Conserv Genet (2007) 8:641–657
DOI 10.1007/s10592-006-9210-1

ORIGINAL PAPER

Conservation phylogenetics of the Asian box turtles (Geoemydidae, *Cuora*): mitochondrial introgression, numts, and inferences from multiple nuclear loci

Phillip Q. Spinks · H. Bradley Shaffer

Received: 1 May 2006 / Accepted: 30 August 2006 / Published online: 10 November 2006

Abstract Asian box turtles (genus *Cuora*, family Geoemydidae) comprise a clade of 12 aquatic and semiaquatic nominate species distributed across southern China and Southeast Asia. Over the last two decades, turtles throughout Asia have been harvested at an unsustainable rate to satisfy demands for food, traditional Chinese medicine, and the pet trade. Consequently, all species of *Cuora* were recently placed on the IUCN Red List, nine are currently listed as critically endangered by the IUCN, and all species are listed in Appendix II of CITES. We compiled a 67-specimen mitochondrial (~1,650 base pairs (bp) from two mitochondrial genes) and a 40-specimen nuclear-plus-mitochondrial (~3,900 bp total, three nuclear introns plus an additional ~860 bp mitochondrial gene fragment) DNA data set to reconstruct the phylogeny of *Cuora* species and to assess genetic diversity and species boundaries for several of the most problematic taxa. Our sampling included 23 *C. trifasciata*, 17 *C. zhoui* and 1–4 individuals of the remaining 10 species of *Cuora*. Maximum likelihood, maximum parsimony and Bayesian analyses all recovered similar, well resolved trees. Within the *Cuora* clade, mitochondrial and nuclear sequence data indicated that both *C. zhoui* and *C. mccordi* represent old lineages with little or no history of interspecific gene flow, suggesting that they are good genealogical species. Based on mtDNA, *Cuora pani* was paraphyletic and *C. trifasciata* was composed of two highly divergent lineages that were not each other's closest relatives; both cases of non-monophyly were due to a mtDNA sequence that was widespread and identical in *C. aurocapitata, C. pani* and *C. trifasciata*. However, when combined with nuclear DNA results, our data indicate that *C. trifasciata* is a single, monophyletic taxon, and that mitochondrial introgression and nuclear-mitochondrial pseudogenes have led to a complex pattern of mitochondrial gene relationships that does not reflect species-level histories. Our results imply that captive ''assurance colonies'' of both *C. trifasciata* and *C. pani* should be genotyped to identify pure, non-hybrid members of both taxa, and we recommend that introgressed and pure taxa be managed as independent entities until the full evolutionary histories of these critically endangered turtles are resolved.

Keywords *Cuora* · *Mauremys* · Geoemydidae · Bataguridae · R35 · RELN · HNF-1α nuclear intron · Numt · Pseudogene · Introgression

P. Q. Spinks (✉) · H. B. Shaffer
Section of Evolution and Ecology, University of California, Davis, CA 95616, USA
e-mail: pqspinks@ucdavis.edu

H. B. Shaffer
Center for Population Biology, University of California, Davis, CA 95616, USA

Introduction

The Asian box turtles (genus *Cuora*, Family Geoemydidae) are a clade of 12 small-to-medium-sized aquatic and semiaquatic turtles distributed throughout southern China and Southeast Asia (Ernst and Barbour 1989; Iverson 1992; see also http://www. emys.geo. orst.edu). Currently, the entire genus is of conservation concern and is a focal group for the Asian Turtle Crisis (van Dijk et al. 2000). Over the last two decades, turtles throughout Asia have been harvested at an

unsustainable rate to satisfy the enormous, growing demand for turtles for human consumption, traditional Chinese medicine and the pet trade (Lau and Haitao 2000; van Dijk 2000; Parham et al. 2001). Demand for some species has become so great that turtles can fetch extremely high prices in Chinese markets. For example, *C. trifasciata* sell for up to $1,000 USD per kilogram due to the scientifically unverified, but widely believed medicinal qualities of its shell, and as this species has become less common, collection pressure on other species has increased (Lau and Haitao 2000; Lau et al. 2000; Parham et al. 2001). As a consequence, most species of *Cuora* are listed as critically endangered (exceptions are *amboinensis* = vulnerable and *flavomarginata* and *mouhotii* = endangered) by the International Union for Conservation of Nature and Natural Resources (IUCN 2006) and all *Cuora* species are listed in Appendix II of the Convention on International Trade in Endangered Species of Wild Flora and Fauna (CITES) (UNEP-WCMC 2005).

In response to the Asian turtle crisis, conservation organizations such as Conservation International (CI, http://www.conservation.org), and the Turtle Survival Alliance (TSA, http://www.turtlesurvival.org) as well as private individuals are taking emergency measures to slow the extirpation and extinction of the world's chelonian species, especially those from Southeast Asia. These efforts vary, and depending on the conservation status of a particular species, range from in situ habitat protection and headstarting programs for species that retain some viable habitat and populations, to ex situ captive populations which serve as "assurance colonies" for eventual release back into nature. However, most Asian turtle conservation and management efforts are hindered by a lack of basic biological information, since little is generally known about the natural history or patterns of genetic diversity within these species.

The conservation situation may be particularly acute for several recently described taxa because they are extremely rare, their geographic origins are unknown, and some are apparently of hybrid origin. For example, ten new geoemydid species (14 total, but four were synonyms) have been described from China in the last two decades (Kou 1989; Parham et al. 2001), and virtually nothing is known about the distribution or habitat requirements of these species. All were described from specimens culled from food and pet markets and most have yet to be found in the wild (Lau and Haitao 2000; van Dijk 2000; Parham et al. 2001). In addition, some of these recently described species are probably human-mediated hybrids produced in turtle farms in China and Southeast Asia (Table 1) since they are morphologically intermediate between known species and have displayed characteristics of hybrid taxa in recent genetic analyses (Artner et al. 1998; Fritz and Obst 1997; Parham et al. 2001; Wink et al. 2001; Spinks et al. 2004; Stuart and Parham 2004, 2006).

Table 1 Problematic *Cuora* and other recently described taxa

Cuora aurocapitata (Luo and Zhong 1988), and *Cuora pani* (Song 1984). These recently described taxa are highly similar in morphology and mitochondrial DNA (references 1, 2, 3, 4, 5 below), but both appear to be valid taxa based on nDNA sequence data and some mtDNA sequence data (this study)

Cuora mccordi (Ernst 1988). Known only from pet trade specimens, but appears to be a valid taxon (3, 4, 5, this study)

"Cuora serrata" Described as a subspecies of *C. galbinifrons* by Iverson and McCord (1992), and elevated to species by Fritz and Obst (1997). "*Cuora serrata*" are hybrids between *C. bourreti* or *C. galbinifrons*, and *C. mouhotii*, but some may be natural hybrids (4, 5, 6, 7)

Cuora trifasciata (Bell, 1825). All previous molecular analyses have recovered *C. trifasciata* as sister to *C. pani/aurocapitata*—the *C. trifasciata* species complex of Parham et al. (2004). However, results of the current study indicate hybridization/introgression among *C. aurocapitata, pani*, and *trifasciata*

Cuora zhoui (Zhao, Zhou and Ye 1990). Known only from pet trade specimens, but appears to be a valid taxon (2, 3, 4, this study)

"Mauremys iversoni" (Pritchard and McCord, 1991) are recent anthropogenic hybrids between *Mauremys mutica* and *C. trifasciata* (4, 7, 8, 9)

Mauremys pritchardi (McCord 1997) appear to be hybrids between *Mauremys mutica* or *M. annamensis* and *M. reevesii*, but the issue is confounded by the uncertainty surrounding the validity of *M. cf. annamensis*. *Mauremys pritchardi* might also represent a natural hybrid taxon (4, 8, 9, 10)

"Ocadia glyphistoma" (McCord and Iverson, 1994) are recent anthropogenic hybrids between *Mauremys sinensis* and *M. cf. annamensis* (4, 10)

"Ocadia philippeni" (McCord and Iverson, 1992) are recent anthropogenic hybrids between *Mauremys sinensis* and *C. trifasciata* (4, 10)

"Sacalia pseudocellata" (Iverson and McCord, 1992) are recent anthropogenic hybrids between *C. trifasciata* and *Sacalia quadriocellata* (4, 10)

Known hybrids are enclosed in quotations following Parham et al. (2001)

(1) McCord and Iverson 1991, (2) Honda et al. 2002a, (3) Parham et al. 2004, (4) Spinks et al. 2004, (5) Stuart and Parham 2004, (6) Shi et al. 2005, (7) Parham et al. 2001, (8) Wink et al. 2001, (9) Feldman and Parham 2004, (10) Stuart and Parham 2006

Four of the recently described species including *C. aurocapitata, C. mccordi, C. pani* and *C. zhoui* do not appear to be hybrid taxa (Table 1). They do not appear to be morphologically intermediate between known taxa, and in recent genetic analyses both *C. mccordi* and *C. zhoui* were reciprocally monophyletic and nested within *Cuora* (Honda 2002a; Parham et al. 2001, 2004; Spinks et al. 2004). *Cuora mccordi* and *C. zhoui* were described from market specimens and have not been found in the wild (Parham et al. 2001). And, while *C. aurocapitata* and *C. pani* have been found in the wild (Luo and Zhong 1988; Parham and Li 1999; Song 2001), there is considerable overlap in diagnostic characters between these two taxa, suggesting that *C. aurocapitata* might best be considered a subspecies of *C. pani* (*C. p. aurocapitata*) (McCord and Iverson 1991). Results from recent molecular analyses suggest a close relationship among *C. aurocapitata, C. pani, C. trifasciata* and *C. zhoui* with *C. mccordi* only distantly related to these taxa (Honda et al. 2002a; Parham et al. 2004; Spinks et al. 2004). However, all of the preceding analyses were based on small sample sizes (1–3 individuals/species) and exclusively on mitochondrial (mt)DNA. Given the potential for human-mediated hybridization throughout this complex of turtles, strong conclusions about the validity of *C. aurocapitata C. mccordi, C. pani* and *C. zhoui*, their divergence from *C. trifasciata*, and their relationships to other *Cuora* species should be viewed as tentative at best, pending further analyses with larger, field-collected material and additional nuclear (n)DNA markers.

Recently, we were provided with blood samples from a relatively large group of live turtles including 15 *C. trifasciata* from Kadoorie Farms and Botanical Gardens (KFBG) in Hong Kong, China, and 14 *C. zhoui* including six that are breeding stock from the Allwëtter Zoo (Germany). Our goals in the current study were to use this relatively deep taxon sampling (given the rarity of these species) to provide new insights into the genetic diversity within *Cuora*. We emphasized the *C. trifasciata* species complex (sensu Parham et al. 2004), since all previous work has recognized a close, poorly resolved set of relationships among the three nominate species in the complex, and considerable uncertainty about the validity of *C. aurocapitata* (McCord and Iverson 1991; Honda et al. 2002a; Stuart and Parham 2004). Particularly given their endangered status (all three species are listed as Critically Endangered by the IUCN), there is great urgency to determine both the validity of the currently recognized taxa and the possibility that additional unrecognized lineage diversity may exist in the global captive population.

Materials and methods

Taxon sampling

Our analysis included all 12 species of *Cuora* and two outgroups. Our samples include 15 live *C. trifasciata* from KFBG, 10 of which were field-collected. Of these, five were collected in the central New Territories, two in the northeastern New Territories, and three from the south side of Hong Kong Island (Fig. 1, Appendix). All three individuals collected from Hong Kong Island might be translocated or escaped pets, given that two were found in developed coastal areas and one was collected near a large reservoir, rather than the forested streamside habitat that characterizes the species (Mell 1922, cited in Ernst and Barbour 1989; Stuart et al. 2001) (Fig. 1). Other live individuals lacking locality data include four *C. trifasciata* from the Fort Worth Zoo, and two turtles from a private collection. We also included GenBank sequences from two individuals (Appendix). Both the KFBG and Fort Worth Zoo turtles are part of ''assurance colonies'' for this endangered species that are maintained as breeding stocks; the presumption is that they represent a single genetic lineage. We also have a relatively large sampling of *C. zhoui* (17) including the Allwëtter Zoo stock, eight live individuals from private collections, and GenBank sequences from three individuals. Photo vouchers for most non-museum voucher samples are available from the Shaffer lab website: http://www2.eve.ucdavis.edu/shafferlab. For outgroups, we used *M. mutica* (two individuals) because *Mauremys* is the sister clade to *Cuora* (Honda et al. 2002b; Spinks et al. 2004). In total, our ingroup consisted of 65 individuals: 23 *C. trifasciata*, 17 *C. zhoui* and 1–4 of each of the remaining species. *Cuora yunnanensis* which has been considered to be extinct by the IUCN in spite of the possibility that two live individual have recently been found in China (IUCN 2006; http://www.iucnredlist.org/search/details.php?species=5957) is based solely on mtDNA sequence data recovered from a single museum specimen (Parham et al. 2004 see Appendix). The sequences included here that were generated by Stuart and Parham (2004) represent a fairly broad geographic taxon sampling of known-locality *Cuora* specimens including *amboinensis* (Laos), *flavomarginata* (Anhui Province, China), *galbinifrons* (Vietnam and Laos), and *mouhotii* (Hainan Province, China). Our analysis extends the geographic coverage for this genus since we include known-locality specimens of *trifasciata* from Hong Kong (mainland China) and an *amboinensis* from the Philippines (Luzon Bay, Appendix).

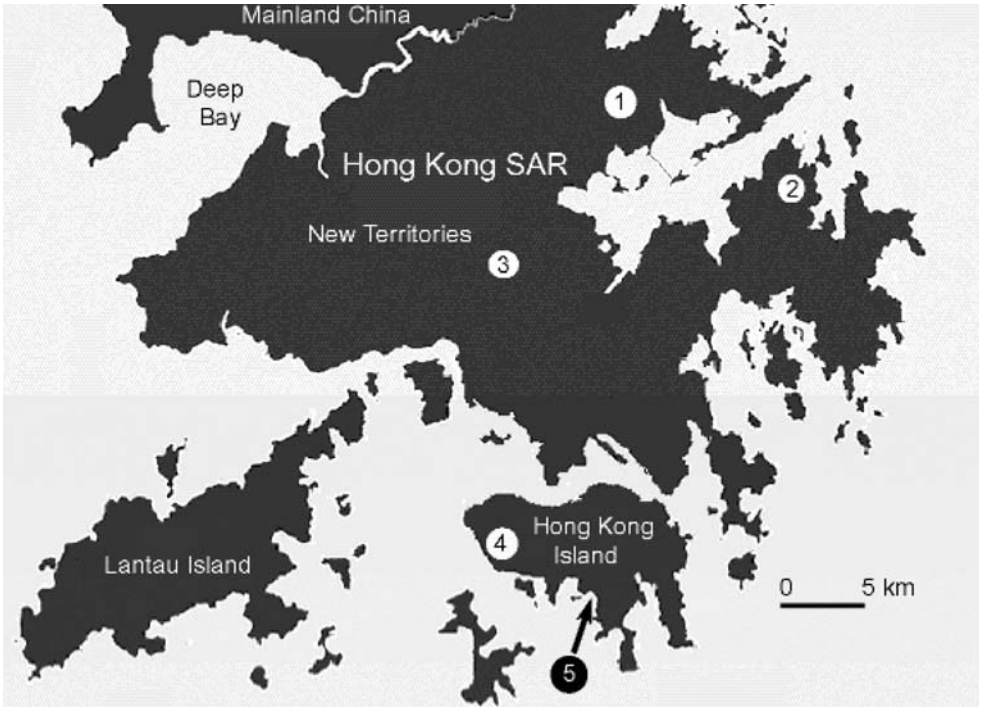

Fig. 1 Map showing Hong Kong Island and Hong Kong Special Administrative Region (SAR). Due to threats from poachers, exact collection localities were not provided to us. Numbers on map indicate general collection localities for some *C. trifasciata*, and HBS indicates tissue samples from these localities that are in the collection of H. Bradley Shaffer, University of California Davis. Site #1, northeast New Territories, HBS41825, HBS41839, HBS41826. Site #2, eastern New Territories, HBS41831, HBS41829. Site #3, central New Territories, HBS41827, HBS41828. Site #4, southwest Hong Kong Island, HBS41832. Site #5, southern Hong Kong Island, HBS41833, HBS41829

DNA sampling

mtDNA data

The majority of our mtDNA sequence data includes sequences generated for this study, but also includes some previously reported GenBank sequences (Appendix). Our mtDNA data set is composed of the cytochrome *c* oxidase subunit I (COI) gene, and the nicotinamide adenine dinucleotide dehydrogenase subunit 4 (ND4) gene plus the flanking Histidine and Serine tRNAs and part of the Leucine tRNA (hereafter collectively referred to as ND4) for 67 individuals (sequences from two individuals were excluded, see below). In addition, we collected sequences for the nicotinamide adenine dinucleotide dehydrogenase subunit 1 (ND1) gene for a subset of 40 individuals (see below). For the new sequence data generated for this study, DNA was extracted from muscle or blood using a salt extraction protocol (Sambrook and Russell 2001). Partial sequences were collected for the COI, and ND4 genes using 15–20 μl volume Taq-mediated PCR reactions, and reaction conditions and primers L-turtCOIc + H-turtCOIc for the COI gene and L-ND4 + H-leu for ND4 (Stuart and Parham 2004, Table 2).

Sequence chromatograms were examined using SeqEd v1.0.3 (Applied Biosystems). Some COI and ND4 sequence chromatograms displayed sequence heterogeneity (double peaks) at several positions within either gene fragment (Fig. 2) indicating the presence of multiple templates for these reactions (Primmer et al. 2002). Multiple templates could be due to (1) contaminated samples, (2) the presence of nuclear mitochondrial pseudogenes (numts), (3) mitochondrial heteroplasmy, or a combination of these conditions. We re-extracted and re-sequenced a subset of these individuals, but still observed the same multiple peaks in the sequence chromatograms. Thus, contamination (at least post tissue-collection) did not appear to be the problem. In order to generate authentic mtDNA, we followed Thalman et al. (2004) and used long-range PCR to amplify large fragments of the mitochondrial genome containing the COI and ND4 fragments (Fig. 2). We used these large fragments as templates in successive PCR reactions to avoid sequencing numts under the assumption that large PCR fragments of the expected length (based on the complete mitochondrial genome of *Mauremys* (formerly *Chinemys*) *reevesii*, GenBank accession # AY676201) represented authentic mtDNA. For long-range PCR, we used Platinum® *Taq* DNA polymerase High Fidelity (Invitrogen) with supplied reagents and manufacturers reaction conditions. For individuals with putative numts, we amplified one or two large fragments per individual (Appendix) including: (1) an ≈11 kb fragment here termed LR4 that spans the mtDNA genome from COI to the control region (CR) (Fig. 2, forward primer = L-turtCOIc, reverse

Table 2 Primers and annealing temperatures used in this analysis

Primer name	Sequence	Locus	Temperature
L-turtCOIc[a]	5′-TACCTGTGATTTTAACCCGTTGAT-3′	COI	60°C
H-turtCOIc[a]	5′-TGGTGGGCTCATACAATAAAGC-3′	COI	60°C
CytbG[b]	5′-AACCATCGTTGTWATCAACTAC-3′	Cytochrome *b*	58°C
THR-8[b]	5′-GGTTTACAAGACCAATGCTT-3′	Cytochrome *b*	58°C
ND1F[c]	5′-GGMTAYATACAACTTCGAAAAGG-3′	ND1	60°C
ND1R[c]	5′-GGTTTTAGCCTCTATTATTCACCC-3′	ND1	60°C
ND2R[c]	5′-GAGGTTCTATCTCTTGTTTGGGGC-3′	ND2	55°C
L-ND4[a]	5′-GTAGAAGCCCCAATCGCAG-3′	ND4	60°C
H-Leu[a]	5′-ATTACTTTTACTTGGATTTGCACCA-3′	ND4	60°C
DES-2[d]	5′-GGATTTAGGGGTTTGACGAGAAT-3′	Control region	55°C
12SXLF[c]	5′-GATTAGATACCCCACTATGCTTAG-3′	12S	55°C
HNFAL-F[e]	5′-GCAGCCCTCTACACCTGGTA-3′	HNF-1α intron 2	62°C
HNFAL-R[e]	5′-CAATATCCCCTGACCAGCAT-3′	HNF-1α intron 2	62°C
R35Ex1[f]	5′-ACGATTCTCGCTGATTCTTGC-3′	R35 intron 1	58°C
R35Ex2[f]	5′-GCAGAAAACTGAATGTCTCAAAGG-3′	R35 intron 1	58°C
R35In1CF[c]	5′-TTKVTGBAATKTATGGRRAG-3′	R35 intron 1	58°C
R35In1CR[c]	5′-CTYYCCATAMATTVCABMAA-3′	R35 intron 1	58°C
RELN61F[c]	5′-TGAAAGAGTCACTGAAATAAACTGGGAAAC-3′	Reelin intron 61	61°C
RELN61R[c]	5′-GCCATGTAATTCCATTATTTACACTG-3′	Reelin intron 61	61°C

[a] Stuart and Parham (2004)

[b] Spinks et al. (2004)

[c] This analysis

[d] Starkey et al. (2003)

[e] Primmer et al. (2002)

[f] Fujita et al. (2004)

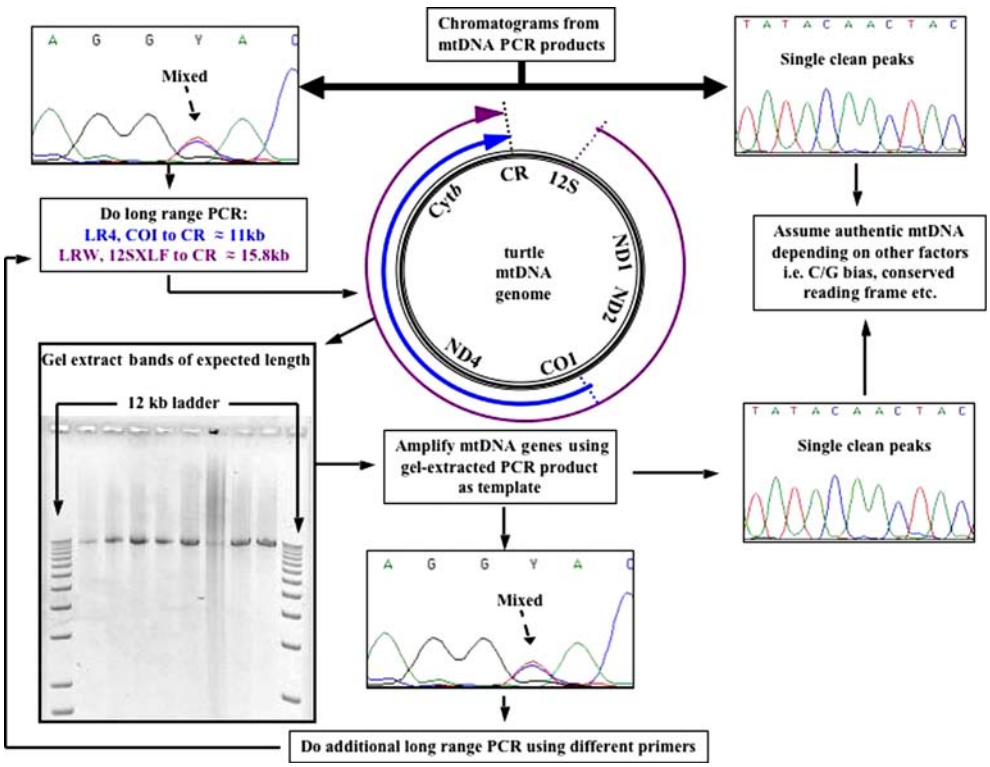

Fig. 2 Schematic diagram of strategy used to obtain true mtDNA sequences when numts are encountered (following Thalman et al. 2004). The relative positions of the mitochondrial genes sequenced for this analysis or used as priming sites are indicated on the mtDNA genome illustration

primer = DES-2), and (2) a ≈15.8 kb fragment (LRW) spanning most of the mtDNA genome from 12S to the CR (12SXLF, DES-2) (all primers in Table 2). We amplified the LR4 fragment for 25 individuals, but the

LRW fragment was amplified for three individuals only (Appendix). We then gel-extracted the resulting bands of expected length plus some shorter bands if present, and cleaned the PCR products using the QIAquick® Gel Extraction Kit (QIAGEN). We used these fragments as templates for additional rounds of PCR using the COI and ND4 primers. All PCR products were then sequenced in both directions on ABI 3730 automated sequencers at the UC Davis Division of Biological Sciences sequencing facility (http://www.dnaseq.ucdavis.edu/), and sequences were aligned by eye in PAUP* 4.0b10 (Swofford 2002). Our assumption was that sequence chromatograms derived from long-range templates of the expected length that displayed single clean peaks would represent authentic mtDNA; if numts and mtDNA were coamplified and sequenced, the chromatograms would display sites with multiple peaks (Primmer et al. 2002). Sequences derived from the long-range PCR fragments that were shorter than expected represent candidate numts under the assumption that gene order and genome size are conserved within turtle mtDNA (Parham et al. 2006).

As a final test for true mtDNA, we sequenced an additional mtDNA gene from PCR reactions utilizing *genomic* DNA templates under the hypothesis that numts might not be present for all mitochondrial genes. We amplified (from genomic template) and sequenced a fragment of the mitochondrial ND1 gene, using primers that we developed based on an alignment of the complete mitochondrial genomes of two turtles, *M. reveesii* and *Chrysemys picta* (GenBank accession Nos. AY676201 and AF069423 respectively) (primers and annealing temperature are listed in Table 2). Using these primers, we sequenced a representative subset of 40 taxa for ND1 (Appendix). All sequencing chromatograms for the ND1 sequences were clean with no indication of multiple templates except for *C. bourreti* 1. Chromatograms for this individual were mixed indicating a numt for this mtDNA gene. For this individual only, we amplified a ≈2.18 kb mitochondrial fragment (which spans a region including portions of the ND1 and ND2 genes) using our ND1 forward primer and an ND2 reverse primer developed for this analysis (Table 2). We then gel-extracted the appropriate-sized band, and used it as template in an additional PCR reaction. Chromatograms from sequence data generated from this PCR product contained single peaks with no indication of multiple copies so we included it in our ND1 data set.

Nuclear DNA data

To bring a nuclear perspective to our work, we subsampled 40 individuals from the mtDNA analysis including at least two representatives of most species and sequenced them for three nuclear introns: intron 1 of the fingerprint protein 35 (R35, ~1200 bp in length) (Fujita et al. 2004), intron 2 of the hepatocyte nuclear factor 1α (HNF-1α, ~900 bp) (Horlein et al. 1993; Primmer et al. 2002), and intron 61 of the Reelin gene (RELN ~1,200 bp, see below). While this is the first application of nucleotide sequence data from the HNF-1α and RELN loci to the phylogenetics of turtles, R35 has been used for inter and intraspecific analyses of turtles (Engstrom et al. 2004; Fujita et al. 2004; Spinks and Shaffer 2005; Near et al. 2005) including geoemydids (Spinks et al. 2004). Genetic divergence within the *Cuora* clade is relatively shallow based on mtDNA data (Honda et al. 2002a; Parham et al. 2004; Spinks et al. 2004), which may imply even less among-taxon variability in the nuclear genome (Hudson & Coyne 2002). Based on this previous work, we anticipated that there might be sufficient resolution in our nuclear loci to provide independent evidence of genealogical species boundaries in *Cuora*, but not necessarily with strong statistical support.

For generating nDNA data, we used the same PCR conditions as for mtDNA. R35 primers were from Fujita et al. (2004) and HNF-1α primers were from Primmer et al. (2002). The RELN primers were developed by PQS for intron 61 of the Reelin gene based on a characterization of the mouse Reelin gene (Royaux et al. 1997). Primers and annealing temperatures are listed in Table 2. PCR products were separated on 1–2% agarose gels, stained with ethidium bromide, and visualized on a UV light box. For some individuals, we observed weak secondary bands in the gels. In these cases, we ran 50μl-volume PCR reactions through 1% agarose gels and gel-extracted the resulting bands of expected length using the QIAquick® Gel Extraction Kit (QIAGEN) and manufacturer's protocol. These gel-extracted PCR products were then sequenced in both directions using the amplification primers.

Patterns from the sequencing chromatograms indicated that in all three loci some individuals were heterozygous for length polymorphisms which usually corrupt sequence reads downstream of the indel site (see Bhangale et al. 2005, Fig. 1B). Several individuals were heterozygous for length polymorphisms at the R35 locus so we designed internal forward and reverse primers (Table 2), and used these primers in combination with the external primers to sequence most of this intron for the putative length-polymorphic individuals.

Phylogenetic analyses

Sequences were aligned, checked for nucleotide ambiguities, and translated for coding regions with MacClade

4.06 (Maddison and Maddison 2003). We performed maximum parsimony (MP), maximum likelihood (ML) and Bayesian analyses on three separate data partitions: (1) combined COI and ND4 sequences for all 67 taxa plus redundant sequences derived from both long-range PCR products and putative numt sequences from nine individuals (Appendix), (2) combined COI, ND4 and ND1 sequences from 40 individuals, and 3) combined nDNA sequences from 40 individuals.

Maximum parsimony and ML analyses were performed using PAUP* 4.0b10 (Swofford 2002) with ten random stepwise heuristic searches. For MP we used tree bisection-reconnection (TBR) branch swapping. For the ML analysis we used subtree pruning-regrafting (SPR) branch swapping, and model parameters that were estimated using Modeltest 3.06 (Posada and Crandall 1998) and selected under the Akiake Information Criterion (AIC). We bootstrapped each data set with 100 pseudoreplicates (Felsenstein 1985).

Mixed-model analyses (Yang 1996) are increasingly used in phylogenetic reconstructions to account for evolutionary heterogeneity among different genes/data partitions. Using the combined data partitions with mixed model analyses has the advantages of incorporating the maximum amount of information into an analysis while simultaneously accounting for the different evolutionary characteristics of the various data partitions (Brandley et al. 2005). Therefore, we divided our mtDNA and nDNA data into discreet partitions, and used MrBayes V3.1.1 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003) to perform mixed model Bayesian analyses for the nuclear and mitochondrial data sets. The nDNA data were divided into three separate partitions (one for each intron), but we partitioned the mtNDA data by codon position (three partitions/mtDNA gene) with an additional partition for the complete Histidine and Serine tRNAs and partial Leucine tRNA. Model parameters were selected for each partition (see TreeBase S1625) via AIC using Modeltest V3.06PPC (Posada & Crandall 1998). We ran each Bayesian analyses with two replicates and four chains for 10^7 generations, and sampled the chains every 10^3 generations. We determined that stationarity had been reached when the potential scale reduction factor (PSRF) equaled 1, and when-log likelihood (-lnL) scores plotted against generation time reached a stationary value.

Results

COI and ND4 mtDNA sequence data

In all cases where double peaks (indicative of mixed samples) were observed in the mtDNA sequencing chromatograms, we employed long-range PCR to distinguish mitochondrial from numt copies (Fig. 2). The mitochondrial genome is circular and non-recombining while numts are part of the nuclear genome, and are therefore linear and able to recombine. In addition, numts are presumably functionless, and can accumulate length polymorphism due to insertion and deletion events that do not occur in functional, coding mtDNA. Thus, long-range PCR reactions will normally amplify only the targeted mitochondrial fragment, and these fragments will be of predictable lengths since the mitochondrial genome of turtles is highly conserved (Parham et al. 2006). Bands that are longer or shorter than the targeted mitochondrial fragment represent numts that have been length modified and can be avoided by gel-extracting the targeted fragment (Fig. 2).

All long-range PCR reactions produced bands of the expected length (≈ 11 kb for LR4, and ≈15.8 kb for LRW). Some reactions also produced smaller bands (≈8 kb for LR4, and ≈6 kb for LRW), and we sequenced COI and ND4 from the ≈8 kb LR4, and ≈6 kb LRW fragment for seven and two individuals respectively (Fig. 3). Almost all sequencing chromatograms from PCR reactions utilizing the ≈11 kb LR4 template showed single clean peaks and, therefore, no indication of mixed templates in these PCR reactions. However, sequencing chromatograms derived from the ≈11 kb LR4 template for two *C. trifasciata* (#22 and #23, Appendix) showed double peaks at several positions indicating the presence of numts for these individuals. Based on visual examination of the chromatograms, these sequences appear to be mixtures of clade ''A'' haplotypes (see below) and numt sequences from Node 1 (Fig. 3). However, because it was impossible to identify the mtDNA orthologs with certainty, we excluded these two individuals from the analysis. Chromatograms from PCR reactions utilizing the ≈15.8 kb LRW fragment all showed single clean peaks and were assumed to represent true mtDNA.

All sequences including those from putative authentic and numt sequences displayed characteristics typical of authentic mtDNA. For example, the protein coding reading frame was conserved in both the COI gene and in the coding region of the ND4 fragment, and base composition was biased, showing the typical mtDNA deficiency of guanine nucleotides (A = 31%, C = 26%, G = 15%, T = 28% averaged across both genes). In addition, the putative authentic mtDNA sequence data generated from multiple PCR reactions and templates from a particular individual were always the same. Cytochrome oxidase I

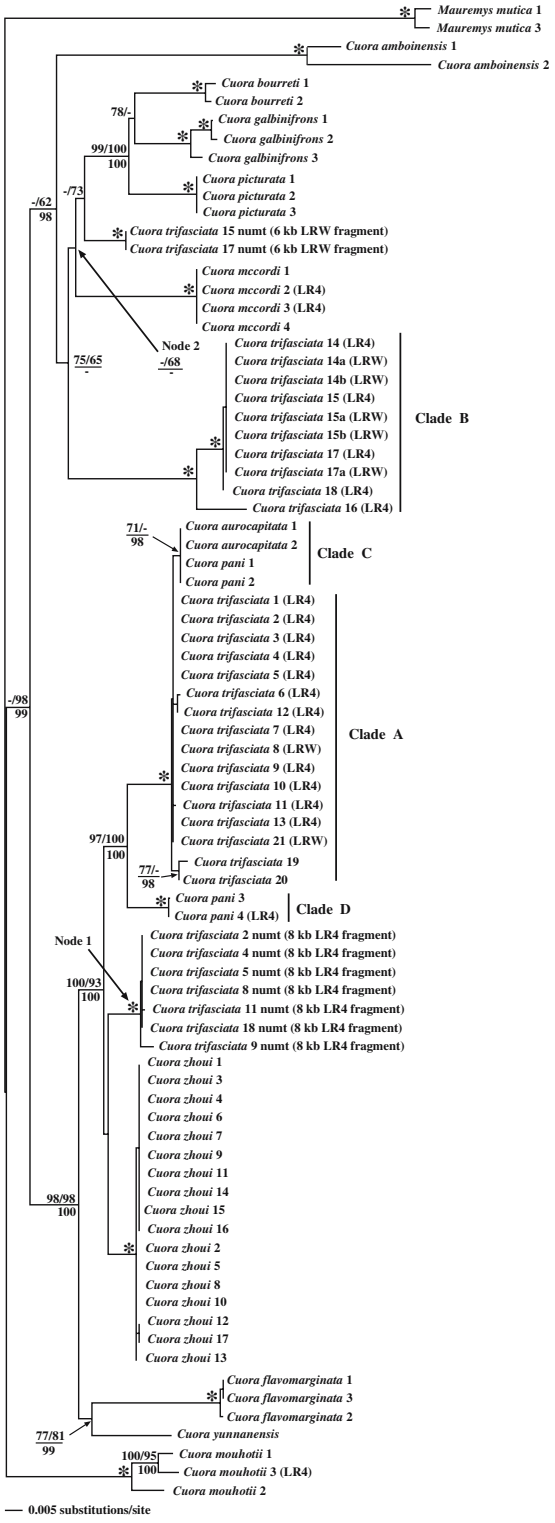

Fig. 3 Maximum likelihood tree based on the 81-OTU combined COI and ND4 data set (1,651 bp). Estimated model parameters conform to the GTR + G + I model of sequence evolution. –ln L = 6684.6669, rate matrix: A–C = 1, A–G = 27.7948 A–T = 1, C–G = 1, C–T = 21.9712, and G–T = 1. Base frequencies: A = 0.32, C = 0.26, G = 0.14, and T = 0.28. Proportion of invariable sites (I) = 0.4412, and γ-shape parameter = 0.6291. Numbers above branches are MP/ML bootstrap proportions ≥ 60%, and Bayesian Posterior Probabilities ≥ 95% are below branches. * denotes nodes that received 100% support values from all three analytical methods. Included in this tree are sequences generated from nine putative numts as well as redundant sequences from multiple different long-range PCR templates. For example, there are three sequences for *C. trifasciata* 14 including sequences generated from the LR4 template, *C. trifasciata* 14 (LR4), and sequences generated using two different LRW templates, *C. trifasciata* 14a (LRW), and *C. trifasciata* 14b (LRW) (see text)

sequences amplified using the ≈11 kb LR4 and ≈15.8 kb LRW fragments from *C. trifasciata* #s 14, 15 and 17 for example were identical within each individual as were their ND4 sequences (Fig. 3, Appendix).

81-OTU mitochondrial COI and ND4 phylogeny

Altogether we generated up to 831 bp of COI and 820 bp of ND4 sequence data for 61 operational taxonomic units (OTU) including sequences from 47 individuals, sequences from nine putative numts and five redundant sequences (sequences from different long-range PCR templates of the same individual). Our sequences, combined with GenBank sequences from 20 individuals, resulted in a mtDNA data set composed of 1,651 bp for 81 OTUs. Excluding outgroups, maximum uncorrected "P" sequence distance was 10.4%. Of the 1651 characters 1189 were constant, and 403 were parsimony-informative. The ML analyses recovered two trees that were not significantly different from one another (Shimodaira-Hasegawa [SH] test, P = 0.457). Figure 3 is the best ML tree (i.e. tree with the lowest–lnL score) with MP/ML bootstrap proportions ≥ 60% and Bayesian Posterior Probabilities (BPP) ≥ 95% as indicated.

All analyses of this data set identified two highly divergent clades of *C. trifasciata* which we labeled clade "A" and "B". Clade "A" has been identified in several recent mtDNA analyses (Honda 2002a; Parham et al. 2004; Spinks et al. 2004; Stuart and Parham 2004), whereas clade "B" represents a strongly supported, novel discovery. Clade "B" is the sister group to *C. mccordi* plus the *C. galbinifrons* species complex, and is further defined by a single nucleotide indel at the ND4/tRNA^{His} boundary (TreeBase #S1625). Of the five individuals that fell into clade "B", one (#16) was "field" collected on the south side of Hong Kong Island. However, this individual was collected from an extremely unnatural habitat near Repulse Bay, and therefore might have been translocated (Fig. 1, Appendix). The remaining 16 *C. trifasciata* sequences

fell into a second, distantly-related clade ''A'' (Fig. 3). Clade ''A'' was also well supported (100%, all methods), and shows a close, somewhat complex relationship with *C. pani* and *C. aurocapitata*. The putative numt sequences from clade ''A'' individuals (Node 1, Fig. 3) were sister to *C. zhoui*, and except for the numt sequence from *C. trifasciata* #18 (see nuclear results below), the clade ''B'' numt sequences were sister to the *C. galbinifrons* species complex.

Cuora pani was also recovered as two distinct lineages. Two *C. pani* and both *C. aurocapitata* (which together comprise clade ''C'', Fig. 3) were identical at the mtDNA sequence level and very similar (average uncorrected ''P'' pairwise sequence divergence = 0.31%) to *C. trifasciata* clade ''A''. However, two other *C. pani* (clade ''D'') comprised a distinct sister group to the (*aurocapitata* + *pani* clade ''C'' + *C. trifasciata* clade ''A'') group, rendering *pani* paraphyletic with respect to the other two species. Based on cursory morphological examination, individuals from the two mtDNA-based clades of *C. trifasciata* and *C. pani* do not show marked intraspecific variation (P. Crow, personal communication) but detailed morphometric analyses have yet to be performed.

Since 20 of our terminals (COI and ND4 only) were represented by GenBank sequences from Parham et al. (2004), our respective mtDNA analyses shared some similarities including (1) *C. zhoui* monophyly with strong support (100%, all methods), even with our increased sampling of 14 individuals (Parham et al. had three individuals), and (2) *Cuora mccordi* monophyly with strong support (100%), even with the addition of three new specimens in our analysis. *Cuora yunnanensis* was represented by a single individual, and so cannot formally be tested for monophyly although it was well differentiated, at least for the COI and ND4 mtDNA genes, potentially indicating monophyly.

The topology recovered from our combined data was quite different than that recovered by Parham et al. (2004), but we were unable to compare the respective topologies in a statistical framework due to the uncertainty surrounding species determinations for the *C. pani* and *C. trifasciata* specimens. Nevertheless, our increased taxon and gene sampling revealed additional lineages of *C. trifasciata* and *C. pani* which might greatly change our understanding of the phylogenetics and conservation genetics of *Cuora* (see below).

ND1 data and combined mtDNA data phylogeny

Our ND1 sequence data set was composed of up to 816 bp for 40 individuals. Within this data set, 627 bp were constant, 163 were parsimony-informative, and maximum uncorrected ''P'' sequence divergence of the ingroup was 9.60%. In order to compare ND1 versus COI + ND4 phylogenies, we first generated an ML analysis from the 40-taxon ND1 data partition; this ML analysis recovered a single tree (not shown). The phylogeny derived from the ND1 sequences recovered the same topology as in Fig. 3 with respect to the divergent positions of the clade ''A'' and ''B'' *C. trifasciata* (providing further evidence that the phylogenetic positions of these clades were based on authentic mtDNA data and not numts).

Next we performed ML analyses on a pruned COI + ND4 data set that contained the same 40 taxa as in the ND1 analysis. ML trees from the two data partitions (ND1, COI + ND4) were then compared using the SH test. The phylogeny recovered from the ND1 sequences was not statistically significantly different from the COI + ND4 ML phylogeny (SH test, P = 0.568). In addition, we performed an incongruence length difference test (partition homogeneity test implemented in PAUP* 4.0b10 [Swofford 2002]) on the ND1 vs COI + ND4 data partitions and determined that the data partitions were not incongruent ($P = 0.65$). Therefore, we combined the ND1 and COI + ND4 sequences for a final mtDNA analysis.

Our combined ND1, COI, and ND4 data set was composed of up to 2,467 bp for 40 individuals. Of the 2,467 characters, 1,818 were constant and 555 were parsimony informative. Figure 4 shows the ML phylogeny with MP/ML bootstrap proportions ≥ 60% and BPP ≥ 95% as indicated. We also used this combined mtDNA sequence data set to test for species monophyly of *C. pani* and *C. trifasciata* since these species were recovered as paraphyletic in all of our previous analyses. We constructed separate constraint trees such that *C. pani* was monophyletic or *C. trifasciata* was monophyletic, and then compared these trees (concurrently) to the unconstrained tree of Fig. 4 using the SH test. The hypothesis that *C. pani* and *C. trifasciata* were monophyletic based on mtDNA data was rejected ($P = 0.002$ and $P = 0.000$, respectively).

Nuclear DNA sequence data and phylogeny

For our nDNA sequence data, we collected up to 3,102 bp for 40 individuals including 864 bp of HNF-1α, 1,090 bp of R35 intron 1, and 1,148 bp of intron 61 of the RELN gene. However, despite extensive PCR optimization, we were unable to generate R35 or RELN sequence data for *C. pani* 1 or 2. Of the 3,102 characters, 2,863 were invariant while 150 were parsimony-informative. Maximum uncorrected ''P'' sequence divergence

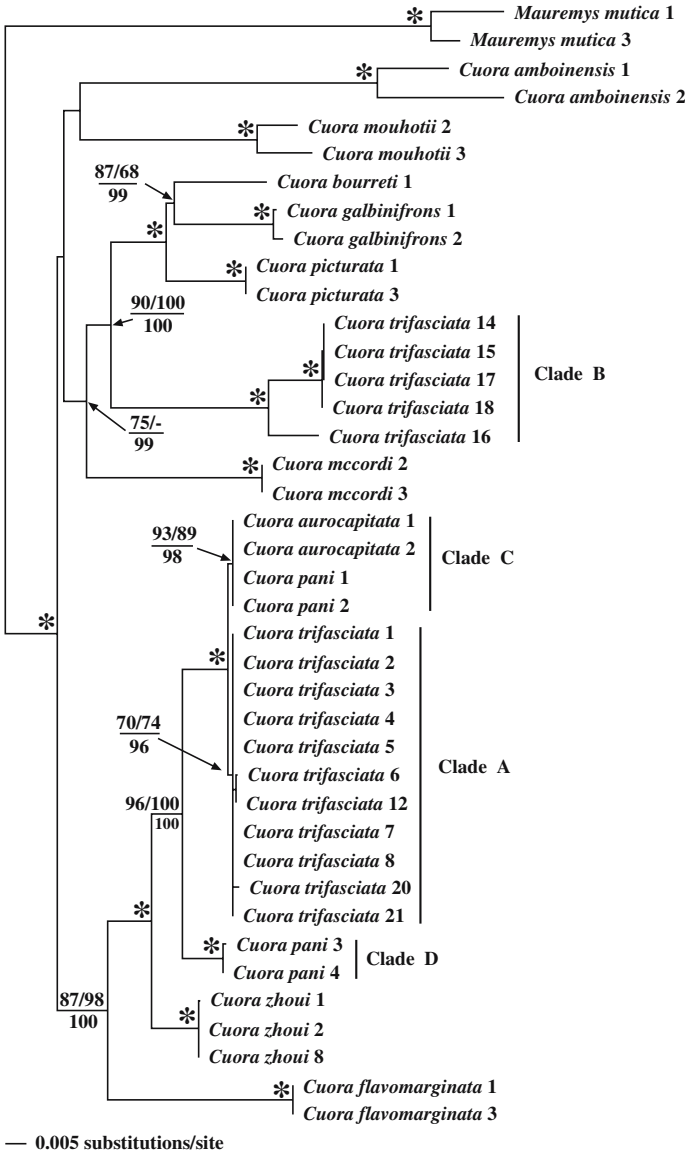

Fig. 4 Maximum likelihood tree based on the 40-taxon COI, ND4 and ND1 data set (2,467 bp). Estimated model parameters conform to the GTR + I + G model of sequence evolution. – ln L = 9033.58772, rate matrix: A–C = 1, A–G = 19.5706, A–T = 0.4183, C–G = 0.4183, C–T = 14.2555, and G–T = 1. Base frequencies: A = 0.32, C = 0.28, G = 0.13, and T = 0.28. Proportion of invariable sites (I) = 0.5665, and γ-Shape parameter = 1.1888. Numbers above branches are MP/ML bootstrap proportions ≥ 60% and Bayesian posterior probabilities ≥ 95% are below branches. * denotes nodes that received 100% support values from all three analytical methods

within the ingroup was 7%. Maximum likelihood (ML) analysis of these sequences recovered four trees that were not significantly different from one another (SH test, $P \geq 0.516$). Figure 5 is the best ML tree with bootstrap proportions ≥ 60% and BPP ≥ 95% as indicated. Although few of the deeper nodes were well supported in this analysis, two critical results are evident: the monophyly of all morphologically identified *C. trifasciata*, and the reciprocal monophyly of *C. aurocapitata* and *C. pani* 3 & 4. The picture is clouded, however, by the divergent positions of *C. pani* 1 & 2, but this result is most likely due to missing data (no R35 or RELN) for these taxa. It is also noteworthy that all *C. trifasciata* clade ''A'' individuals (with the exception of #18) fall into a well-supported subclade, although this is not the case for clade ''B'' individuals.

Discussion

Numts, introgression, and the true placement of *C. trifasciata*

How could the clades ''A'' and ''B'' of *C. trifasciata* be so genetically divergent at the mitochondrial locus, but not so at the nuclear level? One obvious possibility is that the phylogenetic position of one of the *C. trifasciata* clades is actually based on numts. After all, we have strong evidence that numts do exist in these animals (Stuart and Parham 2004), and the similarity of our phylogenetic pattern to some that are thought to be based on numts (i.e., Sorenson and Quinn 1998; Bates et al. 2004) motivated our thorough sequencing effort to investigate this possibility. Using

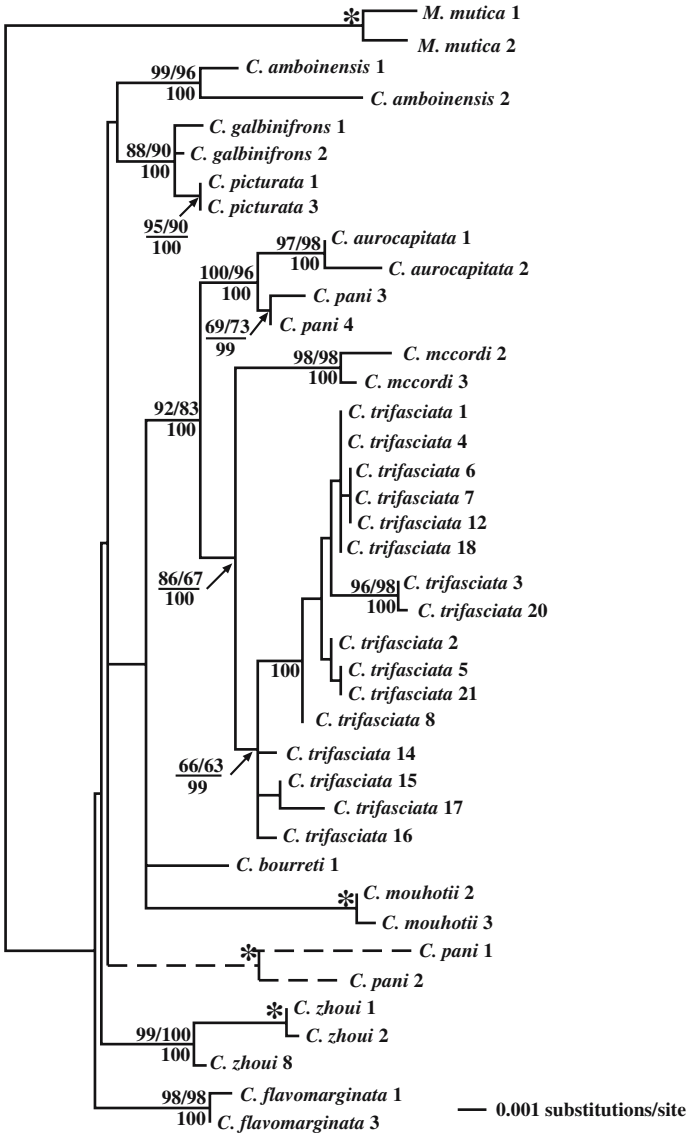

Fig. 5 Maximum likelihood tree based on the 40-taxon nDNA data set (3,102 bp). Estimated model parameters conform to the HKY + G + I model of sequence evolution. $-\ln L$ = 6324.75505, Transition/transversion ratio = 1.6465. Base frequencies: A = 0.30, C = 0.19, G = 0.20, and T = 0.31. Proportion of invariable sites (I) = 0.6188, and γ-shape parameter = 0.9211. Numbers above branches are MP/ML bootstrap proportions ≥ 60% and below branches are Bayesian Posterior Probabilities ≥ 95%. * indicates nodes with support values of 100% for all three analytical methods

our long-range PCR strategy, we consistently recovered identical sequences from clade ''B'' individuals from multiple PCR reactions of different large fragments. In fact, the LRW fragment was about 15800 bp in length, encompassing most (95%) of the mitochondrial genome. If these fragments are numts then they would be >1 kb longer than the longest reported chordate numt (http://www.pseudogene.net), a possible, but seemingly unlikely, outcome. Furthermore, sequence data from an additional mtDNA gene (ND1) was cleanly amplified and sequenced from *genomic* template, and revealed the same topology as recovered from the COI and ND4 genes, a result we would not expect if the clade ''A'' and ''B'' haplotypes were the result of numts or heteroplasmy.

Another possibility is that we were preferentially amplifying a numt gene copy (in clade ''B'') to the exclusion of the true mtDNA copy. This argument seems highly unlikely. The ratio of mtDNA to nDNA might be lower in blood-based extractions (most of our samples) compared to other soft tissue like muscle, but even in blood-based DNA extractions, there are many more copies of a particular mitochondrial gene than there are of any single-copy nuclear locus. Individual mitochondria typically contain multiple copies of their genome (Robin and Wong 1998), and circulating cells like peripheral blood mononuclear cells (PBMCs) usually contain numerous mitochondria. Thus, most PBMCs will contain abundant copies of the mitochondrial genome. For example, the red blood cell (RBC) to PBMC ratio is about 10:1 for the green sea turtle (*Chelonia mydas)*, with an average of 351×10^3 RBCs and 37×10^3 PBMCs per μl of blood (Wood and Ebanks 1984). We were unable to locate an estimate

for the number of mitochondrial genome copies/PBMC for turtles, but estimates for humans range from 264 copies/cell (Casula et al. 2005) to 409 copies/cell (Gahan et al. 2001). Thus, it is reasonable to expect that the average turtle PBMC will probably contain several hundred copies of the mitochondrial genome, in which case blood-based DNA extractions will provide ample copies of mtDNA genes. For instance, assuming there are 264 mitochondrial genomic copies/PMBC, then one μl of turtle blood might contain 37×10^3(PBMCs) $\times$ 264 (copies/cell) = 9.77×10^6 mtDNA gene copies compared to [351×10^3(RBCs) + 37×10^3(PBMCs)] $\times$ 2 = 7.76×10^5 nDNA gene copies (assuming two alleles/nuclear locus), a 12.5:1 mtDNA:nDNA ratio. Thus, there would have to be numerous insertions/duplications of large-sized individual numts for numts to equal or exceed paralogous mtDNA gene copies in a typical blood-based extraction. Post insertion duplication of small to medium-sized numts is common in some mammals (Hazkani-Covo et al. (2003) found 82 numts in humans for example) but all were <9 kb. On the other hand, Pereira and Baker (2004) found only 13 numts for a reptile (chicken) the largest of which was ≈1.7 kb. Thus, high numbers of extremely large numts would appear to be uncommon in most vertebrates, so even with a lower mtDNA:nDNA ratio compared to other tissue, mitochondrial gene copies will outnumber paralogous numts. Therefore, the likelihood that the clade ''B'' individuals actually contain clade ''A'' mtDNA haplotypes (and vice-versa), but we always failed to amplify them, is negligible.

Based on consistent results from our long-range PCR template sequencing, and the phylogeny from our ND1 data that was based on PCR products amplified directly from genomic DNA (except for *C. bourreti* #1), we conclude that the clade ''A'' and ''B'' data represents authentic mtDNA. What, then, could explain the existence of these two highly divergent clades of *C. trifasciata*? Although we were hampered by the lack of reliable field-collected material and distributional data for *C. aurocapitata*, *C. pani* and *C. trifasciata* to help explain the phylogenetic patterns, we propose mitochondrial introgression as the most plausible explanation. Given the propensity of *Cuora* species to hybridize with each other and with other genera (Artner et al. 1998; Parham et al. 2001; Wink et al. 2001; Spinks et al. 2004; Stuart and Parham 2006) hybridization or introgression is a plausible explanation.

All workers to date have recovered a close relationship among the *C. trifasciata* species complex based on mtDNA sequences (Honda et al. 2002a; Parham et al. 2004; Spinks et al. 2004; Stuart and Parham 2004), and most workers have assumed that this complex is a natural group consisting of *C. trifasciata, C. aurocapitata*,and *C. pani*. However, an intriguing possibility is that clade ''B'' represents the ''real'' phylogenetic position of *C. trifasciata* while the position of clade ''A'' is the result of introgression of *C. aurocapitata* or *C. pani* mtDNA into *C. trifasciata* (one of five hypotheses put forth by Stuart and Parham 2004, in the absence of any knowledge of clade ''B''). This could explain the mtDNA patterns i.e. divergent haplotypes for clade ''B'' individuals and extremely similar mtDNA haplotypes among clade ''A'' *C. trifasciata*, *C. aurocapitata*, and *C. pani*. At present, these three species are allopatric based on known localities (http://www.emys.geo.orst.edu/). However, our wild-caught *C. trifasciata* (those with reliable locality data) contain clade ''A'' haplotypes. Thus, the available data support natural hybridization between these taxa, perhaps indicating historical sympatric or parapatric distributions among *C. trifasciata* and one or both of *C. aurocapitata* or *C. pani*. Interestingly, all clade ''A'' individuals were also recovered as a monophyletic subclade of *C. trifasciata* based on nuclear DNA, suggesting that a true historical barrier to gene flow has existed within the species. Under our mitochondrial introgression interpretation, this barrier interrupted the flow of mitochondrial hybrid genes and non-hybrid nuclear genes within the species.

Another possibility is that lineage ''A'' is true *C. trifasciata*, and lineage ''B'' is the result of hybridization. If members of clade ''B'' were recent hybrids among any *Cuora* species then they should share mtDNA haplotypes with those species of *Cuora*. However, the lineage ''B'' individuals are a minimum of 3.64% sequence divergent from any other individual of *Cuora*—they do not appear to be recent, within-*Cuora* hybrids. Alternatively, if they were hybrids between *Cuora* species and some other geoemydid genus, they should have non-*Cuora* mtDNA haplotypes and should fall outside of the genus. We included all known *Cuora* species in our analysis, and clade ''B'' falls within *Cuora* on a relatively long branch. The monophyly of *Cuora* has previously been confirmed in the context of all recognized genera, and most species, of the family Geoemydidae (Spinks et al. 2004), so it is unlikely that some non-*Cuora* species is phylogenetically nested within the genus. In addition, using methods and data from Spinks et al. (2004) we sequenced the cyt*b* gene from five lineage ''A'' and three lineage ''B'' individuals (using the ≈11 kb LR4 templates) (GenBank Accession numbers in Appendix), and compared these sequences to sequences from most other geoemydid

species. Our cyt*b* results confirmed both the monophyly of *Cuora* and the divergent positions of clades ''A'' and ''B'' within *Cuora* (not shown).

Conservation, assurance colonies, and captive management of Cuora

The *C. trifasciata* in this study were sampled from ''assurance colonies'' for the species. Assurance colonies seek to create ''a network of linked captive management and breeding programmes... The primary goal of these Assurance Colonies will be to guarantee the survival of species which are currently threatened in the wild and to maximize future options for the recovery of threatened wild populations, through captive-breeding and re-introduction programmes.'' (IUCN, http://www.iucn.org/themes/ssc/sgs/sgprofiles/tsasg.htm). However, our genetic analysis indicates that the material in these colonies may represent two distinct groups: a small group of ''pure'' *C. trifasciata* (clade ''B''), and a larger group of introgressed individuals (clade ''A'') that are collectively being managed as a single species. We recommend that all captive *C. trifasciata* be genotyped immediately and managed as two separate entities until the taxonomic position of lineages ''A'' and ''B'' is fully resolved.

The same issues are at stake for the newly-identified lineages of *C. pani* in this study. *Cuora pani* lineage ''C'' are sequences derived from Parham et al. (2004), and they show the sequence identity reported by those authors with *C. aurocapitata* (Fig. 3). However, *C. pani* lineage ''D'' are much more divergent from the (*C. aurocapitata* + *C. trifasciata* lineage ''A'') clade, in keeping with the result based a different individual, and different mitochondrial gene (primarily on cyt*b*) presented in Spinks et al. (2004). For *C. pani*, the picture is clouded by the divergent positions of *C. pani* 1 & 2 in the nuclear DNA data (Fig. 5). However, given that *C. pani* 1 & 2 are unique in our entire analysis in missing most of the nDNA data, we do not feel confident in their placement based on nuclear DNA. Based on the mitochondrial sequence identity of *C. pani* 1 & 2 with *C. aurocapitata* and *C. trifasciata* clade ''A'' (Figs. 3, 4), and the divergence of *C. pani* 1 & 2 from *C. pani* 3 & 4 for both mt and nDNA, we feel that the strongest current interpretation is that *C. pani* is a distinct taxon from *C. aurocapitata* and *C. trifasciata*, and that *C. pani* 3 & 4 are pure, non-hybrid animals. *Cuora pani* 1 & 2, however, may well represent hybrid mtDNA animals. Under this hypothesis, both mitochondrial and nuclear data place non-hybrid *C. pani* and *C. aurocapitata* as reciprocally monophyletic sister species.

Our sampling of live *C. zhoui* probably represents about 28% (Meier 1999 cited in CITES Proposal 11.36) of the known individuals of this species. We found virtually no mitochondrial variation among the entire group, but we did find variation at the nuclear level. Apparently, the population size of this species is small and the lack of mitochondrial genetic diversity revealed in our analyses might indicate that this species, or at least the material from which our samples were derived, has recently experienced a genetic bottleneck. Efforts should be made to genotype all available *C. zhoui* to determine the extent of genetic variation within captive animals in order to most effectively manage the captive breeding programs.

The survival of Asia's chelonian fauna in the wild will ultimately depend on curbing the demand for turtles and turtle products. Ex situ conservation actions like captive breeding and assurance colonies are emergency measures meant to maintain species until the demand for turtles decreases. However, in most cases we know very little about these species. Hybridization, whether human-mediated or natural, is clearly widespread within and among *Cuora* species, as is the retention of numts in their genomes. Our relatively large taxon sampling revealed probable introgression of *C. aurocapitata* mitochondria into both *C. trifasciata* and *C. pani*, further highlighting the importance of genetic analyses for effective ex situ conservation and management of these turtles. In addition, great care must be taken to identify orthologous from paralogous gene copies, including numts, to gain realistic insights into the genetic architecture of both captive and wild specimens.

Acknowledgments Harald Artner, Elmar Meier, and Hans-Dieter Philippen are thanked for much assistance to PQS and for facilitating sample collections. For tissue samples, we thank H. Artner, James Barzyk, Rafe Brown (University of Kansas Natural History Museum), Carla Cicero (Museum of Vertebrate Zoology), Gary Ades and Paul Crow (Kadoorie Farms and Botanical Gardens), Rick Hudson (Fort Worth Zoo), Elmar Meier, and Alan Resetar (Field Museum of Natural History). Jim Parham, Dave Starkey, Bryan Stuart, and an anonymous reviewer provided helpful comments on the manuscript. This material is based upon work supported by the National Science Foundation under grants No. DEB-0213155, DEB-0516475 and DEB-0507916, and the UC Davis Agricultural Experiment Station.

Note added in proof While our manuscript was in press, a new species of *Cuora* was described based on phenotypic variation and mitochondrial (mt) DNA sequence data [Blanck T, McCord WP, Le M (2006) On the variability of *Cuora trifasciata* (Bell, 1825). Edition Chimaira, Frankfurt.]. Blanck et al. (2006) identify a divergent mt clade of *C. trifasciata* which they described as a new species ''*Cuora cyclornata*''. However, based on our results reported here, our ongoing research, and a preliminary analysis of their mitochondrial data (provided by

Minh Le to PQ Spinks) we tentatively conclude that the mitochondrial haplotypes of ''*C. cyclornata*'' represent our clade ''B'' (i.e. non-introgressed *C. trifasciata* haplotypes). Thus, recognition of ''*C. cyclornata*'' is most likely unwarranted, since the genetic data in support of it are entirely mtDNA, and most likely represent non-introgressed *C. trifasciata* haplotypes. Therefore, we suggest that ''*C. cyclornata*'' should be considered a junior synonym of *C. trifasciata*.

Appendix

Species, sample identification, tissue type and GenBank accession numbers for all samples used in this study

Species	Tissue ID	Tissue type	COI	GenBank accession numbers				
				ND4	ND1	HNF-1alpha	R35	RELN
Cuora amboinensis 1	FMNH 255262	Soft tissue	AY357738	AY364609	EF011318	EF011278	EF011428	EF011234
Cuora amboinensis 2	KUNHM RMB4500	Soft tissue	EF011465	EF011357	EF011319	EF011279	EF011429	EF011235
Cuora aurocapitata 1	FMNH 261570	Soft tissue	AY357740	AY364606	EF011320	EF011280	EF011430	EF011236
Cuora aurocapitata 2	MVZ 234642	Soft tissue	EF011466	EF011358	EF011321	EF011281	EF011431	EF011237
Cuora bourreti 1	FMNH 261574	Soft tissue	AY357757	AY364618	EF011322	EF011282	EF011432	EF011238
Cuora bourreti 2	FMNH 261577	NA	AY357751	AY364624				
Cuora flavomarginata 1	MVZ 230464	Soft tissue	AY357739	AY364610	EF011323	EF011283	EF011433	EF011239
Cuora flavomarginata 2	HBS 41894	Blood	EF011467	EF011359				
Cuora flavomarginata 3	HBS 41895	Blood	EF011468	EF011360	EF011324	EF011284	EF011434	EF011240
Cuora galbinifrons 1	FMNH 255694	Soft tissue	AY357742	AY364612	EF011325	EF011285	EF011435	EF011241
Cuora galbinifrons 2	FMNH 256544	Soft tissue	AY357748	AY364615	EF011326	EF011286	EF011436	EF011242
Cuora galbinifrons 3	HBS 41888	Soft tissue	EF011469	EF011361				
Cuora mccordi 1	FMNH 261571	NA	AY357737	AY364608				
Cuora mccordi2	HBS 41892	Blood	EF011470	EF011362	EF011327	EF011287	EF011437	EF011243
Cuora mccordi 3	HBS 41893	Blood	EF011471	EF011363	EF011328	EF011288	EF011438	EF011244
Cuora mccordi 4	HBS 41882	Blood	EF011472	EF011364				
Cuora mouhotii 1	MVZ 230482	NA	AF348274	AF348287				
Cuora mouhotii 2	HBS 41868	Blood	EF011473	EF011365	EF011329	EF011289	EF011439	
Cuora mouhotii 3	HBS 41865	Blood	EF011474	EF011366	EF011330	EF011290	EF011440	
Cuora pani 1	MVZ 230512	Soft tissue	AY357741	AY364607	EF011331	EF011291		
Cuora pani 2	MVZ 230513	Soft tissue	AY590457	AY590461	EF011332	EF011292		
Cuora pani 3	HBS 41890	Soft tissue	EF011475	EF011367	EF011333	EF011293	EF011441	EF011245
Cuora pani4	HBS 41896	Blood	EF011476	EF011368	EF011334	EF011294	EF011442	EF011246
Cuora picturata 1	FMNH 261575	Soft tissue	AY357760	AY364628	EF011335	EF011295	EF011443	EF011247
Cuora picturata 2	FMNH 261576	NA	AY357761	AY364629				
Cuora picturata 3	HBS 38448	Soft tissue	EF011477	EF011369	EF011336	EF011296	EF011444	EF011248
***Cuora trifasciata* 1**	KFBG 021325338	Blood	EF011478	EF011370	EF011337	EF011297	EF011445	EF011249
Cuora trifasciata2	KFBG 029122546	Blood	EF011479	EF011371	EF011338	EF011298	EF011446	EF011250
***Cuora trifasciata* 3**	KFBG 030566034	Blood	EF011480	EF011372	EF011339	EF011299	EF011447	EF011251
Cuora trifasciata4	KFBG 042041021	Blood	EF011481	EF011373	EF011340	EF011300	EF011448	EF011252
Cuora trifasciata5	KFBG 042095629	Blood	EF011482	EF011374	EF011341	EF011301	EF011449	EF011253
Cuora trifasciata6	KFBG 042100272	Blood	EF011483	EF011375	EF011342	EF011302	EF011450	EF011254
Cuora trifasciata7	KFBG 042110519	Blood	EF011484	EF011376	EF011343	EF011303	EF011451	EF011255
Cuora trifasciata8	KFBG 042371302	Blood	EF011485	EF011377	EF011344	EF011304	EF011452	EF011256
Cuora trifasciata9	FWZ 994001	Blood	EF011486	EF011378				
Cuora trifasciata10	FWZ 994002	Blood	EF011487	EF011379				
Cuora trifasciata11	FWZ 994003	Blood	EF011488	EF011380				
Cuora trifasciata12	FWZ 994004	Blood	EF011489	EF011381	EF011345	EF011305	EF011453	EF011257
***Cuora trifasciata* 13**	HBS 41889	Soft tissue	EF011490	EF011382				
***Cuora trifasciata* 14**[a]	KFBG 025888102	Blood	EF011491	EF011383	EF011346	EF011306	EF011454	EF011258
***Cuora trifasciata* 15**[a]	KFBG 030556581	Blood	EF011494	EF011386	EF011347	EF011307	EF011455	EF011259
Cuora trifasciata16	KFBG 030552847	Blood	EF011497	EF011389	EF011348	EF011308	EF011456	EF011260
Cuora trifasciata17[a]	KFBG 030567630	Blood	EF011498	EF011390	EF011349	EF011309	EF011457	EF011261
Cuora trifasciata18	HBS 41891	Blood	EF011500	EF011392	EF011350	EF011310	EF011458	EF011262
Cuora trifasciata 19	MVZ 230467	NA	AF348271	AF348296				
Cuora trifasciata 20	MVZ 230636	NA	AF348270	AF348297	EF011351	EF011311	EF011459	EF011263
Cuora trifasciata21	KFBG 030278065	Blood	EF011501	EF011393	EF011352	EF011312	EF011460	EF011264
Cuora trifasciata 22	KFBG 029338867	Blood						
Cuora trifasciata 23	KFBG 042327328	Blood						
Cuora yunnanensis	MNHN 1907.10	NA	AY590460	AY572868				

continued

Species	Tissue ID	Tissue type	COI	GenBank accession numbers				
				ND4	ND1	HNF-1alpha	R35	RELN
Cuora zhoui 15	MTD T 949	NA	AY590458	AY590462				
Cuora zhoui 16	MTD T 1074	NA	AY593968	AY572865				
Cuora zhoui 17	MTD T 1075	NA	AY593969	AY572866				
Cuora zhoui 1	HBS 41855	Blood	EF011502	EF011394	EF011353	EF011313	EF011461	EF011265
Cuora zhoui 2	HBS 41856	Blood	EF011503	EF011395	EF011354	EF011314	EF011462	EF011266
Cuora zhoui 3	HBS 41857	Blood	EF011504	EF011396				
Cuora zhoui 4	HBS 41858	Blood	EF011505	EF011397				
Cuora zhoui 5	HBS 41859	Blood	EF011506	EF011398				
Cuora zhoui 6	HBS 41860	Blood	EF011507	EF011399				
Cuora zhoui 7	HBS 41861	Blood	EF011508	EF011400				
Cuora zhoui 8	HBS 41862	Blood	EF011509	EF011401	EF011355	EF011315	EF011463	EF011267
Cuora zhoui 9	HBS 41871	Blood	EF011510	EF011402				
Cuora zhoui 10	HBS 41872	Blood	EF011511	EF011403				
***Cuora zhoui* 11**	HBS 41873	Blood	EF011512	EF011404				
Cuora zhoui 12	HBS 41874	Blood	EF011513	EF011405				
Cuora zhoui 13	HBS 41875	Blood	EF011514	EF011406				
***Cuora zhoui* 14**	HBS 41877	Blood	EF011515	EF011407				
Mauremys mutica 1	MVZ 230476	Soft tissue	AF348262	AF348278	EF011316	EF011276	EF011426	EF011232
Mauremys mutica 3	MVZ 230487	Soft tissue	EF011464	EF011356	EF011317	EF011277	EF011427	EF011233

Tissues were either blood or other soft tissue (i.e. muscle or tail tip). NA indicates sequences that were generated elsewhere. COI and ND4 sequences from specimens in **Bold** initially displayed sequence heterogeneity and were subsequently sequenced from the LR4 long-range PCR template. Sequences downloaded from GenBank are underlined. All ND1 sequences were generated using genomic template except for *C. bourreti* 1 which was generated from a long range PCR product (see text). GenBank accession numbers for the cytochrome *b* and putative numt sequences are EF011268-EF011275, and EF011408-EF011425 respectively. GenBank accession numbers for CO1 and ND4 sequences generated from additional LR4 and LRW templates include COI: EF011492, EF011493, EF011495, EF011496, EF011499, and ND4: EF011384, EF011385, EF011387, EF011388, EF011391

FMNH Field Museum of Natural History (Chicago, USA), *FWZ* Fort Worth Zoo (Fort Worth, USA), *HBS* tissue collection of H. Bradley Shaffer (University of California Davis, USA), *KFBG* Kadoorie Farms and Botanical Gardens (Hong Kong, China), *KUNHM* Kansas University Natural History Museum (Lawrence, USA), *MVZ* Museum of Vertebrate Zoology (Berkeley, USA), *MTD T* Museum für Tierkunde (Dresden, Germany)

[a] Specimens that were also sequenced from the LRW long-range PCR template (see also Fig. 3 and text)

References

Artner H, Becker H, Jost, U (1998) Erstbericht über Haltung und Nachzucht der Japanischen Sumpfschildkröte *Mauremys japonica* (Temminck and Schlegel, 1835). Emys 5:5–22

Bates JM, Bowie RCK, Willard DE et al (2004) A need for continued collecting of avian voucher specimens in Africa: why blood is not enough. Ostrich 74:187–191

Bhangale TR, Rieder MJ, Livingston RJ, Nickerson DA (2005) Comprehensive identification and characterization of diallelic insertion-deletion polymorphisms in 330 human candidate genes. Hum Mol Gen 14:59–69

Brandley MC, Schmitz A, Reeder TW (2005) Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst Biol 54:373–390

Casula M, Bosboom-Dobbelaer I, Smolders K et al (2005) Infection with HIV-1 induces a decrease in mtDNA J. Infect Dis 191:1468–1471

van Dijk PP (2000) The status of turtles in Asia. In: van Dijk PP, Stuart S, Rhodin A (eds) 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. Chelonian Research Foundation. Lunenberg, pp 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. Chelonian Research Foundation, Lunenberg

Engstrom TN, Shaffer HB, McCord WP (2004) Multiple data sets, high homoplasy, and the phylogeny of softshell turtles (Testudines: Trioinychidae). Syst Biol 53:693–710

Ernst CH, Barbour RW (1989) Turtles of the world. Smithsonian Institution Press, Washington

Feldman CR, Parham JF (2004) Molecular systematics of Old World stripe-necked turtles (Testudines: *Mauremys*). Asiat Herpetol Res 10:28–37

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fritz U, Obst FJ (1997). Zum taxonomischen Status von *Cuora galbinifrons serrata* Iverson and McCord, 1992 und *Pyxidea mouhotii* (Gray, 1862) (Reptilia: Testudines: Bataguridae). Zool Abh Staatl Mus Tierk Dresden 49:261–279

Fujita MF, Engstrom TN, Starkey DE, Shaffer HB (2004) Turtle phylogeny: insights from a novel nuclear intron. Mol Phylogenet Evol 31:1031–1040

Gahan ME, Miller F, Lewin SR, Cherry CL et al (2001) Quantification of mitochondrial DNA in peripheral blood

mononuclear cells and subcutaneous fat using real-time polymerase chain reaction. J Clin Virol 22:241–247
Hazkani-Covo E, Sorek R, Graur D (2003) Evolutionary dynamics of large numts in the human genome: rarity of independent insertions and abundance of post-insertion duplications. J Mol Evol 56:169–174
Honda M, Yasukawa Y, Hirayama R, Ota H (2002a) Phylogenetic relationships of the Asian box turtles of the genus *Cuora* sensu lato (Reptilia: Bataguridae) inferred from mitochondrial DNA sequences. Zool Sci 17:1305–1312
Honda M, Yasukawa Y, Ota H (2002b) Phylogeny of the Eurasian freshwater turtles of the genus *Mauremys* Gray 1869 (Testudines) with special reference to a close affinity of *Mauremys japonica* with *Chinemys reevesii*. J Zool Syst Evol Res 40:195–200
Horlein AK, Grajer H, Igo-Kemenes T (1993) Genomic structure of the POU-related hepatic transcription factor HNF-1 alpha. Biol Chem Hoppe-Seyler 374:419–425
Hudson RR, Coyne JA (2002) Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565
Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754–755
IUCN (2006) 2006 IUCN red list of threatened species. http://www.iucnredlist.org. Downloaded on 25 June 2006
Iverson JB (1992) A revised checklist with distribution maps of the turtles of the world. Privately printed. Richmond, IN
Kou Z (1989) *Cyclemys* from Yunnan, a description of a new species and a new record to China (Testudinata: Emydidae) In: Current Herpetology in East Asia, Kyoto, pp 193–197
Lau M, Haitao S (2000) Conservation and trade of terrestrial and freshwater turtles and tortoises in the People's Republic of China. In: van Dijk P, Stuart S, Rhodin A (eds) 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. Chelonian Research Foundation, Lunenberg, pp 30–38
Lau M, Chan B, Crow P, Ades G (2000) Trade and conservation of turtles and tortoises in the Hong Kong Special Administrative Region, People's Republic of China. In: van Dijk P, Stuart S, Rhodin A (eds) 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. Chelonian Research Foundation, Lunenberg, pp 39–44
Luo B, Zong Y (1988). A new species of *Cuora–Cuora aurocapitata*. Acta Herpetologica Sinica 3:13–15
Maddison DR, Maddison WP (2003) MacClade 4: analysis of phylogeny and character evolution, version 4.06. Sinauer, Sunderland
McCord WP, Iverson JB (1991) A new box turtle of the genus *Cuora* (Testudines: Emydidae) with taxonomic notes and a key to the species. Herpetologica 47:407–420
Near TJ, Meylan PA, Shaffer HB (2005) Assessing concordance of fossil calibration points in molecular clock studies: an example using turtles. Am Nat 165:137–146
Parham JF, Li D (1999) A new locality for *Cuora pani* song 1984 with comments on its known range. Asiat Herpetol Res 8:111–113
Parham JF, Simison BW, Kozak KH et al (2001) New Chinese turtles: endangered or invalid? A reassessment of two species using mitochondrial DNA, allozyme electrophoresis and known-locality specimens. Anim Conserv 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. Proc R Soc Lond B 271:S391–S394
Parham JF, Feldman CR, Boore JR (2006) 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 Evol Biol 6:11. DOI:10.1186/1471–2148–6-11
Pereira SL, Baker AJ (2004) Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics. BMC Evol Biol 4:17
Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818
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. Mol Ecol 11:603–612
Robin ED, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. J Cell Physiol 136:507–513
Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574
Royaux I, de Rouvroit CL, D'Arcangelo G et al (1997) Genomic organization of the mouse *Reelin* gene. Genomics 46:240–250
Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
Shi H, Parham JF, Simison WB et al (2005) A report on the hybridization between two species of threatened Asian box turtles (Testudines: *Cuora*) in the wild on Hainan Island (China) with comments on the origin of '*serrata*'-like turtles. Amphibia-Reptilia 26:377–381
Song M (2001) Translation: a new species of the turtle genus *Cuora* (Testudoformes: Testudinidae). Asiat Herpetol Res 9:142–144
Sorenson MD, Quinn TW (1998) Numts: a challenge for avian systematics and population biology. Auk 115: 214–221
Spinks PQ, Shaffer HB (2005) Rangewide molecular analysis of the western pond turtle (*Emys marmorata*): cryptic variation, isolation by distance, and their conservation implications. Mol Ecol 14:2047–2064
Spinks PQ, Shaffer HB, Iverson JB, McCord WP (2004) Phylogenetic hypotheses for the turtle family Geoemydidae. Mol Phylogenet Evol 32:164–182
Starkey DE, Shaffer HB, Burke RR et al (2003) Molecular systematics, phylogeography, and the effects of Pleistocene glaciation in the painted turtle (*Chrysemys picta*) complex. Evolution 57:119–128
Stuart BL, Parham JF (2006) Recent hybrid origin of three rare Chinese turtles. Conserv Genet (in press)
Stuart BL, Parham JF (2004) Molecular phylogeny of the critically endangered Indochinese box turtle (*Cuora galbinifrons*). Mol Phylogenet Evol 31:164–177
Stuart BL, van Dijk PP, Hendrie DB (2001) Photographic guide to the turtles of Thailand, Laos, Vietnam and Cambodia. Wildlife Conservation Society, Phnom Penh
Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland
Thalman O, Hebler J, Poinar HN et al (2004) Unreliable mtDNA due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. Mol Ecol 13: 321–335

UNEP-WCMC. 6 October, 2005. UNEP-WCMC Species Database: CITES Listed Species on the World Wide Web. http://www.sea.unep-wcmc.org

Wink M, Guiking D, Fritz U (2001) Molecular evidence for hybrid origins of *Mauremys iversoni* Pritchard and McCord, 1991, and *Mauremys pritchardi* McCord, 1997 (Reptilia: Testudines: Bataguridae). Zool Abh Staatl Mus Tierk Dresden 52:41–49

Wood FE, Ebanks GK (1984) Blood cytology and hematology of the green sea turtle, *Chelonia mydas*. Herpetologica 40:331–336

Yang Z (1996) Maximum-likelihood models for combined analysis of multiple sequence data. J Mol Evol 42:587–596