported as the sister group to the remaining deirochelyines in the current analyses while Spinks et al. (2009) recovered Chrysemys as sister to the remaining deirochelyines but without strong support. Likewise, Trachemys was paraphyletic but without support in the analysis of Spinks et al. (2009) while Trachemys is monophyletic with strong support in the current analysis.

Relationships among the Emydinae are identical to our previous analyses (e.g., Spinks et al., 2009), including the placement of the problematic Clemmys guttata is the sister group to an Emys + Terrapene clade, and Emys marmorata + Emys pallida as the sister group to the remaining Emys (including Emys blandingii, Fig. 2).

### 3.4. Concatenated mtDNA + nuDNA phylogeny

Our combined mtDNA + nuDNA data set consisted of up to 23,330 bp from four mitochondrial genes and 30 nuclear loci. This matrix contained five missing sequences and ~1.5% missing data (Dryad doi: 10.5061/dryad.j47h4). The phylogeny recovered from analyses of these data was completely resolved and supported...
with BPP = 1.0 at all nodes (Fig. 3), but relationships among the subfamily Deirochelyinae are largely incongruent with results from previous analyses (i.e. Spinks et al., 2009; Wiens et al., 2010) except for the placement of Trachemys as sister to Graptemys + Malaclemys (recovered here and by Wiens et al., 2010, although Wiens et al. (2010) recovered Trachemys as paraphyletic with respect to Graptemys and Malaclemys, see their Fig. 3). Relationships among the Emydinae, however are mostly consistent with previous analyses (e.g., Spinks et al., 2009; Wiens et al., 2010) except for the unstable position of C. guttata. In the combined mtDNA + nuDNA analyses, we recover C. guttata as the sister group to Terrapene, but C. guttata has also been recovered as the sister group to Emys + Terrapene (Spinks et al., 2009) or as most closely related to E. marmorata + E. pallida (Wiens et al., 2010). In addition, we continue to recover E. marmorata + E. pallida as sister to E. blandingii + the European species (E. orbicularis + E. trinacris) clade (Fig. 3), in contrast to the reciprocal monophyly of the North American (E. blandingii, E. marmorata, E. pallida) and European taxa (E. orbicularis, E. trinacris) based on analyses of mtDNA only (Fig. 1).

3.5. Multilocus species tree reconstruction

Despite extensive effort, we were unable to obtain a stable estimate of the posterior distribution from the BEAST analyses. We employed several strategies in an attempt to obtain these estimates including (1) simplifying substitution models for most partitions, (2) simplifying clock models (i.e. using strict molecular clocks) for several partitions, and (3) increasing the number of
generations/analysis. Some analyses appeared to reach stationarity for most parameters, but in these analyses the ESS for some parameters remained <200, and the estimated species tree was highly incongruent with respect to the vast majority of previous phylogenetic analyses of the Emydidae (not shown). Thus, we regard the distribution of trees from this analysis as inaccurate, probably resulting from inadequate mixing of the MCMC and an ultimate failure of the analysis to converge.

Results from the MP-EST analyses, however, returned trees that were much more consistent with other analyses and our current understanding of emydid relationships. We recovered a fully resolved and well-supported Deirochelyinae, but relationships among emydine genera were unresolved using this approach (Fig. 4).

3.6. Divergence time analyses

We recovered the origin of crown Emysternia (node 1) in the Eocene at ~52.49 Ma, crown Emydidae (node 2) in the Eocene (~41.79 Ma) and the subfamilies Deirochelyinae (node 10) and Emydinae (node 3) each diverged roughly synchronously in the Oligocene (~28.64 Ma and ~31.08 Ma, respectively; all node numbers refer to Fig. 5). The Emydinae consists of three deeply divergent and relatively ancient lineages comprised of Glyptemys (node 8, ~17.35 Ma), Terrapene + Clemmys (node 5, ~22.38 Ma) and Emys (node 6, ~19.29 Ma). The phylogenetic diversification pattern of Deirochelyinae is more pectinate, with sequential diversification (from basal to terminal) of Deirochelys (node 10, ~31.08 Ma) followed by Chrysemys (node 11, ~24.46 Ma), Pseudemys (node 12,
Fig. 4. Consensus tree with bootstrap support values for the 31-gene dataset estimated using MP-EST and the 3100 bootstrap gene trees from the MrBayes analyses. The branch lengths for each species are not estimable due to a lack of topological variance among gene tree triplets (see the MP-EST v1.4 manual). See Appendix S1 for specimen information.

~20.91 Ma), and the *Graptemys* + *Malaclemys* + *Trachemys* clade (node 13, ~15.6 Ma). In sharp contrast to the deepest nodes, the more shallow nodes tend to have relatively narrow 95% highest posterior density (HPD) values, consistent with the interpretation that most emydid species arose within the last 6 Ma during the Neogene (Fig. 5).

4. Discussion

The Emydidae has been the focus of extensive ecological and evolutionary research that relies on accurate species delimitation and phylogeny reconstruction, but assessing the tempo, mode and rate of evolutionary diversification has been hindered by the unsettled phylogeny of the group. Our analyses clearly indicate that many of the relationships among emydid genera are falling into place, as indicated by concordance among data sets and strong statistical support within analyses. A few intergeneric relationships, and a number of groupings within genera, remain unresolved, and constitute important areas for future research.

Consensus and congruence across the emydid tree vary widely among the phylogenies presented here and several previous analyses. For example, phylogenies generated from morphological characters recover the traditional Deirochelyinae/Emydinae division of the family, but are otherwise often incongruent with those based on molecular characters (Stephens and Wiens, 2003). In a similar vein, phylogenies generated from analyses of up to seven concatenated nuDNA loci are inconsistent with one another and often depend on the loci analyzed and the depth of taxon sampling (Spinks and Shaffer, 2009; Spinks et al., 2009; Wiens et al., 2010).

Several of the most problematic phylogenetic areas of the emydid tree at the generic level have centered on the subfamily Emydinae. Among those, perhaps the most consistently intractable has
been the shifting placement of the spotted turtle C. guttata. Our current work indicates that this is a function of very short internodes near the root of the Emydinae (Figs. 2 and 3); given the conflicting placements of C. guttata based on mtDNA and nuDNA, we remain uncertain concerning its final placement as the sister group of Terrapene (Fig. 3, mtDNA + nuDNA) or of all emydines except Glyptemys (nuDNA only, Fig. 2). The resolution of another persistent challenge in emydine phylogeny, the monophyly of Emys (in the sense of Feldman and Parham, 2002), is strongly supported by our expanded mitochondrial and nuclear data sets, as are the relationships among the five contained species (identical in Figs. 2 and 3). Finally, although relevant branch lengths at the base of Emydinae are short, we consistently resolve Glyptemys as the sister-group to the remaining Emydinae (Figs. 2 and 3). Formal species tree analyses would constitute strong corroborative evidence on the phylogeny of emydids, particularly for the Emydinae given our near-exhaustive taxon sampling. Unfortunately, species tree analyses either failed to converge (+BEAST) or failed to resolve relationships among many genera (MP-EST). Recalcitrant nodes with low support values (Fig. 4) are often indicative of relatively short speciation intervals characterized by sequence data with few informative mutations (Lanier and Knowles, 2015), suggesting that resolution of the emydid phylogeny using species tree methods might require additional data, or may prove impervious to species tree methods. On the other hand, analyses of the combined mtDNA + nuDNA data provide a well-supported genus-level phylogeny that is often in agreement with previous analyses (e.g., Angelczyk et al., 2011; Feldman and Parham, 2002; Spinks and Shaffer, 2009). Given these results, we view the tree from analysis of our combined mtDNA + nuDNA data (Fig. 3) as the most reliable currently available estimate of emydid phylogeny, and we use this phylogeny and resulting divergence time estimates to help illuminate the tempo of diversification within Emydidae and to inform emydid taxonomy.

4.1. Divergence time estimates

Estimating divergence times for the primary lineages of crown Testudines has gained increased attention in recent years (e.g., Dornburg et al., 2011; Fritz et al., 2011a; Heath, 2012; Lourenço et al., 2012; Naro-Maciel et al., 2008; Near et al., 2005; Shaffer et al., 1997; Spinks and Shaffer, 2009; Sterli et al., 2013; Werneburg et al., 2015). In general, divergence time estimates for all but the most recent nodes in the current analysis are characterized by wide 95% HPD values, demonstrating the uncertainty surrounding many of these divergence time estimates (Fig. 5, Table 1). On the other hand, for comparable nodes there is a great deal of similarity among the various analyses that are based on
different combinations of molecular markers, fossil constraints, taxon sampling and methodological approach (Tables 1 and 2). Thus, although there is uncertainty, there is also growing consensus for some divergence times within the Emydidae and we consider these estimates to be useful for comparisons with morphology-based estimates of phylogeny and temporal diversification.

Mean divergence time estimates for Emy sternia are inconsistent among results from the current study and previous analyses ranging from \(~52\) Ma to \(90\) Ma (Table 1). On the other hand, estimates for the origin of crown Emydidae are relatively consistent ranging from \(41.79\) to \(56.2\) (Table 1). In addition, divergence times for several more inclusive emydid clades are consistent across analyses and in line with the palaeontological record. For example, the Emydidae originated in North America and are probably derived from lindholmemydid-like ancestors that were common in Asia during the Cretaceous–Paleocene (Claude and Tong, 2004). The earliest fossil taxa that can be confidently placed within the emydid crown group come from the Eocene (\(~55\) Ma [Holroyd et al., 2001; Sukhanov, 2000]), and divergence time estimates place the origin of the Emydidae well within this timeframe (Table 1). Within the Emydidae, our analysis places the crown ages of Deirochelyinae and Emydinae in the Oligocene (\(31.08\) and \(28.84\) Ma, respectively), and we recovered the most recent common ancestor of Graptemys + Malaclemys + Trachemys at \(15.6\) Ma which is nearly identical to the estimates generated by Fritz et al. (2011a) and Joyce et al. (2013) (\(15.13\) and \(14.51\), respectively). Within the Emydinae, divergence times generated here are very similar to those reported in Spinks and Shaffer (2009). Given that the nucleotide data utilized in Spinks and Shaffer (2009) were included in the current analysis, this might be expected, although the fossil constraints employed in Spinks and Shaffer (2009) were different from those used here and may be unreliable (Parham and Irmis, 2008) (Table 2).

A particularly important node age recovered here is the relatively recent age for one of the largest radiations of living turtles. Three deirochelyine genera including Graptemys (14 species), Pseudemys (8 species) and Trachemys (\(~16\) species) together comprise \(~11\%\) of extant turtle species richness, and our results place the origin of this large clade (node 12, Fig. 5) at \(~20.91\) Ma just prior to the mid Miocene climatic optimum (\(15–17\) Ma, Zachos et al., 2001). All deirochelyine taxa are highly aquatic freshwater species while the Emydinae contain aquatic, semi-aquatic, and terrestrial lineages (Ernst and Barbour, 1989). Although our age estimates have a broad posterior distribution, the relatively warm wet temperatures during the mid Miocene climatic optimum might have supported aquatic turtle diversification and thus could help explain the extensive diversification within Deirochelyinae. We should note, however, that if this explanation applies across emydids, it would also predict more extensive species diversification within Emydys.

At the species level, our sampling is relatively sparse, but our current results suggest that many species diverged within the Pliocene (\(~5.3–2.6\) Ma), a pattern found in disparate North American taxa (e.g., Avise et al., 1992; Bryson et al., 2012; Derkarabetian et al., 2011; Moreno-Letelier et al., 2014). Additional rangewide sampling is necessary to determine whether there have been synchronous bouts of speciation that characterize disparate emydid lineages.

4.2. Taxonomic notes

Delimiting species and reconstructing resultant species trees for the deirochelyine genera Graptemys, Pseudemys and Trachemys continue to be the most challenging aspects of emydid systematics and taxonomy and accurate species delimitation in these groups will require analyses based on extensive within-taxon sampling and much larger data sets. Within Emydinae, several taxonomic issues have received a considerable amount of attention in the recent literature, and our results can be brought to bear on these issues. However, we emphasize that a stable taxonomy requires more than just adequate data and clear phylogenetic resolution; it also requires that the community adopt a uniform position on which of the available alternatives is most attractive (or least offensive). We offer our views here, in the hopes that it will help our community come to resolution on a stable, long-lasting taxonomy.

4.2.1. The genus Emydys and the classification of E. blandingii

The North American species blandingii was originally described and assigned to the genus Emydys by Holbrook (1838). However, Loveridge and Williams (1957), based on their interpretations of morphological characters, suggested that blandingii was more closely related to the southeastern US Deirochelys than to the Eurasian Emydys and placed blandingii in the monotypic genus Emydioidea. However, a critical and overlooked issue is that the phylogenetic and taxonomic revisions proposed by Loveridge and Williams (1957) were based on their hypothesized phylogeny for some cryptodirans that was not based on any formal phylogenetic analysis, and their phylogeny is dramatically different than the modern consensus understanding of these relationships (Fig. 2 in Loveridge and Williams, 1957). Bramble (1974) reanalyzed the

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Node</td>
<td>Mean 95% HPD</td>
<td>Mean 95% HPD</td>
<td>Mean 95% HPD</td>
<td>Mean 95% HPD</td>
</tr>
<tr>
<td>Emydernia</td>
<td>1</td>
<td>52.49 32.83–88.69</td>
<td>– –</td>
<td>82.43 65.58–96.53</td>
<td>– – – –</td>
</tr>
<tr>
<td>Emydinae</td>
<td>2</td>
<td>41.79 32–71.61</td>
<td>– –</td>
<td>44.15 32.14–55.47</td>
<td>– – – –</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>28.84 17.98–51.13</td>
<td>– –</td>
<td>29.4 25.2–37.7</td>
<td>– – – –</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>25.44 15.61–45.03</td>
<td>– –</td>
<td>– – – –</td>
<td>– – – –</td>
</tr>
<tr>
<td>Clemmys + Terrapene</td>
<td>5</td>
<td>22.38 13.33–39.8</td>
<td>– –</td>
<td>– – – –</td>
<td>– – – –</td>
</tr>
<tr>
<td>Emys</td>
<td>6</td>
<td>19.29 10.87–34.62</td>
<td>– –</td>
<td>– – – –</td>
<td>– – – –</td>
</tr>
<tr>
<td>Glyptemys</td>
<td>7</td>
<td>16.35 8.61–29.74</td>
<td>– –</td>
<td>– – – –</td>
<td>– – – –</td>
</tr>
<tr>
<td>Terrapene</td>
<td>8</td>
<td>17.35 8.04–32.79</td>
<td>– –</td>
<td>– – – –</td>
<td>– – – –</td>
</tr>
<tr>
<td>Deirochelyinae</td>
<td>9</td>
<td>14.25 8.29–25.59</td>
<td>– –</td>
<td>– – – –</td>
<td>– – – –</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>22.5 13.33–39.8</td>
<td>– –</td>
<td>– – – –</td>
<td>– – – –</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>16.31 8.61–29.74</td>
<td>– –</td>
<td>– – – –</td>
<td>– – – –</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>14.51 9.7–27.64</td>
<td>– –</td>
<td>– – – –</td>
<td>– – – –</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>13.61 8.18–24.1</td>
<td>10.04 6.48–13.85</td>
<td>– – – –</td>
<td>– – – –</td>
</tr>
</tbody>
</table>

\(^4\) Indicates values estimated from chronograms.
Table 2. Showing type of molecular data and number of loci, number of fossil constraints employed, software utilized and taxon sampling used in four recent divergence time analyses including P. megacephalum the analyses of Fritz et al. (2011) who did not include P. megacephalum. This analysis Fritz et al. (2011) Joyce et al. (2013) Lourenço et al. (2012) Spinks and Shaffer (2009)

<table>
<thead>
<tr>
<th>Data sampling</th>
<th>mtDNA</th>
<th>nuDNA</th>
<th>mtDNA</th>
<th>nuDNA</th>
<th>mtDNA</th>
<th>nuDNA</th>
<th>mtDNA</th>
<th>nuDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td># molecular makers</td>
<td>4</td>
<td>30</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td># fossil constraints</td>
<td>3</td>
<td>2</td>
<td>22</td>
<td>7</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taxon sampling</td>
<td>41 taxa, all emydid genera</td>
<td>25 taxa, all dierocheline lineages</td>
<td>37 taxa, most major turtle lineages</td>
<td>16 taxa, all emydine genera</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.2.2. The genus Emys and the classification of western pond turtles

Baird and Girard (1852) described E. marmorata (western pond turtle) and assigned this species to the genus Emys. However, Agassiz (1857, page 444) reassigned E. marmorata and three additional emydine species each to monotypic genera solely because he felt that these species were "significantly different" from one another. Given our current understanding and a more modern approach to taxonomy, the actions of Agassiz (1857) were unnecessary because there were no taxonomic issues that needed to be corrected (i.e. paraphly). Later, Strauch (1890, cited in Bettelheim et al., 2005) reassigned marmorata to the genus Clemmys, which at that time contained C. guttata, Clemmys insculpta, and Clemmys muhlenbergii. However, relatively early molecular work indicated that the taxonomic rearrangements of Strauch (1890) resulted in a grossly non-monophyletic Clemmys (Bickham et al., 1996; Burke et al., 1996; Feldman and Parham, 2002), a result strongly supported by our more extensive molecular analyses (Figs. 2 and 3). Additional taxonomic arrangements were subsequently necessary to reestablish monophyletic groups, and most current authors have assigned insculpta, and muhlenbergii to Glyptemys and returning marmorata to either Emys (Bickham et al., 1996; Burke et al., 1996; Feldman and Parham, 2002), or to Actinemys (Holman and Fritz, 2001). Assigning insculpta and muhlenbergii to Glyptemys has been widely accepted, but the generic allocation of E. marmorata (and E. pallida, see below) remains unsettled (see also Fritz et al., 2011b for a recent review).

Based on morphological evidence, Seeliger (1945) recognized two subspecies within the western pond turtle including Emys marmorata marmorata and E. m. pallida. Recently, Spinks et al. (2014) used an extensive, 89-locus, 925-individual rangewide molecular analysis of the complex, elevated both to species level (E. marmorata and E. pallida, respectively), and clarified both the range of each taxon and the very limited regions of introgression between the two. Proponents of returning E. marmorata and E. pallida to Actinemys generally favor this arrangement based on the perceived distant relationship between marmorata/pallida and the remaining emydine taxa. For example, Bury and Germano (2008, pg. 001.2) state: “Holman and Fritz (2001) and Stephens and Wiens (2003) believe that the western pond turtle is not closely related to any extant species and should be placed in its own genus, Actinemys.” However, this argument is based on a subjective definition of “closely related” and the premise that some degree of morphological or genetic divergence constitutes generic level differentiation. However, Fritz et al. (2007, pg. 419) state that E. marmorata and E. blandinii are “closely related” to E. orbicularis, Germaino and Rathbun (2008, pg. 188) emphasize that E. marmorata is “closely related” to other North American emydids, and Stephens and Wiens (2008, pg. 78) suggest that all emydids are “closely related”. Whether taxa are “closely related” based on morphological, genetic or other data sources is clearly subjective, and therefore provides an inadequate framework for a stable taxonomy. Across groups, “closely-related” becomes even more subjec-
tive. For example, turtles in general are distantly related compared to birds, but both turtles and birds are closely related with respect to tardigrades.

We support the recognition of a more inclusive *Emys* (including *blandingii*, *marmorata orbitularis*, *pallida*, and *trinacris*) based on two lines of support. First, the group is now demonstrably monophyletic based on phylogenetic analysis of mtDNA (*Bickham et al., 1996; Burke et al., 1996; Feldman and Parham, 2002; Spinks and Shaffer, 2009; Stephens and Wiens, 2003; Wiens et al., 2010; this study), combined morphological and molecular data (*Stephens and Wiens, 2003*), extensive nuDNA sequence data (*Spinks and Shaffer, 2009; Spinks et al., 2009; this study, but see *Wiens et al., 2010*), and combined mtDNA + nuDNA sequence data (*Wiens et al., 2010; this study*). *Fritz et al. (2011b)* argued that evidence for the monophyly of the group was not particularly strong, but the data presented here provide the additional data that *Fritz et al. (2011b)* suggested is necessary to strongly support taxonomic decisions within the Emydinae. Monophyly alone cannot tell us whether this clade of five species should be placed in a single genus, but this arrangement is consistent with a classification based on monophyly. We also note that the age of an inclusive five-species *Emys* (~19 my), *Terrapene* (~14 my) and *Glyptemys* (~17 my) are all roughly consistent, rendering an Emydinae comprised of three temporally co-equal genera, plus one monotypic outlier of uncertain relationships (*C. guttata*). Although equality of age or other aspects of divergence are not necessary for generic delimitation, it is convenient for comparative analyses. The combination of demonstrable monophyly, similar crown ages, and a return to earlier generic allocations lead us to support this four-genus concept of Emydinae.

4.2.3. North American box turtles: *Terrapene*

Recent analyses of the North American box turtles (*Terrapene*) reveal a case where taxonomic revisions have been carried out before the data necessary for a stable phylogeny and species delimitation were firmly in hand. *Terrapene* is widespread across North America from Arizona through northern and central Mexico east to the Atlantic coast and north to Canada (*Turtle Taxonomy Working Group, 2014*). The genus currently consists of four species including *Terrapene carolina*, *Terrapene coahuila*, *Terrapene nelsoni*, and *Terrapene ornata* several of which contain one or more subspecies (*Turtle Taxonomy Working Group, 2014*). *Martin et al. (2013)* analyzed a modest molecular data set including one mtDNA gene (*CYTB*) and one nuDNA marker (*GAPDH*) for all *Terrapene* species and subspecies, and based on these results combined the *T. carolina* subspecies *mexicana* and *trinquis* into *mexicana* and elevated this newly-constructed group to full species status as *Terrapene mexicana*. Their results were inconclusive for *Terrapene carolina bauri*, which they recognize as distinct but of uncertain species status (either a part of *carolina* or its own distinct species), *Spinks et al. (2009)* lacked relevant samples of several taxa including *mexicana*, but did recover *Terrapene carolina trinquis* as potentially monophyletic but essentially identical to *T. c. bauri* based on mtDNA (*CYTB*) and seven nuclear loci. *Butler et al. (2011)* analyzed mtDNA for all species of *Terrapene*, plus a relatively large taxon sampling of the *T. carolina* subspecies. Unlike *Martin et al. (2013)*, *Butler et al. (2011)* found that *T. carolina* was paraphyletic with respect to *T. ornata*. Further complicating the issue, microsatellite data reported by *Cureton et al. (2011)* found indications of hybridization and introgression between *T. carolina* and *T. ornata*.

In the current analysis, our sparse taxon sampling of *T. carolina* was variably paraphyletic with respect to both *T. c. trinquis* and *T. ornata* based on analyses of mtDNA and nuDNA (*Figs. 1–3 and 5*). Given the potential for current and historical hybridization within this group and recent analyses indicating that the phylogeny of *Terrapene* is clearly not yet stable, we agree with *Fritz and Havas (2014)* that “*T. mexicana*” should be treated as a junior synonym of *T. c. triunguis*. Species delimitation and relationships within this widespread, declining genus should be a high priority target for future systematics research utilizing both comprehensive population-level sampling and a genomic approach that has sufficient power to detect admixture. Until such work is completed, we recommend that the traditional view of a four-species *Terrapene* taxonomy, including *T. carolina*, *T. ornate*, *T. nelsoni* and *T. coahuila* (*Turtle Taxonomy Working Group, 2014*) be maintained and recognized.

5. Conclusions

Ecologically and demographically, the Emydidae represent the best-studied group of turtles, and a well-supported phylogeny is essential to better understand patterns of diversification in the group. Our analyses provide a temporal framework for further analyses of paleontology and historical biogeography for this important clade. There remain several unanswered questions regarding the species content of several genera, including *Graptemys*, *Pseudemys*, *Terrapene*, and *Trachemys*, which will undoubtedly require extensive taxon and data sampling coupled with sophisticated species delimitation methods for resolution. Our analyses clarify relationships among most emydida genera, and provide a well-supported phylogeny for a stable taxonomy in the Emydinae.

Acknowledgements

We thank Matthew Bettleheim for clarification on the early taxonomic history of *Emys marmorata*. For tissue samples we thank Jim McGuire and Carol Spencer (Museum of Vertebrate Zoology, University of California, Berkeley), Jose Rosado (Museum of Comparative Zoology, Harvard University), Alan Resetar (Field Museum of Natural History), and Kevin de Queiroz (United States National Museum). We also thank Dan Holland, Cesar Ayres, John Iverson, Matt Aresco, Rich Glor, Paul Moler, Raymond Farrell, Robert Zappalorti, Suzanne McGaugh, Tamás Molnár, and Chris Tabaka for additional samples. This work was supported by funding from the National Science Foundation (DEB-0817042; DEB-1239961; DEB-1457832).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2016.07.007.

References

Butler, J.M., Aresco, M., Austin, J.D., 2011. Morphological and molecular evidence indicates that box turtle (Terrapene carolina major) is not a distinct evolutionary lineage in the Florida Panhandle. J. Biol. Linn. Soc. 102, 889–901.


