The Linnean Society of London BIOLOGICAL Journal of the Linnean Society

*Biological Journal of the Linnean Society*, 2010, **99**, 445–461. With 3 figures

# Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation

JOHN J. WIENS[1]*, CAITLIN A. KUCZYNSKI[1] and PATRICK R. STEPHENS[2]

[1]*Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794-5245, USA*
[2]*Odum School of Ecology, University of Georgia, Athens, GA 30602, USA*

*Received 1 June 2009; accepted for publication 4 August 2009*

Do phylogenies and branch lengths based on mitochondrial DNA (mtDNA) provide a reasonable approximation to those based on multiple nuclear loci? In the present study, we show widespread discordance between phylogenies based on mtDNA (two genes) and nuclear DNA (nucDNA; six loci) in a phylogenetic analysis of the turtle family Emydidae. We also find an unusual type of discordance involving the unexpected homogeneity of mtDNA sequences across species within genera. Of the 36 clades in the combined nucDNA phylogeny, 24 are contradicted by the mtDNA phylogeny, and six are strongly contested by each data set. Two genera (*Graptemys*, *Pseudemys*) show remarkably low mtDNA divergence among species, whereas the combined nuclear data show deep divergences and (for *Pseudemys*) strongly supported clades. These latter results suggest that the mitochondrial data alone are highly misleading about the rate of speciation in these genera and also about the species status of endangered *Graptemys* and *Pseudemys* species. In addition, despite a strongly supported phylogeny from the combined nuclear genes, we find extensive discordance between this tree and individual nuclear gene trees. Overall, the results obtained illustrate the potential dangers of making inferences about phylogeny, speciation, divergence times, and conservation from mtDNA data alone (or even from single nuclear genes), and suggest the benefits of using large numbers of unlinked nuclear loci.  *Biological Journal of the Linnean Society*, 2010, **99**, 445–461.

ADDITIONAL KEYWORDS: branch lengths – congruence – reptiles – systematics.

## INTRODUCTION

New genomic resources now make it possible to address phylogeny using vast numbers of nuclear loci (e.g. Rokas *et al.*, 2003; Takezaki *et al.*, 2004; Philippe *et al.*, 2005; Hallstrom *et al.*, 2007; Wiens *et al.*, 2008). Multilocus phylogenies from nuclear genes may be advantageous in that misleading signals from individual genes at individual nodes may be overcome when data from multiple unlinked loci are combined (Wiens, 1998; Rokas *et al.*, 2003; but see Degnan & Rosenberg, 2006; Kubatko & Degnan, 2007). This misleading signal may arise from discordance between gene and species trees that is a result of incomplete lineage sorting of ancestral polymorphisms, introgression of genes between species, or paralogy (review in Maddison, 1997).

In theory, a phylogeny based on a single locus [e.g. mitochondrial DNA (mtDNA) or a single nuclear gene] may provide a reasonable estimate of the overall species tree, or at least the estimated tree based on multiple unlinked loci. Surprisingly, this assumption is rarely tested empirically, despite extensive debate about the pros and cons of mtDNA in phylogenetics (Rubinoff & Holland, 2005). Many phylogenetic and phylogeographic studies continue to be published using mtDNA data alone (e.g. Hillis & Wilcox, 2005; Kozak *et al.*, 2005; Hyman, Ho & Jermiin, 2007; Klicka, Burns & Spellman, 2007; Lemmon *et al.*, 2007; Wilson, Schrodl & Halanych, 2009), despite various cautionary papers (e.g. Shaw, 2002; Ballard & Whitlock, 2004; Ballard & Rand, 2005; Rubinoff & Holland, 2005). Furthermore,

*Corresponding author. E-mail: wiensj@life.bio.sunysb.edu

phylogenies that combine data from mtDNA and nuclear DNA (nucDNA) may still be dominated by the phylogenetic signal from mtDNA (e.g. Wiens, Engstrom & Chippindale, 2006), given that mtDNA datasets may have more variable characters than nucDNA datasets as a result of the generally faster evolutionary rate of mitochondrial genes. Phylogenies based on a limited number of nuclear loci may also be problematic, given that nuclear genes are also subject to paralogy and introgression, and may be even more susceptible to incomplete lineage sorting than mtDNA data, as a result of their larger effective population size (e.g. Moore, 1995).

In the present study, we document extensive discordance between phylogenies from mtDNA and combined nuclear loci in emydid turtles. We also document an unusual type of discordance in which mitochondrial genes show strikingly reduced variability in certain clades (relative to other mitochondrial genes). This pattern has influenced previous studies of emydid phylogeny (e.g. Lamb *et al.*, 1994), and may have important implications for studies of conservation and evolutionary biology. We also show that phylogenies from single nuclear genes can show striking discordance with the tree from combined nucDNA (e.g. Rokas *et al.*, 2003), even among closely-related species.

Emydid turtles consist of 12 genera (following Stephens & Wiens, 2003a) and approximately 41–47 currently recognized species (Turtle Taxonomy Working Group, 2007; Uetz, 2008). The overall number of species depends largely on which subspecies of *Trachemys scripta* are recognized as distinct species (e.g. Seidel, 2002). Emydids occur from North America to South America and into Europe, although most species and genera occur in eastern North America (Ernst & Barbour, 1989; Uetz, 2008). Emydids include many familiar North American turtles, such as the red-eared slider (*Trachemys scripta elegans*), eastern box turtle (*Terrapene carolina*), and painted turtle (*Chrysemys picta*). Traditionally, emydids have been divided into two subfamilies (e.g. Ernst, Lovich & Barbour, 1994), the aquatic Deirochelyinae (*Chrysemys*, *Deirochelys*, *Graptemys*, *Malaclemys*, *Pseudemys*, and *Trachemys*) and the terrestrial to aquatic Emydinae (*Actinemys*, *Clemmys*, *Emydoidea*, *Emys*, *Glyptemys*, and *Terrapene*).

Several previous studies have addressed the relationships among emydid turtles. However, so far, most have been limited in their taxon sampling. Previous studies have included analyses of mtDNA in *Emys* (Lenk *et al.*, 1999), *Chrysemys* (Starkey *et al.*, 2003), *Graptemys* (Lamb *et al.*, 1994), and Emydinae (Bickham *et al.*, 1996; Feldman & Parham, 2002). Bickham *et al.* (1996) analysed all emydid genera using 16S ribosomal mtDNA sequences, but included only 15 species. Stephens & Wiens (2003a) included morphological data and mtDNA data for almost all emydid species (including mtDNA data from previous studies) but lacked mtDNA data for many species. Subsequently, Stephens & Wiens (2008) added new mtDNA sequences and data from a nuclear gene (R35, RNA fingerprint protein 35; Fujita *et al.*, 2004), but included only 16 of approximately 43 emydid species (see also Stephens & Wiens, 2009). Spinks *et al.* (2009) and Spinks & Shaffer (2009) examined the relationships among emydid genera using mtDNA and nucDNA, but included less than half of the described species (19 and 12, respectively). In the present study, we analyse data from two mitochondrial genes and six nuclear genes for most of the approximately 43 emydid species, along with our previously published morphological data.

## MATERIAL AND METHODS

We obtained tissue or blood samples from almost all emydid species. We also included representatives of some subspecies, which may represent distinct species (as in *T. scripta*; Stephens & Wiens, 2003a). We included a geoemydid (*Cuora flavomarginata*) as our primary outgroup, given that recent analyses place geoemydids in the sister group to emydids (Krenz *et al.*, 2005). We used a kinosternid (*Sternotherus odoratus*) as a more distant outgroup. Information on tissue sources and vouchers are given in the Supporting Information (Appendix S1).

We selected two mitochondrial genes and six nuclear genes for use in these analyses. The two mitochondrial genes, cytochrome *b* (cyt *b*) and NADH dehydrogenase subunit 4 (ND4), have been used extensively in emydids, and appear to have considerable resolving power at all taxonomic levels (e.g. Lamb *et al.*, 1994; Feldman & Parham, 2002; Stephens & Wiens, 2008). The selection of nuclear genes was more difficult because many show little informative variation in emydids. First, we used the R35 intron, which is highly variable within turtles (Fujita *et al.*, 2004) and was previously sequenced in several emydids (Stephens & Wiens, 2008). We also tried primers for many nuclear exons developed for studies of higher-level squamate phylogeny (Townsend *et al.*, 2008), although we found that among the genes that amplified successfully (AHR, GPR37, NGFB, TNS3), only NGFB (nerve growth factor, beta polypeptide) appeared to be variable enough to be useful. Finally, we tested many additional nuclear introns, and selected four which could be amplified and sequenced across most genera and contained useful levels of variation [ETS (ets oncogene): Lyons *et al.*, 1997; GAPD (glyceraldehyde-3-phosphate dehydrogenase): Friesen *et al.*, 1997; ODC (ornithine decarboxylase): Friesen *et al.*, 1999;

**Table 1.** PCR primer sequences used for amplification and sequencing in this study

| Gene | Primer name | Primer sequence (5′ to 3′) | Source |
|---|---|---|---|
| Cyt-b | E-gludg-F | TGATTTGAARAACCAYCGTTG | T. Engstrom (pers. comm.) |
| | CB649-R | GGGTGGAATGGGATTTTGTC | Engstrom, Shaffer & McCord (2004) |
| ND4 | ND4-F | CACCTATGACTACCAAAAGCTCATGTAGAAGC | Arevalo, Davis & Sites (1994) |
| | Hist-R | CCT ATT TTT AGA GCC ACA GTC TAA TG | Engstrom *et al.* (2004) |
| NGFB | NGFB-f2 | GATTATAGCGTTTCTGATYGGC | Townsend *et al.* (2008) |
| | NGFB-r2 | CAAAGGTGTGTGTWGTGGTGC | Townsend *et al.* (2008) |
| ETS | ETS2-F | AGCTGTGGCAGTTTCTTCTG | Dolman & Phillips (2004) |
| | ETS2-R | CGGCTCAGCTTCTCGTAG | Dolman & Phillips (2004) |
| | ETS-F2 | TCTTCATGGCTGAGATGCTACAAGT | Present study |
| | ETS-R2 | TGCCGCTGGGAGAGCTAATGGTGA | Present study |
| GAPD | GapdL890 | ACCTTTAATGCGGGTGCTGGCATTGC | Dolman & Phillips (2004) |
| | Gapdl950 | CATCAAGTCCACAACACGGTTGCTGTA | Dolman & Phillips (2004) |
| ODC | OD-F | GACTCCAAAGCAGTTTGTCGTCTCAGTGT | Friesen *et al.* (1999) |
| | OD-R | TCTTCAGAGCCAGGGAAGCCACCACCAAT | Friesen *et al.* (1999) |
| | ODC-rvs3 | ARTATTGGGTCRACTATCAAAGGAT | Present study |
| | ODC-rvs2 | ATTGGTYRTAAGATTTAGTAAGTCT | Present study |
| R35 | R35-EX1 | ACGATTCTCGCTGATTCTTGC | Fujita *et al.* (2004) |
| | R35-EX2 | GCAGAAAACTGAATGTCTCAAAGG | Fujita *et al.* (2004) |
| Vim | VimSeq3F | TTGAAGAAACTTCATGAGGAGGTA | A. Pyron (pers. comm.) |
| | VimAmp5R | TTCCTTTAAGRGCATCMACTTCAC | A. Pyron (pers. comm.) |
| | Vim-3-f-new2 | AGATTCAGGAACAACACATCCAAA | Present study |
| | Vim-5-r-new1 | ATTAGCTTCCTGTTTGGCCTGACGTA | Present study |
| | Vim-5-r-new2 | TTTGTCTGCGGTACTCATTAGCTTCC | Present study |
| | Vim-5K2-RVS | GTTGAACAAGCCAGATTTACAGTTRGTAT- | Present study |
| | Vim-K3- FWD | TTGATGTGGATGTTGCTAAACCAGATCTCAC | Present study |

Vim (vimentin); A. Pyron, pers. comm.]. We used standard methods of DNA extraction and polymerase chain reaction (PCR) amplification, and purified PCR products were sequenced using an ABI 3100 automated sequencer. Primer sequences are provided in Table 1. The length and variability of each gene (along with data partitions and models, see below) is provided in Table 2. GenBank numbers are provided in the Supporting Information (Appendix S2). The data obtained for cyt *b*, ND4, and R35 were supplemented by sequences available on GenBank (e.g. Feldman & Parham, 2002; Fujita *et al.*, 2004; Spinks *et al.*, 2004; Near, Meylan & Shaffer, 2005; Stuart & Parham, 2007).

Sequences from each gene were initially analysed using parsimony (see below) to identify potential contaminants, as indicated by different species having identical sequences. Potential contaminant sequences were resequenced, and only high-quality sequences were used (i.e. few or no ambiguous bases). However, sequences were not excluded on the basis of incongruence with other genes or previous taxonomy, in order to avoid biasing our estimates of congruence among genes.

Given the lack of insertions and deletions, alignment of protein-coding sequences was straightforward and was carried out by eye. MacClade, version 4.0 (Maddison & Maddison, 2000) was used to translate sequences to amino acids to confirm the lack of stop codons. Most introns contained some apparent insertions or deletions (indels). Alignments for introns were performed using MUSCLE (Edgar, 2004). We coded inferred indels for inclusion in phylogenetic analyses using the approach of Simmons & Ochoterena (2000). In general, this involved treating the presence or absence of insertions and deletions as binary characters. In some cases, the lengths of indels apparently varied among species, and were treated as ordered multistate characters (e.g. no deletion near position 248 = state 0; deletion near position 248 is present and four bases long = state 1; deletion near position 248 is present and seven bases long = state 2). Given the limited variation in the nuclear genes, the indel characters substantially increased the number of potentially informative characters (Table 2).

Phylogenetic analyses were conducted primarily using Bayesian analysis and parsimony (see below). Current software packages for maximum likelihood

**Table 2.** Genes used in phylogenetic analyses, showing lengths (after alignment), number of variable and parsimony-informative characters (excluding indel characters and outgroup taxa), number of indel characters, best-fitting model of sequence evolution, and total number of taxa included

| Gene | Length | Variable characters | Parsimony-informative characters | Indel characters | Best-fitting model | Taxa |
|---|---|---|---|---|---|---|
| MtDNA | | | | | | |
| Cyt-b | 648 | 236 | 164 | 0 | HKY+I+Γ | 41 |
| ND4 | 616 | 225 | 165 | 0 | HKY+I+Γ | 39 |
| NucDNA (protein-coding) | | | | | | |
| NGFB | 536 | 32 | 14 | 0 | HKY+Γ | 37 |
| NucDNA (introns) | | | | | | |
| ETS | 758 | 129 | 47 | 10 | GTR+Γ | 30 |
| GAPD | 433 | 26 | 62 | 4 | HKY+Γ | 38 |
| ODC | 552 | 26 | 59 | 5 | HKY+Γ | 30 |
| R35 | 946 | 63 | 28 | 0 | HKY+Γ | 26 |
| Vim | 741 | 202 | 76 | 15 | GTR+Γ | 40 |

did not allow inclusion of indel characters, and were not used. We conducted separate analyses of each gene and then performed combined analyses of all mitochondrial genes together and all nuclear genes together. We also performed combined analyses of nucDNA and mtDNA data, and molecular and morphological data (see below).

For Bayesian analyses, we first identified the best-fitting model for each gene, using the Akaike information criterion in MrModelTest, version 2.0 (Nylander, 2004). Indels were coded as 'standard' characters using the Mk model (Lewis, 2001), with a separate parameter for rate variation among characters (Γ).

Separate Bayesian analyses were conducted to determine if partitions within genes were supported (Brandley, Schmitz & Reeder, 2005). Codon positions were treated as separate partitions for cyt *b* and ND4. The harmonic mean of the log-likelihoods of the post burn-in trees, both with and without partitions within each gene, were compared using the Bayes factor (Nylander *et al.*, 2004; see also Brown & Lemmon, 2007). Given the few variable characters in NGFB (Table 2), partitions were not tested for this gene. No partitions were used within introns.

Bayesian analyses were performed using MrBayes, version 3.1.2 (Huelsenbeck & Ronquist, 2001). Two replicate searches were performed on each data set, each using four chains and default priors. Analyses for single genes used $2.0 \times 10^6$ generations each, sampling every 1000 generations, whereas analyses for combined genes used $4.0 \times 10^6$. Stationarity was identified based on plots of log-likelihoods over time and the standard deviation of split frequencies between parallel searches. Stationarity was achieved within the first 10% of generations for each analysis, and thus this cut-off was used. The phylogeny and branch lengths were estimated from the majority-rule consensus of the pooled post burn-in trees from the two replicates. Clades with posterior probabilities ($Pp$) $\geq 0.95$ were considered as strongly supported (Wilcox *et al.*, 2002; Alfaro, Zoller & Lutzoni, 2003; Erixon *et al.*, 2003; Huelsenbeck & Rannala, 2004).

The shortest parsimony trees were found using heuristic searches with 500 random taxon-addition sequence replicates, using PAUP* 4.0b10 (Swofford, 2002). Support was evaluated using nonparametric bootstrapping (Felsenstein, 1985), with 200 bootstrap pseudoreplicates and ten random taxon-addition sequence replicates per bootstrap pseudoreplicate. Clades with bootstrap values $\geq 70\%$ were considered strongly supported, following Hillis & Bull (1993 (but see also their extensive caveats). Parsimony trees were generally similar to Bayesian trees (see Supporting Information, Appendix S4), and only Bayesian results are presented.

The combined mtDNA genes and combined nucDNA genes were analysed. However, some taxa proved difficult to amplify for a given gene despite repeated attempts (particularly for the nuclear introns) and were coded as having missing data ('?') in the combined analyses. Recent studies based on simulations (Wiens, 2003; Philippe *et al.*, 2004; Wiens & Moen, 2008) and empirical data (Driskell *et al.*, 2004; Philippe *et al.*, 2004; Wiens *et al.*, 2005) suggest that such 'incomplete taxa' can be accurately placed in phylogenies, regardless of their missing data cells, if the overall number of characters in the analysis is relatively high. For the combined mtDNA and nucDNA sequence data (5230 characters total), taxa were missing (on average) 21.2% of the characters, with a range among species of 2.8–72.7%.

Finding incongruence between genes raises the questions of whether data should be combined and how the final tree should be estimated. We follow the general approach outlined by Wiens (1998). When gene trees give different answers for a given node, a critical question is whether the discordant clades are strongly supported by each gene (e.g. Bayesian $Pp \geq 0.95$; bootstrap $\geq 70\%$). If they are strongly supported, this pattern may indicate discordant gene histories or some other systematic error. If the conflict is strongly supported by only one gene (or neither), then the conflict may simply be explained by stochastic sampling of characters (e.g. Bull *et al.*, 1993). If there is strongly supported conflict between two genes, then the combined analysis may not yield the correct answer for that node. However, if multiple, unlinked genes are sampled, then the combined analysis may reflect the true species phylogeny, given the assumption that deviations between the gene trees and species tree will not occur in the same way in a majority of unlinked genes. However, we acknowledge that, under some circumstances, many genes may converge on an incorrect answer and mislead a combined analysis (e.g. Degnan & Rosenberg, 2006; Edwards, Liu & Pearl, 2007; Kubatko & Degnan, 2007), and that incongruence may be common on short branches (Wiens *et al.*, 2008). Unfortunately, alternative methods designed to estimate species trees without concatenation (e.g. Edwards *et al.*, 2007) would be problematic to apply here as a result of incomplete overlap in taxon sampling between genes. In the present study, we focus primarily on comparing the results of the combined nucDNA data and the combined mtDNA data, and we emphasize the strongly supported conflicts between these combined datasets.

We also summarized patterns of discordance among nuclear genes. Given that no genes include exactly the same sampling of taxa, we used the combined nucDNA tree as a reference tree and evaluated which genes yielded trees that were strictly concordant or discordant with each node of this tree. We considered a given gene as concordant with a given clade from the combined nucDNA data, even if the gene did not include every known species within that clade, as long as the species that were sampled formed a monophyletic group consistent with that clade. However, some cases were still ambiguous as a result of missing data in certain taxa (e.g. the clade of *T. s. elegans* and *Trachemys gaigeae* cannot be evaluated for genes lacking data for *T. gaigeae*). We acknowledge that there are limitations to this approach. For example, a gene will be counted as discordant with a clade from the combined nucDNA data if a single species does not appear in that clade. In other words, the fact that all the other species are placed in the clade is ignored. Similarly, a single 'misplaced' taxon may have cascading effects on several nodes. Overall, our approach for assessing congruence is conservative, but may be overly conservative in some cases. Other approaches to assessing concordance would be complicated by the differences in taxon sampling between genes and the large number of genes and taxa.

We also tested for a relationship between the lengths of each branch in the combined, Bayesian nucDNA tree and the proportion of genes that are concordant with that branch (following Wiens *et al.*, 2008), based on the separate Bayesian analyses of the nuclear genes. We expect greater incongruence on shorter branches, assuming that incomplete lineage sorting will cause more incongruence when there is a shorter time between splitting events (e.g. Maddison, 1997). For a given branch, genes were only considered as concordant or discordant if they contained sufficient taxa (see above). We tested for this relationship using Spearmann's rank correlation, implemented in Statview.

We found unusually short branch lengths in the mtDNA genes for *Graptemys* and *Pseudemys*, relative to branch lengths in other emydid genera (for mtDNA) and for these same genera for the nuclear genes. To quantify these differences, we summarized the average branch lengths within each of these genera (including both internal and terminal branches) for the combined nucDNA data and separately for the mtDNA data. For comparison, we did the same for *Glyptemys*, *Terrapene*, the *Emys* + *Emydoidea* clade, and *Trachemys*. For the mtDNA data for *Trachemys*, we only included the *Trachemys* that formed a clade (i.e. *Trachemys scripta elegans*, *Trachemys scripta emolli*, *Trachemys scripta venusta*). For all clades, we excluded the branch representing the ancestor of the group. Note that this analysis does not assume any equivalence among clades ranked as genera, given that it is the average of the terminal and internal branches for a set of species that are being compared.

Finally, we conducted some analyses with the morphological data set of Stephens & Wiens (2003a), together with the combined nucDNA and mtDNA data. However, analyses including the mtDNA are potentially compromised by the problems described below. In previous morphological analyses (Stephens & Wiens, 2003a), intraspecific variation was coded using frequency methods (i.e. for discrete characters; Wiens, 1995) and gap-weighting (i.e. for quantitative characters; Thiele, 1993). However, these approaches were problematic for Bayesian analyses because they require large numbers of ordered states (MrBayes currently allows no more than five ordered states) and weighting between characters to maintain equivalence (i.e. scaling methods; Wiens, 2001) was difficult. Therefore, these characters were recoded as binary following Stephens & Wiens (2008). Overall, the mor-

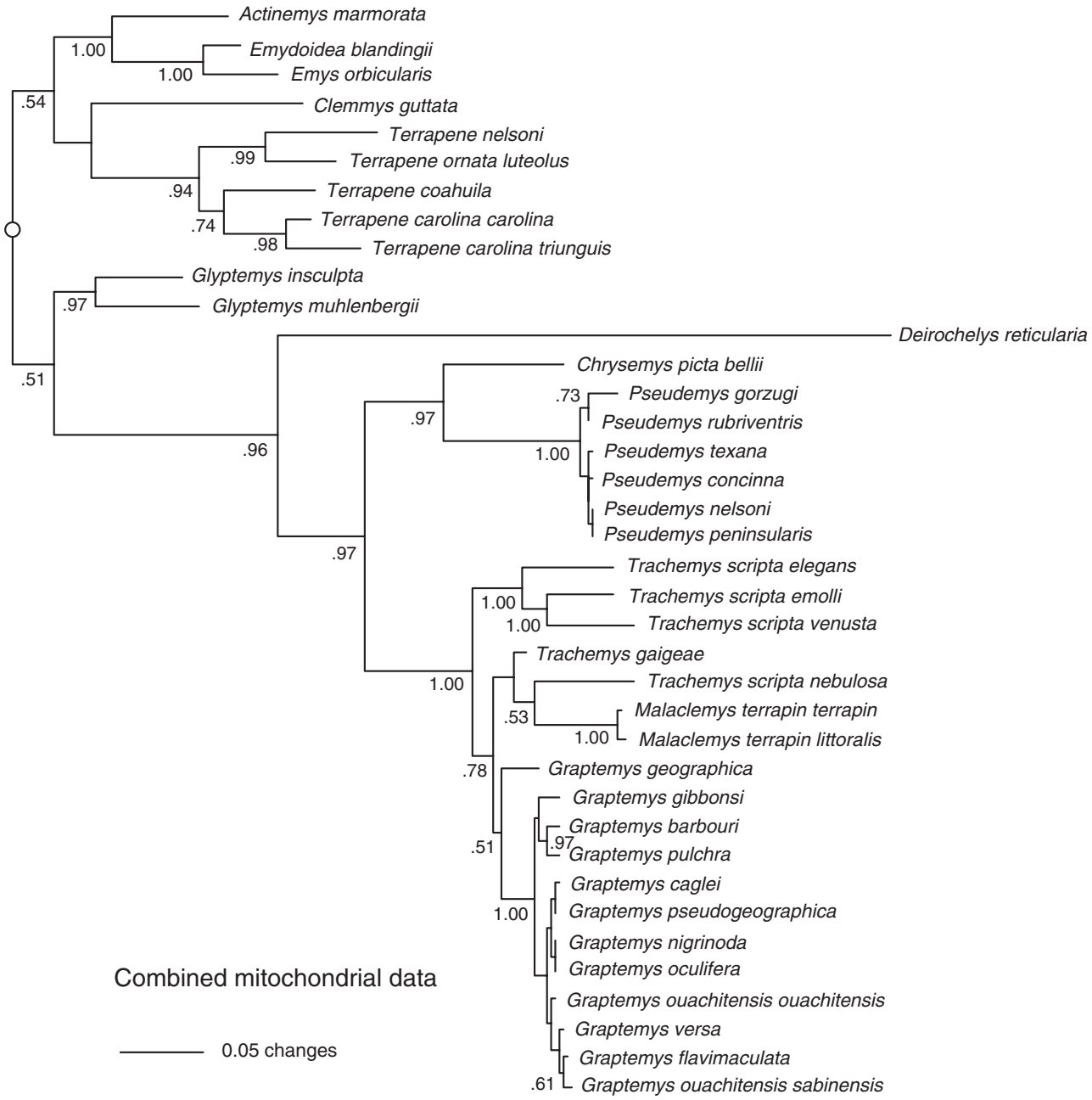


**Figure 1.** Phylogeny of emydid turtles based on a combined, partitioned Bayesian analysis of mitochondrial DNA sequences from the cytochrome *b* and ND4 genes. Numbers adjacent to nodes indicate Bayesian posterior probabilities ($Pp$) $\geq 0.50$. The outgroup taxa are excluded to facilitate presentation of branch lengths, and the root is indicated with an open circle.

phological data matrix consists of 245 characters, of which 207 are parsimony-informative for the taxa included in the present study. In Bayesian analyses, morphological characters were analysed using the Mk model (Lewis, 2001), with a parameter ($\Gamma$) for rate heterogeneity among characters. The morphological data matrix is included in the Supporting Information (Appendix S3).

## RESULTS

Trees from Bayesian analyses of the combined mtDNA and nucDNA are summarized in Figs 1, 2. Trees from individual mitochondrial and nuclear genes and parsimony trees for the combined nucDNA and combined mtDNA data are shown in the Supporting Information (Appendices S4, S5). The Bayesian

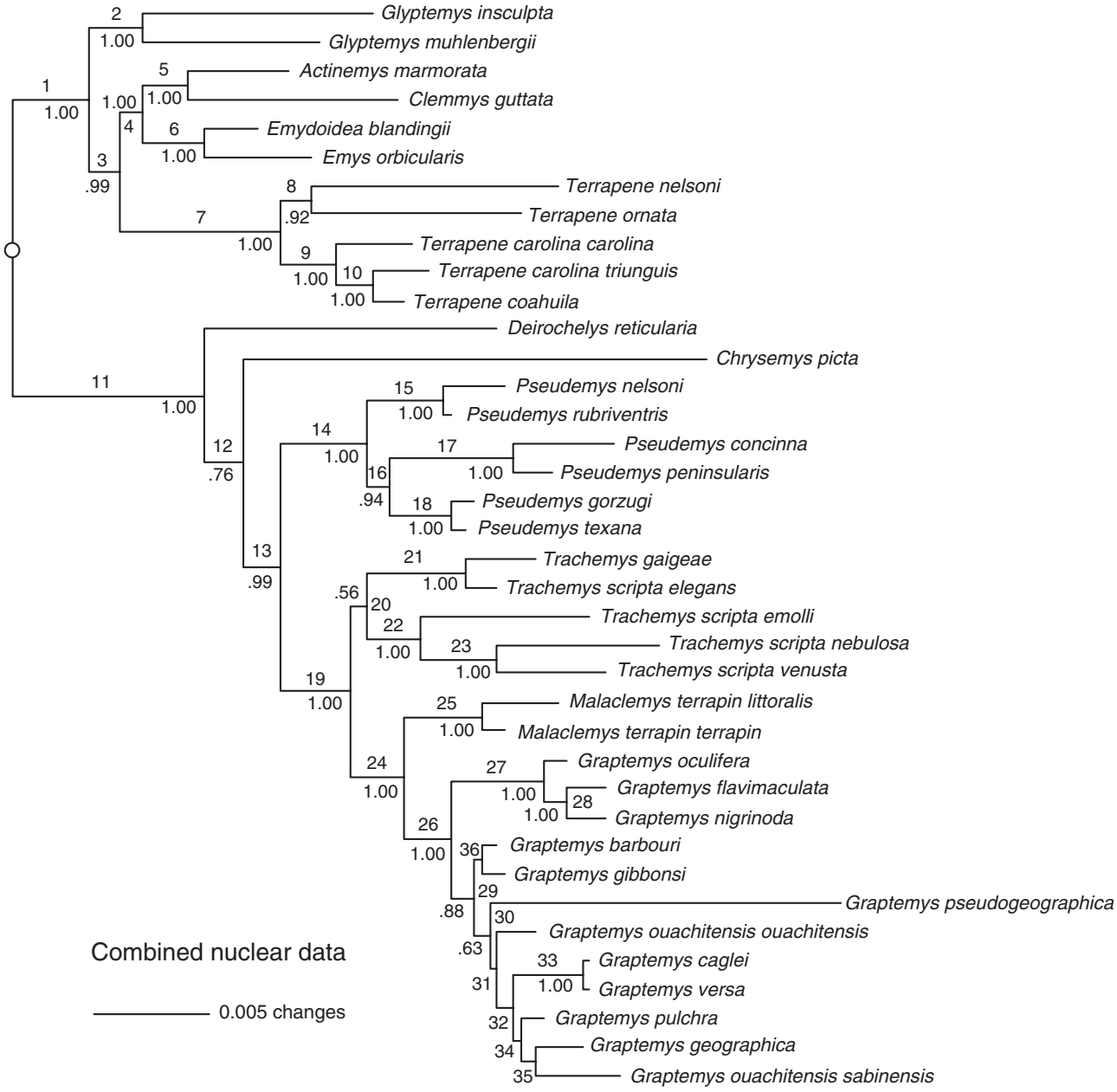


**Figure 2.** Phylogeny of emydid turtles based on a combined, partitioned Bayesian analysis of DNA sequences from six nuclear loci (coding: NGFB; intron: ETS, GAPD, ODC, R35, Vim). Probabilities adjacent to nodes indicate Bayesian posterior probabilities ($Pp$) $\geq 0.50$; integers correspond to clade numbers in Table 3 (which lists the genes that separately support or reject each clade). The outgroup taxa are excluded to facilitate presentation of branch lengths, although the root is indicated with an open circle.

tree from the combined nucDNA is contradicted by the mtDNA tree for 24 of 36 nodes. However, many of these conflicts are either weakly supported by the nucDNA data, weakly supported by the mtDNA data, or both. For example, there are several unusual patterns that are weakly supported by mtDNA data but strongly rejected by the nucDNA data, such as non-monophyly of Emydinae (*Actinemys*, *Clemmys*, *Emydoidea*, *Emys*, and *Terrapene*).

Six clades are both conflicting and strongly supported by the combined nucDNA data and combined mtDNA data, suggesting possible deviations between gene and species phylogenies. (1) *Actinemys* and *Clemmys guttata* are strongly supported as sister

**Table 3.** Average branch lengths within select clades of emydid turtles showing the similarity of mitochondrial DNA branch lengths in *Graptemys* and *Pseudemys* to those in nuclear genes

| Clade | Mitochondrial DNA | Nuclear DNA |
|---|---|---|
| *Glyptemys* | 0.056 | 0.009 |
| *Emydoidea* + *Emys* | 0.033 | 0.004 |
| *Terrapene* | 0.039 | 0.004 |
| *Pseudemys* | 0.004 | 0.002 |
| *Trachemys* | 0.040 | 0.004 |
| *Graptemys* | 0.005 | 0.002 |

taxa by the nucDNA data, whereas the mtDNA places *Actinemys* as sister group of the *Emys* + *Emydoidea* clade. (2) *Terrapene carolina triunguis* is strongly supported as sister taxon to *Terrapene coahuila* by nucDNA data, but *T. c. triunguis* is sister taxon to *Terrapene carolina carolina* in the mtDNA tree. (3) A clade including *Pseudemys*, *Trachemys*, *Malaclemys*, and *Graptemys* (and excluding *Chrysemys* and *Deirochelys*) is strongly supported by nucDNA data, but *Pseudemys* is strongly supported as the sister taxon to *Chrysemys* by the mtDNA data. (4) NucDNA data strongly support a clade uniting *Trachemys gaigeae* and *T. s. elegans*, whereas mtDNA data strongly support a clade of *Trachemys* excluding *T. gaigeae*. (5) NucDNA data strongly support a clade of *T. s. emolli*, *T. s. nebulosa*, and *T. s. venusta*, whereas mtDNA data strongly exclude *T. s. nebulosa*. (6) Similarly, *T. s. nebulosa* and *T. s. venusta* are strongly supported as sister taxa by nucDNA data, whereas mtDNA data strongly exclude *T. s. nebulosa* from this clade.

A particularly unusual pattern of discordance involves the branch lengths in the genera *Pseudemys* and *Graptemys*. On the basis of the combined nucDNA data (Fig. 2), the species of *Pseudemys* sampled are highly divergent, and relationships among them are strongly supported. In the mtDNA tree (Fig. 1), the divergences among species are very shallow relative to those in most of the rest of the tree, and no internal nodes within the genus are strongly supported. A similar pattern of shallow mtDNA divergence occurs in *Graptemys*, although the nucDNA tree is not as well supported as in *Pseudemys*, and some nodes in the mtDNA tree are strongly supported. The unusual branch lengths in *Graptemys* and *Pseudemys* are most obvious when branch lengths within genera are compared between mtDNA and nucDNA trees (Table 3). Clearly, branch lengths for mtDNA in *Graptemys* and *Pseudemys* are more similar to those in the nuclear genes than they are to those in other genera for the mtDNA, although branch lengths in these two genera are not unusually short in the nucDNA tree.

Despite the fact that the phylogeny from the combined nucDNA data is generally well-supported (Fig. 2), many of the separately analysed nuclear genes are discordant with parts of this phylogeny (Table 4). On average, only 38% of the nuclear genes are strictly concordant with any given clade from the combined nucDNA phylogeny. Four clades in the combined analysis are not supported by any of the separately analysed nuclear genes. Furthermore, for 56% of the 36 clades (Table 4) at least one nuclear gene is both discordant with the combined nucDNA tree and strongly supported. We find a strong relationship between branch lengths in the combined nucDNA tree and the proportion of genes congruent with the branch (i.e. greater congruence on longer branches), using Spearman's rank correlation ($r_s = 0.607$; $P = 0.0003$).

Conflicts between individual nuclear gene trees and the combined nucDNA phylogeny are not evenly distributed among genes (Tables 4, 5). Although all nuclear gene trees are discordant with the combined nucDNA phylogeny for at least 33% of their nodes, some genes disagree for most nodes (> 70% for ETS and NGFB), and show strongly supported incongruence with the combined nucDNA for many nodes (18 and 12 for ETS and NGFB, respectively). However, three genes (GAPD, R35, Vim) show no strongly supported conflict with the combined nucDNA phylogeny. Most genes show strong support for a similar number of nodes in the combined nucDNA tree (five to ten). Overall, the combined nucDNA tree is most concordant with GAPD and Vim.

The tree from Bayesian analysis of the combined nucDNA and mtDNA is shown in Figure 3. The tree from combined molecular and morphological data is similar (see Supporting Information, Appendix S5). Corresponding parsimony trees, and trees from morphology and the combined nucDNA and morphology are also shown in the Supporting Information (Appendix S6). The combined-data Bayesian tree (Fig. 3) resolves many conflicts between mtDNA and nucDNA in favour of nucDNA, including: (1) monophyly of emydines; (2) most relationships within *Pseudemys*; (3) most relationships within *Trachemys*; (4) the Texas endemic clade (*Graptemys caglei* + *Graptemys versa*) of *Graptemys*; and (5) the sawback clade (*Graptemys flavimaculata*, *Graptemys nigrinoda*, *Graptemys oculifera*) within *Graptemys*. However, many conflicts are also resolved in favour of mtDNA, including: (1) placement of *Actinemys* as sister group of *Emydoidea* + *Emys*; (2) placement of *Chrysemys* with *Pseudemys*; (3) monophyly of the two sampled *T. carolina* subspecies; (4) basal placement of *Graptemys geographica* within *Graptemys*; and (5) the *pulchra* group within *Graptemys* (including *Graptemys barbouri*, *Graptemys gibbonsi*, *Graptemys pulchra*). In

**Table 4.** Congruence and discordance of trees from separately analysed nuclear genes with specific clades (Fig. 2) of the phylogeny inferred from the combined nuclear data

| Clade number | Genes supporting | Genes rejecting | % supporting |
|---|---|---|---|
| 1* | ETS*, NGFB, Vim* | GAPD, ODC, R35 | 50 |
| 2* | GAPD, NGFB, ODC, Vim | ETS* | 80 |
| 3* | ETS* | GAPD, NGFB*, ODC, R35, Vim | 17 |
| 4* | GAPD, ODC | ETS, NGFB*, R35, Vim | 33 |
| 5* | GAPD, Vim* | ETS*, NGFB*, ODC*, R35 | 33 |
| 6* | GAPD, R35*, Vim | NGFB*, ODC | 60 |
| 7* | GAPD, ODC, R35* | ETS*, NGFB, Vim | 50 |
| 8 | ODC, Vim | GAPD, NGFB | 50 |
| 9* | GAPD*, Vim* | ETS*, NGFB, ODC, R35 | 33 |
| 10* | Vim | GAPD, NGFB, ODC | 25 |
| 11* | ETS*, GAPD*, NGFB*, R35, Vim* | ODC | 83 |
| 12 | ETS, GAPD, Vim | NGFB, ODC, R35 | 50 |
| 13* | | ETS, GAPD, NGFB, ODC, R35, Vim | 0 |
| 14* | ETS, GAPD, Vim* | NGFB, ODC, R35 | 50 |
| 15* | GAPD*, Vim* | ETS, NGFB, ODC, R35 | 33 |
| 16 | GAPD*, Vim | ETS*, NGFB, ODC, R35 | 33 |
| 17* | GAPD, Vim* | ETS, NGFB | 50 |
| 18* | ETS*, GAPD*, NGFB*, ODC, Vim* | | 100 |
| 19* | GAPD*, Vim | ETS*, NGFB, ODC, R35 | 33 |
| 20 | | ETS*, GAPD, NGFB, ODC, R35, Vim | 0 |
| 21* | GAPD* | Vim | 50 |
| 22* | Vim | ETS*, GAPD, NGFB, ODC | 20 |
| 23* | ODC*, R35* | ETS*, GAPD, NGFB, Vim | 33 |
| 24* | ETS, ODC, R35* | GAPD, NGFB, Vim | 50 |
| 25* | ETS*, GAPD, Vim* | NGFB, R35 | 60 |
| 26* | R35* | ETS*, GAPD, NGFB, ODC, Vim | 17 |
| 27* | Vim* | ETS*, GAPD, NGFB*, R35 | 20 |
| 28* | GAPD, ODC | ETS*, NGFB*, Vim | 40 |
| 29 | Vim | ETS*, GAPD, NGFB*, ODC*, R35 | 17 |
| 30 | Vim | ETS*, GAPD, NGFB*, ODC*, R35 | 17 |
| 31 | Vim | ETS*, GAPD, NGFB*, ODC*, R35 | 17 |
| 32 | Vim | ETS*, GAPD, NGFB*, ODC*, R35 | 17 |
| 33* | ETS*, GAPD, Vim | NGFB | 75 |
| 34 | | ETS*, GAPD, NGFB*, Vim | 0 |
| 35 | | ETS*, GAPD, NGFB*, Vim | 0 |
| 36 | ODC, Vim | GAPD, NGFB | 50 |

Genes marked with asterisks indicate that a given clade is strongly supported or rejected by that gene. Asterisks by clade numbers indicate that the clade is strongly supported in the combined nucDNA data.

addition, some relationships are largely unique to the combined analysis (and yet are well supported), including (1) placement of *Clemmys* as sister group to *Actinemys* + *Emys* + *Emydoidea* and (2) basal placement of *Pseudemys gorzugi* + *Pseudemys texana* clade within *Pseudemys*. Overall, the combined-data trees resolve some conflicts that were weakly supported by either mtDNA or nucDNA (e.g. monophyly of emydines, many relationships within *Graptemys* and *Pseudemys*). However, the resolution of some of the strongly supported conflicts in the combined-data tree may be arbitrary, and may be influenced by potentially misleading phylogenetic signal in the mtDNA data (e.g. placement of *Actinemys*, non-monophyly of *Trachemys*).

## DISCUSSION

### Causes of discordance between mtDNA and nucDNA trees

Our results from emydid turtles show widespread discordance between trees from mtDNA and nucDNA, in both topology and branch lengths. The results

**Table 5.** Summary of concordance and discordance of trees from separately analysed nuclear genes with the phylogeny inferred from the combined nuclear data, based on the number of clades in the trees from individual genes that support or reject clades in the tree from the combined nuclear data

| Gene | Supporting | Rejecting | Strongly supporting | Strongly rejecting |
|---|---|---|---|---|
| ETS | 9 | 22 | 5 | 18 |
| GAPD | 18 | 18 | 7 | 0 |
| NGFB | 4 | 31 | 2 | 12 |
| ODC | 9 | 20 | 2 | 5 |
| R35 | 6 | 17 | 6 | 0 |
| Vim | 24 | 12 | 10 | 0 |
| mtDNA | 12 | 24 | 8 | 13 |

Levels for the mitochondrial DNA tree are included for comparison.

appear to illustrate several potential dangers of deriving phylogenies from mtDNA data alone (although the frequency and taxonomic extent of these problems remain uncertain) and, conversely, demonstrate the potential benefits of a multi-locus nuclear approach.

Why might nuclear and mtDNA phylogenies disagree? In general, phylogenies from different genes may disagree because one or both of the gene trees have been incorrectly reconstructed (e.g. as a result of sampling too few characters or due to long-branch attraction) or because one or both of the gene trees differ from the species tree (e.g. as a result of paralogy, incomplete lineage sorting, or introgression).

mtDNA may be problematic at higher-taxonomic levels as a result of higher rates of change, leading to greater homoplasy and the potential for long-branch attraction (e.g. Wiens & Hollingsworth, 2000; Weisrock, Harmon & Larson, 2005). This issue may be illustrated in emydids by the failure of the mtDNA data to support the monophyly of the subfamily Emydinae. Although this traditionally recognized subfamily is strongly supported by morphological data (Stephens & Wiens, 2003a) and combined nucDNA (Fig. 2), the mtDNA data weakly place *Glyptemys* with deirochelyines. A similar pattern is present in the ND4 data alone, where the outgroup root appears to fall within emydines (not shown). Similarly, analyses of cyt *b* fail to support monophyly of Deirochelyinae, a clade strongly supported by morphological and nucDNA data.

We also found many conflicts within these major clades. These conflicts appear less likely to involve long-branch attraction (because the branches are considerably shorter), and more likely to involve discrepancies between gene and species trees. However, we lack a specific mechanism to explain many of these conflicts. For example, there is no obvious signature of introgression between species in different genera.

Nevertheless, we speculate that some conflicts, and the unusual branch lengths in *Graptemys* and *Pseudemys*, might be associated with putatively mitochondrial genes having been transferred to the nuclear genome (numts; Zhang & Hewitt, 1996). Comparisons of branch lengths show that the two putatively mitochondrial genes in *Graptemys* and *Pseudemys* are evolving at a rate that is more similar to that of the nuclear genes that we sequenced in these same genera than to the mitochondrial genes in other genera (Table 3). However, we do not know whether there are or were copies of these genes in the mitochondrial genome as well in these taxa.

Other processes that might explain the pattern of short branch lengths within these genera appear much less likely. For example, extensive and recent introgression might homogenize mitochondrial genes among species and thereby reduce branch lengths. However, many sampled taxa in *Pseudemys* are entirely allopatric from all or most congeners (e.g. *Pseudemys gorzugi*, *Pseudemys nelsoni*, *Pseudemys rubriventris*, *Pseudemys texana*; Ernst *et al.*, 1994; Conant & Collins, 1998) and some are geographically distant (e.g. New England versus western Texas), making recent, homogenizing introgression across the genus appear unlikely. Furthermore, such a pattern of homogenizing introgression is absent in the nuclear genes (Fig. 2). Similarly, many species of *Graptemys* are allopatric with respect to each other, and many are very morphologically distinct, again making extensive and recent introgression seem unlikely (Ernst *et al.*, 1994; Conant & Collins, 1998). In addition, homogenizing gene flow among the species in these genera would require extensive overland dispersal by these largely aquatic organisms, which appears very unlikely. We know of no dramatic shifts in the ecology or generation time that could explain the ten-fold variation in branch lengths among closely-related genera (Ernst *et al.*, 1994), nor any molecular mechanism that would cause such a dramatically reduced rate of molecular evolution. A previous study suggested that there is low mitochondrial

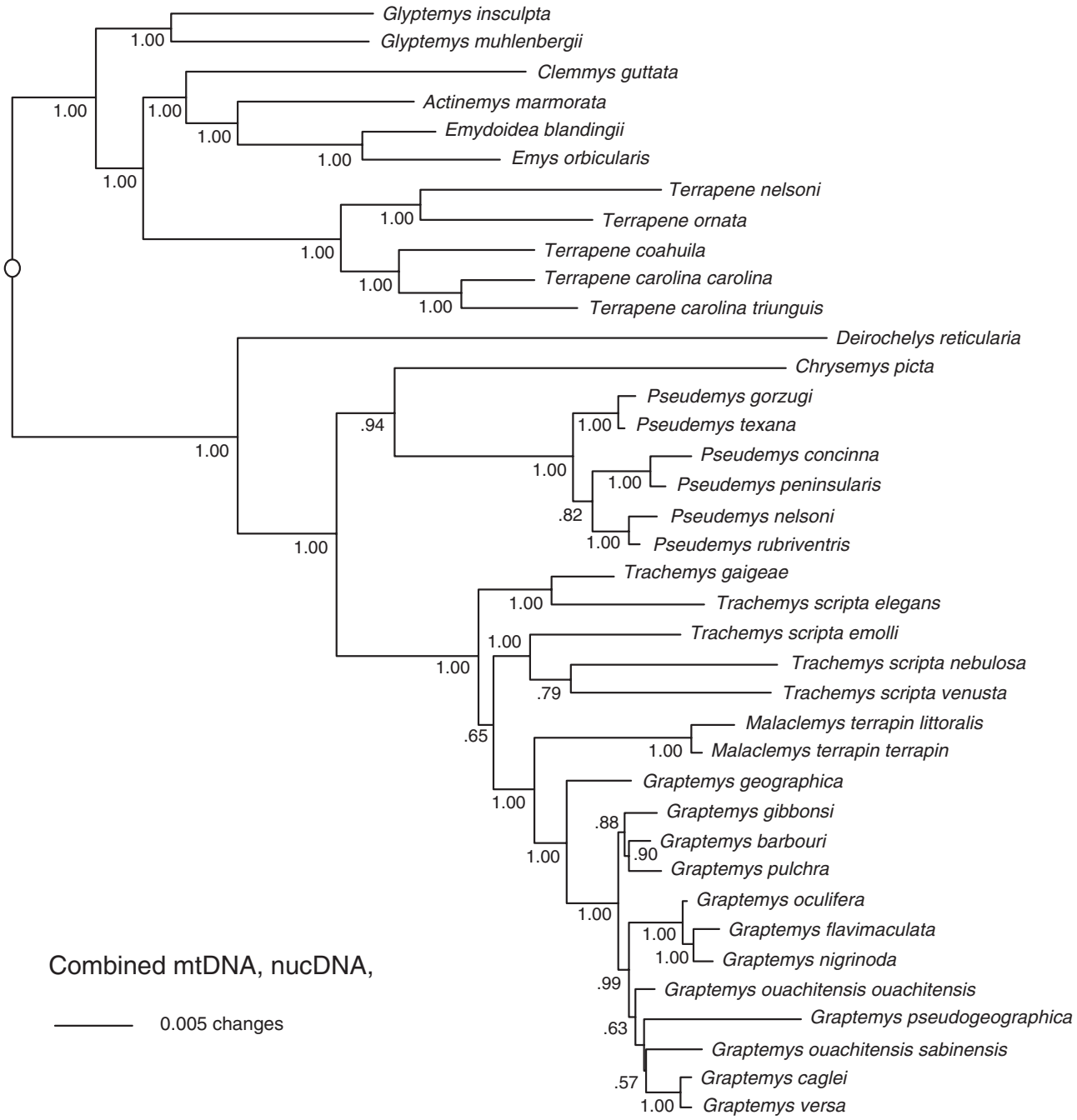


**Figure 3.** Phylogeny of emydid turtles based on a combined, partitioned Bayesian analysis of mitochondrial DNA (mtDNA) and nuclear DNA (nucDNA). Numbers adjacent to nodes indicate Bayesian posterior probabilities ($Pp$) $\geq 0.50$. The outgroup taxa are excluded to facilitate presentation of branch lengths, and the root is indicated with an open circle.

variability in turtles in general relative to other vertebrates (Avise *et al.*, 1992), possibly as a result of low metabolic rates or long generation times, although this would not explain the contrasting patterns among emydid genera. Given that sequences are similar but not identical among species within these genera, contamination or other laboratory error also seems unlikely. The problem occurs regardless of whether samples are from blood or tissue (J. J. Wiens, unpubl. data). Furthermore, a previous study of *Graptemys* phylogeny also found unusually limited divergence in their mitochondrial genes (Lamb *et al.*,

1994). Finally, at least one recent study has revealed evidence for the presence of numts in a clade of turtles that is closely related to emydids (geoemydids; Spinks & Shaffer, 2007).

Two pieces of information make these putative numts in emydids seem particularly surprising. First, we found no stop codons or indels in these two genes. This raises the question of whether these genes continue to function in the nuclear genome, and whether mitochondrial copies still persist. Second, it is surprising that both genes have been transported to the nuclear genome, and in the ancestors of each of these two genera. Alternately, it may be that our primers only amplify the nuclear copies in these genera (although it seems strange that this would happen coincidentally in both genera in both genes). This phenomenon would be an interesting subject for future study.

Similar problems with one or both of these genes may be present in other taxa in our putative mtDNA data, in addition to *Graptemys* and *Pseudemys*. In both cyt *b* and ND4, the two samples of *Malaclemys* are nearly identical despite considerable geographic distance separating them (Maryland versus Gulf Coast). The genera *Emydoidea* and *Emys* are nearly identical in cyt *b* but highly divergent in ND4. Given that *Emydoidea* occurs in North America and *Emys* in Europe (Ernst & Barbour, 1989), recent divergence or gene flow seems highly unlikely (and should be reflected in ND4 if present). The two subspecies of *Terrapene carolina* included (*T. c. carolina* and *T. c. triunguis*) are highly divergent in ND4 sequences (and do not appear as sister taxa) but are nearly identical for cyt *b* (again the pattern occurring in only one gene would seem to argue against mitochondrial introgression between these taxa). When the two mitochondrial genes are combined, these two taxa appear strongly supported as sister taxa, but not in the combined nucDNA. Thus, this may explain at least one strongly supported conflict between nucDNA and mtDNA in emydids. In general, the combination of numts with seemingly true mitochondrial genes may explain other instances of incongruence in our data, but these cases are less obvious.

A recent study (Spinks & Shaffer, 2009) has suggested that conflicts between mtDNA and nucDNA over relationships among *Actinemys*, *Emys*, and *Emydoidea* are explained by mitochondrial introgression between *Actinemys* and *Emydoidea*. However, the authors admit that the patterns they discuss could also have arisen via incomplete lineage sorting, and they present no conclusive evidence to support or refute either hypothesis. Furthermore, the analyses of ND4 in the present study found no evidence of a close relationship between *Actinemys* and *Emydoidea* as would be expected if mitochondrial introgression between these genera had occurred in the past (although many emydine relationships from ND4 are weakly supported).

### Implications of misleading branch lengths

The phenomenon of potentially misleading mitochondrial branch lengths that we observe in *Graptemys* and *Pseudemys* has many implications. For example, looking at the tree from the mtDNA data, it might reasonably be inferred that the species of *Graptemys* and *Pseudemys* are very recent, and underwent very rapid speciation and morphological divergence. Indeed, a previous study of *Graptemys* (Lamb *et al.*, 1994) noted remarkably low genetic divergence in mtDNA among these species, and suggested that these species may have all diverged very rapidly in the Pleistocene (see also Stephens & Wiens, 2003b). However, the nuclear data do not appear to support this assertion (especially not for *Pseudemys*). Similarly, a study in birds (Price *et al.*, 1998) incorrectly inferred rapid radiation within the warbler genus *Dendroica* based on inadvertently including slow-evolving nuclear copies of a putative mitochondrial gene (Price *et al.*, 2000).

There is also the possibility that species with almost identical mtDNA sequences might not be considered distinct. Clearly, this would be a mistake and might even lead to distinct species that are vulnerable to extinction losing their protected status, especially if they are not well differentiated morphologically. In fact, *G. flavimaculata*, *G. oculifera*, and *Pseudemys alabamensis* are listed as endangered on the IUCN Red List of Threatened Species (http://www.iucnredlist.org/). In summary, the results obtained from emydid turtles appear to reinforce the importance of inferring phylogeny and branch lengths from multiple nuclear loci in addition to mtDNA (see also review in Rubinoff & Holland, 2005). However, the extent and frequency of these potential problems will remain uncertain without further quantitative comparisons of trees from mtDNA and multiple nuclear loci.

### Discordance among nuclear genes

Our results also suggest the benefits of using a large number of nuclear loci, and the dangers of using only one or two. We found that different loci vary dramatically in their concordance with the combined nucDNA tree. For example, some loci (e.g. ETS, NGFB) differ for most nodes (> 70%) and with strong support in many cases (> 50% of nodes for ETS). These results suggest that basing a phylogenetic hypothesis on a single nuclear locus may lead to reconstructing many incorrect branches, possibly with strong support. On the positive side, three of the six loci showed no

strongly supported conflicts with the combined nucDNA tree. We also found that the congruence among genes increases for longer branches (Wiens *et al.*, 2008), which is a pattern suggesting that some incongruence is a result of incomplete lineage sorting. Other potential causes of discordance include introgression and paralogy, although we did not find obvious signatures for either problem in our data. For example, we found no pairs of species that are strongly placed together by some nuclear genes but strongly placed elsewhere by others (which might indicate recent introgression). Nor did we find a clade of seemingly distantly-related species embedded within any of the gene trees (which might indicate ancient gene duplication and paralogy). However, cases of ancient introgression or recent paralogy may be more difficult to detect, and may explain some conflict in our data.

## Progress in emydid turtle phylogeny and areas for future research

The present study not only resolves many outstanding issues in emydid phylogeny, but also suggests the need for additional work. Within subfamily Emydinae, generic and species-level relationships are strongly resolved by the nucDNA data, and many clades are now supported by both mtDNA, nucDNA, and combined analyses. These include the monophyly and basal position of *Glyptemys*, the *Emys* + *Emydoidea* clade, and the monophyly and most interspecific relationships within *Terrapene*. However, the position of *Actinemys marmorata* is strongly disputed by nucDNA and mtDNA, and by different nuclear genes. Placement of this species is also unresolved by Spinks & Shaffer (2009); likelihood analyses of their mtDNA and nucDNA data each give conflicting but weakly supported resolutions, and the species is in a large basal polytomy in their Bayesian species-tree (BEST) analysis. Additional analyses, including more nuclear loci, may be necessary to resolve its relationships. Although Spinks & Shaffer (2009) argue that *Actinemys* and *Emydoidea* should be placed within *Emys*, there is no phylogenetic justification for this change, and it leads to unnecessary instability in the long-standing generic names *Emys* and *Emydoidea*. Most importantly, our analyses of the combined nucDNA data show *Emys* (*sensu* Spinks & Shaffer, 2009) to be non-monophyletic, given that *Actinemys* clusters with *Clemmys guttata* (Fig. 2) with very strong support.

The present study also strongly resolves most generic-level relationships within Deirochelyinae. Placement of *Deirochelys* as sister to all other deirochelyines is concordant between nucDNA, mtDNA, and combined-data trees. However, placement of *Chrysemys* is strongly disputed and analyses of the separate nuclear genes are suprisingly ambiguous, despite strong support in the combined analysis (see also Spinks *et al.*, 2009). Relationships within *Pseudemys* are generally well-resolved by nucDNA and combined-data analyses (despite one conflict between them), but two species (*P. alabamensis*, *P. suwanniensis*) should be included in future analyses. The clade of *Trachemys* + *Malaclemys* + *Graptemys* is strongly supported by nucDNA, mtDNA, and combined analyses. The clade of *Malaclemys* + *Graptemys* is supported by nucDNA and combined-data analyses.

Monophyly of *Trachemys* is weakly supported by nucDNA and weakly rejected by mtDNA and combined-data analyses. Relationships of *Trachemys* are in need of further study, in order to resolve the monophyly of the genus and include more species (e.g. the West Indian species *T. decorata*, *T. decussata*, *T. stejnegeri*, and *T. terrapen*, and the South American species *T. adiutrix* and *T. dorbigni*). More geographic sampling is also needed within *T. scripta*, which contains 12 subspecies and extends from North America to South America (Uetz, 2008), and may contain many distinct species (Seidel, 2002).

Relationships within *Graptemys* are also in need of additional study. The nucDNA strongly support monophyly of the genus (weakly supported by mtDNA), and show strong support for the traditional sawback clade (*G. flavimaculata*, *G. nigrinoda*, *G. oculifera*) and the clade of Texas endemics (*G. caglei*, *G. versa*). The mtDNA strongly support placing *G. geographica* as sister taxon to all other species, and the *pulchra* group clade (including *G. barbouri*, *G. gibbonsi*, *G. pulchra*) as sister taxon to the remaining species, as found by Lamb *et al.* (1994). However, these relationships are not supported by the nucDNA. The combined-data trees include these strongly-supported clades from both the mtDNA and nucDNA trees, and may represent the best estimate of *Graptemys* relationships so far. Nevertheless, some relationships remain poorly supported (e.g. placement of *Graptemys pseudogeographica* and the seemingly polyphyletic *Graptemys ouachitensis*). Fully resolving relationships within *Graptemys* may require the addition of many more nuclear loci, especially ones that are sufficiently fast-evolving to be variable among the most closely-related species.

## Conclusions

Our results for emydid turtles show extensive discordance between mtDNA and nucDNA phylogenies. In addition to conflicts over phylogenetic placement of species, we also found extreme discordance in estimated branch lengths in at least two genera

(*Graptemys*, *Pseudemys*). The pattern of unusually short branch lengths in the mtDNA for these genera may be misleading for analyses of species limits, conservation, and estimates of divergence times, as well as rates of speciation and morphological divergence. The results also reveal extensive discordance between the trees from the nuclear genes. Thus, analyses of a single nuclear gene may give phylogenetic estimates that differ extensively (and with strong support) from the phylogeny based on multiple nuclear loci. Overall, the results obtained reinforce the need for using multiple nuclear loci to resolve organismal phylogenies, in addition to mtDNA. Fortunately, analyses that include multiple nuclear loci should become increasingly tractable as more genomes are sequenced and large numbers of nuclear markers are developed for phylogenetics using genomic resources (Townsend *et al.*, 2008).

## ACKNOWLEDGEMENTS

For their generous assistance in obtaining blood and tissue samples we thank E. Callender, M. Forstner, J. Howeth, C. Ivanyi and the Arizona-Sonora Desert Museum, S. Lee and J. Pramuk of the Bronx Zoo, P. Lindeman, T. Mann and the Mississippi Museum of Natural Science, C. Moritz and the Museum of Vertebrate Zoology (Berkeley), E. Myers, and W. M. Roosenburg. We thank A. Pyron and F. Burbrink for generously providing primer sequences for the Vimentin gene. For comments on the manuscript, we thank four anonymous reviewers. Molecular laboratory work was supported by an NSF Dissertation Improvement Grant DEB 0412793 to P.R.S. and J.J.W. and by EF 0334923 to J.J.W.

## REFERENCES

**Alfaro ME, Zoller S, Lutzoni F. 2003.** Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution* **20:** 255–266.

**Arevalo E, Davis SK, Sites JW. 1994.** Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosome races of the *Sceloporus grammicus* complex (Phrynosomatidae) in central Mexico. *Systematic Biology* **43:** 387–418.

**Avise JC, Bowen BW, Lamb T, Meylan AB, Bermingham E. 1992.** Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in the Testudines. *Molecular Biology and Evolution* **9:** 457–473.

**Ballard JWO, Rand DM. 2005.** The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology, Evolution, and Systematics* **36:** 621–642.

**Ballard JWO, Whitlock MC. 2004.** The incomplete history of mitochondria. *Molecular Ecology* **13:** 729–744.

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

**Brandley MC, Schmitz A, Reeder TW. 2005.** Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. *Systematic Biology* **54:** 373–390.

**Brown JM, Lemmon AR. 2007.** The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Systematic Biology* **56:** 643–655.

**Bull JJ, Huelsenbeck JP, Cunningham CW, Swofford DL, Waddell PJ. 1993.** Partitioning and combining data in phylogenetic analysis. *Systematic Biology* **4:** 384–379.

**Conant R, Collins JT. 1998.** *A field guide to reptiles and amphibians. Eastern and Central North America*, 3rd edn. Boston, MA: Houghton Mifflin Company.

**Degnan JH, Rosenberg NA. 2006.** Discordance of species trees with their most likely gene trees. *PLoS Genetics* **2:** 762–768.

**Dolman G, Phillips B. 2004.** Single copy nuclear DNA markers characterized for comparative phylogeography in Australian wet tropics rainforest skinks. *Molecular Ecology Notes* **4:** 185–187.

**Driskell AC, Ané C, Burleigh JG, McMahon MM, O'Meara BC, Sanderson MJ. 2004.** Prospects for building the tree of life from large sequence databases. *Science* **306:** 1172–1174.

**Edgar RC. 2004.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792–1797.

**Edwards SV, Liu L, Pearl DK. 2007.** High resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 5936–5941.

**Engstrom TN, Shaffer HB, McCord WP. 2004.** Multiple datasets, high homoplasy and the phylogeny of softshell turtles. *Systematic Biology* **53:** 693–710.

**Erixon P, Svennblad B, Britton T, Oxelman B. 2003.** Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic Biology* **52:** 665–673.

**Ernst CH, Barbour RW. 1989.** *Turtles of the world*. Washington, DC: Smithsonian Institution Press.

**Ernst CH, Lovich JE, Barbour RW. 1994.** *Turtles of the United States and Canada*. Washington, DC: Smithsonian Institution Press.

**Feldman CR, Parham JF. 2002.** Molecular phylogenetics of emydine turtles: taxonomic revision and the evolution of shell kinesis. *Molecular Phylogenetics and Evolution* **22:** 388–398.

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

**Friesen VL, Congdon BC, Kidd MG, Birt TP. 1999.** PCR

primers for the amplification of five nuclear introns in vertebrates. *Molecular Ecology* **8:** 2147–2149.

**Friesen VL, Congdon BC, Walsh HE, Birt TP. 1997.** Intron variation in marbled murrelets detected using analyses of single-stranded conformational polymorphisms. *Molecular Ecology* **6:** 1047–1058.

**Fujita MK, Engstrom TN, Starkey DE, Shaffer HB. 2004.** Turtle phylogeny: insights from a novel nuclear intron. *Molecular Phylogenetics and Evolution* **31:** 1031–1040.

**Hallstrom BM, Kullberg M, Nilsson MA, Janke A. 2007.** Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups. *Molecular Biology and Evolution* **24:** 2059–2068.

**Hillis DM, Bull JJ. 1993.** An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42:** 182–192.

**Hillis DM, Wilcox TP. 2005.** Phylogeny of the New World true frogs (*Rana*). *Molecular Phylogenetics and Evolution* **34:** 299–314.

**Huelsenbeck JP, Rannala B. 2004.** Frequentist properties of Bayesian posterior probabilities. *Systematic Biology* **53:** 904–913.

**Huelsenbeck JP, Ronquist F. 2001.** MrBayes: Bayesian inference of phylogeny. *Bioinformatics* **17:** 754–755.

**Hyman IT, Ho SYW, Jermiin LS. 2007.** Molecular phylogeny of Australian Helicarionidae, Euconulidae and related groups (Gastropoda: Pulmonata: Stylommatophora) based on mitochondrial DNA. *Molecular Phylogenetics and Evolution* **45:** 792–812.

**Klicka J, Burns K, Spellman GM. 2007.** Defining a monophyletic Cardinalini: a molecular perspective. *Molecular Phylogenetics and Evolution* **45:** 1014–1032.

**Kozak KH, Larson A, Bonett R, Harmon LJ. 2005.** Phylogenetic analysis of ecomorphological divergence, community structure, and diversification rates in dusky salamanders (Plethodontidae: *Desmognathus*). *Evolution* **59:** 2000–2016.

**Krenz JG, Naylor GJP, Shaffer HB, Janzen FJ. 2005.** Molecular phylogenetics and evolution of turtles. *Molecular Phylogenetics and Evolution* **37:** 178–191.

**Kubatko LS, Degnan JH. 2007.** Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* **56:** 17–24.

**Lamb T, Lydeard C, Walker RB, Gibbons JW. 1994.** Molecular systematics of map turtles (*Graptemys*): a comparison of mitochondrial restriction site versus sequence data. *Systematic Biology* **43:** 543–559.

**Lemmon EM, Lemmon AR, Collins JT, Lee-Yaw JA, Cannatella DC. 2007.** Phylogeny-based delimitation of species boundaries and contact zones in the trilling chorus frogs (*Pseudacris*). *Molecular Phylogenetics and Evolution* **44:** 1068–1082.

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

**Lewis PO. 2001.** A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50:** 913–925.

**Lyons LA, Laughlin TF, Copeland NG, Jenkins NA, Womack JE, O'Brien SJ. 1997.** Comparative anchor tagged sequences (CATS) for integrative mapping of mammalian genomes. *Nature Genetics* **15:** 47–56.

**Maddison DR, Maddison WP. 2000.** *MacClade*, Version 4.0. Sunderland, MA: Sinauer Associates.

**Maddison WP. 1997.** Gene trees in species trees. *Systematic Biology* **46:** 523–536.

**Moore WS. 1995.** Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* **49:** 718–726.

**Near TJ, Meylan PA, Shaffer HB. 2005.** Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. *American Naturalist* **165:** 137–146.

**Nylander JAA. 2004.** *MrModeltest*, Version 2.0. Software distributed by the author. Uppsala University: Evolutionary Biology Centre. Available at: http://www.ebc.uu.se/systzoo/staff/nylander.html

**Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL. 2004.** Bayesian phylogenetic analysis of combined data. *Systematic Biology* **53:** 47–67.

**Philippe H, Lartillot N, Brinkmann H. 2005.** Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa and Protostomia. *Molecular Biology and Evolution* **22:** 1246–1253.

**Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, Casane D. 2004.** Phylogenomics of eukaryotes: impact of missing data on large alignments. *Molecular Biology and Evolution* **21:** 1740–1752.

**Price T, Gibbs HL, de Sousa L, Richman AD. 1998.** Different timing of the adaptive radiations of North American and Asian warblers. *Proceedings of the Royal Society of London Series B, Biological Sciences* **265:** 1969–1975.

**Price T, Lovette IJ, Bermingham E, Gibbs HL, Richman AD. 2000.** The imprint of history on communities of North American and Asian warblers. *American Naturalist* **156:** 354–367.

**Rokas A, Williams BL, King N, Carroll SB. 2003.** Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425:** 798–804.

**Rubinoff D, Holland BS. 2005.** Between two extremes: mitochondrial DNA is neither the Panacea nor the Nemesis of phylogenetic and taxonomic inference. *Systematic Biology* **54:** 952–961.

**Seidel ME. 2002.** Taxonomic observations on extant species and subspecies of slider turtles, genus *Trachemys*. *Journal of Herpetology* **36:** 285–292.

**Shaw KL. 2002.** Conflict between mitochondrial and nuclear DNA phylogenies of a recent species radiation: what mitochondrial DNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 16122–16127.

**Simmons MP, Ochoterena H. 2000.** Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* **49:** 369–381.

**Spinks PQ, Shaffer HB. 2007.** Conservation phylogenetics of the Asian box turtles (Geoemydidae, *Cuora*): mitochondrial introgression, numts, and inferences from multiple nuclear loci. *Conservation Genetics* **8:** 641–657.

**Spinks PQ, Shaffer HB. 2009.** Conflicting mitochondrial and nuclear phylogenies for the widely disjunct *Emys* (Testudines: Emydidae) species complex, and what they tell us about biogeography and hybridization. *Systematic Biology* **58:** 1–20.

**Spinks PQ, Shaffer HB, Iverson JB, McCord WP. 2004.** Phylogenetic hypotheses for the turtle family Geoemydidae. *Molecular Phylogenetics and Evolution* **32:** 164–182.

**Spinks PQ, Thomson RC, Lovely GA, Shaffer HB. 2009.** Assessing what is needed to resolve a molecular phylogeny: simulations and empirical data from emydid turtles. *BMC Evolutionary Biology* **9:** 56.

**Starkey DE, Shaffer HB, Burke RR, Forstner MRJ, Iverson JB, Janzen FJ, Rhodin AGJ, Ultsch GR. 2003.** Molecular systematics, phylogeography, and the effects of Pleistocene glaciation in the painted turtle (*Chrysemys picta*) complex. *Evolution* **57:** 119–128.

**Stephens PR, Wiens JJ. 2003a.** Ecological diversification and phylogeny of emydid turtles. *Biological Journal of the Linnean Society* **79:** 577–610.

**Stephens PR, Wiens JJ. 2003b.** Explaining species richness from continents to communities: the time-for-speciation effect in emydid turtles. *American Naturalist* **161:** 112–128.

**Stephens PR, Wiens JJ. 2008.** Testing for evolutionary tradeoffs in a phylogenetic context: ecological diversification and evolution of locomotor performance in emydid turtles. *Journal of Evolutionary Biology* **21:** 77–87.

**Stephens PR, Wiens JJ. 2009.** Evolution of sexual size dimorphisms in emydid turtles: ecological dimorphism, Rensch's rule, and sympatric divergence. *Evolution* **63:** 910–925.

**Stuart BL, Parham JF. 2007.** Recent hybrid origin of three rare Chinese turtles. *Conservation Genetics* **8:** 169–175.

**Swofford DL. 2002.** *PAUP*: phylogenetic analysis using parsimony**, Version 4.0b10. Sunderland, MA: Sinauer Associates.

**Takezaki N, Figueroa F, Zaleska-Rutczynska Z, Takahata N, Klein J. 2004.** The phylogenetic relationship of tetrapod, coelacanth, and lungfish revealed by the sequences of 44 nuclear genes. *Molecular Biology and Evolution* **21:** 1512–1524.

**Thiele K. 1993.** The Holy Grail of the perfect character: The cladistic treatment of morphometric data. *Cladistics* **9:** 275–304.

**Townsend TM, Alegre ER, Kelley ST, Wiens JJ, Reeder TW. 2008.** Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: an example from squamate reptiles. *Molecular Phylogenetics and Evolution* **47:** 129–142.

**Turtle Taxonomy Working Group. 2007.** An annotated list of modern turtle terminal taxa, with comments on areas of taxonomic instability and recent change. *Chelonian Research Monographs* **4:** 173–199.

**Uetz P. 2008.** *The TIGR reptile database*. Available at: http://www.reptile-database.org. Accessed June 26, 2008.

**Weisrock DW, Harmon LJ, Larson A. 2005.** Resolving the deep phylogenetic relationships among salamander families: analyses of mitochondrial and nuclear genomic data. *Systematic Biology* **54:** 758–777.

**Wiens JJ. 1995.** Polymorphic characters in phylogenetic systematics. *Systematic Biology* **44:** 482–500.

**Wiens JJ. 1998.** Combining data sets with different phylogenetic histories. *Systematic Biology* **47:** 568–581.

**Wiens JJ. 2001.** Character analysis in morphological phylogenetics: problems and solutions. *Systematic Biology* **50:** 689–699.

**Wiens JJ. 2003.** Missing data, incomplete taxa, and phylogenetic accuracy. *Systematic Biology* **52:** 528–538.

**Wiens JJ, Engstrom TN, Chippindale PT. 2006.** Rapid diversification, incomplete isolation, and the 'speciation clock' in North American salamanders (genus *Plethodon*): testing the hybrid swarm hypothesis of rapid radiation. *Evolution* **60:** 2585–2603.

**Wiens JJ, Fetzner JW, Parkinson CL, Reeder TW. 2005.** Hylid frog phylogeny and sampling strategies for speciose clades. *Systematic Biology* **54:** 719–748.

**Wiens JJ, Hollingsworth BD. 2000.** War of the iguanas: conflicting molecular and morphological phylogenies and long-branch attraction in iguanid lizards. *Systematic Biology* **49:** 143–159.

**Wiens JJ, Kuczynski CA, Smith SA, Mulcahy D, Sites JW Jr, Townsend TM, Reeder TW. 2008.** Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. *Systematic Biology* **57:** 420–431.

**Wiens JJ, Moen DS. 2008.** Missing data and the accuracy of Bayesian phylogenetics. *Journal of Systematics and Evolution* **46:** 307–314.

**Wilcox TP, Zwickl DJ, Heath TA, Hillis DM. 2002.** Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Molecular Phylogenetics and Evolution* **25:** 361–371.

**Wilson NG, Schrodl M, Halanych KM. 2009.** Ocean barriers and glaciation: evidence for explosive radiation of mitochondrial lineages in the Antarctic sea slug *Doris kerguelenensis* (Mollusca, Nudibranchia). *Molecular Ecology* **18:** 965–984.

**Zhang D-X, Hewitt GM. 1996.** Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution* **11:** 247–251.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Appendix S1.** Sources and vouchers for turtles used for molecular data in this study.
**Appendix S2.** GenBank numbers for sequences used in the phylogenetic analyses.
**Appendix S3.** Morphological data matrix used in some of the phylogenetic analyses.
**Appendix S4.** Trees from separate Bayesian analyses of each of the mitochondrial and nuclear genes.
**Appendix S5.** Trees from parsimony analyses of the combined mtDNA data, combined nucDNA data, combined mtDNA and nucDNA, and combined mtDNA, nucDNA, and morphological data.
**Appendix S6.** Trees from Bayesian analysis of the combined nucDNA, mtDNA, and morphological data, Bayesian analysis of the combined nucDNA and morphological data, parsimony analysis of the combined nucDNA and morphological data, Bayesian analysis of the morphological data alone, and parsimony analysis of the morphological data alone.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.