



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

PHILLIP Q. SPINKS^{1*}, ROBERT C. THOMSON¹, BILL HUGHES², BRAD MOXLEY³,
RAFE BROWN⁴, ARVIN DIEMOS⁵ 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.

© 2012 The Linnean Society of London, *Zoological Journal of the Linnean Society*, 2012, **164**, 811–824.
doi: 10.1111/j.1096-3642.2011.00788.x

ADDITIONAL KEYWORDS: Asian turtle crisis – confiscation – extinction – species delimitation.

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; Buhmann *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 crassicolis* 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 Geomydidae, 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, non-governmental organizations (NGOs), veterinarians, and private individuals for long-term care, with the ultimate goal of establishing or enhancing assurance colonies (Hudson & Buhmann, 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).



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.

One of the great challenges in working with taxa like *H. 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 &

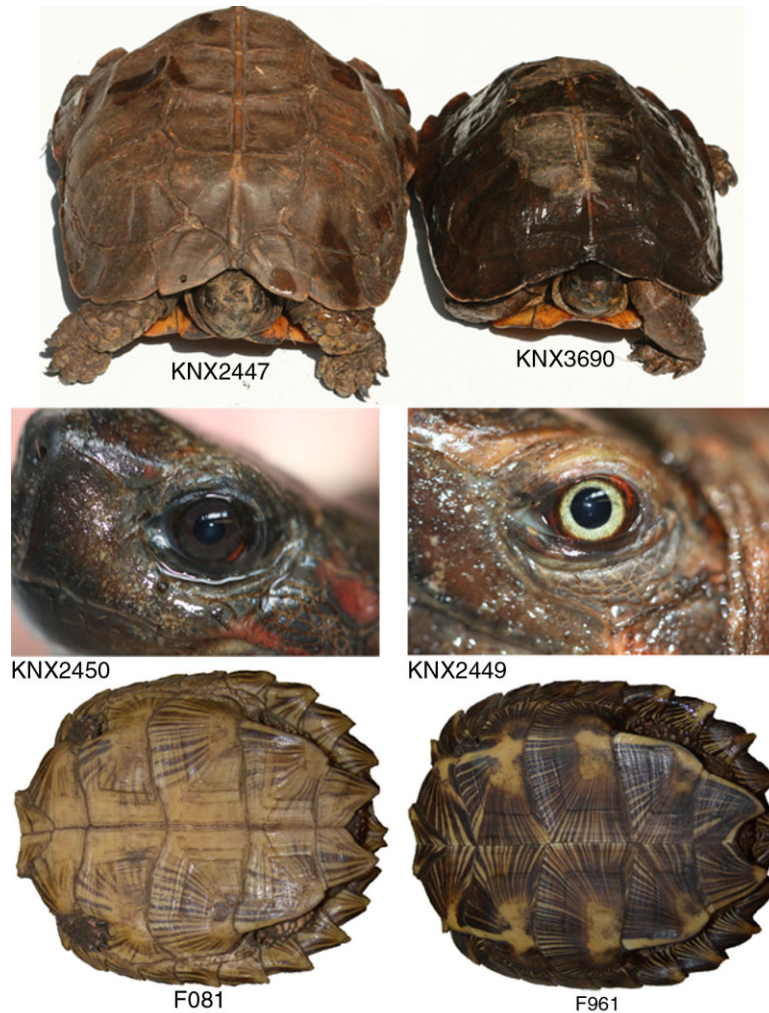


Figure 2. Side-by-side comparisons showing variation in relative size, iris colour, and light versus dark shell coloration among samples of *Heosemys spinosa*. The images show some of the variation present among six individual turtles, but are not meant to indicate that there are two discrete groups of turtles in this collection.

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

mtDNA data set, including ten *H. spinosa*, two *H. anandalii*, 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); HMGB2 (Backström, Fagerberg & Ellegren, 2008; Barley *et al.*, 2010); HNFL and TGFB2 (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 5×10^6 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 (Rambaut & 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 (θ) and a $\Gamma(2, 2000)$ prior on the age of the root in the species tree (τ_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

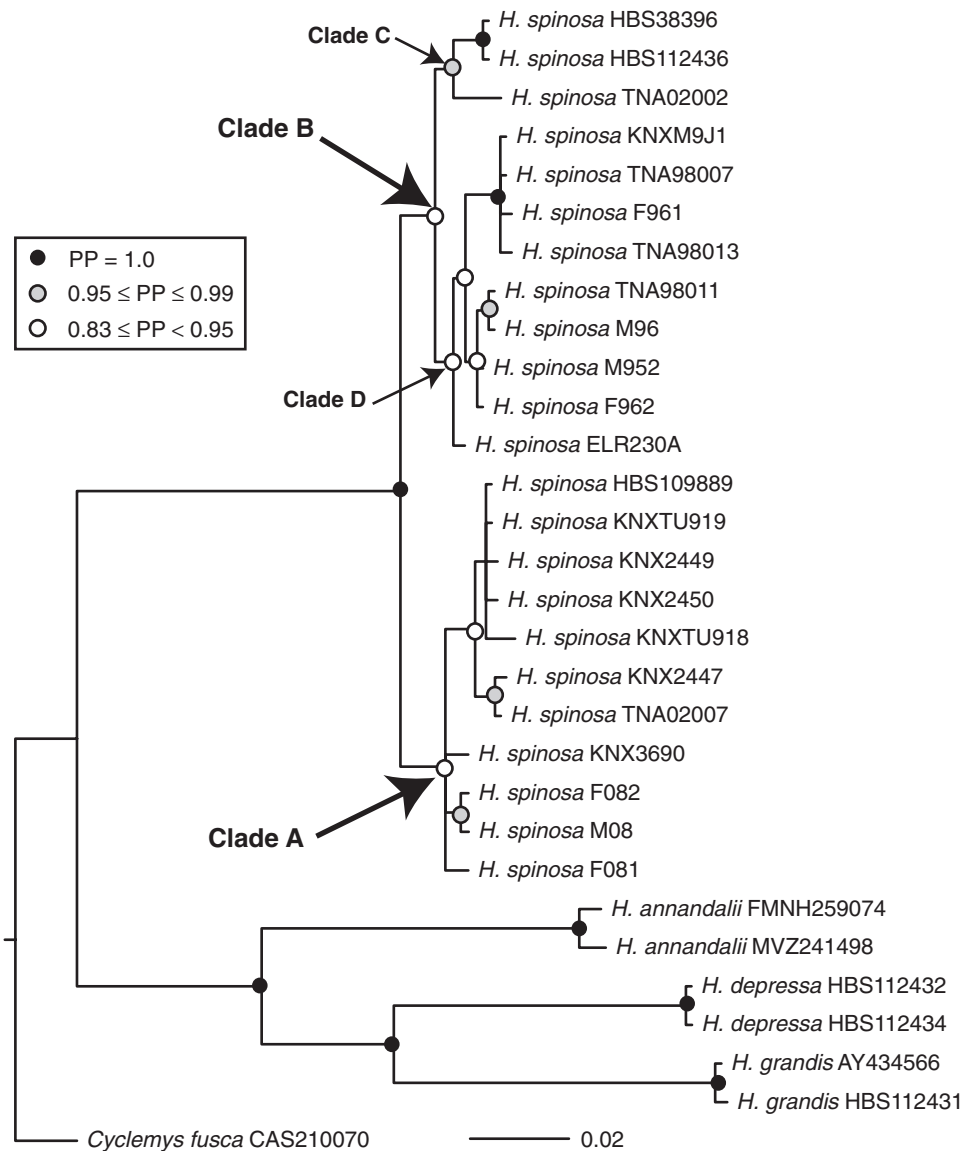


Figure 3. *Heosemys* mitochondrial phylogeny: majority-rule consensus tree of the posterior distribution of trees from the Bayesian analyses of the cytochrome *b* data set (971 bp). Model of molecular evolution: GTR + Γ . Support values (posterior probabilities, PPs) as indicated.

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

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 variation at the second codon position (not shown). 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 $-\ln L$ 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 we 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 INDELLIGENT 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

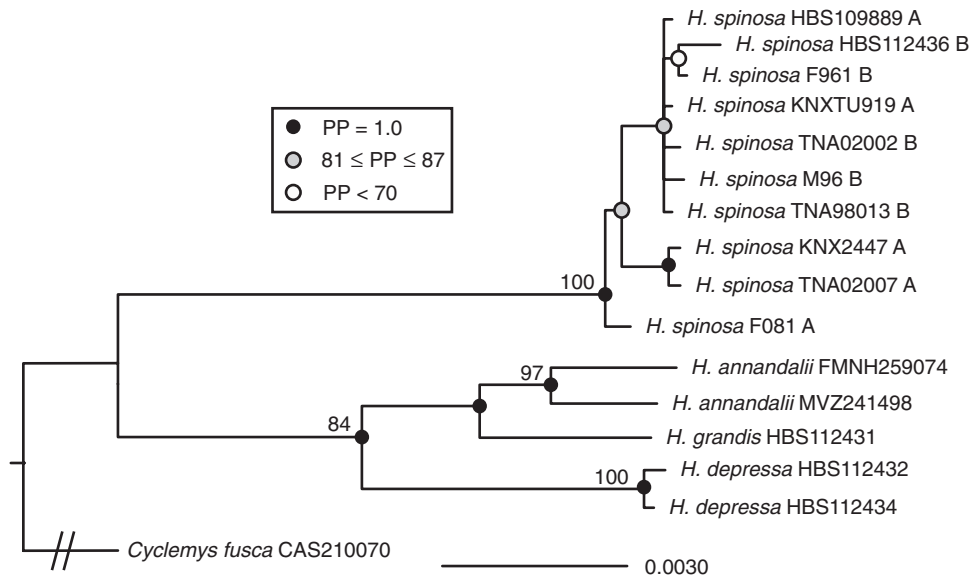


Figure 4. Majority-rule consensus tree of the posterior distribution of trees from the Bayesian analyses of the concatenated nuclear loci data (11 loci, 7541 bp). $-\ln L = 12\,653.11$. Models of molecular evolution for each loci: GTR (AHR, AIING, BMP2, HMGB2, PAX, R35, and TFGB2); HKY (HNFL, TB01, and 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.

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 $-\ln L$ 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 $\Gamma(2, 2000)$ priors, but was not supported under the $\Gamma(2, 200)$ 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

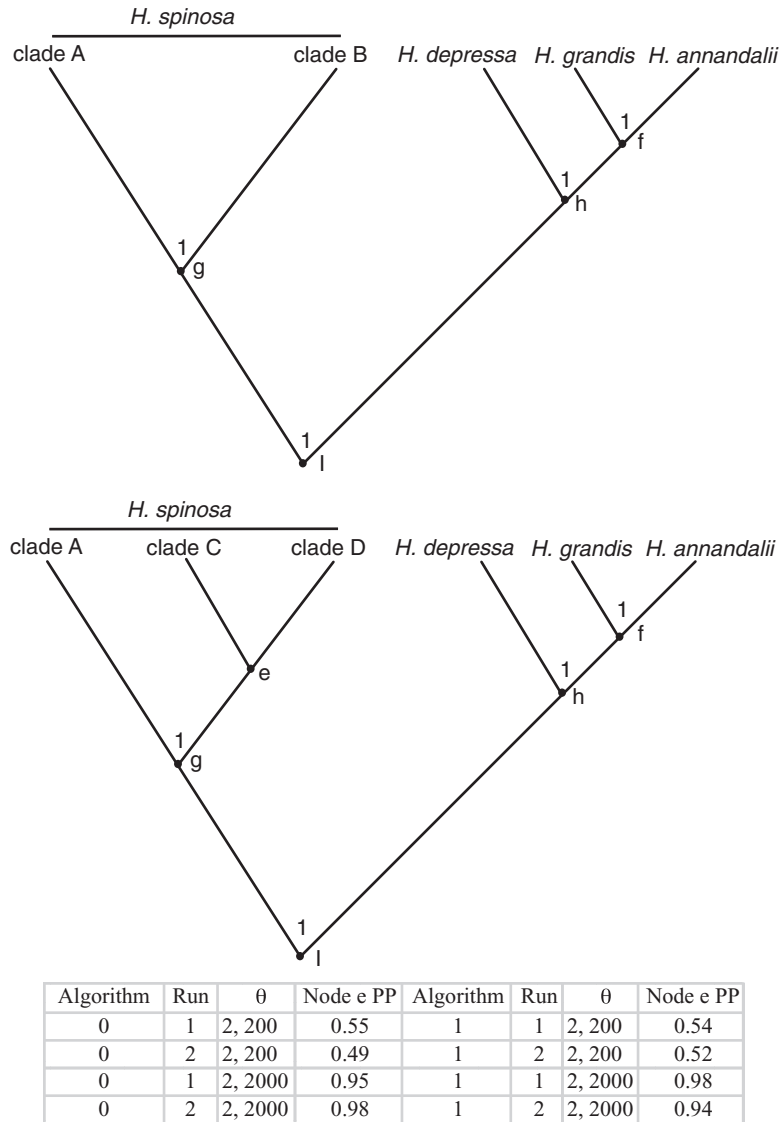


Figure 5. Lineage diversification results for *Heosemys*, with *Heosemys spinosa* split into two (A and B, top panel) or three (A, C, D, bottom panel) clades. Estimates were generated using priors for θ of $\Gamma(2, 200)$ and $\Gamma(2, 2000)$, whereas priors for τ were $\Gamma(2, 2000)$ for all analyses. 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 θ and τ , and $-\log$ likelihood scores for each analysis are provided in Table S1.

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 & Alcalá, 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 (Aratrakorn, Thunhikorn & Donald, 2006; Soh, Sodhi & Lim, 2006; Brühl & Eltz, 2009; Danielsen *et al.*, 2009). Secretive and cryptic

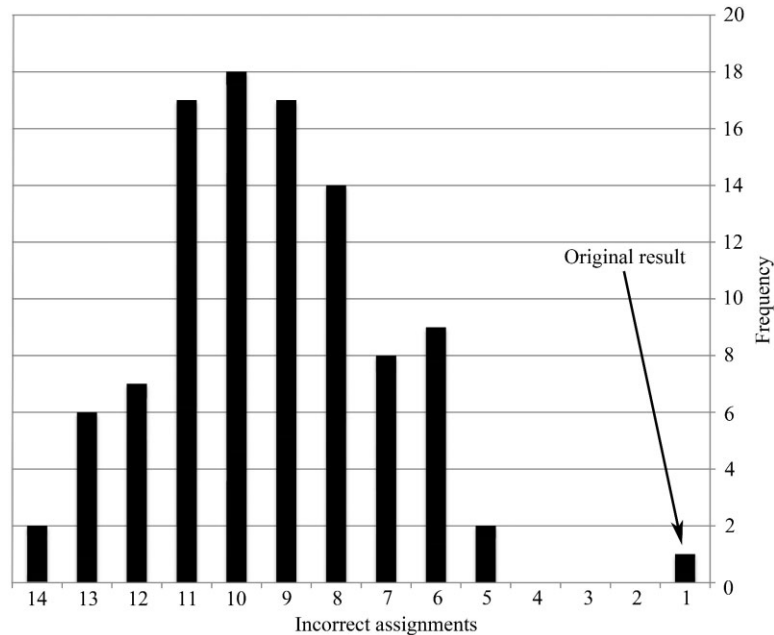


Figure 6. Incorrect classification frequencies for the discriminant function analysis (DFA) of 19 *Heosemys spinosa* specimens, based on membership of the two mitochondrial DNA clades (A and B, see Fig. 3), and for 100 randomly created groups of nine and ten individuals.

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

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. *Heosemys 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

- Ades G, Crow P. 2002.** The Asian turtle rescue operation: temporary holding and placement at Kadoorie Farm and Botanic Garden, Hong Kong. *Turtle and Tortoise Newsletter* **6**: 2–7.
- Anderson J. 1875.** Description of some new Asiatic mammals and Chelonia. *Annals and Magazine of Natural History London* **16**: 282–285.
- Aratrakorn S, Thunhikorn S, Donald PF. 2006.** Changes in bird communities following conversion of lowland forest to oil palm and rubber plantations in southern Thailand. *Bird Conservation International* **16**: 71–82.
- Backström N, Fagerberg S, Ellegren H. 2008.** Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Molecular Ecology* **17**: 964–980.
- Bain R, Biju SD, Brown R, Das I, Diesmos A, Dutta S, Gower D, Inger R, Iskandar D, Kaneko Y, Lau MW, Meegaskumbura M, Ohler A, Pethiyagoda R, Stuart B, Wilkinson M, Xie F. 2008.** Amphibians of the Indomalayan Realm. In: Stuart SN, Hoffmann M, Chanson J, Cox NA, Berridge R, Ramani P, Young BE, eds. *Threatened amphibians of the world*. Barcelona: Lynx Ediciones, with IUCN – The World Conservation Union, Conservation International, and NatureServe, 74–79.
- Barley A, Spinks PQ, Thomson RC, Shaffer HB. 2010.** Fourteen nuclear genes provide phylogenetic resolution for difficult nodes in the turtle tree of life. *Molecular Phylogenetics and Evolution* **55**: 1189–1194.
- Bhargale TR, Rieder MJ, Livingston RJ, Nickerson DA. 2005.** Comprehensive identification and characterization of diallelic insertion-deletion polymorphisms in 330 human candidate genes. *Human Molecular Genetics* **14**: 59–69.
- Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007.** Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* **22**: 148–155.
- Boulenger GA. 1903.** Report on the batrachians and reptiles. In: Annandale N, Robinson HC, eds. *Fasciculi Malayaenseis, anthropological and zoological results of an expedition to Perak and the Siamese Malay States. 1901–1902, undertaken by Nelson Annandale and Herbert C. Robinson. Vol. 1, zoology*. Liverpool, UK: University Press, 131–176.
- Brown JH, Lomolino MV. 1998.** *Biogeography*, 2nd edn. Sunderland, MA: Sinauer.
- Brown RM, Diesmos AC. 2009.** Philippines biology. In: Gillespie R, Clague D, eds. *Encyclopedia of islands*. Berkeley, CA: University of California Press, 723–732.
- Brown RM, Diesmos AC, Alcalá AC. 2008.** Philippine amphibian biodiversity is increasing in leaps and bounds. In: Stuart SN, Hoffmann M, Chanson JS, Cox NA, Berridge R, Ramani P, Young BE, eds. *Threatened amphibians of the world*. Barcelona: Lynx Ediciones, with IUCN – The World Conservation Union, Conservation International, and NatureServe, 82–83.
- Brown RM, Stuart BL.** In press. *Patterns of biodiversity discovery through time: an historical analysis of defining*

- amphibian species boundaries in the Southeast Asian mainland and island archipelagos*. Invited chapter contribution, Southeast Asian Gateway Evolution, Royal Holloway, Egham, UK: University of London.
- Brühl CA, Eltz T. 2009.** Fuelling the biodiversity crisis: species loss of ground-dwelling forest ants in oil palm plantations in Sabah, Malaysia (Borneo). *Biodiversity Conservation* **2**: 519–529.
- Buhlmann KA, Akre TSB, Iverson JB, Karapatakis D, Mittermeier RA, Georges A, Rhodin AGJ, van Dijk PP, Gibbons JW. 2009.** A global analysis of tortoise and freshwater turtle distributions with identification of priority conservation areas. *Chelonian Conservation Biology* **8**: 116–149.
- Cheung SM, Dudgeon D. 2006.** Quantifying the Asian turtle crisis: market surveys in southern China, 2000–2003. *Aquatic Conservation: Marine and Freshwater Ecosystems* **16**: 751–770.
- Danielsen F, Beukema H, Burgess ND, Parish F, Brühl CA, Donald PF, Murdiyarto D, Phalan B, Reijnders L, Struebig M, Fitzherbert EB. 2009.** Biofuel plantations on forested lands: double jeopardy for biodiversity and climate. *Conservation Biology* **23**: 348–358.
- Daugherty CH, Cree A, Hay JM, Thompson MB. 1990.** Neglected taxonomy and continuing extinctions of tuatara (*Sphenodon*). *Nature* **347**: 177–179.
- Diesmos AC, Parham JF, Stuart BL, Brown RM. 2005.** The phylogenetic position of the recently rediscovered Philippine forest turtle (Bataguridae: *Heosemys leytensis*). *Proceedings of the California Academy of Sciences* **56**: 31–41.
- van Dijk PP. 2000.** The status of turtles in Asia. In: van Dijk PP, Stuart BL, Rhodin A, eds. 2000. Asian turtle trade: proceedings of a workshop on conservation and trade of freshwater turtles and tortoises in Asia. Phnom Penh, Cambodia, December 1999. Chelonian Research Monographs 2. Lunenburg, MA: Chelonian Research Foundation, 39–44.
- van Dijk PP, Stuart BL, Rhodin AGJ, eds. 2000.** Asian turtle trade proceedings of a workshop on conservation and trade of freshwater turtles and tortoises in Asia Phnom Penh, Cambodia, 1–4 December 1999. Chelonian Research Monographs 2. Lunenburg, MA: Chelonian Research Foundation.
- Dmitriev DA, Rakitov RA. 2008.** Decoding of superimposed traces produced by direct sequencing of heterozygous indels. *PLoS Computational Biology* **4**: e1000113. doi:10.1371/journal.pcbi.1000113.
- Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A. 2010.** *Geneious v5.3*. Available at: <http://www.geneious.com>
- Edgar RC. 2004.** MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.
- Fong JJ, Parham JF, Stuart BL, Carter RL. 2007.** A genetic survey of heavily exploited, endangered turtles: caveats on the conservation value of trade animals. *Animal Conservation* **10**: 452–460.
- Fritz U, Guicking D, Auer M, Sommer RS, Wink M, Hundsdoerfer AK. 2008.** Diversity of the Southeast Asian leaf turtle genus *Cyclemys*: how many leaves on its tree of life? *Zoologica Scripta* **37**: 367–390.
- Fujita MK, Engstrom TN, Starkey DE, Shaffer HB. 2004.** Turtle phylogeny: insights from a novel nuclear intron. *Molecular Phylogenetics and Evolution* **31**: 1031–1040.
- Gong SP, Chow AT, Fong JJ, Shi H. 2009.** The chelonian trade in the largest pet market in China: scale, scope and impact on turtle conservation. *Oryx* **43**: 213–216.
- Gray JE. 1831.** *Synopsis reptilium or short descriptions of the species of reptiles. Part 1. Cataphracts, tortoises, crocodiles, and enaliosaurians*. London, UK: Treuttel, Wurtz & Co.
- Gray JE. 1860.** On some new species of mammalia and tortoises from Cambodia. *Annals and Magazine of Natural History, London* **2**: 217–218.
- Guindon S, Gascuel O. 2003.** A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**: 696–704.
- Hey J, Waples RS, Arnold ML, Butlin RK, Harrison RG. 2003.** Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology and Evolution* **18**: 597–603.
- Hirayama R. 1984.** Cladistic analysis of batagurine turtles (Batagurinae: Emydinae: Testudinoidea): a preliminary result. *Studia Geologica Salmanticensia, vol Especial 1. Studia Palaeocheloniologica* **1**: 141–157.
- Hudson R, Buhlmann K. 2002.** Turtle rescue – turtle survival alliance executive summary. *Turtle and Tortoise Newsletter* **6**: 12–14.
- Hudson R, Coyne JA. 2002.** Mathematical consequences of the genealogical species concept. *Evolution* **56**: 1557–1565.
- Huelsenbeck JP, Ronquist F. 2001.** MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.
- Inger RF, Stuart BL, Iskandar DT. 2009.** Systematics of a widespread Southeast Asian frog, *Rana chalconota* (Amphibia: Anura: Ranidae). *Zoological Journal of the Linnean Society* **155**: 123–147.
- IUCN. 2011.** IUCN red list of threatened species. Version 2011.1. Available at: <http://www.iucnredlist.org> Downloaded on 16 June 2011.
- Ives I, Spinks PQ, Shaffer HB. 2007.** Morphological and genetic variation in the endangered Sulawesi tortoise *Indotestudo forstenii*: evidence of distinct lineages? *Conservation Genetics* **9**: 709–713.
- Kimball RT, Braun EL, Barker FK, Bowie RCK, Braun MJ, Chojnowski JL, Hackett SJ, Han KL, Harshman J, Heimer-Torres V, Holznagel W, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Reddy S, Sheldon FH, Smith JV, Witt CC, Yuri T. 2009.** A well-tested set of primers to amplify regions spread across the avian genome. *Molecular Phylogenetics and Evolution* **30**: 654–660.
- Knowles LL, Carstens BC. 2007.** Delimiting species without monophyletic gene trees. *Systematic Biology* **56**: 887–895.
- Leaché AD, Fujita MK. 2010.** Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). *Proceedings of the Royal Society of London B* **277**: 3071–3077.

- Lemmon AR, Brown JM, Stanger-Hall K, Moriarty Lemmon E. 2009.** The effect of ambiguous data on phylogenetic estimates obtained by maximum likelihood and bayesian inference. *Systematic Biology* **58**: 130–145.
- Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB, Kent J. 2000.** Biodiversity hotspots for conservation priorities. *Nature* **403**: 853–858.
- 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, allozyme electrophoresis and known-locality specimens. *Animal Conservation* **4**: 357–367.
- Posada D. 2008.** jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**: 1253–1256.
- Primmer CR, Borge T, Lindell J, Sætre GP. 2002.** Single-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>
- Rannala B, Yang Z. 2003.** Bayes estimation of species 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.
- Sambrook J, Russell DW. 2001.** *Molecular cloning: a laboratory manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Shaffer HB, Thomson RC. 2007.** Delimiting species in recent radiations. *Systematic Biology* **56**: 896–906.
- Shi H, Parham JF, Lau M, Tien-Hsi C. 2007.** Farming endangered turtles to extinction in China. *Conservation Biology* **21**: 5–6.
- 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 Peninsular Malaysia. *Biology Conservation* **129**: 149–166.
- Spinks PQ, Shaffer HB. 2007.** Conservation phylogenetics of the Asian box turtles (Geoemydidae, *Cuora*): mitochondrial introgression, numts, and inferences from multiple nuclear loci. *Conservation Genetics* **8**: 641–657.
- Spinks PQ, Shaffer HB, Iverson JB, McCord WP. 2004.** Phylogenetic hypotheses for the turtle family Geoemydidae. *Molecular Phylogenetics and Evolution* **32**: 164–182.
- Spinks PQ, Thomson RC, Barley AJ, Newman CE, Shaffer HB. 2010.** Testing avian, squamate, and mammalian 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 mixed models. *Bioinformatics* **22**: 2688–2690.
- Stamatakis A. 2006b.** Phylogenetic models of rate heterogeneity: a high performance computing perspective. Proceedings of 20th IEEE/ACM International Parallel and Distributed Processing Symposium (IPDPS2006), High Performance Computational Biology Workshop, Proceedings on CD, Rhodes, Greece.
- Stephens M, Donnelly P. 2003.** A comparison of Bayesian methods for haplotype reconstruction. *American Journal of Human Genetics* **73**: 1162–1169.
- Stephens M, Smith NJ, Donnelly P. 2001.** A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* **68**: 978–989.
- 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, Inger RF, Voris HK. 2006.** High level of cryptic species diversity revealed by sympatric lineages of Southeast Asian forest frogs. *Biology Letters* **2**: 470–474.
- Stuart BL, Parham JF. 2004.** Molecular phylogeny of the critically endangered Indochinese box turtle (*Cuora galbinifrons*). *Molecular Phylogenetics and Evolution* **31**: 164–177.
- 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 Phylogenetics and Evolution* **49**: 514–525.
- Townsend TM, Alegre RE, Kelley ST, Weins JJ, Reeder TW. 2007.** Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: an example from squamate reptiles. *Molecular Phylogenetics and Evolution* **47**: 129–142.
- Vetter H, van Dijk PP. 2006.** *Terralog, turtles of the world, Vol. 4. East and South Asia*. Frankfurt am Main and Rodgau: Chimaira and Aqualog.
- Wink M, Guicking D, Fritz U. 2000.** Molecular evidence for 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.

SUPPORTING INFORMATION

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.