Cryptic variation and the tragedy of unrecognized taxa: the case of international trade in the spiny turtle *Heosemys spinosa* (Testudines: Geoemydidae)

PHILLIP Q. SPINKS1*, ROBERT C. THOMSON1, BILL HUGHES2, BRAD MOXLEY3, RAFE BROWN4, ARVIN DIESMOS5 and H. BRADLEY SHAFFER1

1Department of Evolution and Ecology and Center for Population Biology, University of California, Davis, CA 95616, USA
2Tennessee Aquarium, Chattanooga, TN 37402, USA
3Knoxville Zoological Gardens, Knoxville, TN 37914, USA
4Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045, USA
5Herpetology Section, Zoology Division, National Museum of the Philippines, Ermita 1000, Manila, Philippines

Received 22 July 2011; revised 1 September 2011; accepted for publication 26 September 2011

Loss of habitat and human exploitation have driven many turtle species to the brink of extinction, particularly in many parts of southern Asia. The spiny turtle (*Heosemys spinosa*) is a terrestrial species distributed throughout the Sundaland region of South-East Asia. Despite international legislative protection, *H. spinosa* continues to be illegally collected for the food and traditional medicine markets of China. Given its widespread distribution, taxonomists have reasonably questioned whether *H. spinosa* truly represents a single evolutionary lineage or multiple undiagnosed species. Recently, a large and illegal shipment of rare, wild-caught *H. spinosa* was confiscated in Hong Kong, China, and the turtles were eventually distributed to several zoos and academic collections. Based on analyses of these individuals, along with additional individuals from the pet trade and museum collections, we found concordant genetic and phenotypic variation, indicating that two distinct types of *H. spinosa* exist in this collection of turtles. Further characterization of this variation will require field surveys and the collection of additional morphological and genetic data from specimens of known geographic provenance. However, our data indicate that this highly exploited, endangered species may contain additional cryptic taxa, and emphasize the critical need for systematic evaluation of species before unrecognized variation is lost forever.


INTRODUCTION

Conservation efforts can focus on a variety of organizational levels, from biomes, to habitats, to the processes that generate or maintain natural variation. However, when species are the units of conservation, efforts are necessarily focused on preserving distinct lineages. The extent of variation within most species, and thus the boundaries of species themselves, is often imperfectly known, even in well-known vertebrate taxa, which hinders the most basic conservation efforts (Daugherty *et al.*, 1990). The accurate and precise identification of biodiversity is thus a key goal of both conservation phylogenetics and systematics research. However, the identification of discrete

*Corresponding author. E-mail: pqspinks@ucdavis.edu*
lineages, including species, remains a frustratingly difficult empirical challenge that continues to hinder both taxonomy and conservation biology, sometimes in the face of considerable effort and data (Hey et al., 2003; Knowles & Carstens, 2007; Shaffer & Thomson, 2007).

Quantifying biodiversity takes on added urgency for threatened or endangered taxa, a problem that is of particular importance for turtles (Testudines). As a clade, turtles are one of the most threatened vertebrate groups (Ives, Spinks & Shaffer, 2007; Buhlmann et al., 2009; IUCN, 2011). Especially imperilled are the species found in Asia, where virtually all taxa are heavily exploited by humans (van Dijk, Stuart & Rhodin, 2000; Shi, Parham & Lau M. Tien-Hsi, 2007). Habitat loss and over-collection for the Asian food market, the traditional Chinese medicine (TCM) trade, as well as the global pet trade have driven many species to the brink of extinction (van Dijk, 2000). Here, we focus on cryptic diversity in the spiny turtle (Heosemys spinosa Gray, 1831), a terrestrial species distributed throughout Sundaland from peninsular Thailand, Malaysia, and southern Myanmar, across Indonesia to the Philippines and Singapore (Fig. 1).

PHYLOGENETICS OF HEOSEMYS AND BIOGEOGRAPHY OF H. SPINOSA

Heosemys currently comprises four species: Heosemys annandalii Boulenger, 1903, Heosemys depressa Anderson, 1875, Heosemys grandis Gray, 1860, and H. spinosa (Diesmos et al., 2005). Taylor (1920) assigned a fifth species (Heosemys leytensis Taylor, 1920) to the genus, but Diesmos et al. (2005) demonstrated that H. leytensis is closely related to Siebenrockiella crassicollis Gray 1831, and therefore relegated H. leytensis to Siebenrockiella. Although composed of only four currently recognized species, phylogenetic relationships among Heosemys species have been challenging to resolve. Based on a combined analysis of mitochondrial (cytochrome b) and nuclear (R35) sequence data across a broad taxonomic sampling of Geoemydidae, Spinks et al. (2004) recovered H. annandalii as sister to (H. grandis + H. depressa), with H. spinosa as sister to these three, although only the monophyly of Heosemys and the sister relationship of (H. grandis + H. depressa) were well supported. Diesmos et al. (2005) used data from Spinks et al. (2004) to determine the phylogenetic position of H. leytensis, but in the Diesmos et al. (2005) analysis, H. spinosa was recovered as sister to (H. grandis + H. depressa), with H. annandalii recovered as sister to the rest of the clade. However, as in the analysis of Spinks et al. (2004), Diesmos et al. (2005) found no statistical support for the phylogenetic positions of H. annandalii and H. spinosa. Thus, although the monophyly of Heosemys is well supported, relationships within it remain largely unresolved.

At the species level, little is known about geographic variation of H. spinosa. Vetter & van Dijk (2006) suggested that the species is composed of two distinct forms, including a ‘mainland form’ restricted to Malaysia, Thailand, and southern Myanmar, and an ‘insular form’ found throughout the remainder of the species’ range in Indonesia and the Philippines. However, Vetter & van Dijk (2006) provided no diagnoses for their two forms, and it is not clear what might differentiate mainland versus insular populations of H. spinosa. Given its disjunct distribution across a biogeographically complex and heterogeneous region, our a priori expectation was that this species might harbour significant intraspecific morphological and/or genetic variation, including unrecognized cryptic species.

Primarily as a result of unsustainable levels of exploitation and habitat destruction, H. spinosa is rare in the wild (possibly except for populations in Singapore and Brunei; IUCN, 2011), and has been classified as an endangered species (IUCN, 2011). However, H. spinosa (and many other threatened and endangered turtles) continue to be illegally collected for the food/TCM markets of China (Cheung & Dudgeon, 2006; Gong et al., 2009), and large shipments of live turtles or turtle parts are routinely confiscated in Asian ports (see http://www.traffic.org). In December of 2001, Hong Kong customs officials seized an exceptionally large illegal shipment of approximately four and a half tons of turtles, consisting of ~7500 live and ~2000 dead or dying turtles. These animals had been collected from the wild throughout South-East Asia and Indonesia, and were destined for the commercial markets of southern China (Ades & Crow, 2002). Live turtles from this confiscation were initially sent to Kadoorie Farms and Botanic Gardens, located in the Hong Kong Special Administrative Region (SAR), for triage, and even after an enormous community effort, more than a third eventually perished (Ades & Crow, 2002). Most survivors were later flown to the USA and Europe, and were distributed among zoos, nongovernmental organizations (NGOs), veterinarians, and private individuals for long-term care, with the ultimate goal of establishing or enhancing assurance colonies (Hudson & Buhlmann, 2002). The Tennessee Aquarium, Knoxville Zoo, and several of the authors received a total of five confiscated H. spinosa from this seizure, and comparisons among these and additional H. spinosa from reptile dealers revealed notable intraspecific morphological variation. For example, individuals could be sorted into two class sizes, with one being more than twice the adult mass of the other (Fig. 2; Appendix S1).
One of the great challenges in working with taxa like Heosemys spinosa is that wild-caught material of known provenance is often rare, making it extremely difficult to conduct standard phylogeographic and systematic studies (Parham & Li, 1999; Fong et al., 2007). However, genetic samples from existing specimens may shed light on levels and patterns of variation that exist in nature, and suggest strategies for captive breeding efforts and the need for field efforts to better understand the geography of lineage or species distributions (e.g. Fong et al., 2007). Because H. spinosa are difficult to breed in captivity, it is almost certainly the case that most or all captive specimens are wild caught, but lack reliable locality data. Variation among these captive individuals is pronounced, including variation in iris colour. For example, the iris colour in our samples varies from mostly white to dark brown and grey, whereas the shell colour varies from relatively light to dark (Fig. 2). As there are no locality data for the confiscated or pet trade H. spinosa, we cannot determine if these phenotypic classes are geographically structured, or if they are coincident with the hypothesized mainland and insular forms proposed by Vetter &

Figure 1. Map of the Sundaland region showing historical collection localities for Heosemys spinosa as purple (dark grey in the print version) circles (coordinates from http://emys.geo.orst.edu). The arrow indicates the collection locality for sample ELR230A (deposited at the National Museum of the Philippines) from the Island of Tawi Tawi, Philippines.
van Dijk (2006). However, using our relatively extensive genetic sampling, we can reconstruct a molecular phylogeny for *Heosemys* and use it as a framework to investigate variation within *H. spinosa*.

In this paper, our goals are to estimate a molecular phylogeny for *Heosemys* and assess genetic variation for a subsample of captive *H. spinosa* to determine if the divergent phenotypes correlate with genetic variation. However, we included sequence data from a turtle collected from the island of Tawi Tawi (ELR230A, deposited in the National Museum of the Philippines) at the extreme south-western edge of the Philippine Archipelago (Fig. 1). We use morphological analyses as well as analyses of mitochondrial and nuclear gene sequences in conjunction with a coalescent-based method to reconstruct the phylogeny for *Heosemys*, and assess genetic variation within our morphologically diverse set of turtles.

**MATERIAL AND METHODS**

**TAXON AND MOLECULAR DATA SAMPLING**

We assembled a mitochondrial DNA (mtDNA) cytochrome *b* (*cyt b*) gene sequence data set composed of four sequences downloaded from GenBank plus the 26 sequences generated here: 20 *H. spinosa*, two *H. annandalii*, two *H. depressa*, one *H. grandis*, and one *Cyclemys fusca* (Fritz et al., 2008), a closely related out-group to *Heosemys* (Hirayama, 1984; Spinks et al., 2004). We also assembled an 11-locus nuclear DNA (nuDNA) data set for 16 individuals subsampled to span the variation present in the

---

**Figure 2.** Side-by-side comparisons showing variation in relative size, iris colour, and light versus dark shell coloration among samples of *Heosemys spinosa*. The images show some of the variation present among six individual turtles, but are not meant to indicate that there two discrete groups of turtles in this collection.
mtDNA data set, including ten *H. spinosa*, two *H. an-nandali*, two *H. depressa*, and one *H. grandis*, plus the out-group.

**Molecular methods**

Tissue samples consisted of blood or other soft tissue. We stored blood samples in lysis buffer (100 mM NaCl, 100 mM Tris, 50 mM EDTA) and soft tissue samples in 95% ethanol (EtOH). All samples are maintained at −80 °C. We extracted DNA using a standard salt extraction protocol (Sambrook & Russell, 2001), and generated the cyt b sequences using the primers and PCR conditions from Spinks *et al.* (2004). Our nuDNA data set consisted of data from 11 loci including: AHR, BMP2, and NGF (Townsend *et al.*, 2007); HMBG2 (Backström, Fagerberg & Ellegren, 2008; Barley *et al.*, 2010); HNFL and TGFβ2 (Primmer *et al.*, 2002); NB22519 (Backström *et al.*, 2008); PAXIP1 (Kimball *et al.*, 2009); R35 (Fujita *et al.*, 2004); and TB01 and TB29 (Thomson *et al.*, 2008). The polymerase chain reaction (PCR) conditions for HMBG2 can be found in Barley *et al.* (2010), and see Spinks *et al.* (2010) for the PCR conditions for the remaining nuclear loci. PCR products were cleaned and sequenced in both directions by Beckman Coulter Genomics using Sanger sequencing (http://www.beckmangenomics.com).

**Phylogenetic analyses of mtDNA and nuDNA**

We edited and aligned all sequences using MUSCLE (Edgar, 2004), and checked coding loci for pseudogenes using GENEIOUS 5.1 (Drummond *et al.*, 2010). Gene trees were reconstructed under Bayesian inference (BI) using MRBAYES 3.1.1 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). In addition, we performed two separate analyses on the mtDNA, including one analysis where the data were partitioned by codon, and one analysis on the unpartitioned data set. Given the large number of possible partitioning schemes available for the nuclear data and the relatively small level of divergence in our nuclear genes, we analysed our nuclear data as partitioned by locus only (11 partitions). We also analysed each nuclear locus individually. Models of molecular evolution for parameter estimation were selected using the JMODELTEST 0.1.1, using Akaike's information criterion (AIC) (Guindon & Gascuel, 2003; Posada, 2008). As MRBAYES does not incorporate some commonly used models of sequence evolution, we used models similar to those selected via JMODELTEST, but implemented in MRBAYES. The Bayesian analyses consisted of two replicates and four chains for $5 \times 10^6$ generations. Chains were sampled every 1000 generations, with the first 25% of samples discarded as burn-in. We checked that the runs had reached stationarity by ensuring that the potential scale reduction factor equalled 1, the average standard deviation of split frequencies between independent runs approached 0, and by visually examining the Markov chain Monte Carlo (MCMC) output to ensure that all chains were sampling from the same target distribution in TRACER (Ramaut & Drummond, 2007). Finally, we performed a partitioned-model maximum likelihood (ML) analysis on the concatenated nuDNA data set using RaxML (Stamatakis, 2006a), and we performed bootstrap analysis with 1000 pseudoreplicates. For this analysis, the data were partitioned by locus, and the GTR+CAT model (Stamatakis, 2006b) was applied to each partition.

**Species delimitation**

We tested for the presence of two distinct lineages within our *H. spinosa* samples using BPP 2.0 (Rannala & Yang, 2003; Yang & Rannala, 2010). BPP is a Bayesian modelling approach that accounts for uncertainties in the ancestral coalescent process and unknown gene trees by integrating over uncertainty in the gene trees, while explicitly modelling lineage sorting using the coalescent process (Yang & Rannala, 2010). BPP requires a user-provided guide tree, and tests whether the proposed species in that tree are supported by the molecular data. We used a $\Gamma(2, 200)$ prior on the population coalescent parameters ($\theta$) and a $\Gamma(2, 2000)$ prior on the age of the root in the species tree ($\tau_0$), whereas the other divergence time parameters were estimated under a uniform Dirichlet prior (Yang & Rannala, 2010). In addition, we varied the priors, using $\theta \sim \Gamma(2, 2000)$ and $\tau_0 \sim \Gamma(2, 2000)$, which assumes relatively large ancestral populations sizes but relatively shallow divergences among species. The latter combination of priors favours models containing fewer lineages (Yang & Rannala, 2010), and is therefore conservative with respect to the recognition of taxa. We used PHASE 2.1.1 (Stephens, Smith & Donnelly, 2001; Stephens & Donnelly, 2003) to reconstruct haplotype data for each of our 11 nuclear loci, and used these data and the majority-rule consensus tree from the posterior distribution of trees from the Bayesian analyses of the concatenated nuDNA data as the guide tree in BPP 2.0 to test the mtDNA-based substructure within *H. spinosa* (see Fig. 3). Initially, we ran the MCMC analyses for 5000 generations using algorithm 0, and adjusted ‘fine tune’ variables in the control file until the acceptance proportions were close to 0.3 and were contained in the interval (0.15, 0.70). Once adequate acceptance proportions were attained, we ran four final analyses (two using algorithm 0 and two using algorithm 1) for 100 000
generations (sampling every five generations) using different starting seeds to confirm consistency among runs.

**Morphometrics**

We measured the lengths and widths of the plastron and carapace, overall height of the shell, and mass for most of the *H. spinosa* included here (Appendix S1). To test the hypothesis that the two major mtDNA clades (A and B, see below) represent two distinct morphotypes of *H. spinosa*, we first removed the overall effect of body size by regressing each variable on carapace length, and used the resulting residuals as size-independent shape measurements. We then performed a two-group discriminant function analysis (DFA) on the residuals grouped by mtDNA clade (i.e. clade A versus clade B), and scored the number of incorrect assignments of individuals to groups as a measure of group differentiation. Because DFA can sometimes return spurious results when the number of individuals per group is small and the number of variables per individual is large, we did not test clade C versus clade D (see below). In addition, we used a permutation test (100 replicates) to determine if our morphological data performed better at

**Figure 3.** *Heosemys* mitochondrial phylogeny: majority-rule consensus tree of the posterior distribution of trees from the Bayesian analyses of the cytochrome *b* data set (971 bp). Model of molecular evolution: GTR + G. Support values (posterior probabilities, PPs) as indicated.
discriminating the two mitochondrial clade members compared with random sets of turtles from our data set. For this procedure, each turtle was randomly assigned to one of two groups (always with nine individuals in one group and ten in the other), and we tallied the distribution of misclassifications to determine how the DFA performed at discriminating randomly assembled groups compared with the mtDNA clade A versus clade B grouping. The DFA analysis and permutation test were carried out in R (http://www.r-project.org).

RESULTS

MITOCHONDRIAL PHYLOGENY

Sequence chromatograms from four *H. spinosa* cyt b sequences exhibited multiple peaks in several positions, indicating possible contamination of our DNA templates and/or the presence of nuclear mitochondrial pseudogenes (numts). Therefore, we excluded these four sequences from the analysis. Our final mtDNA sequence data set was composed of up to 971 bp of cyt b data for each of 30 turtles, including 26 sequences generated for this analysis (GenBank accession numbers JN797961-JN797986), plus four downloaded from GenBank (AM931654, AY434566, AY434578, and EU787022). The data set was nearly complete, with only 2% missing data. Parameter estimates from the posterior distribution of several parameters from the mtDNA partitioned model Bayesian analyses failed to converge because of a lack of parameters from the mtDNA partitioned model Bayesian analyses. Because of this, we simplified the model by collapsing the partitions and report the results from this unpartitioned analysis. Analyses of the unpartitioned data the partitions and report the results from this unpartitioned analysis. Because of this, we simplified the model by collapsing the partitions and report the results from this unpartitioned analysis. Analyses of the unpartitioned data set recovered a tree with a mean -lnL score of 5321.92 (Fig. 3). All species were recovered as monophyletic with strong support [Bayesian posterior probabilities (PP) = 1], as were relationships among *Heosemys* (PP = 1). The mtDNA phylogeny that we recovered was also topologically equivalent to the tree recovered in Spinks et al. (2004), in that *H. spinosa* was sister to the remaining *Heosemys* (Fig. 3). In addition, we recovered a deep but weakly supported bifurcation among *H. spinosa*, which we refer to as clades A and B (Fig. 3). Clade B was further subdivided into two major reciprocally monophyletic clades (C and D; Fig. 3).

NUCLEAR DNA PHYLOGENIES AND THE SPECIES TREE

Our nuDNA data set consisted of up to 7541 bp generated from 11 nuclear loci for each individual, except for our lone *H. grandis*, where were unable to generate data from NB22519 and PAXIP1. The remaining sequences were mostly complete, with only 4% missing data. However, patterns from the sequencing chromatograms indicated the presence of heterozygous length polymorphisms in the NB22519 and PAXIP1 loci (two and three *H. spinosa*, respectively), which corrupt sequence reads downstream of the indel site (Bhangale et al., 2005). To correct for this, we used INDELIGENT 1.2 (Dmitriev & Rakitov, 2008) to reconstruct nucleotide sequences (available at http://ctap.inhs.uiuc.edu/dmitriev/indel.asp). In total, we generated 167 new nuDNA sequences (GenBank numbers JN797882-JN798074, JN806257), and included five GenBank sequences (AM931701, EU787078, EU787161, EU787220, and EU787268). Unfortunately, the sample from Tawi Tawi Island (*H. spinosa* ELR230A) was degraded: we were only able to generate sequence data from TB29 for this sample. Therefore, *H. spinosa* ELR230A was excluded from the analyses of the concatenated nuDNA data, although this sample was included in the single gene TB29 analysis (Fig. S1).

Individual gene trees were topologically similar, and recovered *H. spinosa* as monophyletic at all loci (strongly so for nine of the 11 loci), whereas *H. annandalii* and *H. depressa* were monophyletic, with strong support at six and nine out of ten loci, respectively (Fig. S1). We were unable to assess the monophyly of *H. grandis* because we only had one sample of this species (the additional *H. grandis* mtDNA sequence was downloaded from GenBank).

The nuclear phylogeny recovered from analyses of the 11 concatenated loci was novel with respect to our mitochondrial results and those of previous analyses (i.e. Spinks et al., 2004; Diesmos et al., 2005) in some respects, and was similar in others. As in the mtDNA analysis, all *Heosemys* species were recovered as monophyletic with strong support (PP = 1.0), and *H. spinosa* fell out as the sister group to all other *Heosemys* (but with weak support; Fig. 4). However, the nuDNA recovered *H. depressa* as a well-supported sister group to (*H. annandalii + *H. grandis*) (PP = 1.0), rather than the well-supported (*H. depressa + *H. grandis*) relationship recovered from the mtDNA (Fig. 3). Missing data can lead to biased estimates of phylogeny (Lemmon et al., 2009), and we were unable to generate data from two loci (NB22519 and PAXIP1) for our *H. grandis* sample. To check for these possible biases, we reanalysed the concatenated nuDNA data with NB22519 and PAXIP1 excluded (not shown), and recovered the same topology as in Figure 4. Thus, the novel phylogeny recovered from analyses of the concatenated 11-locus nuDNA data set (Fig. 4) is well
supported, and does not appear to be biased by missing data.

**HYPOTHESIS TESTING, MOLECULAR DATA**

The hypothesis tested here is the fully resolved tree where *H. spinosa* were constrained to mtDNA clades A–D, conditional on a fixed topology with *H. spinosa* as sister to (*H. depressa* (*H. grandis* + *H. annandalii*)). Specifically, we used BPP to test two hypotheses: (1) a tree that split *H. spinosa* into two more inclusive clades (clades A and B); or (2) a more resolved tree that split *H. spinosa* into clades A, C and D (i.e. nodes g and e in Fig. 5 were better supported than a tree that collapsed node g, node e, or both of these nodes). Under all combinations of priors and algorithms that we tested, the posterior probability of the tree with *H. spinosa* split into clades A and B = 1.0, which is consistent with the presence of these two lineages within *H. spinosa*. In addition, the −lnL scores from all runs were virtually identical, as were analyses using the same set of priors but different starting seeds (Table S1). Results from analyses employing different algorithms varied for some parameter estimates, most notably θ for the *H. annandalii*–*H. grandis* divergence, and the basal divergence among all species (Appendix S1). Thus, the pattern of divergence among samples of *H. spinosa* revealed by the mitochondrial data (i.e. clade A versus clade B; Fig. 3) was consistent with the nuclear data under the multispecies coalescent model employed by BPP (Fig. 5). However, support for the subdivision among clade B samples was uncertain. Under either algorithm (i.e. 0 or 1) the split between clades C and D was supported when using the Γ(2, 200) priors, but was not supported under the Γ(2, 200) priors (Fig. 5).

**HYPOTHESIS TESTING, MORPHOMETRIC DATA**

Using the multivariate morphometric data, the hypothesis being tested is the division of *H. spinosa* samples into two major groups (mtDNA clades A and B). Results from the DFA and permutation tests again strongly supported the recognition of these two groups. Based on the DFA, eight out of nine *H. spinosa* with clade-A mitotypes were correctly assigned to clade A, and ten out of ten clade-B turtles were correctly assigned to clade B. The permutation test strongly supported the interpretation that grouping turtles by mtDNA clade is significantly better than grouping randomly (P = 0.01), as the number of correctly classified turtles from the initial DFA analysis (18/19), was greater than all 100 random groups of ten and nine specimens (Fig. 6). Among those 100 random groupings, between five and 14 out of 19 individuals were misclassified, suggesting that the

---

Figure 4. Majority-rule consensus tree of the posterior distribution of trees from the Bayesian analyses of the concatenated nuclear loci data (11 loci, 7541 bp). −lnL = 12 653.11. Models of molecular evolution for each loci: GTR (AHR, AIING, BMP2, HMGB2, PAX, R35, and TFG2); HKY (HNFL, TB01, and TB29), and HKY + Γ (NB22519). Bayesian support values (posterior probabilities, PPs) as indicated. Numbers above branches are maximum likelihood (ML) bootstrap support values. Mitochondrial clade membership is indicated for each sample. The out-group branch is not drawn to scale.
mtDNA-based groups, with only a single misclassified individual, are highly significantly different from random groups (Fig. 6).

**DISCUSSION**

In general, our knowledge of biodiversity remains inadequate, especially in regions with high biodiversity, a situation that some view as a ‘glorious opportunity’ because ‘there are so many species waiting to be discovered’ (Brown & Lomolino, 1998). On the other hand, one could view our inadequate taxonomic knowledge with apprehension because many species may become extinct before they can be discovered and described. Sundaland is one of the ‘hottest’ biodiversity hot spots (Myers et al., 2000), and much of the biodiversity in this region is not yet described (Brown, Diesmos & Alcala, 2008; Stuart & Bain, 2008; Brown & Diesmos, 2009). Furthermore, enormous conservation risks are present in this area because of pressure from urban development and the intense exploitation of natural resources (Aratrkorn, Thunhikorn & Donald, 2006; Soh, Sodhi & Lim, 2006; Brühl & Eltz, 2009; Danielsen et al., 2009). Secretive and cryptic

---

**Figure 5.** Lineage diversification results for *Heosemys*, with *Heosemys spinosa* split into two (A and B, top panel) or three (A, C, D, bottom panel) clades. Estimates were generated using priors for $\theta$ of $\Gamma(2, 200)$ and $\Gamma(2, 2000)$, whereas priors for $\tau$ were $\Gamma(2, 2000)$ for all analyses. Numbers above nodes are posterior probability (PP) support values. PP support values for node e (all runs) are provided below the tree. Means and standard deviations for $\theta$ and $\tau$, and $-\log$ likelihood scores for each analysis are provided in Table S1.
taxa are of particular concern because quantifying cryptic biodiversity in these organisms is more difficult than for large, obvious, or well-studied taxa (Stuart, Inger & Voris, 2006; Bickford et al., 2007). For example, amphibians are often relatively inconspicuous and secretive, and recent studies suggest that the amphibian species diversity throughout South-East Asia/Sundaland is greatly underestimated (Stuart et al., 2006; Bain et al., 2008; Inger, Stuart & Iskandar, 2009; Brown & Stuart, in press). Even relatively charismatic and well-studied taxa, such as turtles, are often poorly known. For example, a new species of Asian leaf turtle, *Cyclemys enigmatica* Fritz et al. 2008 was recently described from the southern Malay Peninsula and Greater Sunda Islands (Fritz et al., 2008), an area that has suffered catastrophic losses of habitat to agriculture (Aratrakorn et al., 2006; Sodhi & Brook, 2006). Describing species and lineage diversity is clearly an important, continuing process in this region.

Taken together, our analyses suggest that significant variation is present within the *H. spinosa* lineage. However, because of inherent limitations with the available genetic samples, we are unable to fully characterize this variation. Our samples lack locality data, and the danger posed by incorrectly interpreting putative diversity in samples of unknown provenance has been amply demonstrated in turtles (Wink, Guicking & Fritz, 2000; Parham et al., 2001; Stuart & Parham, 2004, 2007; Spinks et al., 2004; Fong et al., 2007; Spinks & Shaffer, 2007). We thus tread a delicate line in this work: on the one hand, it is critically important to recognize potentially imperilled biodiversity in endangered taxa, but the lack of field-collected, museum-vouchered specimens, combined with previous examples of newly described Asian turtles that turned out to represent anthropogenically produced hybrids, demands caution before interpreting variation as representing undescribed lineages.

Both mtDNA and nuDNA analyses confirm that the currently recognized four species of *Heosemys* are justifiable based on gene tree monophyly. In addition, the mtDNA phylogeny and results from the BPP analysis of 11 nuclear genes hint at the presence of two major clades (the ‘species’ of Leaché & Fujita, 2010) within the group (i.e. clades A and B), but nuDNA gene trees do not recover this split. Our morphometric data further support the reality of the two major mtDNA clades, and the reanalysis of random sets of turtles from this study confirm that this morphological result is not caused by chance or a biased morphometric study. Taken together, these data are consistent with either of two straightforward explanations. The first is that at least two evolutionary lineages exist within *H. spinosa*, but that the split is recent enough that nuDNA gene tree monophyly has not yet evolved. The time required for nuclear gene tree monophyly to develop is substantial (Hudson & Coyne, 2002). This is particularly true in

![Figure 6](image-url)
turtles, which characteristically have long generation times. The second is that our sample of *H. spinosa* is drawn from a single, largely admixed lineage that maintains a relatively high level of haplotypic and morphological diversity. Given that both the BPP and morphometric analyses confirm the groupings based on the mtDNA gene tree, we provisionally favour the interpretation that true lineages exist within this wide-ranging species. In particular, the BPP analysis, which was based solely on nuDNA data, suggests that the pattern and timing of lineage sorting in this data set is at least consistent with the presence of two lineages. However, the BPP analysis necessarily relies on a very specific model (panmictic populations within species-level lineages, no gene flow following divergence, and clock-like molecular evolution), and these may be seriously violated in our samples.

For many turtle species, conservation has reached a crisis level, and we must do the best job possible with the available material to understand both lineage boundaries and their geographic distributions to enable effective conservation and management. *Hosemys spinosa* exemplifies the plight of many species in that basic taxonomic and biogeographic data are lacking, and much of the material available for scientific research comes from illegally collected turtles lacking locality data (Parham & Li, 1999; Fong et al., 2007). Our work suggests that two lineages, perhaps representing separate conservation targets, may exist in this heavily exploited species. Adding additional captive and confiscated specimens may further resolve the boundaries of these putative lineages, or may lead to a refutation of their reality. Even a few specimens from known geographic localities would immensely help to align these lineages to geographic regions. Ultimately, a relatively comprehensive set of known locality samples spanning the range of the species will be necessary to fully describe and quantify diversity within this charismatic and imperilled lineage. However, our work does establish the hypothesis, perhaps consistent with Vetter & van Dijk (2006), that multiple lineages may exist and require conservation actions in this species.

ACKNOWLEDGEMENTS

We thank Chris Tabaka for tissue samples, Samantha Price and Chris Searcy for statistical advice, Cathy Newman for assistance with molecular lab work, and two anonymous reviewers for comments that improved this article. This research was supported by grants from the US National Science Foundation (DEB-0817042 to HBS and 0743491 to RMB); RMB and ACD acknowledge the Philippine Department of the Environment and Natural Resources for issuing research and export permits.

REFERENCES


Brown RM, Stuart BL. In press. Patterns of biodiversity discovery through time: an historical analysis of defining...


Lemmon AR, Brown JM, Stanger-Hall K, Moriarty
Lemmon E. 2009. The effect of ambiguous data on phylo-

Myers N, Mittermeier RA, Mittermeier CG, Fonseca
GAB, Kent J. 2000. Biodiversity hotspots for conservation

Parham JF, Li D. 1999. A new locality of Cuora pani Song
1984 with comments on its known range. *Asiatic Herpetology
Research* 8: 111–113.

Parham JF, Simison WB, Kozak KH, Feldman CR, Shi
HT. 2001. New Chinese turtles: endangered or invalid? A
reassessment of two species using mitochondrial DNA,
alloyze electrophoresis and known-locality specimens.


nucleotide polymorphism characterization in species with
limited available sequence information: high nucleotide
diversity revealed in the avian genome. *Molecular Ecology*
11: 603–612.

Rambaut A, Drummond AJ. 2007. *Tracer v1.4*. Available at:
http://beast.bio.ed.ac.uk/Tracer

divergence times and ancestral population sizes using
DNA sequences from multiple loci. *Genetics* 164: 1645–
1656.

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

Spring Harbor Laboratory Press.

Shaffer HB, Thomson RC. 2007. Delimiting species in

endangered turtles to extinction in China. *Conservation

Sodhi NS, Brook BW. 2006. *Southeast Asian biodiversity in
crisis*. Cambridge, UK: Cambridge University Press.

Soh MCK, Sodhi NS, Lim SLH. 2006. High sensitivity of
montane bird communities to habitat disturbance in Pen-
insular Malaysia. *Biological Conservation* 129: 149–166.

Spinks PQ, Shaffer HB. 2007. Conservation phylogenetics of
the Asian box turtles (Geomydidae, *Cuora*): mitochondrial
introgression, nufts, and inferences from multiple nuclear

Phylogenetic hypotheses for the turtle family Geemydidae.
*Molecular Phylogenetics and Evolution* 32: 164–182.

Spinks PQ, Thomson RC, Barley AJ, Newman CE,
Shaffer HB. 2010. Testing avian, squamate, and mamma-
lian nuclear markers for cross amplification in turtles.
*Conservation Genetics Resources* 2: 127–129.

Stamatakis A. 2006a. RAxML-VI-HPC: maximum likelihood-
based phylogenetic analyses with thousands of taxa and

Stamatakis A. 2006b. Phylogenetic models of rate heteroge-
neity: a high performance computing perspective. Proceed-
ings of 20th IEEE/ACM International Parallel and Distributed
Processing Symposium (IPDPS2006), High Per-
formance Computational Biology Workshop, Proceedings on
CD, Rhodes, Greece.

methods for haplotype reconstruction. *American Journal of
Human Genetics* 73: 1162–1169.

method for haplotype reconstruction from population data.

Stuart BL, Bain RH. 2008. Amphibian species discovery in
mainland Southeast Asia. In: Stuart SN, Hoffmann M,
Chanson JS, Cox NA, Berridge RJ, Ramani P, Young BE,
eds. *Threatened amphibians of the world*. Barcelona:
Lynx Ediciones, with IUCN – The World Conservation
Union, Conservation International, and NatureServe,
83–84.

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

Taylor EH. 1920. Philippine turtles. *Philippine Journal of
Science* 16: 111–144.

Thomson RC, Shedlock AM, Edwards SV, Shaffer HB.
2008. Developing markers for multilocus phylogenetics in
non-model organisms: a test case with turtles. *Molecular

Townsend TM, Alegre RE, Kelley ST, Reeder
TW. 2007. Rapid development of multiple nuclear loci for
phylogenetic analysis using genomic resources: an example
from squamate reptiles. *Molecular Phylogenetics and Evo-
Iution* 47: 129–142.

Vetter H, van Dijk PP. 2006. Teralog, turtles of the world,
Vol. 4. East and South Asia. Frankfurt am Main and
Rodgau: Chimaira and Aqualog.

hybrid origin of *Mauremys iversoni* Pritchard et McCord,
1992 and *Mauremys pritchardi* McCord, 1997 (Reptilia:
Testudines: Bataguridae). *Zoologische Abhandlungen,
Dresden* 51: 41–49.

Yang Z, Rannala B. 2010. Bayesian species delimitation
using multilocus sequence data. *Proceedings of the National
Academy of Sciences of the United States of America* 107:
9264–9269.
Additional supporting information may be found in the online version of this article:

**Figure S1.** Maximum clade credibility trees from Bayesian analyses of single loci. The out-group was removed to simplify the presentation. Support values (posterior probability, PP) are as indicated.

**Table S1.** Mean and standard deviation (SD) for population size (θ) and divergence time estimates (τ) from the BPP analyses. Analyses were run with *Heosemys spinosa* samples split into clades A and B (upper table), and with samples split into clades A, C and D (lower table); Hann, *Heosemys annandalii*; Hdep, *Heosemys depressa*; Hgrand, *Heosemys grandis*; Hspin, *Heosemys spinosa*. Lower case letters in column headers correspond to the nodes in Figure 5. NA indicates values that are unavailable because PP support values were < 0.50.

**Appendix S1.** Table showing collection, individual ID, and mtDNA clade membership for *Heosemys spinosa*. Also shown are shell measurements (mm) and mass (g); CL, carapace length; CW, carapace width; H, overall shell height; M, mass; PL, plastron length; PW, plastron width. Mitochondrial sequences were not generated for F95, TNA98006, TNA98010, and TNA98012 because of putative nuclear mitochondrial pseudogenes (numts). *Turtles from the 2001 Hong Kong confiscation.*

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