Sequence-based molecular phylogenetics and phylogeography of the American box turtles (Terrapene spp.) with support from DNA barcoding

Bradley T. Martin a,*, Neil P. Bernstein b, Roger D. Birkhead c, Jim F. Koukl a, Steven M. Mussmann d, John S. Placyk Jr. a

aDepartment of Biology, The University of Texas at Tyler, Tyler, TX 75799, USA
bDepartment of Natural and Applied Sciences, Mount Mercy University, Cedar Rapids, IA 52402, USA
cAlabama Science in Motion, Auburn University, Auburn, AL 36849, USA
dDepartment of Biological Sciences, University of Arkansas, Fayetteville, AR 72701, USA

ARTICLE INFO

Article history:
Received 31 December 2012
Accepted 8 March 2013
Available online 20 March 2013

Keywords:
Box turtle
DNA barcoding
Molecular phylogenetics
Phylogeography
Terrapene
Emydidae

ABSTRACT

The classification of the American box turtles (Terrapene spp.) has remained enigmatic to systematists. Previous comprehensive phylogenetic studies focused primarily on morphology. The goal of this study was to re-assess the classification of Terrapene spp. by obtaining DNA sequence data from a broad geographic range and from all four recognized species and 11 subspecies within the genus. Tissue samples were obtained for all taxa except for Terrapene nelsoni klauberi. DNA was extracted, and the mitochondrial DNA (mtDNA) cytochrome b (Cytb) and nuclear DNA (nucDNA) glyceraldehyde-3-phosphate-dehydrogenase (GAPD) genes were amplified via polymerase chain reaction and sequenced. In addition, the mtDNA gene commonly used for DNA barcoding (cytochrome oxidase c subunit I; COI) was amplified and sequenced to calculate pairwise percent DNA sequence divergence comparisons for each Terrapene taxon. The sequence data were analyzed using maximum likelihood and Bayesian phylogenetic inference, a molecular clock, AMOVAs, SAMOVAs, haplotype networks, and pairwise percent sequence divergence comparisons. Terrapene carolina mexicana and T. c. yucatanana formed a monophyletic clade with T. c. triunguis, and this clade was paraphyletic to the rest of T. carolina. Terrapene ornata ornata and T. o. luteola lacked distinction phylogenetically, and Terrapene nelsoni was confirmed to be the sister taxon of T. ornata. Terrapene c. major, T. c. bauri, and Terrapene coahuila were not well resolved for some of the analyses. The DNA barcoding results indicated that all taxa were different species (>2% sequence divergence) except for T. c. triunguis – T. c. mexicana and T. o. ornata – T. o. luteola. The results suggest that T. c. triunguis should be elevated to species status (Terrapene mexicana), and mexicana and yucatanana should be included in this group as subspecies. In addition, T. o. ornata and T. o. luteola should not be considered separate subspecies. The DNA barcoding data support these recommended taxonomic revisions. Because conservation efforts are typically species-based, these results will be important for facilitating successful conservation management strategies.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The American box turtle genus, Terrapene (Chelonia, Emydidae), includes four primarily terrestrial New World species, the eastern (T. Carolina; Linnaeus, 1758), ornate (T. ornata; Agassiz, 1857), spotted (T. nelsoni; Stejneger, 1925), and Coahuilan (T. coahuila; Schmidt and Owens, 1944) box turtles, characterized, in part, by their ability to tightly close their plastron via a single moveable hinge (plastral shell kinesis). With the exception of T. coahuila, each species is further divided into multiple subspecies including the eastern (T. c. carolina), three-toed (T. c. triunguis), Mexican (T. c. mexicana), Yucatan (T. c. yucatanana), Gulf Coast (T. c. major), Florida (T. c. bauri), and the purportedly extinct giant (T. c. putnamii) box turtles for T. carolina; the ornate (T. o. ornata) and desert (T. o. luteola) box turtles for T. ornata; and the northern spotted (T. n. klauberi) and southern spotted (T. n. nelsoni) box turtles for T. nelsoni. While their current classification is generally accepted based on morphological data and geographic distributions, the use of molecular data in understanding the evolutionary history of the group has been limited, as the systematic studies that have included Terrapene have (1) focused on higher level intergeneric classification, (2) failed to include all taxa within Terrapene (e.g., all species and subspecies), or (3) been limited in sample sizes and/or geographic sampling (Bickham et al., 1996; Butler et al., 2011; Feldman and
Parham, 2002; Herrmann and Rosen, 2009; Spinks and Shaffer, 2009; Stephens and Wiens, 2003; Wiens et al., 2010). When Terrapene has been the focus of more detailed molecular taxonomic scrutiny, issues inherent in the genus have been problematic. Furthermore, although the importance of using morphological data in phylogenetic analyses has recently been emphasized (Losos et al., 2012), issues specific to Terrapene have also complicated the use of morphology for assessing their classification.

First, paraphyly within Terrapene carolina has made phylogenetic inference of this species particularly problematic with Stephens and Wiens (2003) stating that T. carolina “might also consist of multiple species.” Second, a significant amount of inter- and intraspecific morphological variation along with the overlap of characters has made phylogenetic inference based on morphology, which is what the current classification is based on (e.g., Minx, 1996), less than useful. Third, intergradation between sympatric taxa imposes significant difficulties in resolving specific and subspecific relationships (Butler et al., 2011; Carr, 1940, 1952; Conant and Collins, 1991; Milstead, 1969). Therefore, a more thorough molecular phylogenetic investigation is warranted to resolve these issues and to assess the validity of the current classification. The lack of such information is compounded by the problematic conservation status of members of this genus.

Terrapene populations are declining throughout their range, in part, due to habitat loss resulting from increasing urbanization, collection from the wild for the pet trade, and changes in predator pressures (Dodd, 2001). The 2011 International Union for Conservation of Nature (IUCN) Red List classifies T. carolina as Vulnerable, T. ornata as Near Threatened, T. coahuila as Endangered with a Very High Risk of Extinction, and T. nelsoni as Data Deficient (although it was listed as Threatened on the 2006 Red List). In the United States, the various subspecies of the endemic T. carolina and T. ornata are state listed as Species of Special Concern in New Hampshire, Connecticut, Michigan, Texas, and Massachusetts, Protected in Indiana and Kansas, Threatened in Iowa, and Endangered in Maine, Wisconsin, and Illinois. Thus, understanding Terrapene evolutionary history is particularly urgent, as conservation efforts are often species-based. Our research will be important to both those interested in the evolutionary history of the Terrapene, as well as those that might be attempting to conserve them.

The goals of this research are to (1) resolve the evolutionary history of the Terrapene genus by assessing their classification using molecular phylogenetic and genetic barcoding data and (2) assess the population structure within Terrapene to evaluate the phylogeography of the genus.

2. Methods

Tissue samples in the form of toenails, shell shavings and fragments, bone fragments, scutes, muscle tissue, feces, and blood were collected/obtained for all four extant species and all but one of the ten currently recognized subspecies (i.e., T. n. klauberi; Appendix A). Geographic distributions based on Dodd (2001) and morphological identification were used to identify collected specimens. At least three individuals of each subspecies were chosen from every US state within their range to obtain a sufficient sample size and to include wide geographic sampling (Fig. 1). In addition, at least three tissue samples were used in the data analyses for the Mexican species and subspecies (except for T. c. yucatana and T. n. carolina).

Fig. 1. Sampling localities for the mitochondrial DNA (mtDNA) cytochrome b gene and the nuclear DNA (nucDNA) glyceraldehyde-3-phosphate dehydrogenase (GAPD) gene. The hatched region in the central USA indicates a sympatric zone with more than one subspecies present as proposed by Dodd (2001), and the hatching in the Gulf Coast region indicates a potential intergradation zone as indicated by Carr (1952).
nelsoni due to the very limited amount of tissue samples available for these taxa).

2.1. DNA extractions, PCR, and sequencing

Genomic DNA was extracted from tissue samples with the illustration™ Tissue & cells genomicPrep Mini Spin Kit (GE Healthcare). One mitochondrial DNA (mtDNA) gene (cytochrome b (Cytb) and one nuclear DNA (nucDNA) gene (glyceraldehyde-3-phosphate dehydrogenase (GAPD)) were then amplified and sequenced. For Cytb, the entire 1097 base pair (bp) gene along with part of the adjacent tRNA-threonine (tRNA-thr) gene was amplified and sequenced using the forward primer CybG and the reverse primer THR-8 (Engstrom et al., 2007; Spinks et al., 2004). A 430-bp region of the GAPD gene, including intron 11 and partial coding region, was amplified with each PCR. The following parameters were used for the Cytb DNA amplification: 35 cycles of 1 min denaturing at 94°C, 1 min annealing at 51°C, and 2 min DNA elongation at 72°C. GAPD PCR parameters were as follows: initial denaturation for 5 min at 94°C followed by 35 cycles of 30 s denaturing at 94°C, 1 min annealing at 63°C, and 1 min 30 s extension at 72°C. Amplification of the PCR product was verified via gel electrophoresis and the amplified DNA was purified with the EZ.N.A. Cycle Pure Kit (OMEGA biotek). Purified DNA was concentrated to the level recommended by Eurofins MWG Operon (20–40 ng/µL) and shipped to Eurofins MWG Operon for sequencing reactions using BigDye® Terminator v 3.1 Cycle Sequencing kits (Applied Biosystems).

2.2. Sequence analysis and phylogenetic inference

DNA was sequenced on an ABI 3730xl DNA sequencer at Eurofins MWG Operon and manually proofread and edited using Sequencer 4.9 (Gene Codes Corporation). Sequence alignments were conducted in Clustal X 2.0.11 (Thompson et al., 1997). Final editing was done using MacClade 4.08 (Maddison and Maddison, 1989). When available, existing GenBank sequences were included in the analyses (Table 1). The nucDNA GAPD sequence alignment contained IUPAC ambiguity characters designating heterozygous sites, and because some phylogenetic inference software does not recognize ambiguity characters we used the program HaploType Inference by Parsimony (HAPAR) to infer haplotypes from genotypic data (Wang and Xu, 2003). HAPAR uses maximum parsimony to resolve heterozygous characters into each possible genotype and collapses identical genotypes to find the minimum number of haploid haplotypes. HAPAR analysis for the GAPD gene was used for all analyses except for the analyses of molecular variance (AMOVAs) and the spatial analyses of molecular variance (SAMOVAs), which are able to read IUPAC ambiguity characters (Excoffier et al., 2005).

Tajima’s D and Fu and Li’s D′ and F′ tests for neutrality were conducted for each gene using DnaSp v 5.10.01 (Librado and Rozas, 2009) to confirm that natural selection did not significantly influence the phylogenetic data and that the inferred phylogeny largely reflects the background rate of mutation (Fu and Li, 1993; Tajima, 1989). Phylogenies were inferred via maximum likelihood (ML: Felsenstein, 1981), and Bayesian inference (BI: Larget and Simon, 1999; Rannala and Yang, 1996; Smouse and Li, 1989; Yang and Rannala, 1997) methods. PhyML 3.0 was used to generate ML trees (Guindon et al., 2010), and BEAST v1.6.2 was used to infer BI trees (Drummond and Rambaut, 2007). Non-parametric bootstrap resampling (Felsenstein, 1985) was employed to quantify the statistical support for ML phylogenies and the Markov chain Monte Carlo (MCMC) method was used to infer confidence values for BI (Mau, 1996; Mau and Newton, 1997; Mau et al., 1999; Rannala and Yang, 1996; Yang and Rannala, 1997). One thousand non-parametric bootstrap replications were used to generate ML trees (Pattengale et al., 2010). Bootstrap support values above 70% were considered well-supported (Hillis and Bull, 1993). The Cytb BI analysis was run for 3.0 × 10⁶ MCMC generations using default temperatures and with sampling trees occurring every 100 generations, while the GAPD BI analysis was run for 3.0 × 10⁶ MCMC generations and with sampling trees occurring every 1,000 generations. The aforementioned 3.0 × 10⁶ and 3.0 × 10⁷ MCMC generations were chosen to make effective sample sizes (ESS) >200 for each individual parameter in the analysis, as determined by Tracer v1.5, a program recommended by Drummond and Rambaut (2007) to analyze BEAST v1.6.2 output files. The likelihood scores were monitored during each analysis until stabilization, and the samples obtained prior to stabilization (7500 for both Cytb and GAPD) were discarded as burn-in (Parham et al., 2006). Nodes having a Bayesian posterior probability (BPP) >0.95 were considered well-supported (Huelsenbeck and Ronquist, 2001). For all analyses, jModelTest 0.1 was used to determine substitution model parameters using the Akaike Information Criterion corrected for small sample size (AICc; Posada, 2008). ML and BI analyses were conducted for Cytb using the TmP2uF + I + G substitution model, with I = 0.4450, G = 0.6160, and the sample size = 1097. The rate matrix parameters were as follows: AC = 3.2798, AG = 21.7809, AT = 3.2798, CG = 1.0000, CT = 21.7809, and GT = 1.0000. The base frequencies were set to 0.3059 (A), 0.3131 (C), 0.1213 (G), and 0.2598 (T). GAPD ML and BI analyses were conducted using the TmP2uF + I substitution model, with I = 0.8480 and the sample size = 430. The GAPD rate

<table>
<thead>
<tr>
<th>Taxon</th>
<th>GenBank ID</th>
<th>GAPD</th>
<th>COI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrapene carolina</td>
<td>carolina</td>
<td>AF258871</td>
<td>GQ896138</td>
</tr>
<tr>
<td>Terrapene carolina</td>
<td>triangula</td>
<td>FJT70616</td>
<td>GQ896139</td>
</tr>
<tr>
<td>Terrapene coahuila</td>
<td>-</td>
<td>AF258872</td>
<td>GQ896140</td>
</tr>
<tr>
<td>Terrapene ornata</td>
<td>ornata</td>
<td>GQ896203</td>
<td>GQ896142</td>
</tr>
<tr>
<td>Terrapene ornata</td>
<td>tuculola</td>
<td>AF258874</td>
<td>N/A</td>
</tr>
<tr>
<td>Terrapene nelsoni</td>
<td>nelsoni</td>
<td>AF258873</td>
<td>GQ896141</td>
</tr>
<tr>
<td>Clemmys guttata</td>
<td>-</td>
<td>FJT70591</td>
<td>GQ896113</td>
</tr>
<tr>
<td>Glyptemys insculpta</td>
<td>-</td>
<td>AF258876</td>
<td></td>
</tr>
<tr>
<td>Glyptemys muhlenbergii</td>
<td>-</td>
<td>AF258875</td>
<td></td>
</tr>
<tr>
<td>Emys orbicularis</td>
<td>-</td>
<td>AF258868</td>
<td></td>
</tr>
<tr>
<td>Emys marmorata</td>
<td>-</td>
<td>AF258867</td>
<td></td>
</tr>
<tr>
<td>Emydidae blandingii</td>
<td>-</td>
<td>AF258869</td>
<td></td>
</tr>
</tbody>
</table>

*Table 1* GenBank Accession numbers used in this study for the cytochrome b (Cytb), glyceraldehyde-3-phosphate dehydrogenase (GAPD), and cytochrome oxidase c subunit 1 (COI) genes.
matrix parameters were as follows: AC = 4.1405, AG = 8.7269, AT = 4.1405, CG = 1.0000, CT = 8.7269, and GT = 1.0000. The base frequencies were set to 0.2089 (A), 0.2012 (C), 0.3000 (G), and 0.2900 (T). Each phylogeny was rooted with a published GenBank sequence from Clemmys (Table 1), which is considered the sister genus to Terrapene (Bickham et al., 1996; Bramble, 1974; Feldman and Parham, 2002; Stephens and Wiens, 2003). The Clemmys Gen-Bank reference sequence for GAPD contained IUPAC ambiguity characters and was separated into two haplotypes after HAPAR analysis. Zero-length branches for all genes were collapsed into unique haplotypes using Collapse 1.2 to reduce clutter and computation time.

2.2.1. Combined mtDNA and nucDNA phylogenetic analysis

The Cytb and GAPD DNA sequence data were concatenated into a single dataset to infer a combined mtDNA and nucDNA phylogram. The combined dataset contained 172 sequences condensed into 121 haplotypes. Prior to phylogenetic analysis, an Incongruence Length Difference (ILD; Farris et al., 1994; Mickevich and Farris, 1998) test with 100 replicates was performed using PAUP v4.0b10 (Swoford, 2003) to assess whether the topologies of the two trees were congruent. Each gene was partitioned separately for the combined analysis, and the model parameters for each partition were kept the same as for the individual analyses.

2.3. Haplotype networks

Haplotype networks are useful for visualizing sequence variation within species or among closely related congeners because (1) there could potentially be a high number of mutational variants and (2) reversion to ancestral haplotypes is possible in recently diverged taxa (Crandon, 1994; Posada and Crandon, 2001). Because these conditions could apply to some taxa within Terrapene, haplotype networks may provide some insight into intraspecific and interspecific population structuring for this genus.

Using the 95% statistical parsimony procedure (Templeton, 1998), TCS 1.13 (Templeton et al., 1992) was used to estimate a genealogy. For clarity, reticulation loops were removed a posteriori based on coalescent theory (Crandon, 1994). Haplotype bubbles were sized relative to the number of sequences within each haplotype. The haplotype bubbles connected by branches differed by one mutational step, and smaller bubbles were placed on the branches to represent missing intermediate steps.

2.4. AMOVA and SAMOVA analyses

Spatial Analysis of Molecular Variance (SAMOVA) was performed using SAMOVA 1.0 (Dupanloup et al., 2002) and Analysis of Molecular Variance (AMOVA; Excoffier et al., 1992) was conducted using ARLEQUIN v. 3.11 (Excoffier et al., 2005) to examine population structures. Tests of AMOVA based on the mtDNA and nucDNA phylogenies consisted of five total groups and were as follows: (1) T. c. carolina – T. c. major – T. c. bauri, (2) T. c. triunguis – T. c. mexicana – T. c. yucatan, (4) T. o. ornata – T. o. luteola, and (5) T. nelsoni. Third, T. coahuila was apportioned into a group along with T. c. carolina and T. c. major to evaluate whether a greater population structure would be observed by grouping T. c. carolina with the T. c. major group. Fourth, T. c. bauri was combined with T. c. carolina – T. c. major to evaluate whether a greater amount of population structure would be observed by grouping T. c. bauri with the T. c. carolina group.

2.5. Cytb molecular clock analysis

A relaxed, uncorrelated lognormal molecular clock was placed on the Cytb sequence data using the BEAST v1.6.2 software package to estimate divergence times (Drummond and Rambaut, 2007). A molecular clock was not inferred for GAPD due to a generally lower resolution of the ML and Bayesian phylogenograms. Clemmys guttata, Glyptemys muhlenbergii, G. insculpta, Emys orbicularis, and E. marmorata were used to root the tree, as previous data indicated that these genera are most closely related to Terrapene within Emydidae (Bickham et al., 1996; Bramble, 1974; Feldman and Parham, 2002; Stephens and Wiens, 2003). The molecular clock analysis was conducted using the Yule Process speciation tree prior and a GTR + I + G substitution model with the parameters and tree priors equal to those for the previously mentioned Cytb ML and BI phylogenetic analyses, as determined by jModelTest 0.1 (Posada, 2008). Fossil data and previously published divergence estimates were used to calibrate the molecular clock. The root was calibrated to 25.4 million years ago (mya), with the standard deviation (SD) = 2.01, as the most recent common ancestor (MRCA) to Emydinae, and the MRCA to Glyptemys was calibrated to 17.0 mya, with the SD = 2.90 (Spinks and Shaffer, 2009). Based on fossil data, the
MRCA to *T. ornata* was calibrated at 12.5 mya with the SD = 1.00, and the MRCA to *T. carolina* was calibrated to 10.0 mya with the SD = 1.00 (Holman and Fritz, 2005; Spinks and Shaffer, 2009). The molecular clock analysis was conducted using a Markov chain of \(7.5 \times 10^6\) generations, with sampling occurring every 1000 generations. As with the previously mentioned BI trees, \(7.5 \times 10^6\) generations were chosen to bring the ESS > 200 for all parameters and tree priors in the analysis when analyzed using Tracer v1.5 (Drummond and Rambaut, 2007). The number of generations excluded from the analysis as burn-in was chosen based on visual inspection for stabilization of the log likelihood values.

2.6. Pairwise percent sequence divergences

2.6.1. Cytb and GAPD

Pairwise Jukes-Cantor DNA sequence divergences corrected for population comparisons (Jukes and Cantor, 1969) were calculated with DnaSP v5.10.01 (Librado and Rozas, 2009) and used to...
calculate the percent sequence divergence between taxonomic groups. Interspecific mtDNA sequences for most Emydine turtles typically vary between 4% and 6%, with a mean of 5% for the mtDNA Cytb gene, and interspecific nuclear DNA sequences typically vary between ~0.2–4%, with a mean of ~1% in most freshwater turtles and tortoises, depending on the gene (Feldman and Parham, 2002; FitzSimmons and Hart, 2007). Therefore these average values were used as references to compare the species and subspecies within Terrapene.

2.6.2. DNA barcoding

A 650 base pair region of the mtDNA cytochrome oxidase c subunit 1 (COI) gene was amplified and sequenced for DNA barcoding purposes. The 20 μL COI PCR reactions consisted of the same volumetric ratios as with the Cytb and GAPD genes (see Section 2.1.). Two primer sets were used to amplify and sequence the selected COI region. The first consisted of the forward primer L-turtCOI and the reverse primer H-turtCOIb (Stuart and Parham, 2004), and the second consisted of the forward primer VF2_t1 (Ivanova et al., 2007). The following PCR parameters were used with the L-turtCOI and H-turtCOIb primers: an initial denaturation of 95 °C for 5 min; 35 cycles of 95 °C for 45 s, 55 °C for 45 s, 72 °C for 45 s; and a final extension of 72 °C for 6 min. PCR parameters for the VF2_t1 and FR1d_t1 primers were as follows: an initial denaturation of 94 °C for 2 min; 35 cycles of 94 °C for 30 s, 52 °C for 40 s, and 72 °C for 1 min; and a final extension of 72 °C for 10 min.

Pairwise Kimura 2-parameter (K2P) divergence comparisons were performed in Mega 5.05 (Tamura et al., 2011). K2P comparisons were utilized over Jukes and Cantor divergence comparisons to be consistent in comparisons with the Barcode of Life Data Systems (BOLD) database (www.boldsystems.org). Reid et al. (2011) indicated that interspecific COI sequences for many organisms (including turtles) are divergent by >2% and intraspecific percent sequence divergences are often <2%. This 2% sequence divergence was used as a reference value for Terrapene pairwise DNA sequence comparisons to evaluate their taxonomic status. As an a posteriori analysis based on the Cytb results, which showed two groups composed primarily of T. c. major. T. c. major was divided into two separate groups, with one consisting solely of T. c. major and one consisting of T. c. major and T. coahuila. This analysis was performed to assess whether T. c. major consisted of more than one phylogenetic lineage.

3. Results

3.1. Phylogenetic analyses

3.1.1. Cytochrome b

The mtDNA Cytb phylogram contains 253 sequences distributed into 103 haplotypes (Fig. 2). Out of the 1097 characters, 254 are variable, and 191 are parsimoniously informative. Tajima’s D, Fu and Li’s D’ and F’ tests indicate that Cytb is not significantly being influenced by natural selection and that it largely reflects the background rate of mutation (Tajima’s D: −0.07205, P > 0.10; Fu and Li’s D’: −1.14188, P > 0.10; Fu and Li’s F’: −0.80538, P > 0.10). The phylogenetic analysis split Terrapene into eastern and western clades. The eastern clade contains T. carolina and T. couhila, and the western clade contains T. ornata and T. nelsoni, which agrees with Minx (1996). However, several relationships within each clade diverge from his phylogenetic hypotheses. Specifically, the relationships within the eastern “trichotomy” (i.e., the T. c.
carolina, T. c. bauri, and the T. c. triunguis – T. c. yucatana – T. c. mexicana clades) and between T. o. ornata and T. o. luteola are not apparent in the molecular phylogenies.

3.1.1.1. T. c. carolina – T. c. triunguis and their associated taxa. The Cytb phylogenetic data suggests that the T. c. triunguis – T. c. yucatana – T. c. mexicana clade is paraphyletic to T. carolina. Rather, there appear to be two monophyletic clades within the larger eastern clade: (1) one including T. c. major, T. c. bauri, T. c. carolina, and T. coahuila and (2) one including T. c. mexicana, T. c. yucatana, and T. c. triunguis.

3.1.1.2. T. c. major – T. coahuila. Terrapene c. major is polyphyletic within T. carolina, as it is found in three different clades: (1) one including T. c. major, T. c. bauri, T. c. carolina, and T. coahuila and (2) one including T. c. mexicana, T. c. yucatana, and T. c. triunguis.

3.1.1.3. Western clade. Within the western clade, Terrapene o. ornata and T. o. luteola form a monophyletic clade. Terrapene o. ornata and T. o. luteola lack the population structure that Herrmann and Rosen (2009) suggested (they found unique groupings of T. o. ornata and T. o. luteola). In our dataset, T. o. ornata and T. o. luteola do not share any haplotypes but do not show any apparent pattern of grouping, suggesting that they are lacking distinction phylogenetically. Terrapene nelsoni and T. ornata are sister taxa within the western clade.

3.1.2. Glyceraldehyde-3-phosphat de dehydrogenase

The nucDNA GAPD phylogram contains 33 haplotypes and 201 sequences after resolving the sequences containing IUPAC ambiguity characters into haploid haplotypes using HAPAR (Fig. 3). Thirty-one of the 430 characters are variable, and 13 characters are parsimoniously informative. For GAPD, Tajima’s D, Fu and Li’s D* and F* tests for neutrality indicate that natural selection is not significantly influencing the rate of mutation (Tajima’s D: 1.30376, P > 0.10; Fu and Li’s D*: -2.35865, P > 0.05; Fu and Li’s F*: -2.37390, P > 0.05).

The split of the eastern (T. carolina) and western (T. ornata) clades is well supported with Bayesian posterior probabilities (BPP) but not with bootstrap resampling. In discordance with the mtDNA data set, one individual morphologically identified as a T. c. carolina is found in the western clade.

3.1.2.1. T. c. carolina – T. c. triunguis. Within the eastern clade, the GAPD tree topology agrees with the mtDNA data in terms of the paraphyly between T. c. triunguis – T. c. mexicana – T. c. yucatana and T. carolina, with the exception of three haplotypes (haplotypes 3, 12, and 29; see Appendix A for more information) containing some T. c. carolina individuals being found within the T. c. triunguis clade and one haplotype containing T. c. triunguis being found in the T. c. carolina clade (haplotype 2; Fig. 3). As with the mtDNA phylogeny, T. c. mexicana, and T. c. yucatana are monophyletic.
within the T. c. triunguis clade, although T. c. major and T. coahuila are also present in this clade for the GAPD gene specifically.

3.1.2.2. Taxa associated with T. c. carolina. The GAPD tree indicates that T. c. bauri is polyphyletic as it is present in both the T. c. carolina and T. c. triunguis clades, which is in discordance with the mtDNA tree. In concordance with the mtDNA tree, T. c. major is polyphyletic within T. carolina for the nucDNA GAPD tree, being distributed into three clades: (1) a clade consisting of one haplotype that shares T. c. major and T. coahuila and several haplotypes of T. c. major; (2) a clade in which T. c. major haplotypes are shared with T. c. carolina; and (3) three haplotypes that are grouped with both T. c. triunguis and T. c. carolina.

3.1.2.3. Western clade. Within the western clade, Terrapene nelsoni is sister to T. ornata, but this relationship is not well supported with bootstrap or BPP values. As with the mtDNA tree, T. o. ornata and T. o. luteola appear to lack genetic distinction. They are not grouped in any apparent pattern for either gene, and for GAPD specifically, T. o. luteola share some haplotypes with T. o. ornata.

3.1.3. Combined mtDNA and nucDNA phylogenetic analysis

Cunningham (1997) suggests P-values for the ILD test to assess whether there is a significant improvement (P < 0.01) or reduction (P > 0.001) in phylogenetic accuracy. The ILD test for our combined mtDNA and nucDNA phylogeny indicates that there is not a significant improvement or reduction in phylogenetic accuracy (P = 0.01). Thus, the combined mtDNA and nucDNA phylogeny is not included here.

3.2. Haplotype networks

The Cytb haplotype network consists of 103 haplotypes and is divided into nine subgroups that do not fall within 95% confidence intervals using the statistical parsimony procedure (Fig. 4). Terrapene o. ornata and T. o. luteola are the only taxa that that fall within a 95% confidence interval and are located within the same clade. As is the case with the Cytb phylogram, T. c. major is polyphyletic, being distributed among three clades and sharing haplotypes with T. c. carolina in some cases. Terrapene coahuila is closely associated with T. c. major and is separated from T. c. major by nine missing intermediate steps, suggesting the need for additional sampling for T. coahuila.

The GAPD haplotype network consists of 33 unique haplotypes after removing heterozygous characters and collapsing identical sequences using HAPAR (Fig. 5). The network is split into two subgroups, with one containing T. nelsoni and the other containing the rest of Terrapene. Although several taxa within the GAPD network are difficult to interpret because of polyphyly, five main clades are evident: (1) T. c. carolina, (2) T. c. triunguis, (3) T. c. major – T. coahuila, (4) T. ornata, and (5) T. nelsoni. However it should be noted that some of the aforementioned five groups are polyphyletic for a few individuals. For example, individuals from T. c. carolina are present in haplotypes containing T. o. ornata or T. c. triunguis, and vice versa. Also, T. c. major is present in haplotypes containing T. c. triunguis and T. c. carolina. Some of the aforementioned haplotypes consist of several taxa (e.g., haplotypes 2, 12, and 29; Fig. 5; Appendix A). Finally, T. c. yucatana is present in a clade consisting of T. coahuila and T. c. major, and T. c. bauri is polyphyletic, with one haplotype being found in the T. c. triunguis clade and the other in the T. c. carolina clade.

3.3. Population structure

The AMOVAs and SAMOVAs contained 59 haplotypes for Cytb because they were not subjected to HAPAR analysis, while Cytb contained the same 103 haplotypes. A total of seven AMOVA hypotheses and one SAMOVA analysis were conducted for each
Analyses of molecular variance (AMOVAs) run a priori and a posteriori to the cytochrome b (Cytb) mtDNA and glyceraldehyde-3-phosphate dehydrogenase (GAPD) nucDNA phylogenetic analyses from this study (Figs. 2 and 3). The a priori AMOVAs are defined as follows: The “All subspecies” AMOVA consists of all ten species and subspecies grouped separately; the “Four species” AMOVA is apportioned based on the morphological phylogeny of Måns (1996); and the “East/west clades” AMOVA divides Terrapene into an eastern clade and western clade. The a posteriori AMOVAs are defined as follows: The “T. o. ornata/T. o. luteola combined” AMOVA contains all taxa apportioned separately except for T. o. ornata and T. o. luteola; the “mtDNA and nucDNA phylogenies” AMOVA is based on the Cytb and GAPD molecular phylogenies from this study (Figs. 2 and 3); the “T. coahuila/T. carolina combined” AMOVA follows the “mtDNA and nucDNA phylogenies” AMOVA with the exception of placing T. coahuila into T. carolina; and the “T. c. bauri into unique group” AMOVA follows the “mtDNA and nucDNA phylogenies” with the exception of separating T. c. bauri into its own unique group.

### Table 2

<table>
<thead>
<tr>
<th>Cytb</th>
<th>GAPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>% variation among groups</td>
<td>$\Phi_{CT}$</td>
</tr>
<tr>
<td>East/west clades</td>
<td>62.08</td>
</tr>
<tr>
<td>Four species</td>
<td>63.87</td>
</tr>
<tr>
<td>All subspecies</td>
<td>85.89</td>
</tr>
</tbody>
</table>

### Table 3

Cytochrome b SAMOVA with the populations maximally differentiated into 19 groups. Samples were labeled by scientific name followed by the state from which they were collected. Too = T. o. ornata, Tol = T. o. luteola, Tnels = T. nelsoni, Tcoah = T. coahuila, Tcm = T. c. major, Tcc = T. c. carolina, Tcb = T. c. bauri, Tct = T. c. triunguis, Tcy = T. c. yucatanana, and TcMx = T. c. mexicana.

### Table 4

Glyceraldehyde-3-phosphate dehydrogenase SAMOVA with the populations maximally differentiated into seven groups. Samples were labeled by scientific name followed by the state from which they were collected. Too = T. o. ornata, Tol = T. o. luteola, Tnels = T. nelsoni, Tcoah = T. coahuila, Tcm = T. c. major, Tcc = T. c. carolina, Tcb = T. c. bauri, Tct = T. c. triunguis, Tcy = T. c. yucatanana, and TcMx = T. c. mexicana.

3.4. Molecular clock analysis

The Cytb time-calibrated molecular clock analysis indicated an eastern/western divergence time estimate of ~15 mya (Table 5; Fig. 6). The estimated divergence times for carolina/coahuila, carolina/bauri, and carolina/major splits were ~10.6 mya, ~10.3 mya, and ~7.2 mya, respectively, and the divergence estimates for triunguis/yucatanana and triunguis/mexicana were ~9.2 and ~7.0 mya, respectively. Divergence times for T. o. ornata and T. o. luteola could not be calculated because they are so genetically similar and because the haplotypes containing these taxa are intermixed in no apparent pattern within the T. ornata clade.

3.5. Percent sequence divergences

3.5.1. Cytb and GAPD

Pairwise percent sequence divergences for the mtDNA Cytb gene ranged from 0.583% to 7.34%, while percent divergences for the nucDNA GAPD gene ranged from 0.350% to 3.15% (Table 6). For both Cytb and GAPD, T. c. carolina – T. c. trianguis, T. c. carolina – T. o. ornata, T. c. carolina – T. c. bauri, T. o. ornata – T. n. nelsoni, and T. o. ornata – T. c. trianguis showed relatively high sequence divergences. The comparison between T. c. trianguis and T. c. yucatanana was relatively high for Cytb but not for GAPD. For GAPD, the percent divergence between T. c. carolina and T. coahuila was relatively high; this comparison was low for Cytb. The aforementioned pairwise comparisons showed a percentage of sequence divergence greater than or equal to what Feldman and Parham (2002) indicated as typically representing interspecific relationships within Emydine turtles for the mtDNA Cytb gene and what FitzSimmons and Hart (2007) indicated as interspecific for nucDNA in freshwater turtles and tortoises. The remaining groups showed relatively lower nucleotide divergences.

3.5.2. DNA barcoding

Twenty-four DNA sequences were obtained for the COI gene (Appendix A). COI pairwise comparisons indicate that all taxa are
divergent by more than the reference value of 2% (Reid et al., 2011) except for T. c. triunguis – T. c. mexicana, T. o. ornata – T. o. luteola, and the T. c. major clade that did not include T. coahuila and T. c. carolina (Table 7). Furthermore, the majority of the comparisons are at least ~5% divergent, with some between 6% and 7.5%.

4. Discussion

4.1. Phylogenetic analyses

Our molecular phylogenetic data (Figs. 2 and 3) agree with Minx (1996) in several ways. First, the phylogenies support the monophyly of the eastern and western clades. Second, T. c. mexicana, T. c. yucatan, and T. c. triunguis form a monophyletic clade. Third, the sister relationship of T. ornata and T. nelsoni is supported. However, our data bring several currently hypothesized relationships within Terrapene into question, while leaving some taxa unresolved. It should be noted that many of the clades in the GAPD phylogenetic analysis are well-supported by Bayesian posterior probabilities but not by bootstrap resampling. This situation can occur in datasets with low variability (such as our GAPD dataset), and it has been shown that BI can pick up low phylogenetic signal more effectively than bootstrap resampling (Alfaro et al., 2003).

4.1.1. Eastern clade

Several relationships proposed by Minx (1996) are not supported in the molecular phylogenies. First, the T. c. triunguis – T. c. yucatan – T. c. mexicana clade is paraphyletic to the T. c. carolina – T. c. major clade. Second, with respect to the sister relationship between T. carolina and T. coahuila as hypothesized by morphological and previous molecular data (Auffenberg and Milstead, 1965; Feldman and Parham, 2002; Milstead, 1969; Minx, 1992, 1996; Wiens et al., 2010), the molecular phylogenetic analyses show that T. coahuila is associated with T. carolina, but the relationship between the two lacks resolution. Terrapene coahuila and T. c. major also appear to be closely related, but T. c. major is polyphyletic, also being present within other T. carolina clades (see Section 4.3 for possible explanations of this polyphyly). Third, the relationship of T. c. bauri to the rest of T. carolina is unclear, with the Cytb phylogeny supporting (weakly) the monophyly of T. c. bauri with T. carolina but the GAPD phylogeny supporting the polyphyly of T. c. bauri in both the T. c. carolina and T. c. triunguis clades.

4.1.2. Western clade

Terrapene o. ornata and T. o. luteola are monophyletic within T. ornata but are not grouped in any apparent pattern in this clade. Herrmann and Rosen (2009) found population structuring and a unique clade for T. o. luteola in their haplotype network and molecular phylogeny, respectively. However, for the phylogenies presented here these two taxa are intermixed within the T. ornata clade, and, in some cases, T. o. ornata and T. o. luteola share haplotypes, suggesting they lack subspecific resolution.

4.2. Haplotype networks

The Cytb haplotype network (Fig. 4) was divided into nine subgroups that did not fall within 95% confidence intervals with the statistical parsimony method. The only two taxa that fell within a 95% confidence interval were T. o. ornata and T. o. luteola. Accordingly, this mtDNA network indicates a high amount of population structure at the subspecific level, excluding T. o. luteola. The GAPD haplotype network (Fig. 5) is allocated into just two subgroups, T. nelsoni and all other taxa. The GAPD network indicates high population structure for four main clades: Terrapene c. carolina, T. c. triunguis, T. ornata, and T. nelsoni. Terrapene coahuila is closely associated with some of the T. c. major haplotypes. As with the Cytb and GAPD phylogenies, however, T. c. major is polyphyletic, and there are other T. c. major clades in addition to the one associated with T. coahuila, making the T. coahuila - T. c. major relationship convoluted. The presence of these main clades in the GAPD network and the lack of a division between T. o. ornata and T. o. luteola within the T. ornata subgroup in the Cytb network suggests that (1) T. c. carolina and T. c. triunguis form distinct groups, and as such, their current classification status needs to be amended and (2) T. o. ornata and T. o. luteola are very closely related, and they possibly do not show enough population structure to be considered separate subspecies. There is, however, some polyphyley between T. c. carolina and T. c. triunguis and between T. c. carolina and T. ornata in the GAPD phylogeny and haplotype network, and we discuss possible explanations for this polyphyly below.

4.3. Polyphyly within Terrapene

Polyphyley was present in the nucDNA GAPD gene that is not present in the mtDNA Cytb gene. For example, in the GAPD analysis one T. c. carolina from IL is found within T. ornata; three haplotypes containing T. c. carolina from SC, GA, NC, and VA are present in the T. c. triunguis clade; a haplotype containing several T. c. triunguis individuals are present within the T. c. carolina clade; and three T. c. major haplotypes are found within T. c. triunguis. In addition, several T. c. major individuals from various localities are found within T. c. carolina for both Cytb and GAPD. For the IL T. c. carolina found within the T. ornata clade, it is possible that this individual is a hybrid because it was collected at a locality where the two species are sympatric. Based on discordance between the nucDNA and mtDNA phylogenies, the T. c. major individuals found within T. c. carolina and T. c. triunguis are possibly hybrids as their ranges are close together and often are sympatric. Since T. c. major is thought to have originated from a hybridization event between the extinct T. c. putmani (giant box turtle) and T. c. carolina, based on similarities of morphological features (Auffenberg, 1958, 1959; Dodd, 2001), it is possible that T. c. major represents to some extent the ancestral genetic lineage of the extinct T. c. putmani.

The presence of some polyphyletic individuals between the T. c. triunguis and T. c. carolina clades is more difficult to explain. Specifically, two haplotypes containing T. c. carolina from GA, SC, VA, and NC are present within the T. c. triunguis clade for the haplotype network (haplotypes 12 and 29; Fig. 5), plus one additional haplotype for the phylogram (haplotype 3; Fig. 3). One T. c. triunguis haplotype is also present in the T. c. carolina clade (haplotype 2). The absence of haplotype 3 in the haplotype network makes sense because there was a reticulation loop in the network which we resolved using coalescent theory (Crandall, 1994). As for the other polyphyletic haplotypes, we suggest two possibilities. First, haplotype 12 may represent an ancestral haplotype because it connects T. c. triunguis and T. c. carolina on the haplotype network (Fig. 5). Second, it is possible that incomplete lineage sorting has occurred for these polyphyletic and discordant individuals (Avise et al., 1983; Maddison, 1997; Neigel and Avise, 1986; Rosenberg, 1987; Alfaro et al., 2003).
Incomplete lineage sorting occurs in cases of short internodes coinciding with large effective population sizes. In other words, a speciation event is followed by a large demographic expansion, resulting in ancestral lineages being found within certain clades in a gene tree. Incomplete lineage sorting can occur even in populations that are allopatric (Avise et al., 1983; Neigel and Avise, 1986), and could explain at least some of the polyphyly found within \textit{GAPD}. The individuals from NC and VA in haplotype 12 represent most of the \textit{T. c. carolina} individuals in the haplotype connecting \textit{T. c. carolina} and \textit{T. c. triunguis} on the haplotype network (Fig. 5), with the only other \textit{T. c. carolina} individual in haplotype 12 being from GA. Thus, it is possible that the NC and VA individuals represent an ancestral lineage and/or that incomplete lineage sorting has occurred for these individuals. The polyphyletic \textit{T. c. yucatana} present in the \textit{GAPD} haplotype network could also be a result of incomplete lineage sorting, as Milstead’s (1969) hypothesized that \textit{T. c. yucatana} arose from a hybridization event between \textit{T. c. putnami} and \textit{T. c. triunguis}.

Table 6
Percent divergences calculated using Jukes Cantor nucleotide divergences corrected for comparing populations.

<table>
<thead>
<tr>
<th></th>
<th>Cytochrome b (Cyts)</th>
<th>GAPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{carolina} – \textit{bauri}</td>
<td>6.10</td>
<td>1.11</td>
</tr>
<tr>
<td>\textit{carolina} – \textit{coahuila}</td>
<td>0.583</td>
<td>1.11</td>
</tr>
<tr>
<td>\textit{carolina} – \textit{triunguis}</td>
<td>6.71</td>
<td>0.983</td>
</tr>
<tr>
<td>\textit{triunguis} – \textit{mexicana}</td>
<td>2.56</td>
<td>0.350</td>
</tr>
<tr>
<td>\textit{triunguis} – \textit{yucatana}</td>
<td>6.59</td>
<td>0.642</td>
</tr>
<tr>
<td>\textit{ornata} – \textit{carolina}</td>
<td>6.83</td>
<td>1.39</td>
</tr>
<tr>
<td>\textit{ornata} – \textit{nelsoni}</td>
<td>6.08</td>
<td>3.15</td>
</tr>
<tr>
<td>\textit{ornata} – \textit{luteola}</td>
<td>0.583</td>
<td>0.408</td>
</tr>
<tr>
<td>\textit{ornata} – \textit{triunguis}</td>
<td>7.34</td>
<td>1.20</td>
</tr>
</tbody>
</table>

2002).
It is possible that some of the polyphyly is a result of hybrids being present in an intergradation zone (Dodd, 2001). Butler et al. (2011) similarly identified several individuals from GA that molecularly resembled T. c. carolina but morphologically resembled T. c. triunguis, giving an indication of hybridization. The sampling localities of their GA samples are also near an intergradation zone of T. c. carolina, T. c. major, T. c. triunguis, and T. c. bauri in the Gulf Coast region (Fig. 1). Our SC samples are from the far southwest region of the state, and because of the close proximity of southwest SC and GA, it is possible that this GA hybrid T. c. triunguis – T. c. carolina population has spread to SC. This intergradation zone may also at least in part be responsible for the T. c. major and T. c. bauri polyphyly seen in our dataset. Lastly, the T. c. triunguis individuals in the T. c. carolina clade (haplotype 2; Figs. 3 and 5) consist of the same or nearly the same haplotypes 29 within the T. c. triunguis clade, and the T. c. bauri haplotype in the T. c. carolina clade also contains only one individual that is also present in the larger T. c. bauri haplotype within the T. c. triunguis clade. The HAPAR analysis splits these individuals into two separate T. c. triunguis and T. c. bauri haplotypes, respectively, depending on the resolution of the heterozygous characters. This could have resulted from a low number of parsimoniously informative characters in the GAPD dataset. Despite polyphyly within some clades and some discordance between the mtDNA and nucDNA phylogenies, both phylogenies and haplotype networks are well resolved enough to interpret relationships between T. c. carolina – T. c. triunguis, T. ornata – T. nelsoni, and T. c. triunguis – T. c. mexicana – T. c. yucatana.

4.4. Assessing the population structure of Terrapene

4.4.1. Population structure based on AMOVA analyses

The AMOVAs (Table 3) used to assess the separation or clumping of T. o. ornata and T. o. luteola indicate a low amount of population structure for these taxa. This suggests that T. o. luteola and T. o. ornata be clumped into one group as T. ornata. For both Cytb and GAPD, the lowest amount of between-group population structure is found when assigning just two a priori groups consisting of the eastern and western clades, supporting that the east/west classification is not adequate for describing the population structure within Terrapene.

4.4.1.1. T. carolina population structure. The a priori AMOVA based on the presented mtDNA and nucDNA phylogenies indicates a higher level of population structure than that of the morphological data of Minx (1996). Specifically, a higher amount of population structure is seen when assigning the T. c. triunguis – T. c. mexicana – T. c. yucatana clade into a unique group. This suggests that the current classification needs to be revised by elevating the triunguis – mexicana – yucatana clade to a separate species from T. carolina.

4.4.2. Population structure based on SAMOVA analyses

The Cytb SAMOVA shows maximal genetic apportionment when split into 19 groups (Table 4). However, the GAPD SAMOVA apportions Terrapene into seven groups (Table 5). For GAPD, several taxa share haplotypes in some cases, which reduces the number of groups. For example, a large haplotype consisting mostly of T. c. triunguis also contains a small number of sequences from T. c. yucatana, T. c. mexicana, T. c. major, and one T. c. carolina. This is the only haplotype in which T. c. mexicana is found and is one of only two haplotypes in which T. c. yucatana is found. The Cytb SAMOVA groups T. coahuila into its own unique group, but the GAPD SAMOVA combines T. coahuila with T. c. major from FL. The GAPD SAMOVA places T. c. major, T. c. bauri, T. c. triunguis – T. c. mexicana – T. c. yucatana, T. c. carolina, and T. nelsoni, into unique groups, while T. o. luteola and T. o. ornata are allocated into the same group for both Cytb and GAPD.

The Cytb gene shows much more variation than GAPD. The increased number of unique groups seen for Cytb suggests that there is a barrier to gene flow within taxa. While this barrier may not have been present long enough to cause deep divergences for infraspecific relationships with nucDNA, it is affecting faster evolving genes.

These findings further support (1) separating T. c. triunguis – T. c. mexicana – T. c. yucatana from the rest of T. carolina and classifying them as a separate species, (2) that T. c. bauri may constitute its own species, (3) that T. coahuila is closely associated with some, but not all, T. c. major haplotypes, suggesting multiple T. c. major lineages, and (4) that T. o. ornata and T. o. luteola should be clumped together as T. ornata (without subspecific designations) due to the apparent lack of population structure for these taxa.

4.5. Molecular clock – possible explanations for the speciation of Terrapene

Several historical geological and climatic events may explain the chronogram and estimated divergence times (Table 5; Fig. 6). First, the divergence time for the split between the eastern and western groups occurred ~15 mya. At this time in the middle–late Miocene (Barstovian Age), the climate in central North America (e.g., Kansas and Nebraska) was becoming warmer, and mesic areas were becoming interspersed with grasslands (for a review,
see Axelrod, 1985; Berry, 1918; Chaney and Elias, 1936; Hesse, 1936; Wolfe, 1985). In addition, the earliest known fossil box turtles were found in Barstovian deposits from ~15 mya in Nebraska (Holman, 1987; Holman and Corner, 1985). Because T. o. ornata are typically a more grassland-oriented species and T. carolina typically inhabit mesic woodlands (Dodd, 2001), it makes sense that the divergence between the eastern and western clades would have occurred ~15 mya in the Barstovian Age where savannah-like grasslands were becoming more abundant. The earliest fossils resembling T. ornata were also dated to ~14.5 mya and were found in Barstovian deposits (Holman and Fritz, 2005), which further supports the estimated divergence dates given in Table 5 and Fig. 6.

4.5.1. T. carolina

In comparison, T. carolina tend to inhabit mesic woodlands (Dodd, 2001). Progressing towards the late Miocene (~10 mya), much of eastern North America consisted of deciduous forests that were gradually being separated by emerging grassland in the southeast (Graham, 1965; Webb, 1983; Woodburne, 2004), and the northeast was predominantly deciduous forest (Graham, 1993; Janis et al., 1998; Mai, 1995; for a review, see Manchester, 1999; Tiffney, 1985a, 1985b; Wolfe, 1975). This coincides with the divergence of the T. carolina group because separation from the western group as a result of diverging habitat requirements may have resulted in speciation once box turtles began to migrate eastward. The divergence of the ancestral T. c. carolina lineage ~7 mya in very late Miocene or early Pliocene climatic conditions makes sense geologically due to the woodland habitat seen in the in the northeastern part of the United States and the temperate climate at the time (Woodburne, 2004). Lastly, during the very late Miocene or early Pliocene, climate change caused an increase in provincialism in the North American biomes (Webb, 1977), which could have contributed to the speciation of Terrapene taxa.

4.5.2. T. c. bauri and T. c. major

While T. c. bauri tend to inhabit mesic woodlands, they also are often found in savannah-like biomes (Dodd, 2001). Because Florida developed more mesic habitats over time, it makes sense that T. c. bauri, which could have originated in savannah and salt marsh biomes that were present ~10 mya in northern Florida and the Gulf Coast (Webb, 1977; Woodburne, 2004), would be more adaptive with their habitat preferences. In the middle to late Miocene (~12.5 mya ~5 mya), sea levels were generally receding and sediment from the Appalachian Mountains filled the channel separating central and lower Florida from the Gulf Coast (Randazzo and Jones, 1997). This land connection could have allowed the ancestor to T. c. bauri to migrate southward into useable habitats that were present in southern Florida during the mid-late Miocene and early Pliocene (Wolfe, 1985). It is also possible that after box turtle migrations to peninsular Florida, vicariance events due to the rising and receding of sea levels resulted in the separation and subsequent speciation of an isolated ancestral population of T. c. bauri. Lastly, the MRCA to T. c. major and T. c. carolina diverged ~7 mya, and this relatively recent divergence makes sense due to some non-monophyly being found between the two taxa.

4.5.3. T. c. triunguis, T. c. mexicana, T. c. yucatana

The T. c. yucatana lineage was estimated to have diverged from the T. c. triunguis lineage earlier than T. c. mexicana, which is supported by Milstead’s (1969) explanation of the origin of T. c. mexicana. He postulated that T. c. mexicana originated through an intergrade between T. c. yucatana and T. c. triunguis. It is possible that the MRCA of the T. c. yucatana and T. c. mexicana lineages dispersed southward via coastal drainages along the eastern coast of Mexico that were not affected by the increasing aridity coming from the west (Rosen, 1978). It is also possible that the MRCA to T. c. yucatana was genetically isolated after migrating to the Yucatan Peninsula via the Isthmus of Tehuantepec land bridge in the Pleistocene and Pleistocene from ~8 mya until ~2.5 mya, when fluctuations in sea levels may have caused their isolation (Beard et al., 1982; Bryant et al., 1991; Mulcahy and Mendelson, 2000; Mulcahy et al., 2006). Rosen (1978) also speculated that volcanic activity in the Pliocene could have isolated the Yucatan Peninsula, resulting in the divergence of T. c. yucatana from the ancestral lineage.

4.5.4. Comparisons with published divergence estimates

Divergence estimates for Terrapene have been reported in the literature (Near et al., 2005; Spinks and Shaffer, 2009), but these analyses are mostly focused on intergeneric classification and no studies have been published that comprehensively analyzed the majority of the taxa within the genus. Thus, while we realize that the comparisons of our divergence estimates with geographic and climatic events is speculative, this chronogram provides the first divergence time estimates for most taxa within Terrapene and can help in our understanding of the climatic and geographic processes by which these groups diverged. The pairwise percent divergences also shed light on how divergent the taxa within Terrapene are and provide quantitative evidence of taxonomic classifications that are in need of revisions.

4.6. Percent divergences

4.6.1. Cytb and GAPD

Percent divergences were at interspecific levels between T. c. carolina – T. c. triunguis, T. c. carolina – T. c. bauri, T. c. carolina – T. o. ornata, T. o. ornata – T. c. triunguis, and T. o. ornata – T. nelsoni for both mtDNA and nucDNA (Table 6). For only Cytb, percent divergences were at interspecific levels for T. c. triunguis – T. c. yucatana, and for only GAPD T. c. carolina – T. coahuila were at interspecific levels. Each of these pairwise comparisons was equivalent to or greater than what Feldman and Parham (2002) and FitzSimmons and Hart (2007) considered as representing separate species for the mtDNA Cytb gene in Emydid turtles and nucDNA in freshwater turtles, respectively. In fact, many of these comparisons are in the range for what are typically considered inter-family divergence levels.

4.6.2. DNA barcoding

The DNA barcoding results (Table 7) indicate that all taxa represent unique species except for T. c. triunguis – T. c. mexicana, T. o. ornata – T. o. luteola, and the T. c. major clade that did not include T. coahuila and T. c. carolina. The mutation rate of the COI gene is relatively high, and as such it is useful for delineating congeners and conspecific groups (Cox and Hebert, 2001; Wares and Cunningham, 2001). DNA barcoding has also been shown to be accurate in delineating species from one another (Hebert et al., 2003), which is useful for our dataset because some of the congenic relationships within Terrapene are in question. Our barcoding data agree with all of the other analyses, and in light of this concordance and with the quantitative evidence that the DNA barcoding analysis provides, it is evident that taxonomic revisions are needed within Terrapene.

5. Conclusions

5.1. Recommendations for classification revisions

Phylogenetic analyses, AMOVAs, SAMOVAs, haplotype networks, pairwise percent divergences, and barcoding data all suggest that classification revisions are in order for Terrapene. With
respect to the T. carolina group, all analyses support the splitting of T. c. carolina and T. c. triunguis into separate species. We recommend that Terrapene c. triunguis be elevated to full species status as Terrapene mexicana triunguis. Terrapene c. mexicana and T. c. yucatanensis should be placed within T. mexicana as T. m. mexicana and T. m. yucatanensis. The species should be named T. mexicana because mexicana was the earliest to be described (Gray, 1849). Terrapene c. bauri represents its own unique group from T. c. carolina according to the AMOVA and SAMOVA analyses, percent divergences, and DNA barcoding data, but in the Cytb phylogeny the clade is not well-supported with bootstrap resampling or BPP and the clade is polyphyletic for the GAPD phylogeny. Thus, the phylogenetic data are inconclusive for T. c. bauri. Data are also inconclusive for T. c. major because of polyphyly. Due to the historical affinity of T. c. major and T. c. bauri with T. c. carolina, these taxa should remain in T. carolina pending further analyses.

It is certainly possible that the polyphyly for T. c. major and incongruences between the Cytb and GAPD phylogenies are due to introgression and hybridization. It is also possible that the thought to be extinct giant box turtle (Terrapene carolina putnami; Hay, 1906) lineage still exists and is represented in one of the T. c. major clades, but we were unable to obtain DNA from fossil remains of T. c. putnami to verify this. Terrapene coahuila is closely associated with one of the T. c. major clades but is not well-resolved in the phylogenies. The retention of T. coahuila as the sister species to T. carolina is supported by the Cytb haplotype network, the GAPD percent divergences, the Cytb SAMOVA, both the Cytb and GAPD AMOVAs, and the barcoding data. However, due to the lack of resolution for T. coahuila in the phylogenies and the convoluted association with T. c. major, we recommend that T. coahuila maintain its current specific status and remain the sister clade to T. carolina until additional data are available.

Terrapene o. ornata and T. o. luteola are very closely related and do not appear to be divergent enough to be considered separate subspecies. We recommend that these taxa should be clumped together as T. ornata, which disagrees with Herrmann and Rosen (2009), who found population structuring between T. o. ornata and T. o. luteola. Terrapene nelsoni should remain the sister species to T. ornata.

It is important to note that these data are based on just two mtDNA genes (Cytb and COI) and one nucDNA gene, and data for additional genes would be beneficial. Furthermore, sequencing additional genes could be helpful with resolving some of the more ambiguous clades, such as T. c. bauri, T. c. major, and T. coahuila. It should be noted that while Bayesian inference is more sensitive to low variability than is bootstrap resampling, this is both a blessing and a curse because it is possible that the posterior probabilities for GAPD could have been inflated due to short internodes in the phylogeny (Alfaro et al., 2003). However, if the genetic barcoding data are any indication, all taxa represent different species except for T. o. ornata – T. o. luteola and T. c. triunguis – T. c. mexicana. The barcoding data agree with most of our other analyses in this regard. This dataset provides an important example that multiple types of analyses and multiple molecular markers should be used before implementing any taxonomic revisions. In the case of discordance between analyses, DNA barcoding percent divergences suggested by the literature to distinguish between species for COI should, perhaps, be used to reevaluate the data from other markers.

In conclusion, we recommend splitting Terrapene into five species (1) T. carolina, (2) T. mexicana, (3) T. coahuila, (4) T. ornata, and (5) T. nelsoni. Terrapene carolina should contain three subspecies (T. c. carolina, T. c. bauri, and T. c. major). Terrapene mexicana should contain three subspecies (T. m. mexicana, T. m. triunguis, and T. m. yucatanensis). While T. m. yucatanensis appeared to be a unique species based on the Cytb percent divergences and the barcoding data, there was a limited sample size for this taxon and the nuclear data disagreed; thus, we recommend that yucatanensis remain with T. mexicana. Lastly, T. coahuila and T. ornata should be monotypic. Since we did not obtain any samples for T. n. klauberi evaluations could not be made between the subspecies of T. nelsoni.

5.2. Conservation implications

It is essential to have an understanding of the evolutionary history of a group in need of conservation management, and these data have shed light on some of the evolutionary relationships within Terrapene. Because conservation efforts are typically species-based and tend to ignore subspecies, the splitting of T. carolina (with 6 subspecies) into T. carolina and T. mexicana (with three subspecies each) will be important for future conservation management strategies. In addition, the Cytb SAMOVA analysis indicated restricted gene flow for some intraspecific populations. Having an understanding of the underlying genetics of intraspecific populations is very important for successfully facilitating conservation management strategies. As such, further population-level analyses are warranted to assess the genetic “health” of individual populations. Parameters including effective population sizes, inbreeding coefficients, and levels of heterozygosity can shed light on the conservation status of Terrapene, and while some work in this regard has already been done (e.g., Buchman et al., 2009; Cureton et al., 2009; Howeth et al., 2008; Kuo and Janzen, 2004), a wider range of populations and geographic localities need to be sampled. Most previous work has focused on limited geographic ranges, and it will be useful to have an understanding of the conservation status of each taxon throughout their range. Thus, while these data provide support for revising the Terrapene classification scheme, there is still much work to be done in terms of finer-scale population genetics.

Future research for Terrapene should focus on two areas. First, additional mtDNA and nucDNA genes need to be sequenced to improve sample sizes. This may help to resolve some of the poorly supported clades and to provide further support for those that are supported. Second, finer-scale population genetic analyses should be performed to better assess the conservation status of Terrapene and to provide insight into their underlying genetics so that successful conservation management strategies can be employed. It is our hope that the data within as well as additional data from future work will aid not only in our general understanding of the evolutionary history of Terrapene, but also with their conservation.

Acknowledgments

We thank the University of Texas at Tyler for laboratory space and facilities. We also thank M. Bangs for his advice on analyzing the phylogenetic data and for his comments and critique of this manuscript. The University of Texas at Tyler Office of Sponsored Research, Chelonian Research Foundation, Louis Stokes Alliance for Minority Participation (LSAMP), Department of Natural and Applied Sciences and Faculty Development Program at Mount Mercy University, Natural Heritage & Endangered Species Program of the Massachusetts Division of Fisheries and Wildlife, East Texas Herpetological Society, and National Box Turtle Conservation Committee provided funding for this research. Tissue samples were collected under a University of Texas at Tyler IACUC protocol (No. 2009-02) issued to JSP.

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.2013.03.006.


