MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 956–967 doi: 10.1111/j.1755-0998.2011.03032.x

Comparing and combining distance-based and character-based approaches for barcoding turtles

B. N. REID,* M. LE,†‡§ W. P. McCORD,¶ J. B. IVERSON,** A. GEORGES,†† T. BERGMANN,‡‡ G. AMATO,§§ R. DESALLE§§ and E. NARO-MACIEL§§¶¶
*Department of Forest and Wildlife Ecology, University of Wisconsin, 1630 Linden Drive, Madison, WI 53706, USA, †Center for Natural Resources and Environmental Studies, Vietnam National University, 19 Le Thanh Tong Street, Hanoi, Vietnam, ‡Faculty of Environmental Sciences, Hanoi University of Science, 334 Nguyen Trai Road, Hanoi, Vietnam, §Department of Herpetology, American Museum of Natural History, New York, NY 10024, USA, ¶East Fishkill Animal Hospital, 455, Route 82, Hopewell Junction, NY 12533, USA, **Department of Biology, Earlham College, Richmond, IN 47374, USA, ††Institute for Applied Ecology, University of Canberra, Canberra, ACT 2601, Australia, ‡‡Institute for Animal Ecology and Evolution, Stiftung Tierärztliche Hochschule Hannover, Hannover 30559, Germany, §§Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, NY 10024, USA, ¶¶Biology Department, College of Staten Island, City University of New York, Staten Island, NY 10314, USA*

Abstract

Molecular barcoding can serve as a powerful tool in wildlife forensics and may prove to be a vital aid in conserving organisms that are threatened by illegal wildlife trade, such as turtles (Order Testudines). We produced cytochrome oxidase subunit one (COI) sequences (650 bp) for 174 turtle species and combined these with publicly available sequences for 50 species to produce a data set representative of the breadth of the order. Variability within the barcode region was assessed, and the utility of both distance-based and character-based methods for species identification was evaluated. For species in which genetic material from more than one individual was available (n = 69), intraspecific divergences were 1.3% on average, although divergences greater than the customary 2% barcode threshold occurred within 15 species. High intraspecific divergences could indicate species with a high degree of internal genetic structure or possibly even cryptic species, although introgression is also probable in some of these taxa. Divergences between species of the same genus were 6.4% on average; however, 49 species were <2% divergent from congeners. Low levels of interspecific divergence could be caused by recent evolutionary radiations coupled with the low rates of mtDNA evolution previously observed in turtles. Complementing distance-based barcoding with character-based methods for identifying diagnostic sets of nucleotides provided better resolution in several cases where distance-based methods failed to distinguish species. An online identification engine was created to provide character-based identifications. This study constitutes the first comprehensive barcoding effort for this seriously threatened order.

Keywords: barcoding, Characteristic Attribute Organization System, species identification, turtles

Received 23 December 2010; revision received 14 April 2011; accepted 22 April 2011

Introduction

Turtles (order Testudines) are highly endangered as a group, with 42% of extant species classified as threatened and 10% classified as critically endangered by the IUCN (Buhlmann *et al.* 2009). Turtles face a similar battery of threats compared with other endangered taxa, including the effects of habitat loss, invasive species, pollution, disease and climate change; however, human overexploitation represents an especially acute threat to the survival of most threatened turtle species (van Dijk *et al.* 2000; Gibbons *et al.* 2000). The turtle trade is at its most intense in China and Southeast Asia, where over 10 million individuals per year are traded as meat, pets or ingredients in traditional remedies (Turtle Conservation Fund 2002). It is important to note, however, that the Asian turtle market handles species from around the world (Cheung & Dudgeon 2006; Nijman & Shepherd 2007), with globalization of trade increasing as native Asian species become increasingly scarce.

The forensic applications of DNA barcoding have great potential as a means for quantifying and regulating trade in endangered turtle species (Ogden *et al.* 2009;

Correspondence: Brendan N. Reid, Fax: (608) 262 9922;
E-mail: bnreid@wisc.edu

Alacs *et al.* 2010). Previous studies have shown that, given a comprehensive sequence database, COI can serve as a reliable forensic marker for identifying unknown zoological material to the species level (Dawnay *et al.* 2007). The forensic applications proposed for barcoding run the gamut from identifying fish species in commercial markets (Costa & Carvalho 2007) to investigating bird–airplane collisions (Dove *et al.* 2008). Recently, barcoding has been shown to be a reliable means of identifying material in the bushmeat trade (Eaton *et al.* 2010). Despite the promise of utilizing DNA barcoding as a tool for their conservation, turtles have been underrepresented in the global barcoding effort. Prior to the initiation of this research, sequences from only 52 species had been deposited in the Barcode of Life Datasystems database (BOLD, accessed 26 February 2009), and the species barcoded were also heavily skewed towards Asian pond turtles (family Geoemydidae) and tortoises (family Testudinidae). Turtles therefore represented a significant gap in the barcode catalogue that we intended to fill.

This report provides novel COI barcode sequences for 174 turtle species. The species barcoded here were chosen because they either appear on the IUCN Red List, indicating that they are species of conservation concern which would probably benefit from the forensic applications of barcoding, or because they belong to clades that are underrepresented within the Testudines with regard to previous barcoding efforts. Publicly available sequences as well as sequences for sea turtles produced in a previous study (Naro-Maciel *et al.* 2010) were added to these novel sequences to better evaluate variability and identification success across the entire order. Distance-based (Hebert *et al.* 2003, 2004) and character-based approaches to barcoding (DeSalle *et al.* 2005; Kelly *et al.* 2007) were both evaluated to determine the effectiveness in distinguishing turtle species. While application of the barcode information gleaned here to quantifying or controlling the wildlife trade is beyond the scope of this report, this information represents a potentially powerful tool for combating the anthropogenic challenges currently faced by turtles on the global scale.

Materials and methods

Taxonomy, sample selection and acquisition

A list of all turtle species on the IUCN Red List (in every category except for 'Extinct') was compiled (IUCN 2009) and cross-referenced against a list of turtle species already present in the BOLD database to produce a master list of red-listed species without barcodes. The IUCN's taxonomic designations were checked against the most widely accepted account of turtle taxonomy (Turtle Taxonomic Working Group 2007) at the time of compilation and revised accordingly. The taxonomy used in this work does not account for several very recent changes in nomenclature (such as the reorganization of several chelid species into the new genus *Myuchelys*; Georges & Thomson 2009). When several alternate genera were listed for a species, the species was assigned to a genus in a way that minimized the total number of genera under consideration. Non-IUCN-listed species from two turtle families (Chelidae and Pelomedusidae) that were underrepresented in the BOLD database were also added to the master list.

Species on this master list that were already represented in the American Museum of Natural History (AMNH)'s collection, either as extracted DNA or frozen tissue, were obtained directly from the museum. Availability of the remaining species was determined by querying the Association of Zoos and Aquariums (AZA)'s zoo holdings database, ISIS (http://www.isis.org) and the museum herpetological collections database HerpNET (http://herpnet.org). Once sources were identified, blood or tissue samples were obtained from a collaborating zoo, museum, university or from the authors' (Georges, Iverson, McCord) collections. In cases where species were protected by national law or listed under one of the appendices of the Convention on International Trade in Endangered Species, care was taken to obtain all relevant permits and observe applicable regulations for the collection of samples and transfer of specimens between institutions. When possible, aliquots of blood or tissue samples obtained from private collections have been deposited into the Ambrose Monell Cryo Collection (AMCC) at the AMNH for future reference. Owing to the nature of the sampling, original collection locality information was unavailable for many samples, including samples obtained from zoo animals and specimens obtained from the pet trade. Where available, voucher numbers and locality information have been uploaded as annotation to the Genbank and BOLD records for the novel sequences presented in this study.

DNA extraction and sequencing

DNA was extracted from blood or tissue using a DNeasy Tissue kit (QIAGEN Inc., Valencia, CA, USA). The COI barcode region was amplified from most species using either turtle-specific or universal primers from previous studies or primers designed in the course of this study (Table 1). PCR conditions for all primer sets except the universal COI-3 primer cocktail were as follows: 95 °C for 5 m; 35 cycles of 95 °C for 45 s, 54 °C for 45 s, 72 °C for 45 s; 72 °C for 6 m; 4 °C indefinitely. PCR for the COI-3 primer cocktail (utilizing primers VF2_t1, FishF2_t1, FishR2_t1 and FR1d_t1) was run according to Ivanova *et al.* 2007 (94 °C for 2 min; 35 cycles of 94 °C for

Table 1 Primers used in this study. 5′ positions are relative to the published mitochondrial sequence for *Chrysemys picta*

Primer name	Sequence	Reference	5′ position
L-turtCOI	5′-ACTCAGCCATCTTACCTGTGATT-3′	Stuart and Parham 2004	5384
L-turtCOIc	5′-TACCTGTGATTTTAACCCGTTGAT-3′	Stuart and Parham 2004	5396
H-turtCOIb	5′-GTTGCAGATGTAAAATAGGCTCG-3′	Stuart and Parham 2004	6327
H-turtCOIc	5′-TGGTGGGCTCATACAATAAAGC-3′	Stuart and Parham 2004	6273
LCO1490	5′-GGTCAACAAATCATAAAGATATTGG-3′	Folmer *et al*. 1994	5423
HCO2198	5′-TAAACTTCAGGGTGACCAAAAAATCA-3′	Folmer *et al*. 1994	6132
VF2_t1	5′-TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC-3′	Ward *et al*. 2005	5426*
FishF2_t1	5′-TGTAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCAC-3′	Ward *et al*. 2005	5426*
FishR2_t1	5′-CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA-3′	Ward *et al*. 2005	6129*
FR1d_t1	5′-CAGGAAACAGCTATGACACCTCAGGGTGTCCGAARAATCARAA-5′	Ivanova *et al*. 2007	6129*
M13F	5′-TGTAAAACGACGGCCCAGT-3′	Messing 1983	n/a
M13R	5′-CAGGAAACAGCTATGAC-3′	Messing 1983	n/a
HturtCOIk[a]	5′-GGTGGGCTCATACAATAAAACC-3′	This study	6272
LturtCOIk[a]	5′-CTACTAACCATAAAGACATCGGTACCC-3′	This study	5426
HturtCOIa[b]	5′-CATACAATGAATCCCAGGAATCCGAT-3′	This study	6264
LturtCOIa[b]	5′-CGCTGACTATTTTCTACTAATC-3′	This study	5413
Fbat2[b]	5′-CTACTAATCATAAAGACATTGG-3′	This study	5426
Rbat1[b]	5′-TAGGCAACTACGTGTGAGATTAT-3′	This study	6180
Fpodo1[c]	5′-CAAACCATAAAGATATTGGCACCC-3′	This study	5429
Rpodo1[c]	5′-GATATTATTGCTCATACTATTCC-3′	This study	6237
Fpelu1[d]	5′-CCCGTTGATTATTCTCCACTAACC-3′	This study	5411
Rpelu1[d]	5′-GATGCTATGGCTCAAACTATTCC-3′	This study	6237
Fpyx1[e]	5′-CTCTACTAACCATAAAGATAT-3′	This study	5424

*Excluding engineered 5′ M13 sequence.
Novel primers with superscript annotations were used for amplifying several species from these specific families: (a) Kinosternidae. (b) Chelidae. (c) Podocnemididae. (d) Pelomedusidae. (e) Testudinidae.

30 s, 52 °C for 40 s and 72 °C for 1 min; 72 °C for 10 min; 4 °C indefinitely). PCR products were cleaned on a BIOMEK automated apparatus using the Ampure system. Cycle sequencing was performed using BigDye reagents (Perkin Elmer, Waltham, MA, USA). Both strands of all PCR products were sequenced with the same primers and used to amplify the products except in the case of COI-3 primer cocktail products, which were sequenced using the M13F and M13R primers. Cycle sequencing PCR was run as follows: 96 °C for 5 m; 35 cycles of 94 °C for 15 s, 50 °C for 15 s, 60 °C for 4 m; 4 °C indefinitely. Cycle sequencing products were ethanol-precipitated and run on an ABI3770 automated sequencer (Applied Biosystems, Foster City, CA, USA).

Sequence variability and distance-based species identification

Novel sequences were assembled and edited in Sequencher (Gene Codes Corporation) and added to a set of publicly available sequences downloaded from BOLD. As nuclear paralogues (numts) have already been detected in several turtle species (Stuart & Parham 2004; Spinks & Shaffer 2007), all sequences were systematically screened to identify numts. Multiple primer pairs were used in most cases to increase the chance of amplifying the true mitochondrial sequence, and all suspected numts (sequences with premature stop codons or frameshift mutations) were expunged from the data set. Sequences were aligned in MEGA 4 (Tamura *et al*. 2007) and trimmed to a region 650 nucleotides in length. The fragment used here begins at base pair 62 of the complete COI sequence (base pair 5453 of the complete *Chrysemys picta* mitochondrial genome), with codon 22 in the translated COI amino sequence being the first complete codon in the fragment. These sites are designated as the first nucleotide and amino acid positions, respectively, in our data set.

Sequence composition and substitution pattern for the entire data set, the number of variable nucleotide and amino acid sites in the data set, and pairwise Kimura 2-parameter (K2P) sequence divergences within groups at multiple taxonomic levels (intraspecific, between species of the same genus and between species

of different genera in the same family) were calculated in MEGA 4. The K2P substitution model rather than a more realistic model was used to calculate distances to allow for repeatability of analyses through the BOLD engine and comparison with canonical distance-based barcoding studies (Hebert *et al.* 2003, 2004). The distribution of pairwise K2P values at each taxonomic level was visualized using a density plot in R (R Foundation for Statistical Computing, Vienna, Austria). Pearson product–moment correlations and Spearman rank correlations between sample size and mean intraspecific distance were also calculated in R to determine whether the number of available samples affected estimates of intraspecific distance.

Two neighbour-joining trees, one for pleurodiran species (side-necked turtles) and one for cryptodiran species (all other turtles), were constructed in MEGA 4 strictly to allow for the visualization of K2P distances for all novel sequences produced in this study. Trees were displayed using the Interactive Tree of Life web service (http://itol.embl.de; Letunic & Bork 2006). Previously published sequences were excluded from these trees because of space considerations. Species were organized into one of four categories (after Hebert *et al.* 2004) based on pairwise K2P distances. The categories used were as follows: Category I (maximum intraspecific distance <2%, minimum interspecific distance >2%), Category II (maximum intraspecific distance ≥2%, minimum interspecific distance >2%), Category III (maximum intraspecific distance <2%, minimum interspecific distance ≤2%) and Category IV (maximum intraspecific distance ≥2%, minimum interspecific distance ≤2%). In species where only one individual was sampled, categories I and II and categories III and IV were conflated as only interspecific distances could be measured.

Character-based analysis and online identification engine

Pure unique identifying characters, defined here as single-nucleotide states that distinguish a species from others in its family, were determined for each family using the Characteristic Attribute Organization System (CAOS; Sarkar *et al.* 2002, 2008; Bergmann *et al.* 2009). When all members of a species share these characters, they are termed 'simple pure characters' (*sensu* Sarkar *et al.* 2002). Characters were identified at the family level to correspond with the previous studies (Kelly *et al.* 2007; Rach *et al.* 2008; Damm *et al.* 2010; Naro-Maciel *et al.* 2010; Yassin *et al.* 2010). A guide tree was first produced using the maximum parsimony module in Phylip (v3.67; Felsenstein 1989) and modified to group individual samples according to current species designations (Turtle Taxonomic Working Group 2007). This guide tree was then incorporated into a NEXUS file containing COI sequence data in MacClade (v4.06; Maddison & Maddison 2000), and the p-gnome script (Rach *et al.* 2008; Sarkar *et al.* 2008) was used to identify characters. The proportion of all species exhibiting within-family identifying characters, as well as the proportion in each family, was calculated. Finally, the number of species exhibiting within-family characters for each of the distance-based categories was evaluated.

An online identification engine ('Project Turtle' in the Ruby-CAOS website, http://boli.uvm.edu/caos-workbench/htdocs/caos.php) was designed to allow for the implementation of the character-based identification method in a manner similar to the user-friendly BOLD interface for distance data. Sequences supplied to the website are first assigned to a family, after which the CAOS-Classifier script in RubyCAOS is employed to establish species identity using the family-level characters described here. If a positive identification is made, the site provides a link to the species description in the Turtles of the World database (http://nlbif.eti.uva.nl/bis/turtles.php); if no identification is possible, a list of possible species is provided.

Results

Taxonomic range and Red List coverage

Information for the taxa included in this study is given in Table S1 (Supporting information). Overall, 220 species from all 14 chelonian families (four of which had no representation in the barcode database before) are represented in the final data set. Of the 204 valid, extant turtle species on the Red List, 35 (17%) had been previously barcoded and another 149 (73%) were barcoded in this study. Owing to the rarity of many of these turtles, multiple samples were not available for all species; however, two or more sequences were available from 69 of the species included in this study.

Barcode fragment variability and distance-based species identification

Approximately half of the nucleotide positions (51.8%) were variable across the data set. Nucleotide composition showed a bias against G consistent with that observed previously in turtles (Spinks *et al.* 2004), and transitions were more frequent than transversions. Approximately two-fifths (40.7%) of amino acid positions were variable (Table 2).

Mean intraspecies K2P divergence across 1403 possible pairwise combinations was 1.3% (Fig. 1). Variance was high, however [standard deviation (SD) = 2.2%], and pairwise intraspecific distances >2% were observed in 15

of the 69 species with $n > 2$. The Pearson and Spearman tests for correlation between sample size and intraspecific divergence gave conflicting results (Pearson's $r = 0.01$, $P = 0.91$; Spearman's rho = 0.26, $P = 0.029$). This indicates a positive relationship between relative (but not absolute) sample size and intraspecific divergence, meaning that although intraspecific distances may be somewhat underestimated in undersampled species there is no linear relationship between sample size and divergence. Mean pairwise divergence between congeneric individuals was 6.4% (SD = 2.6%, Fig. 1). Pairwise K2P differences of <2% were observed between 49 species. Mean intrafamily divergence was 13.6% (SD = 4.3%, Fig. 1). All sequences were uploaded to BOLD and analysed using the BOLD interface, yielding similar results in all cases.

Genus and species groupings for novel sequences on the distance-based trees (Fig. 2) were broadly congruent with the accepted taxonomy (although some accepted genera and species were not monophyletic on the tree). Very low levels of divergence (<1%) were apparent between certain species in some genera (*Elseya*, *Pseudemys*, *Graptemys*, *Trachemys*, *Kinosternon*, *Mesoclemmys*), while very high levels of intraspecies divergence (>4%) were observed in five species (*Kinosternon integrum*, *Elseya novaeguineae*, *Emydura subglobosa*, *Acanthochelys radiolata* and *Amyda cartilaginea*). For species with multiple samples, 43 (62%) were placed in Category I, 9 (13%) were placed in Category II, 11 (16%) were placed in Category III and 6 (9%) were placed in Category IV. For species with one sample, 119 (79%) were placed in Category I/II and 32 (21%) were placed in Category III/IV (Fig. 3).

Table 2 Nucleotide substitution pattern, nucleotide frequencies, and nucleotide and amino acid variability as estimated in MEGA 4. Transitions rates are in bold, while transversion rates are italicized

Maximum composite likelihood estimate of substitution pattern

	A	T	C	G
A	–	*4.58*	*4.37*	**7.58**
T	*4.58*	–	**23**	*2.74*
C	*4.58*	**24.16**	–	*2.74*
G	**12.74**	*4.57*	*4.36*	–

Nucleotide frequencies

A	0.281
T	0.282
C	0.268
G	0.168

Proportion of sites variable

	Variable	Total	% Variable
Nucleotide	337	650	52
Amino acid	88	216	41

Character-based identification

Characteristic Attribute Organization System analysis produced sets of simple identifying characters capable of distinguishing species from all others in their respective families for 155 of the 218 species (71%) in nonmonotypic families. The proportion of species in a given family possessing simple diagnostic traits (Fig. 4) varied from 100% (Cheloniidae, Chelydridae, Pelomedusidae, Podocnemididae) to lower than 60% (Emydidae, Geoemydidae). Example sets of simple identifying characters (in which some characters identified by CAOS are excluded for reasons of space) are shown for the families Podocnemididae (Table 3a) and Trionychidae (Table 3b). Identifying characters could be found in 130 of the 162 species (80%) successfully distinguished by a distance-based threshold (i.e. species in categories I or I/II). Identifying characters were found for 23 of 58 species (40%) in which classification by a distance threshold failed (i.e. species in Categories II, III, III/IV or IV) (Fig. 3).

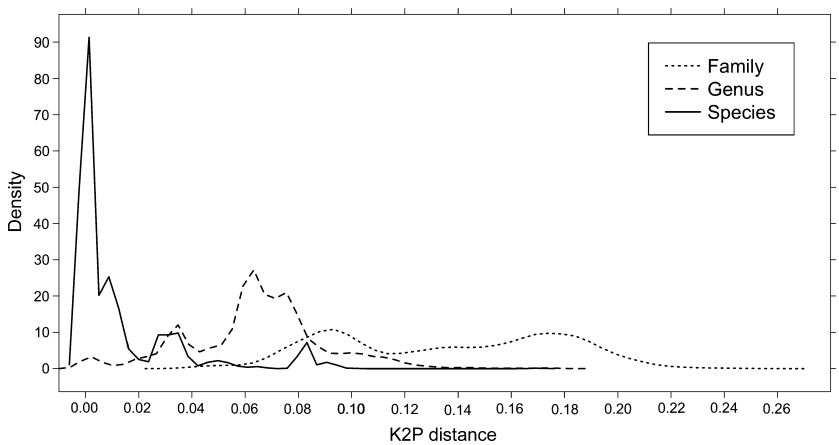

Fig. 1 Density plot of Kimura 2-parameter (K2P) divergences within each taxonomic level.

Fig. 2 Neighbour-joining trees of COI sequences produced in this study, organized by suborder. (a) Pleurodires. (b) Cryptodires.

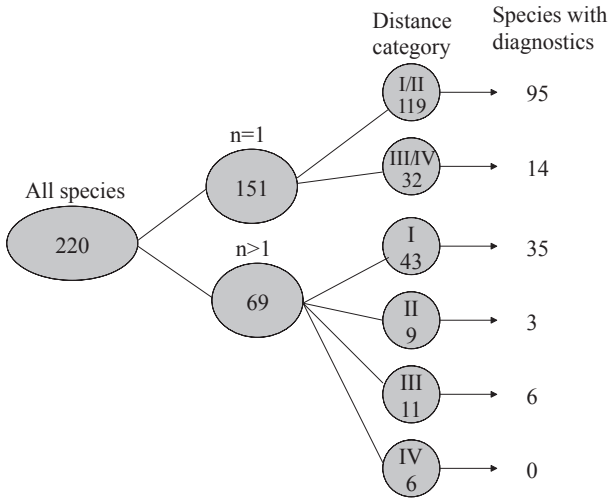

Fig. 3 Number of species in each distance category that exhibit identifying characters at the family level.

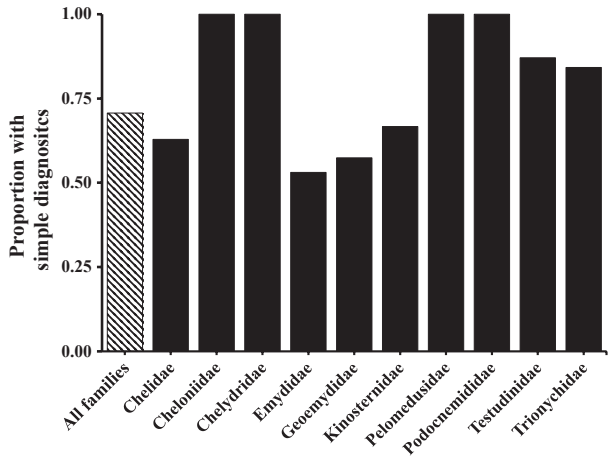

Fig. 4 Proportion of species in the total data set and in each family with identifying characters capable of distinguishing a given species from all others in its family.

Discussion

The barcode sequences assembled here provide a potentially crucial resource for turtle conservation. Barcode records previously existed for only about 50 species; this study more than quadruples that number, allowing approximately two-thirds of extant species to be identified using molecular means and adding entire families to the barcode database that was previously missing. Over the course of the barcoding process, apparent genetic structure was identified in several poorly studied groups, indicating the possible existence of evolutionarily significant units within these putative species that merit further study and possibly extra consideration in conservation efforts. This study also compares distance-based and character-based methods for species identification, and by combining the two highlights a 'third way' for DNA barcoding that may be useful in improving identification efficiency in taxa for which neither distance nor characters are a perfect fit.

While members of the barcoding community have advanced several different methods of distinguishing species using COI sequence information, the distance-based method advanced by Hebert *et al.* (2003) has become and in all probability will remain the standard, workhorse method used in DNA barcoding. Distance-based barcoding uses a 2% divergence (K2P > 0.02) cut-off for vertebrates to determine species identity, implying that individuals should be <2% divergent from members of their own species and more than 2% divergent from members of other species. A maximum of 161 turtle species examined in this study (73%) can be effectively distinguished using this criterion. This is probably an overestimate, as (i) undetected intraspecific divergences >2% may exist in undersampled species and (ii) all closely related species were not sampled for the species examined, leaving open the possibility that some unsampled species could be <2% divergent from the species examined here. In the group of species with more than one individual sampled, the intraspecific divergence criterion was violated about as many times as the interspecific divergence criterion (nine species in Category II vs. 11 species in Category III). As such, raising or lowering the divergence cut-off would probably do little to improve the proportion of species successfully distinguished by a distance-based method.

Species in Category II (high intraspecies divergence) have been targeted as probably examples of cryptic diversity (Hebert *et al.* 2004). Although many of the species identified in this category are rare and/or poorly studied, some evidence points to the existence of cryptic variability within several species. *Elseya novaeguineae*, for example, is regarded as a probably species complex (Georges & Thomson 2009), and the individuals barcoded here fall into three distinct clusters based on COI sequence. *Erymnochelys madagascariensis*, another species that is thought to contain multiple population units (Rafeliarisoa *et al.* 2006), also violated the 2% threshold. In the case of the relatively well-studied species *Cuora galbinifrons*, intraspecific divergences of >2% in the publicly available COI sequences do indeed map to three distinct clades which Stuart & Parham (2004) argued should be granted full species status based on genetic and morphological divergences. This example from the public data seems to support the possibility that these high intraspecific divergences may represent cryptic diversity. However, the controversy surrounding these designations (Turtle Taxonomic Working Group 2007), and

Table 3 Example sets of identifying characters for (a) Podocnemididae and (b) Trionychidae. Simple identifying characters are shaded. Characters providing diagnostic information via the heuristic discussed in the text are boxed.

(a)

	80	89	158	263	308	323	350	368	410	479	527	530	542	545	560
Erymnochelys madagascariensis	A	C	C	C	T	A	A	A	A	A	A	A	T	A	A
Peltocephalus dumerilianus	A	C	G	T	A	C	C	C	A	C	G	C	C	C	T
Podocnemis erythrocephala	T	C	A	A	T	A	C	T	T	C	C	A	C	A	A
Podocnemis expansa	T	A	A	C	T	G	T	T	G	T	C	A	C	A	G
Podocnemis lewyana	C	C	T	T	T	A	C	T	C	C	C	A	A	A	A
Podocnemis sextuberculata	A	T	T	C	T	A	C	T	T	C	C	G	C	A	A
Podocnemis unifilis	G	C	A	C	T	A	C	T	T	C	C	A	C	A	A
Podocnemis vogli	T	C	A	C	C	A	C	T	A	C	C	A	C	T	A

(b)

	5	26	121	218	281	290	323	350	512	521	527	536	545	551	614
Amyda cartilaginea	C	A	T	A	A/T	A	C/T	C	A	A	A	A/T	A	A	T
Chitra chitra	C	A	T	A	T	A	C	C	A	A	A	A	T	A	A
Chitra indica	C	A	T	A	A	A	C	C	G	A	A	A	T	G	A
Cyclanorbis elegans	C	A	T	T	A	G	T	C	T	C	A	A	C	A	G
Cyclanorbis senegalensis	C	A	T	T	A	A	T	A	T	T	G	A	A	T	C
Cycloderma frenatum	C	A	T	T	A	A	C	C	T	A	C	A	A	C	A
Dogania subplana	T	A	T	T	A	A	C	C	A	A	A	A	A	A	C
Lissemys punctata	C	A	T	A	A	T	T	C	C	C	A	A	A	C	A
Lissemys scuttata	C	C	T	A	A	A	T	C	C	C	A	A	A	C	A
Nilssonia formosa	C	A	T	A	G	A	C	C	A	A	A	A	A	A	T
Nilssonia gangeticus	C	A	T	A	A	A	T	T	A	A	A	A	A	A	T
Nilssonia hurum	C	A	T	G	A	A	C	C	A	A	A	A	A	A	T
Palea steindachneri	C	A	T	A	A	C	T	C	A	A	A	G	A	A	T
Pelochelys bibroni	C	A	T	A	A	A	A	C	C	A	A	A	T	A	A
Pelochelys cantori	C	A	T	A	A	A	G	C	C	A	A	A	T	A	A
Pelodiscus sinensis	C	A	C	A	T	C	C	C	A	A	A	A	A	G	T
Rafetus euphraticus	C	A	T	A	A	A	T	C	A	A	A	T	A	G	A
Rafetus swinhoei	C	A	T	A	A	A	T	C	A	G	A	A	A	G	A
Trionyx triunguis	C	A	T	C	A	A	T	C	C	A	A	C	A	A	C

indeed species delimitation based on mitochondrial data alone (Georges & Thomson 2009), reinforces the need for further study including nuclear markers and morphological characteristics to determine the exact nature of this diversity. In some cases, patterns identified in COI match biogeographic patterns that have been documented in better-studied species, suggesting that similar evolutionary processes may have been at play in both. For example, *Kinosternon integrum* is broadly sympatric with the Central American iguanid species *Ctenosaura pectinata*, in which high levels of cryptic diversity as well as secondary contact between closely related species have produced patterns of mtDNA structuring (Zarza *et al.* 2008) similar to those noted here.

Observations of low interspecific differentiation (represented here by species in Category III) have been attributed to hybridization and resulting mitochondrial introgression between species, recent speciation or synonymy (Hebert *et al.* 2004). The frequency of low interspecific divergence in turtles can be attributed to several unique aspects of turtle biology. Evidence from marine turtles in the family Cheloniidae (Karl *et al.* 1995; Lara-Ruiz *et al.* 2006) indicates that some turtle species are still able to hybridize after tens of millions years of separation, and instances of intergenus hybridization have been recorded in other turtle families as well (Parham *et al.* 2001; Buskirk *et al.* 2005). Interspecies and even intergenus hybridization may then be possible, if not necessarily frequent, in the wild for many species. Low rates of both molecular evolution and chromosomal rearrangement in turtles (Bickham 1981; Avise *et al.* 1992) may make this hybridization possible by delaying the evolution of genetic barriers to reproduction. Slower rates of molecular evolution may themselves also be an explanation for low levels of differentiation in species that do not hybridize. Because mitochondrial genes tend to accumulate differences at a rate several-fold slower in turtles than in other vertebrates (Avise *et al.* 1992), species considered 'recent radiations' will probably be nearly identical at COI.

These alternate explanations can be evaluated for some of the well-studied species by using known species

ranges to rule out hybridization events. Most of the *Graptemys* species sequenced here are reciprocally allopatric and isolated in separate river drainages (Lamb *et al.* 1994). Only one species sequenced here (*G. gibbonsi*) has a range wide enough to overlap with those of other species (*G. oculifera* and *G. flavimaculata*), and *G. gibbonsi* is relatively well differentiated from these two species within the genus for the barcode fragment. As such, current hybridization is unlikely between the *Graptemys* species examined here. However, hybridization with the more widely distributed *Graptemys* species (*G. ouachitensis* and *G. pseudogeographica*) remains a possibility. Previous molecular work has identified strikingly low differentiation among *Graptemys* in a coding mitochondrial gene and attributed this to recent (<2.5 million years ago) speciation coupled with low rates of molecular evolution (Lamb *et al.* 1994). Similar explanations for low levels of diversification can be invoked for allopatric species in the recently diversified genera *Trachemys* and *Pseudemys*, although hybridization has been noted between *Pseudemys* species in rare cases (Crenshaw 1965). In the family Emydidae, therefore, slow molecular evolution and recent speciation certainly seem to be major causes of low interspecific diversity, although hybridization cannot be ruled out. However, little is known about divergence times or the likelihood of hybridization for other species exhibiting low levels of divergence, and further research will be necessary before these contributing causes can be fully evaluated.

Hebert *et al.* (2004) identified species in Category IV (high intraspecific divergence, low interspecific divergence) as probably examples of sample misidentification. This interpretation, however, assumes that introgression of mitochondrial haplotypes from species more than 2% divergent is either extremely unlikely or impossible. While this assumption may be valid in other taxa, it is demonstrably false for turtles. Several examples from the public data analysed here bear this out. For *Cuora trifasciata*, a species falling into Category IV in our analysis, introgression has produced several highly differentiated mitochondrial clades within the species, even though individuals form only one nuclear clade (Spinks and Shaffer 2007). Feldman & Parham (2004) hypothesize that introgression with *Mauremys annamensis* is a probably cause of high mitochondrial differentiation within another Category IV species in our analysis, *Mauremys mutica*, and hybridization has been recently noted between *Mauremys reevesi* and *Mauremys sinensis* (Fong & Chen 2010). As such, hybridization cannot be ruled out as an explanation for anomalous divergences within species sequenced in this study falling into Category IV (*Trachemys venusta* and *Emydura subglobosa*).

While distance-based barcoding will probably be effective in discriminating the majority of turtle species, this method seems to fail for a fairly large proportion of species. Character-based barcoding provides an attractive complement to distance-based barcoding, especially in turtles where interspecific divergences are probably to fall below the established threshold in closely related species. Relatively, few studies have been performed to date using character-based barcoding methods (Kelly *et al.* 2007; Rach *et al.* 2008; Damm *et al.* 2010; Naro-Maciel *et al.* 2010; Yassin *et al.* 2010). All have used the CAOS algorithm to determine characters that serve as unique species identifiers. This approach was shown to be more successful for differentiating 19 species within a mollusk genus *(Mopalia*) than distance-based barcoding (Kelly *et al.* 2007). A set of pure characters identified by CAOS, combined with several additional characters to form a compound character, was found to be effective for differentiating 54 of 64 species of Odonata (dragonflies and damselflies; Rach *et al.* 2008). The character-based approach had not previously been attempted on a set of species as large as the one examined in this study.

The efficacy of the simple characters identified by CAOS as species identifiers varied between families. The case of the Podocnemididae represents an extremely successful application of character-based barcoding; all species in the family are represented and each possessed simple identifying character states. Even in *Erymnochelys madagascariensis*, a species that displayed >2% intraspecies divergence, the diagnostic characters could unambiguously differentiate each individual in this species from those of other species. In the case of the Trionychidae, 16 of 19 species could be distinguished by simple characters. However, the remaining three species could be identified using the heuristic method of finding a character that unites them with a group containing only species with simple identifiers (all of which can then be distinguished by these characters). In larger families, the number of species for which characters could be found seemed to decline, possibly because of the increased likelihood of homoplasy and back mutations. As such, splitting families into smaller groups and considering compound characters could increase the success of a character-based method. However, a major caveat for all character-based analysis presented here is that, attributed to limited sample size, these character states may not be fixed.

For the species examined here, combining identifying characters with distance-based methods offers an effective means of increasing the proportion of species that can be successfully identified. Twenty-four species violating the distance threshold possessed identifying characters, meaning that incorporating these characters into the identification process would increase the total proportion of species identified by more than 10%. Identifying characters could be incorporated by a stepwise

process, as shown in Fig. 3, in which species are first identified according to distance-based criteria and then by using identifying characters if ambiguities still remain. The CAOS-based online identification engine described here provides a user-friendly means of carrying out the character-based portion of this approach. However, while characters may aid in species identification, they are not a perfect fix. Species that have extremely similar COI haplotypes, such as those in the genus *Graptemys*, often lacked identifying characters simply because of the lack of available variation in COI. Hybridization and introgression are also serious problems for any mitochondrial identification method. As such, identifying characters provided no resolution for species in Category IV (where introgression was probably an issue). Given the prevalence of introgression among turtle species, the use of a nuclear marker as a supplement to COI-based barcoding methods may be particularly valuable. Promising candidates for a nuclear barcode marker include the following: recombination activation gene 1 (RAG-1; Krenz *et al.* 2005) and the RNA fingerprint protein 35 intron (R35; Fujita *et al.* 2004). Many of the specimens used to generate the novel COI sequences included in this work are currently being sequenced for R35 and RAG-1 as part of separate phylogenetic studies focusing on particular taxa, including the Kinosternidae (Iverson JB, Le M in preparation) and the Australian Chelidae (Georges A, Reid BN, Zhang X, Charlton TR, McCord WP, Le M, in preparation); as such, the utility of both R35 and RAG-1 as complements to the COI-based barcoding presented here will be assessed in the near future.

While this study shows that accepted barcoding paradigms may be insufficient for species identification in some turtle groups, most species can be effectively discriminated by using a combination of existing methods. The existence of a genetic species identification method for turtles can assist in enforcement of existing laws regulating the traffic of turtles and turtle products and in characterizing the extent of trade in species, especially when these species are traded in otherwise unrecognizable forms. Barcoding could also have a number of possible uses in turtle ecology and conservation beyond its obvious utility in controlling wildlife trade. For example, barcoding of gut contents has been used to elucidate trophic interactions that are hard to observe otherwise (Zeale *et al.* 2011). With the addition of turtle sequences to the barcode database, these studies could detect depredation of turtle eggs, which is extremely high for many turtle species and constitutes one of the most important sources of mortality for a group that is otherwise superbly well armoured (Spencer & Thompson 2003). Turtles are in urgent need of protection, and the barcode sequences provided here will provide a useful tool for conservation and management.

Acknowledgments

The work described in this report was supported by a generous grant from the Alfred P. Sloan Foundation. TB was funded by H. Wilhelm Schaumann Stiftung (2009–2011) and DAAD (2010). JBI was funded by the Cope Fund through the Joseph Moore Museum of Natural History. The Sackler Institute for Comparative Genomics at the AMNH provided laboratory space, and Sergios-Orestis Kolokotronis provided valuable technical assistance. Blood and tissue samples were generously provided by the Ambrose Monell Cryo Collection at the AMNH, Dave Collins and the Tennessee Zoo, the Berkeley Museum of Vertebrate Zoology, the University of Canberra, the University of Michigan Museum of Zoology, Tag Engstrom and California State University/Chico, and Victoria Milne and the El Paso Zoo.

References

Alacs EA, Georges A, FitzSimmons NN, Robertson J (2010) DNA detective: a review of molecular approaches to wildlife forensics. *Forensic Science, Medicine and Pathology*, **6**, 180–194.

Amer SA, Kumazawa Y (2009) Complete sequence of the mitochondrial genome of the endangered Nile soft-shelled turtle *Trionyx triunguis*. *Egyptian Journal of Experimental Biology. Zoology*, **5**, 43–50.

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

Bergmann T, Hadrys H, Breves G, Schierwater B (2009) Character-based DNA barcoding: a superior tool for species classification. *Berliner und Münchener Tierärztliche Wochenschrift*, **122**, 446–450.

Bickham JW (1981) Two-hundred-million-year-old chromosomes: deceleration in the rate of karyotypic evolution in turtles. *Science*, **212**, 1291–1293.

Buhlmann KA, Akre TSB, Iverson JB *et al.* (2009) A global analysis of tortoise and freshwater turtle distributions with identification of priority conservation areas. *Chelonian Conservation Biology*, **8**, 116–149.

Buskirk JR, Parham JF, Feldman CR (2005) On the hybridization of two distantly related Asian turtles (*Sacalia* x *Mauremys*). *Salamandra*, **41**, 21–26.

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.

Costa FO, Carvalho GR (2007) The barcode of life initiative: synopsis and prospective societal impacts of DNA barcoding of fish. *Genomics, Society and Policy*, **3**, 29–40.

Crenshaw JW (1965) Serum protein variation in an interspecies hybrid swarm of turtles of the genus *Pseudemys*. *Evolution*, **19**, 1–15.

Damm S, Schierwater B, Hadrys H (2010) An integrative approach to species discovery in odonata: from character-based DNA barcoding to ecology. *Molecular Ecology*, **19**, 3881–3893.

Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS (2007) Validation of the barcoding gene COI for use in forensic genetic species identification. *Forensic Science International*, **173**, 1–6.

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B*, **360**, 1905–1916.

van Dijk PP, Stuart BL, Rhodin AGJ (2000) *Asian Turtle Trade – Proceedings of a Workshop on Conservation and Trade of Freshwater Turtles and Tortoises in Asia*. Chelonian Research Foundation, Lunenburg.

Dove CJ, Rotzel NC, Heacker M, Weigt LA (2008) Using DNA barcodes to identify bird species involved in birdstrikes. *Journal of Wildlife Management*, **72**, 1231–1236.

Eaton MJ, Meyers GL, Kolokotronis S-O, Leslie MS, Martin AP, Amato G (2010) Barcoding bushmeat: molecular identification of Central African

and South American harvested vertebrates. *Conservation Genetics*, **11**, 1389–1404.

Feldman CR, Parham JF (2004) Molecular systematics of old world stripe-necked turtles (Testudines: *Mauremys*). *Asiatic Herpetological Research*, **10**, 28–37.

Felsenstein J (1989) PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics*, **5**, 164–166.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fong JJ, Chen T (2010) DNA evidence for hybridization of wild turtles in Taiwan: possible genetic pollution from trade animals. *Conservation Genetics*, **11**, 2061–2066.

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

Georges A, Thomson S (2009) Diversity of Australasian freshwater turtles, with an annotated synonymy and keys to species. *Zootaxa*, **2496**, 1–37.

Gibbons JW, Scott DE, Ryan TJ *et al.* (2000) The global decline of reptiles, deja vu amphibians. *BioScience*, **50**, 653–666.

He J, Zhou T, Rao D, Zhang Y (2007) Molecular identification and phylogenetic position of *Cuora yunnanensis*. *Chinese Science Bulletin*, **52**, 3305–3309.

Hebert PDN, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society London B*, **270**, S96–S99.

Hebert PDN, Stoeckle MY, Zemlak TS, Frances CM (2004) Identification of birds through DNA barcodes. *PLOS Biology*, **2**, 1657–1663.

International Union for the Conservation of Nature (2009) Red List of Threatened Species. http://www.iucnredlist.org/. Accessed 6 March 2009.

Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, **6**, 998–1002.

Jungt SO, Lee YM, Kartavstev Y, Park IS, Kim DS, Lee JS (2006) The complete mitochondrial genome of the Korean soft-shelled turtle *Pelodiscus sinensis* (Testudines, Trionychidae). *DNA Sequencing*, **17**, 471–483.

Karl SA, Bowen BW, Avise JC (1995) Hybridization among the ancient mariners: characterization of marine turtle hybrids with molecular genetic assays. *Journal of Heredity*, **86**, 262–268.

Kelly RP, Sarkar IN, Eernisse DJ, DeSalle R (2007) DNA barcoding using chitons (genus *Mopalia*). *Molecular Ecology Notes*, **7**, 177–183.

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

Kumazawa Y, Nishida M (1999) Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. *Molecular Biology and Evolution*, **16**, 782–792.

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

Lara-Ruiz P, Lopez GG, Santos FR, Soares LS (2006) Extensive hybridization in hawksbill turtles (*Eretmochelys imbricata*) nesting in Brazil revealed by mtDNA analyses. *Conservation Genetics*, **7**, 773–781.

Letunic I, Bork P (2006) Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, **23**, 127–128.

Maddison DR, Maddison WP (2000) *MacClade 4: Analysis of phylogeny and character evolution. Version 4.0*. Sinauer Associates, Sunderland, Massachusetts.

Messing J (1983) New M13 vectors for cloning. *Methods in Enzymology*, **101**, 20–78.

Mindell DP, Sorenson MD, Dimcheff DE, Hasegawa M, Ast JC, Yuri T (1999) Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *Systematic Biology*, **48**, 138–152.

Naro-Maciel E, Reid B, Fitzsimmons NN, Le M, Desalle R, Amato G (2010) DNA barcodes for globally threatened marine turtles: a novel registry approach to documenting biodiversity. *Molecular Ecology Resources*, **10**, 252–263.

Nijman V, Shepherd CR (2007) Trade in non-native, CITES-listed, wildlife in Asia, as exemplified by the trade in freshwater turtles and tortoises (Chelonidae) in Thailand. *Contributions to Zoology*, **76**, 207–212.

Ogden R, Dawnay N, McEwing R (2009) Wildlife DNA forensics—bridging the gap between conservation genetics and law enforcement. *Endangered Species Research*, **9**, 179–195.

Parham JF, Simison WB, Kozak KH, Feldman CR, Shi H (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.

Parham JF, Stuart BL, Bour R, Fritz U (2004) Evolutionary distinctiveness of the extinct Yunnan box turtle (*Cuora yunnanensis*) revealed by DNA from an old museum specimen. *Proceedings of the Royal Society of London B*, **271**, S391–S394.

Parham JF, Feldman CR, Boore J (2006a) The complete mitochondrial genome of the enigmatic bigheaded turtle (*Platysternon*): description of unusual genomic features and the reconciliation of phylogenetic hypotheses based on mitochondrial and nuclear DNA. *BMC Evolutionary Biology*, **6**, 11.

Parham JF, Macey JR, Papenfuss TJ *et al.* (2006b) The phylogeny of Mediterranean tortoises and their close relatives based on complete mitochondrial genome sequences from museum specimens. *Molecular Phylogenetics and Evolution*, **38**, 50–64.

Rach J, DeSalle R, Sarkar IN, Schierwater B, Hadrys H (2008) Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society of London B*, **275**, 237–247.

Rafeliarisoa T, Shore G, Engberg S, Louis E, Brenneman R (2006) Characterization of 11 microsatellite marker loci in the Malagasy big-headed turtle (*Erymnochelys madagascariensis*). *Molecular Ecology Notes*, **6**, 1228–1230.

Russell RD, Beckenbach AT (2008) Recoding of translation in turtle mitochondrial genomes: programmed frameshift mutations and evidence of a modified genetic code. *Journal of Molecular Evolution*, **67**, 682–695.

Sarkar IN, Thornton J, Planet PJ, Schierwater B, DeSalle R (2002) A systematic method for classification of novel homeoboxes. *Molecular Phylogenetics and Evolution*, **24**, 388–399.

Sarkar IN, Planet PJ, DeSalle R (2008) CAOS software for use in character-based DNA barcoding. *Molecular Ecology Resources*, **8**, 1256–1259.

Spencer RJ, Thompson MB (2003) The significance of predation in nest site selection of turtles: an experimental consideration of macro- and microhabitat preferences. *Oikos*, **102**, 592–600.

Spinks PQ, Shaffer HB (2007) Conservation genetics 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.

Stuart BL, Parham JF (2004) Molecular phylogeny of the critically endangered Indochinese box turtle (*Cuora galbinifrons*). *Molecular Phylogenetics and Evolution*, **31**, 164–177.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Turtle Conservation Fund (2002) *A global action plan for conservation of tortoises and freshwater turtles. Strategy and funding prospectus 2002–2007*. 30 pp. Conservation International and Chelonian Research Foundation, Washington, DC.

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

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B*, **360**, 1847–1857.

Yassin A, Markow TA, Narechania A, O'Grady PM, Desalle R (2010) The genus *Drosophila* as a model for testing tree- and character-based methods of species identification using DNA barcoding. *Molecular Phylogenetics and Evolution*, **57**, 509–517.

Zardoya R, Meyer A (1998) Complete mitochondrial genomes suggests diapsid affinities of turtles. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 14226–14231.

Zarza E, Reynoso VH, Emerson BC (2008) Diversification in the northern neotropics: mitochondrial and nuclear DNA phylogeography of the iguana *Ctenosaura pectinata* and related species. *Molecular Ecology*, **17**, 3259–3275.

Zeale MRK, Butlin RK, Barker GLA, Lees DC, Jones G (2011) Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, **11**, 236–244.

Zhang L, Nie L, Cao C, Zhan Y (2008) The complete mitochondrial genome of the keeled box turtle *Pyxidea mouhotii* and phylogenetic analysis of major turtle groups. *Journal of Genetics and Genomics*, **35**, 33–40.

Data Accessibility

DNA Sequences: Genbank accessions HQ329587–HQ329787; BOLD accessions BENT102-08–BENT335-09. Alignments and trees: TreeBASE accessions S11480.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Descriptive data for all taxa and sequences included in this study. '*N*' indicates the number of individuals sequences for each species; '*H*' indicates the number of haplotypes observed in each species; 'Distance' indicates the species' classification within the distance-based scheme described in the text; 'Diagnostic' indicates the presence ('Y') or absence ('N') of family-level simple identifying characters in the species. References and accession numbers are in bold for novel sequences produced in this study.

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