A molecular phylogeny of tortoises (Testudinidae) based on mitochondrial and nuclear genes

Minh Le a,b,*, Christopher J. Raxworthy a, William P. McCord c, Lisa Mertz a

a Department of Herpetology, Division of Vertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA
b Department of Ecology, Evolution, and Environmental Biology (E3B), Columbia University, 2960 Broadway, New York, NY 10027, USA
c East Fishkill Animal Hospital, 455 Route 82, Hopewell Junction, New York 12533, USA

Received 9 October 2005; revised 1 March 2006; accepted 2 March 2006
Available online 5 May 2006

Abstract

Although tortoises of the family Testudinidae represent a familiar and widely distributed group of turtles, their phylogenetic relationships have remained contentious. In this study, we included 32 testudinid species (all genera and subgenera, and all species of Geochelone, representing 65% of the total familial species diversity), and both mitochondrial (12S rRNA, 16S rRNA, and cyt b) and nuclear (Cmos and Rag2) DNA data with a total of 3387 aligned characters. Using diverse phylogenetic methods (Maximum Parsimony, Maximum Likelihood, and Bayesian Analysis) congruent support is found for a well-resolved phylogeny. The most basal testudinid lineage includes a novel sister relationship between Asian Manouria and North American Gopherus. In addition, this phylogeny supports two other major testudinid clades: Indotestudo + Malacochersus + Testudo; and a diverse clade including Pyxis, Aldabrachelys, Homopus, Chersina, Psammobates, Kinixys, and Geochelone. However, we find Geochelone rampanty polyphyletic, with species distributed in at least four independent clades. Biogeographic analysis based on this phylogeny is consistent with an Asian origin for the family (as supported by the fossil record), but rejects the long-standing hypothesis of South American tortoises originating in North America. By contrast, and of special significance, our results support Africa as the ancestral continental area for all testudinids except Manouria and Gopherus. Based on our systematic findings, we also propose modifications concerning Testudinidae taxonomy.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Testudinidae; Tortoises; Systematics; Taxonomy; Biogeography; 12S; 16S; cyt b; Cmos; Rag2

1. Introduction

The turtle family Testudinidae includes 49 living species of tortoises in 12 extant genera, all of which are completely terrestrial, and which together represent about 18% of the extant world turtle diversity (Ernst and Barbour, 1989; Uetz, 2005). Testudinids are the most broadly distributed non-marine turtle family, occurring on all subpolar continents except Australia, and occupying a wide diversity of habitats varying from rain forests in Southeast Asia and South America, to deserts in North America and Africa. Morphologically, tortoises also show a great diversity of sizes and forms. The largest tortoises, in the Galápagos (Geochelone nigra), can reach 1.5m in length, yet the speckled padloper (Homopus signatus) is mature at only 10cm in length. Most genera have highly domed and rigid carapaces, but hinged-back carapaces occur in the genus Kinixys, and the genus Malacochersus has a remarkable flattened and flexible carapace for occupying rock crevices (Crumly, 1984a; Ernst and Barbour, 1989).

Despite previous research, phylogenetic relationships within the family Testudinidae have remained controversial (Caccone et al., 1999a; Crumly, 1982, 1984a; Gerlach, 2001, 2004; Meylan and Sterrer, 2000; Parham et al., 2006). One of the perceived problems, voiced by several authors, concerns the high level of morphological convergences suspected for some traditionally used
characters (Auffenberg, 1974; Bramble, 1971; Pritchard, 1994). In addition, all testudinid molecular studies to date have been limited in terms of taxonomic sampling because they addressed specific questions of relationship within smaller subsets of the family (with the possible exception of Cunningham’s (2002) study, but this has not yet become available). These studies included relationships within the genus Gopherus (Lamb and Lydeard, 1994), origin and relationships of Malagasy tortoises (Caccone et al., 1999a), origin of Galápagos tortoises (Caccone et al., 1999b), origin of Indian Ocean tortoises (Austin and Arnold, 2001; Austin et al., 2003; Palkovacs et al., 2003), relationships within the genus Indotestudo (Iverson et al., 2001), and relationships within the genus Testudo (Fritz et al., 2005; Parham et al., 2006; van der Kuyl et al., 2002).

Nevertheless, the monophyly of the family has been continuously supported by both molecular and morphological studies; with two groups: Geoemydidae and Emydidae considered sister to testudinids (Crumly, 1984a; Gaufin and Meylan, 1988; Gerlach, 2001; Krenz et al., 2005; Near et al., 2005; Spinks et al., 2004; van der Kuyl et al., 2002).

Within the Testudinidae, establishing the monophyly of the genus Geochelone has proven to be especially problematic. As the largest genus of the family, this group currently includes 10 species, exclusive of five species recently assigned to the genera Manouria and Indotestudo based on the studies of Crumly (1984a,b) and Hooogo and Crumly (1984), and six extant or recently extinct species being placed in the genus Aldabrachelys (Dipsochelys) (Austin et al., 2003; Bour, 1982; Gerlach, 2001). Auffenberg (1974) proposed six subgenera (including Manouria and Indotestudo) for Geochelone based largely on zoogeographic criteria, however, Crumly (1982) was the first to argue against their complete use based on his cladistic analysis using 26 cranial characters. This view has also been corroborated by more recent analyses, which have included more characters and taxa (Crumly, 1984a; Gerlach, 2001). Crumly’s (1984a) results also did not support Geochelone monophyly (Fig. 1). More recent morphological studies have provided conflicting results: Gaufin and Meylan (1988) and Meylan and Sterrer (2000) found support for a monophyletic Geochelone, and Gerlach (2001) found Geochelone to be polyphyletic. In addition, recent molecular studies have shown Malagasy Geochelone paraphyletic with respect to Pyxis (Caccone et al., 1999a; Palkovacs et al., 2002), or provided little resolution between Geochelone and other testudinid genera (van der Kuyl et al., 2002). All of these molecular studies were restricted to mitochondrial genes and only included a subset of Geochelone species.

In terms of their biogeographic history, tortoises offer an interesting group for study because of their almost worldwide distribution, including oceanic islands, e.g., the Galápagos, the East Indies, the West Indies, Madagascar, and the Mascarene Islands. Crumly’s (1984a) study has provided the most detailed discussion of different biogeographic scenarios of testudinids to date. He suggested true tortoises originated in North America, based on the occurrence of the oldest known fossil testudinid, Hadrianus majusculus, from the Early Eocene. Crumly (1984a) also proposed that tortoises from Europe and Africa were more closely related to Asian forms than to North American forms, and that tortoises of the Galápagos, and the Indian Ocean had rafted to these islands. This latter dispersal scenario has been supported by more recent studies (Austin and Arnold, 2001; Caccone et al., 1999b; Palkovacs et al., 2002). However, the origin of the South American tortoises remains a matter of controversy. Most authors have concluded that tortoises invaded tropical South America from North or Central America (Auffenberg, 1971; Gerlach, 2001; Simpson, 1943; Williams, 1950); however, both Simpson (1942) and Crumly (1984a) also suggested Africa as a possible alternative source.

The major objective of our study was to produce a phylogeny of the Testudinidae based on an expanded molecular character dataset, which would be comprehensive at the generic level, and include all the species available to us (the majority of species). We ultimately were able to include species representing all recognized and previously proposed extant genera and subgenera, representing in total 65% of all testudinid species. This broad taxonomic sampling was considered especially important for resolving phylogenetic relationships within the problematic genus Geochelone, and for this group, we were able to include all Geochelone species. To provide a robust estimate of phylogeny, we included a broad diversity of genes, including one mitochondrial protein coding gene (cytb), two mitochondrial ribosomal genes (12S and 16S), and two nuclear genes (Cmos and Rag2), the latter of which have not been previously utilized for resolving relationships between turtles.

2. Materials and methods

2.1. Taxonomic sampling

For our testudinid ingroup, we included 34 taxa (32 species and 2 subspecies) and all genera within the family. We also included all Geochelone species. For two species: Indotestudo forstenii and Gopherus agassizii, for which we were unable to obtain tissue; sequences for these species (reported by Spinks et al., 2004) were obtained from GenBank. A complete list of all tissues, voucher specimens, and sequences are provided in Table 1. This molecular phylogenetic analysis includes the most comprehensive taxon sampling achieved to date for testudinids.

For outgroups, we included two species of Geoemydidae (Rhinoellemmys melanosterna and R. nasuta) and two species of Emydidae (Glyptemys insculpta and Deirochelys reticularia) based on their reported sister relationships to testudinids recovered by both morphological and molecular evidence (Gaufin and Meylan, 1988; Honda et al., 2002; Krenz et al., 2005; Near et al., 2005; Spinks et al., 2004).
2.2. Molecular data

Both mitochondrial and nuclear DNA were utilized in this study to resolve relationships at all phylogenetic levels within the family. Similar approaches have proved successful with other groups exhibiting similar levels of phylogenetic diversity (see Barker et al., 2004; Georges et al., 1999; Hillis et al., 1996; Pereira et al., 2002; Saint et al., 1998). We sequenced three regions of the mitochondrial DNA (mtDNA): the complete cytochrome $b$ sequence, and sections of both the 12S and 16S rRNA genes. For some taxa (see Table 1), we downloaded cyt$\text{b}$ and 12S sequences from GenBank originating from studies by Spinks et al. (2004), Palkovacs et al. (2002), Caccone et al. (1999a,b), and van der Kuyl et al. (2002). We also sequenced two nuclear fragments of the Cmos and Rag2 genes. All primers used for this study (including those developed during the course of this study) are shown in Table 2.

Fig. 1. Crumly's (1984a) preferred hypothesis of relationships for the family Testudinidae, based on 57 morphological characters. In contrast to all the other genera shown here, the genus Geochelone is paraphyletic.
DNA was extracted from tissues and blood samples using the DNeasy kit (Qiagen) following manufacturer’s instructions for animal tissues. PCR volume for mitochondrial genes contained 42.2 μl of 20 mM dNTP, 4 μl of 25 mM MgCl2, 4 μl of each primer, 0.2 μl of Taq polymerase (Promega), and 4 μl of DNA. PCR conditions for these genes were: 95 °C for 1 min; with 42 cycles at 95 °C for 30 s, 45 °C for 45 s, 72 °C for 60 s; and a final extension of 5 min to activate the Taq DNA polymerase. PCR products were visualized using electrophoresis with all PCR conditions similar to those used for mitochondrial genes. PCR products were excised from the gel using a Pasteur pipette, and the gel plug was melted in 300 μl sterile water, 4 μl of DNA or extraction solution). PCR conditions for nuclear genes were the same as above except the first step (95 °C) was carried out in 15 min, and the annealing temperatures were 52 and 58 °C for Rag2 and Cmos, respectively.

PCR products were visualized using electrophoresis through a 2% low melting-point agarose gel (NuSieve GTG, FMC) stained with ethidium bromide. For reamplification reactions, PCR products were excised from the gel using a Pasteur pipette, and the gel plug was melted in 300 μl sterile water at 73 °C for 10 min. The resulting gel-purified product was used as a template in 42.2 μl reamplification reactions with all PCR conditions similar to those used for mitochondrial genes. PCR products were cleaned using PerfectPrep® PCR Cleanup 96 plate (Eppendorf) and cycle sequenced using ABI prism big-dye terminator according to manufacturer's instructions for animal tissues. PCR volume for mitochondrial genes contained 42.2 μl of 20 mM dNTP, 4 μl of 25 mM MgCl2, 4 μl of each primer, 0.2 μl of Taq polymerase (Promega), and 4 μl of DNA. PCR conditions for these genes were: 95 °C for 1 min; with 42 cycles at 95 °C for 30 s, 45 °C for 45 s, 72 °C for 60 s; and a final extension of 5 min to activate the Taq DNA polymerase. PCR products were visualized using electrophoresis with all PCR conditions similar to those used for mitochondrial genes. PCR products were excised from the gel using a Pasteur pipette, and the gel plug was melted in 300 μl sterile water at 73 °C for 10 min. The resulting gel-purified product was used as a template in 42.2 μl reamplification reactions with all PCR conditions similar to those used for mitochondrial genes. PCR products were cleaned using PerfectPrep® PCR Cleanup 96 plate (Eppendorf) and cycle sequenced using ABI prism big-dye terminator according to manufacturer's instructions for animal tissues. 

### Table 1

<table>
<thead>
<tr>
<th>Species names</th>
<th>GenBank No. (12S)</th>
<th>GenBank No. (16S)</th>
<th>GenBank No. (cyt/b)</th>
<th>GenBank No. (Cmos)</th>
<th>GenBank No. (Rag2)</th>
<th>Voucher numbers for this study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chersina angulata</td>
<td>DQ497248</td>
<td>DQ497269</td>
<td>DQ497292</td>
<td>DQ497325</td>
<td>DQ497361</td>
<td>AMCC157819</td>
</tr>
<tr>
<td>Aldabrachelys arnoldii</td>
<td>AY081779</td>
<td>AY081780</td>
<td>DQ497293</td>
<td>DQ497326</td>
<td>DQ497362</td>
<td>YU</td>
</tr>
<tr>
<td>Indotestudo forstenii</td>
<td>DQ497251</td>
<td>DQ497271</td>
<td>DQ497301</td>
<td>DQ497334</td>
<td>DQ497370</td>
<td>AMCC157839</td>
</tr>
<tr>
<td>Indotestudo elongata</td>
<td>DQ497256</td>
<td>DQ497276</td>
<td>DQ497306</td>
<td>DQ497339</td>
<td>DQ497375</td>
<td>AMCC125847</td>
</tr>
<tr>
<td>Gopherus agassizii</td>
<td>AF020882</td>
<td>AF020889</td>
<td>DQ497304</td>
<td>DQ497337</td>
<td>DQ497373</td>
<td>AMCC106886</td>
</tr>
<tr>
<td>Gopherus polyphemus</td>
<td>AY081785</td>
<td>AY081786</td>
<td>DQ497305</td>
<td>DQ497338</td>
<td>DQ497374</td>
<td>AMCC110105</td>
</tr>
<tr>
<td>Homopus signatus</td>
<td>AF020883</td>
<td>AF020890</td>
<td>DQ497307</td>
<td>DQ497339</td>
<td>DQ497375</td>
<td>AMCC125847</td>
</tr>
<tr>
<td>Kinixys belliana</td>
<td>DQ497257</td>
<td>DQ497277</td>
<td>DQ497310</td>
<td>DQ497343</td>
<td>DQ497376</td>
<td>AMCC125847</td>
</tr>
<tr>
<td>Kinixys homeana</td>
<td>DQ497258</td>
<td>DQ497278</td>
<td>DQ497311</td>
<td>DQ497344</td>
<td>DQ497377</td>
<td>AMCC110105</td>
</tr>
<tr>
<td>Malacochersus tornieri</td>
<td>DQ497259</td>
<td>DQ497279</td>
<td>DQ497312</td>
<td>DQ497345</td>
<td>DQ497378</td>
<td>AMCC125859</td>
</tr>
<tr>
<td>Manouria emys emys</td>
<td>DQ497260</td>
<td>DQ497280</td>
<td>DQ497313</td>
<td>DQ497346</td>
<td>DQ497379</td>
<td>AMCC125859</td>
</tr>
<tr>
<td>Manouria emys phayrei</td>
<td>DQ497261</td>
<td>DQ497281</td>
<td>DQ497314</td>
<td>DQ497347</td>
<td>DQ497380</td>
<td>AMCC125859</td>
</tr>
<tr>
<td>Psammobates tentorius</td>
<td>DQ497262</td>
<td>DQ497282</td>
<td>DQ497315</td>
<td>DQ497348</td>
<td>DQ497381</td>
<td>AMCC125859</td>
</tr>
<tr>
<td>Pyxis arachnoidea</td>
<td>AF020880</td>
<td>AF020887</td>
<td>DQ497316</td>
<td>DQ497349</td>
<td>DQ497382</td>
<td>AMCC125859</td>
</tr>
<tr>
<td>Pyxis planicauda</td>
<td>AY081787</td>
<td>AY081788</td>
<td>DQ497317</td>
<td>DQ497350</td>
<td>DQ497383</td>
<td>AMCC125859</td>
</tr>
<tr>
<td>Testudo horsfieldii</td>
<td>AF173330</td>
<td>AF173331</td>
<td>DQ497318</td>
<td>DQ497351</td>
<td>DQ497384</td>
<td>AMCC125859</td>
</tr>
<tr>
<td>Testudo kleinmanni</td>
<td>AB090020</td>
<td>AB090021</td>
<td>DQ497319</td>
<td>DQ497352</td>
<td>DQ497385</td>
<td>AMCC125859</td>
</tr>
<tr>
<td>Glyptemys insculpta</td>
<td>DQ497265</td>
<td>DQ497266</td>
<td>DQ497320</td>
<td>DQ497353</td>
<td>DQ497386</td>
<td>AMCC125859</td>
</tr>
<tr>
<td>Rhinoclemmys nasuta</td>
<td>DQ497268</td>
<td>DQ497269</td>
<td>DQ497321</td>
<td>DQ497354</td>
<td>DQ497387</td>
<td>AMCC125859</td>
</tr>
</tbody>
</table>

All sequences generated by this study have accession numbers: DQ497248 – DQ497396; YU, Yale University Collection; FMNH, Field Museum of Natural History; AMCC, Ambrose Monell Cryo Collection, American Museum of Natural History (http://research.amnh.org/amcc); RAN, Ronald A. Nussbaum Field Series, University of Michigan, Museum of Zoology; KU, Kansas University Natural History Museum. 

* Genbank sequences only.
For maximum likelihood analysis the optimal model for nucleotide evolution was determined using Modeltest V3.7 (Posada and Crandall, 1998). Analyses used a randomly selected starting tree, and heuristic searches with simple taxon addition and the TBR branch swapping algorithm. Support for the likelihood hypothesis was evaluated by bootstrap analysis with 100 replications and simple taxon addition. For Bayesian analyses we used the optimal model determined using Modeltest with parameters estimated by MrBayes Version 3.1. Analyses were conducted with a random starting tree and run for $5 \times 10^6$ generations. Four Markov chains, one cold and three heated (utilizing default heating values), were sampled every 1000 generations. Log-likelihood scores of sample points were plotted against generation time to detect stationarity of the Markov chains. Trees generated prior to stationarity were removed from the final analyses using the burn-in function. Two independent analyses were started simultaneously. The posterior probability values (PP) for all clades in the final majority rule consensus tree are reported. We ran analyses on both combined and partitioned datasets to examine the robustness of the tree topology (Brandley et al., 2005; Nylander et al., 2004). In the partitioned analyses, we divided the data into 11 separate partitions, including 12S, 16S, and the other nine based on gene codon positions (first, second, and third) in cyt$b$, Cmos, and Rag2. Optimal models of molecular evolution for each partition were selected using Modeltest and then assigned to these partitions in MrBayes 3.1 using the command APPLYTO. Model parameters were estimated independently for each data partition using the UNLINK command.

To test alternative hypotheses of relationships, corresponding tree topologies were compared using the Wilcoxon signed-ranks test (Felsenstein, 1985b; Templeton, 1983), to determine if tree length difference could have resulted from chance alone (Larson, 1998). Alternative tree topologies were

---

### Table 2

<table>
<thead>
<tr>
<th>Primer</th>
<th>Position</th>
<th>Sequence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1091 (12S)</td>
<td>491</td>
<td>5’-AAAAAGCTTCAAACCTGGATTAGTACCCCCTCTAT-3’</td>
<td>Kocher et al. (1989)</td>
</tr>
<tr>
<td>H1478 (12S)</td>
<td>947</td>
<td>5’-TAGCTGAGGAGGCTGAGGCGGATGTTG-3’</td>
<td>Kocher et al. (1989)</td>
</tr>
<tr>
<td>AR (16S)</td>
<td>1959</td>
<td>5’-CGGCTTATTATAACAAACAT-3’</td>
<td>Palumbi et al. (1991)</td>
</tr>
<tr>
<td>BR (16S)</td>
<td>2561</td>
<td>5’-CCGGCTGAACTGCAGTACCTG-3’</td>
<td>Palumbi et al. (1991)</td>
</tr>
<tr>
<td>CytbG (cytb)</td>
<td>14368</td>
<td>5’-AACCATCTGTGTATCGTACAT-3’</td>
<td>Spinks et al. (2004)</td>
</tr>
<tr>
<td>GLUDGE (cytb)</td>
<td>14358</td>
<td>5’-TGATCCTGAAACACACGCT-3’</td>
<td>Palumbi et al. (1991)</td>
</tr>
<tr>
<td>CytbSi (cytb)</td>
<td>1501</td>
<td>5’-GGATCAAACACCCAGG-3’</td>
<td>Spinks et al. (2004)</td>
</tr>
<tr>
<td>CytbSr</td>
<td>15030</td>
<td>5’-CGCTGTGATGTTGATCCTG-3’</td>
<td>Spinks et al. (2004)</td>
</tr>
<tr>
<td>THR (cytb)</td>
<td>15593</td>
<td>5’-CTACCTGCAATACACGAC-3’</td>
<td>Spinks et al. (2004)</td>
</tr>
<tr>
<td>THR-S (cytb)</td>
<td>15585</td>
<td>5’-GGTTTACAGGCCAAATGCTT-3’</td>
<td>Spinks et al. (2004)</td>
</tr>
<tr>
<td>CM1 (Cmos)</td>
<td>163</td>
<td>5’-GCGCTGTCCATGACTGGA-3’</td>
<td>Barker et al. (2002)</td>
</tr>
<tr>
<td>CM2 (Cmos)</td>
<td>820</td>
<td>5’-GGGTGTGCGGACACAGGTATAGTGC-3’</td>
<td>Barker et al. (2002)</td>
</tr>
<tr>
<td>Cmos1 (Cmos)</td>
<td>163</td>
<td>5’-GCGCTGTCCATGACTGGAATG-3’</td>
<td>This study</td>
</tr>
<tr>
<td>Cmos3 (Cmos)</td>
<td>812</td>
<td>5’-GTAGATGTCTGTTGATGGA-3’</td>
<td>This study</td>
</tr>
<tr>
<td>F2 (Rag2)</td>
<td>601</td>
<td>5’-CAGGATGAGCCATTTCTCTCCATG-3’</td>
<td>This study</td>
</tr>
<tr>
<td>F2-1 (Rag2)</td>
<td>590</td>
<td>5’-TTCCAGAGCTTACAGGATG-3’</td>
<td>This study</td>
</tr>
<tr>
<td>R2-1 (Rag2)</td>
<td>1312</td>
<td>5’-CAGTTGAATAGAAGGACCCAAATG-3’</td>
<td>This study</td>
</tr>
</tbody>
</table>

---

a Modified from F2R2 (Barker et al., 2004).
b Modified from R2R1 (Barker et al., 2004); Cmos and Rag2 primer positions corresponding to the positions in chicken genomes with GenBank numbers of M19412 and M58531, respectively; primer positions for mitochondrial genes corresponding to the positions in the complete mitochondrial genome of Chrysemys picta (Mindell et al., 1999).
Sequences for all five genes were obtained for each of the 32 taxa with tissues. The final matrix consists of 3387 aligned characters: 12S (408 characters), 16S (583 characters), cyt b (1140 characters), Cmos (602 characters), and Rag2 (654 characters). Five taxa have sequences that include gaps at three loci (cyt b, Cmos, and Rag2). For cyt b, Geochelone pardalis babcockii has a single indel that is three base pairs long corresponding to positions 489–494 of the complete chicken Cmos sequence (GenBank Accession No. M19412). For Rag2, Chersina angulata and Kinixys homeana has a single indel of three base pairs long corresponding to positions 1093–1095 of the partial Pelodiscus sinensis Rag2 sequence (GenBank Accession No. AF369089).

The IDL tests found no significant incongruence between mitochondrial genes (12S vs 16S, $p = 0.07$; cyt b vs 12S, $p = 0.89$; cyt b vs 16S, $p = 0.97$), between nuclear genes (Cmos vs Rag2, $p = 0.16$), and between the nuclear and mitochondrial partitions (nuclear DNA vs mtDNA, $p = 0.7$). Using MP, we also analyzed the data by gene and organelle partitions (see Tables 3 and 4). The analysis indicated that Clades 3, 4, and 5 (see Fig. 4) consistently receive strong support from both mitochondrial and nuclear genes (Table 3). For the following partitions, the strict consensus trees include little phylogenetic resolution: 12S, 16S, Cmos, Rag2, and all the nuclear data. For cyt b, although MP recovered only two equally parsimonious trees, the strict consensus is poorly resolved concerning deeper level relationships between clades including multiple genera. The strict consensus tree using all mtDNA was unresolved for...
The topology of the ML cladogram is congruent with the MP topology for all branches except the following: Malacochersus tornieri sister to either Testudo (BP = 68%, MP) or Indotestudo (BP = 60%, ML), G. radiata sister to either Pyxis (BP = 65%, MP) or Geochelone ymiphora (BP = 82%, ML), and the clade including Kinixys + South American Geochelone (G. carbonaria) group sister to either the clade including the Malagasy and Indian Ocean species, Chersina, Homopus, Psammobates, and G. pardalis (<50%, MP) or the clade including G. elegans, G. platynota, and G. sulcata (87%, ML).

In the combined Bayesian analysis (Fig. 3), −ln L scores reached equilibrium after 19,000 generations while the partitioned Bayesian analysis scores reached stationarity after 27,000 generations in both runs. Phylogenetic relationships supported by the combined and partitioned data are identical to the ML results. Only minor differences in the PP values were found between combined and partitioned Bayesian cladograms: PP values are higher in the partitioned analysis for the following clades: Pyxis and Aldabrachelys (99%), South American testudinids (99%), Kinixys + South American testudinids (86%), Testudo (99%), and A. arnoldi + A. hololissa (74%), and lower for the following: Chersina + Homopus + Psammobates + Geochelone pardalis + Aldabrachelys + Pyxis + Geochelone radiata + G. ymiphora (56%), G. chilensis + G. nigra (93%), and Malacochersus + Indotestudo (95%).

All of our analyses, thus, found three well supported clades of tortoises: (1) Gopherus + Manouria, (2)
Indotestudo + Testudo + Malacochersus, and (3) Geochelone + Pyxis + Aldabrachelys + Homopus + Chersina + Psammobates + Kinixys. It is also evident from these analyses that Geochelone, as currently defined, is rampantly polyphyletic. Our phylogenetic results place Geochelone in at least four independent clades. The first consists of three species, G. elegans, G. platynota, and G. sulcata (G. elegans group). The second clade includes all four South American species (G. carbonaria group), which are sister to Kinixys. G. yniphora and G. radiata from Madagascar belong to a third clade that also includes Pyxis and Aldabrachelys. The final clade includes only a single Geochelone species, G. pardalis, which is sister to the genus Psammobates.

Biogeographic results are summarized by the area cladogram shown in Fig. 4, which conservatively depicts the strict congruence between our MP, ML, and Bayesian tree topologies. The optimized ancestral areas, based on DIVA, are also shown in Fig. 4. The DIVA ancestral area for the family Testudinidae is identified as including three areas: Asia, Africa, and North America, and requires a further seven subsequent dispersals (to South America, Madagascar–Indian Ocean, and Asia). For this three-continent familial ancestral area, Gopherus represents the North American lineage, Manouria represents the Asian lineage, and all other testudinids represent the African lineage. For the incongruent clades between the MP and ML analyses (see Figs. 2 and 3), DIVA ancestral areas are optimized as

Fig. 3. The maximum likelihood (ML) analysis phylogram based on all the data combined, using the GTR+G+I model of molecular evolution (Bayesian tree topology is identical). Base frequency A = 0.313, C = 0.262, G = 0.193, T = 0.230. ML. –ln L = 22007.972; rate matrix: A–C = 3.544, A–G = 13.646, A–T = 3.478, C–G = 0.897, C–T = 42.995, G–T = 1.000; proportion of invariable site (I) = 0.470; gamma distribution shape parameter (G) = 0.553. Total number of rearrangements tried = 13077, score of best tree found = 21992.392. Numbers above and below branches are ML bootstrap values (>50%) and Bayesian posterior probabilities, respectively. See Fig. 2 for taxon distribution. *The sister relationship between A. arnoldi and A. hololissa, cannot be seen due to the short branch length.
follows: *Malacochersus + Testudo*, Africa; *Malacochersus + Indotestudo*, Africa and Asia; *Chersina + Homopus + Psammobates + Geochelone pardalis + Kinixys + the Geochelone carbonaria group + Aldabrachelys + Pyxis + Geochelone radiata + G. yniphora*, Africa; and the *G. elegans* group + *Kinixys + the Geochelone carbonaria group*, Africa.

4. Discussion

4.1. Phylogeny

Concerning the three major clades of testudinids we find in this study, the sister group to all other tortoise species includes *Manouria* and *Gopherus* as sister genera (Figs. 2 and 3). This is the first study to sample broadly across the family and recover this intriguing sister relationship; although Lamb and Lydeard’s (1994) molecular study also found weak support for this sister relationship. Other morphological and molecular studies have consistently reported *Manouria* to be basal to all other tortoises (Crumly, 1984a; Gaffney and Meylan, 1988; Gerlach, 2001; Meylan and Sterrer, 2000; Takahashi et al., 2003; Spinks et al., 2004), with *Gopherus* usually placed as the next most basal lineage for the family (Crumly, 1984a; Spinks et al., 2004). *Manouria* and *Gopherus* are distributed in Asia and North America, respectively, and these two genera show obvious differences in morphological characters, with *Manouria* exhibiting several character states considered to be plesiomorphic, such as the wide anterior edge of the prootic bone (ranging from 80 to 92% of the posterior edge) and the ventrally and dorsally split supracaudal scute (Crumly, 1982). This *Manouria + Gopherus* sister relationship has, to the best of our knowledge, not been previously considered or discussed, and no synapomorphies have been proposed for the two genera (Gaffney and Meylan, 1988; Gerlach, 2001; Meylan and Sterrer, 2000; Takahashi et al., 2003). Winokur and Legler (1975) reported mental glands in *Gopherus*, and Ernst and Barbour (1989), also citing Winokur and Legler, report mental glands in *Manouria*—thus suggesting a potentially homologous character. However, our reading of Winokur and Legler’s (1975) study clearly reveals that for testudinids, these authors only reported mental glands for *Gopherus*. In addition, our own examination of AMNH catalogued material (R-118676 and R-119966), and living specimens, also failed to find mental glands in *Manouria*.

The second major clade found by this study includes *Testudo, Indotestudo, and Malacochersus*. A recent molecular study by Parham et al. (2006) based on the complete mitochondrial genome also recovered this monophyletic group. Their MP, ML, and Bayesian analyses produced conflicting results regarding the relationships of *Malacochersus tornieri* and *Testudo horsfieldii*. Specifically, their MP and Bayesian analyses supported the sister relationship...
between *Malacochersus* and *Indotestudo*, but the ML analysis found *Malacochersus* sister to all other taxa. The position of *T. horsfieldii* was unresolved in the MP analysis (strict consensus tree), but was sister to *T. hermanni* in the other analyses. Another study (based on 1124 bp of cyt b) by Fritz et al. (2005) which included almost all forms of the genus *Testudo* plus *Indotestudo* and *Malacochersus* was also ambiguous regarding relationships among these three genera. Similarly, we also found conflicting branch support for relationships within this clade: the MP analysis give moderate support (BP = 68%) for a *Malacochersus + Testudo* clade, while the ML and Bayesian results supported a *Malacochersus + Indotestudo* clade (ML BP = 60%, PP = 97%). Interestingly, morphological studies have consistently supported *Malacochersus* and *Testudo* as sisters (Crumly, 1984a; Gaffney and Meylan, 1988; Gerlach, 2001). According to Gerlach (2001), the two genera share the character of the processus inferior parietalis meeting the quadrate and partially covering the prootic, while Gaffney and Meylan (1988) found they share four characters: presence of supranasal scales, no contact between the inguinal and femoral scutes, a viewable ventral tip of the processus interfenestralis, and the presence of sutures between this process and the surrounding bones.

Our dataset unambiguously resolves relationships within the three *Indotestudo* species. The sister relationship between *I. travancorica* and *I. elongata* is recovered by MP, ML, and Bayesian analyses with strong support. In previous studies, relationships among these species had remained uncertain. The morphological study by Hoogmoed and Crumly (1984) found *I. travancorica* and *I. forstenii* to be identical, yet the molecular study by Iverson et al. (2001) found strong support for the sister relationship between *I. travancorica* and *I. elongata* using MP, but their ML analysis favored *I. travancorica* and *I. forstenii* as sister taxa. The study by Spinks et al. (2004) also supported this latter hypothesis.

For the third and largest testudinid clade, we find support for five internal clades: (1) *Chersina + Homopus + Psammobates + Geochelone pardalis*, (2) the *Geochelone elegans* group, (3) *Kinixys*, (4) the *Geochelone carbonaria* group, and (5) *Aldabrachelys + Pyxys + Geochelone radiata* and *G. yniphora*. However, relationships between these clades conflict between the MP, ML, and Bayesian analyses, and corresponding branch support is also generally lower (MP BP < 50%, ML BP < 50–87%, and Bayesian PP 57–100%).

Of special significance (and as suspected by previous researchers), we find substantial polyphyly for the problematic genus *Geochelone*. Nevertheless, the phylogenetic results we present here for *Geochelone* were not recovered by previous morphological and molecular studies (Gerlach, 2001; Meylan and Sterrer, 2000; van der Kuyl et al., 2002; Palkovacs et al., 2002). The strongly supported sister relationships we find between *G. pardinus* and *Psammobates*, between *Chersina* and *Homopus*; and the monophyly of *Geochelone elegans* group (*G. elegans*, *G. platynota*, and *G. sulcata*) were not evident in previous morphological studies; and the molecular study by Palkovacs et al. (2002) found support for conflicting relationships concerning *G. sulcata* and *G. elegans* between MP, ML, Bayesian, and neighbor-joining analyses.

Although only weakly supported in terms of branch support (MP and ML BP < 50%, PP 80%), all of our analyses support the intriguing sister relationship between African *Kinixys* and the South American *Geochelone carbonaria* group. This hypothesis of relationship is radical: prior morphological and molecular studies placed members of the *Geochelone carbonaria* group either sister to *Gopherus* (Crumly, 1984a; Gerlach, 2001), or sister to other *Geochelone* species: *G. elegans*, *G. pardinus*, and *G. sulcata* (Crumly, 1984a; Meylan and Sterrer, 2000; Palkovacs et al., 2002; van der Kuyl et al., 2002). Crumly (1984a) did not consider the South American tortoises to be monophyletic (Fig. 1), but other researchers have supported monophyly for a sampled subset of this clade (van der Kuyl et al., 2002), or else otherwise explicitly assumed it (Caccione et al., 1999b; Gaffney and Meylan, 1988; Gerlach, 2001; Meylan and Sterrer, 2000).

The Indian Ocean clade supported by our results: *Aldabrachelys + Pyxys + Geochelone radiata + G. yniphora* is consistent with the molecular results of Austin and Arnold (2001) and Palkovacs et al. (2002). The latter study, like ours, found support for a sister relationship between *G. radiata* and *Pyxys* using MP (65% BP), and an alternative sister relationship of *G. radiata* and *G. yniphora* using Bayesian and ML analyses (ML 82% BP, 100% PP). Morphologically, the sister relationship between *G. yniphora* and *G. radiata* has been frequently supported (e.g., Crumly, 1984a; Gaffney and Meylan, 1988; Gerlach, 2001; Meylan and Sterrer, 2000) and Gerlach (2001) reported two synapomorphies: a ventral ridge on the maxilla–premaxilla suture and keels on the supraoccipital crest. In addition, the close relationship between *Pyxys* and *G. yniphora* + *G. radiata* was also supported by Gerlach with the synapomorphy of an indistinct fenestra postotica. By contrast, other studies (e.g., Gaffney and Meylan, 1988; Meylan and Sterrer, 2000; Takahashi et al., 2003) have varied widely concerning the relationship of *Pyxys* to other testudinid genera. Concerning the Indian Ocean giant tortoises (*Aldabrachelys*), for all five genes we found almost no variation among the three forms: *A. dussumieri*, *A. arnoldi*, and *A. hololissa*. Only two sites showed variation: one for 16S (position 2004) and the other for Cmos (position 349). These results support the conclusions of Palkovacs et al. (2003) and Austin et al. (2003) that all three forms may actually represent a single species and population originating from Aldabra. However, these molecular results need to be evaluated by carefully examining the diagnostic morphological characters used to describe these species in order to confirm that these features do not vary intraspecifically.

### 4.2. Intraspecific divergence

The two samples representing different subspecies of *Geochelone pardinus*: *G. p. pardinus* and *G. p. babcockii* have
an uncorrected pairwise sequence divergence distance of 5.9% for cyt b, the fastest evolving gene in this study with the highest proportion of variable sites (Table 4). These two subspecies are also known to exhibit morphological differences. The carapace shape of G. p. babcockii is distinctly convex dorsally, compared to the flatter G. p. pardalis carapace; and juvenile G. p. babcockii has single carapacial scute spots, while G. p. pardalis has two (Loveridge and Williams, 1957). Furthermore, there is a sexually dimorphic size difference: G. p. pardalis males being slightly larger than females and G. p. babcockii males being much smaller (Broadley, 1989). G. p. pardalis occurs in western South Africa and southern Namibia. G. p. babcockii ranges from Sudan and Ethiopia to Natal and from Cape Province to southwest Africa and southern Angola (Iverson, 1992; Loveridge and Williams, 1957). Our preliminary results suggest that a phylogeographic study of G. pardalis is expected to reveal high genetic divergence within this species, or species complex.

By contrast, the cyt b uncorrected pairwise sequence divergence distance between the two samples we have of Manouria e. emys (Burmese tortoise) and M. e. phayrei (Burmese black tortoise) is just 1.3%. Results showing even less geographic variation for mtDNA within a testudinid species have been recently reported for Geochelone radiata (Leuteritz et al., 2005).

4.3. Biogeographic history

The earliest Testudinidae fossils (not yet described) are reported to be from the lower Paleocene of Asia (Claude and Tong, 2004; Hutchison, 1998; Joyce et al., 2004), however older fossils of the putative ancestor (family Lindholmemydidae) for the superfamily Testudinioidea are known from the upper Cretaceous of Asia (Nessov, 1988; Sukhanov, 2000). The appearance of testudinids fossils by the early Eocene in both Europe (Achilemys cassouleti) and North America (Manouria or Manouria-like fossils and Hadrianus majusculus), and by the late Eocene in Africa (Gigantochersina ammon) (Claude and Tong, 2004; Crumly, 1984a; Holroyd and Parham, 2003; Hutchison, 1996, 1998), indicates that tortoises first originated in Asia, they must have soon dispersed to other continents. The abundance of fossil testudinids also suggests that the fossil record is unlikely to be substantially underestimating the origin time for the family. All evidence to date, thus, supports a late Mesozoic or early Cenozoic origin for the Testudinidae in Asia, with dispersal to North America, Europe, and Africa having been achieved by the Eocene. Our DIVA results also establish the ancestral area for testudinids as including three continents: Asia, North America, and Africa, and finds Manouria and Gopherus as the extant groups representing the Asian and North American lineages, respectively. Remarkably, all other extant testudinid groups are associated with the African ancestral area.

The remarkable biographic basal lineage bifurcation between clades distributed in Asia and North America, which we find here for Manouria and Gopherus, has also been reported in the family Dibamidae (blind skinks) and the skink genus Scincella, which are distributed in the Americas and Asia (Honda et al., 2003; Zug et al., 2001). These Asian–American divergences may have resulted from the well-known disruption of floral and faunal exchange across the Bering Strait in the early Eocene; a biogeographic pattern that has been well documented in other groups (Beard, 2002; Bowen et al., 2002; Maekawa et al., 2005; Sanmartin et al., 2001 and references therein; Tiffney, 1985).

The DIVA ancestral area optimization results provide a striking conclusion: it is most parsimonious to consider all testudinids with the exception of Manouria and Gopherus as having an African ancestral area. The alternative biogeographic scenarios, such as considering the ancestral area for this large group as Asia or the Palearctic, will require making additional assumptions of dispersal or extinction. Our results infer the following dispersals from Africa: Indotestudo and the clade including Geochelone platynota + G. elegans dispersing into Asia, Testudo dispersing into Europe, the clade including Aldabrachelys + Pyxis + Geochelone radiata + G. ypihora dispersing into the Indian Ocean, and the G. carbonaria group dispersing into South America. Africa should therefore be viewed as both a major center of origin and diversification for testudinid species. Three clades found by our study are endemic to Africa: Malacochersus, Kinixys, and the diverse clade including Homopus + Chersina + Psammobates + Geochelone pardalis, and these groups presumably diversified on this continent (the Kinixys belliana population of Madagascar is typically considered as introduced, see Raseliminanana and Vences, 2003).

The dispersal of Testudo and Indotestudo from Africa into Europe and Asia appears to be a novel biogeographic scenario that has not been previously considered. For example, Parham et al. (2006) only discuss dispersal events from a Palearctic center of diversification for Testudo, Indotestudo, and Malacochersus. In addition, Crumly’s (1984a) discussion of biogeographic scenarios implied that African tortoises originated from Asia. However, and by contrast, the Indian Ocean clade including Aldabrachelys + Pyxis + Geochelone radiata + G. ypihora has been previously proposed to have an African origin (Caccone et al., 1999a; Palkovacs et al., 2002), as supported by our DIVA ancestral area optimizations (Fig. 4).

However, perhaps the most unusual biogeographic result concerns the South American + Galápagos tortoises (Geochelone carbonaria group), which we find are members of a larger clade otherwise only including species from Africa, the Indian Ocean, and Asia. To test the previously proposed North American origin for the Geochelone carbonaria clade, based upon a sister relationship with Gopherus (e.g., Gerlach, 2001) we searched for the shortest constraint tree that included these clades as sisters. This constraint tree was 48 steps longer than our most parsimonious hypothesis of relationship, and this step length
difference was significant ($p < 0.001$). By contrast, the consistently recovered sister relationship we found between African *Kimixys* and the *Geochelone carbonaria* clade, and the distribution of the group sister to this clade, which also includes Africa, providing strongest support for an African origin for the South American tortoises (Fig. 4). The DIVA ancestral area optimizations find this entire clade (*Geochelone* + *Pyxis* + *Alldabrachelys* + *Homopus* + *Chersina* + *Psammobates* + *Kimixys*) to have an African ancestral area, thus supporting an early evolutionary history for this diverse tortoise group in this continent.

The dispersal of tortoises from Africa to South America might have been aided by westward sea currents, such as the Equatorial Currents (Fratantoni et al., 2000). These flow between the Congo delta and Maranhao in Brazil and are believed to have been in this direction since the breakup of Gondwana (Houle, 1999; Parrish, 1993). Several vertebrate animal groups are reported to show a pattern of dispersal from Africa to South America across the Atlantic: platyrrhine monkeys and caviomorph rodents between 85 and 35 MYA, (Houle, 1999; Huchon and Douzery, 2001; Mouchaty et al., 2001; Nei et al., 2001; Schrago and Russo, 2003), and *Mabuya* skinks twice during the last 9 million years (Carranza and Arnold, 2003). Tortoises are well known for their ability to float (while extending their heads above water) and survive without food or freshwater for up to six months (Caccone et al., 2002), and oceanic dispersal by testudinid tortoises appears to have occurred many times. For example: *Geochelone* to the West Indies (Williams, 1950, 1952), *Geochelone* to the Galápagos and at least five subsequent dispersals between islands in the archipelago (Caccone et al., 1999b, 2002; Russello et al., 2005), ancestral Malagasy tortoises to Madagascar from mainland Africa (Caccone et al., 1999a), *Alldabrachelys* to Aldabra, Astove, Cosmoledo, Denis, Amirantes, Comoros, Assumption, and the Granitic Seychelles (Austin et al., 2003); *Cylindraspis* (now extinct) to Réunion, Mauritius, and Rodrigues (Austin and Arnold, 2001); and *Hesperotestudo* (now extinct) to Bermuda (Meylan and Sterrer, 2000).

### 4.4. Taxonomic issues

Based on our phylogenetic results, and to reflect monophyletic groupings within the Testudinidae, we propose the following taxonomic changes:

1. The name *Geochelone* (Fitzinger, 1835; type species, *G. elegans*) is retained for just three species: *G. sulcata*, *G. elegans*, and *G. platynota*.

2. The South American and Galápagos Island species (*Geochelone carbonaria* group) are placed in the genus *Chelonooides* (Fitzinger, 1835; type species, *C. carbonaria*), which has been previously recognized as a subgenus (Fitzinger, 1856).

3. *Geochelone pardalis* is placed in the genus *Psammobates* (Fitzinger, 1835).

4. Based on the desirability of maintaining the morphologically distinct and well-supported genera *Pyxis* and *Alldabrachelys*, yet also due to the uncertainty concerning the relationships of *Geochelone yniphora* and *G. radiata* to these genera, we therefore prefer to conservatively place both these species in separate monotypic genera. We recommend the use of the available genus name *Astrochelys* (Gray, 1873; type species, *A. radiata*) for *G. radiata*. However, because there is no available generic name for *G. yniphora*, we thus propose the following new genus:

*Angonoka* new genus

Type species, *Testudo yniphora* Vaillant 1885.

Diagnosis—Testudinid tortoises that have the following characters: (1) gulars fused to form a single scute; (2) gular projection thickened and sexually dimorphic in size, projecting well beyond the carapace edge and curved upward in adult males; (3) highly domed carapace with descending sides; (4) carapace brown, lacking radiating yellow or black lines on scutes; (5) no enlarged tubercles on thighs; (6) tail without an enlarged terminal scale; (7) longitudinal ventral ridge on the maxilla–premaxilla suture; (8) horizontal keels on the supraoccpital crest; (9) pectoral scutes narrow and not contacting the entoplastron.

Content. One species, *Angonoka yniphora* (Vaillant 1885).


Etymology. The generic name “*Angonoka*” is based on the Malagasy word ‘Angonoka’ which is the local name for the type species.

### 5. Conclusions

The broad taxonomic sampling of tortoises we were able to achieve with this study, coupled with the inclusion of both mitochondrial and nuclear genes, has resulted in supported tree topologies with both good resolution, and for many branches, high levels of support. Importantly, these topologies are also largely congruent, regardless of the approach taken concerning phylogenetic analysis. We consider this result to represent a substantial advance in our understanding of testudinid relationships, especially concerning the problematic genus *Geochelone*, which has been in a state of substantial taxonomic confusion for the past 30 years. Although our results find *Geochelone* polyphyletic, representing at least four independent clades, we anticipate these findings will stimulate new interest within these tortoise groups and strengthen future comparative research.

Despite these systematic advances, further phylogenetic analysis will be needed to test the relationships we report here between *Indotestudo*, *Testudo*, and *Malacochersus*, and the relationships of *G. yniphora* and *G. radiata*. The inclusion of additional genes (nuclear and mitochondrial) and morphological data will likely resolve this uncertainty. Importantly, a re-examination of the morphological characters will also provide new insights concerning character evolution for the group, and in addition, aid with the interpretation of the fossil taxa. Unlike most extant reptile families, tortoises have a comparatively rich Cenozoic fossil
Acknowledgments

M.L. thanks Dr. Eleanor Sterling for the support she has provided throughout this project. Both Jeff Groth and Tom Near provided valuable help during this study, and we thank the following individuals for providing valuable tissues: Dr. George Amato (WCS), Drs. Cherri Asa and Randy Junge (Saint Louis Zoo), Dr. Sue Lynch (Chaffee Zoo), Dr. Ronald A. Nussbaum (University of Michigan Museum of Zoology), Eric Palkovacs, Yale University, Dr. Steve Reichling (Memphis Zoo), Dr. Chris Tabaka (Detroit Zoo), Dr. Harold Voris (Field Museum of Natural History), and Mr. Truong Quang Bich, Cuc Phuong National Park, Vietnam. We are also grateful to Dr. John Iverson for reading and providing helpful comments on the early version of the paper and Drs. Brad Shaffer and Gisella Caccone and an anonymous reviewer for suggestions that improved the paper. Margaret Arnold, David Dickey, and Robert J. Pacsocoletto facilitated the sampling of the specimens at the AMNH. Funding for this project was provided by the National Science Foundation Grant DEB 99-84496 awarded to C.J.R., the American Museum of Natural History, the Chelonian Research Foundation, the Department of Ecology, Evolution, and Environmental Biology, Columbia University, and NASA Grant NAG5-8543 to the Center for Biodiversity and Conservation at the American Museum of Natural History. We also acknowledge the generous support provided by the Louis and Dorothy Cullman Program in Molecular Systematic Studies and the Ambrose Monell Foundation.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2006.03.003.

References