SALAMANDRA **59**(4): 336–346
30 October 2023 | ISSN 0036–3375

SALAMANDRA
German Journal of Herpetology

Molecular differentiation and conservation of the Indochinese box turtles *Cuora galbinifrons*, *Cuora bourreti*, and *Cuora picturata*

Flora Ihlow[1,2*], Cäcilia Spitzweg[1*], Melita Vamberger[1*], Lauren Augustine[3,4], Cris Hagen[5], Adam Davis[6], Benjamin Leprince[7], Philipp Wagner[8], Thong Pham Van[7] & Uwe Fritz[1]

[1] Museum of Zoology, Senckenberg Dresden, A. B. Meyer Building, 01109 Dresden, Germany
[2] Chair of Computational Landscape Ecology, Technische Universität Dresden, Helmholtzstr. 10, 01069 Dresden, Germany
[3] Philadelphia Zoo, 3400 West Girard Avenue, Philadelphia, PA, 19104, USA
[4] Smithsonian National Zoological Park, 3001 Connecticut Avenue NW, Washington, DC, 20008, USA
[5] Turtle Survival Alliance, 5900 Core Road, Ste. 504, North Charleston, SC, 29406, USA
[6] Bristol Zoological Society, Bristol, BS10 7TW, UK
[7] Turtle Sanctuary and Conservation Center, rue Béranger, 75003 Paris, France
[8] Allwetterzoo Münster, Sentruper Str. 315, 48161 Münster, Germany

* The first three authors contributed equally to this paper
Corresponding Author: Flora Ihlow, ORCID-ID: 0000-0002-0460-4210, e-mail: research@floraihlow.de

Manuscript received: 24 May 2023
Accepted: 27 July 2023 by Edgar Lehr

Abstract. The Asian box turtles of the *Cuora galbinifrons* complex (*C. galbinifrons*, *C. bourreti*, and *C. picturata*) rank among the most endangered chelonian species in the world. Despite several previous studies, the phylogenetic relationships and species boundaries of this complex remain a matter for dispute due to a shortage of field-collected samples for genetic validation, observed discordance between mitochondrial and nuclear markers, and reported intergradation zones combined with a strong tendency of hybridization in the genus. Here, we re-investigate the relationships and potential hybridization between the species of the *C. galbinifrons* complex based on the most comprehensive dataset to date consisting of 394 morphologically identified specimens (136 *C. galbinifrons*, 200 *C. bourreti*, 49 *C. picturata*, and nine individuals allegedly from Hainan). The turtles mainly came from assurance colonies as well as from zoological and private collections across the USA and Europe. Bayesian and Maximum Likelihood analyses of a concatenated mitochondrial dataset (COI and ND4 plus adjacent tRNA genes) yielded almost identical topologies, supporting three major clades corresponding to *C. galbinifrons*, *C. bourreti*, and *C. picturata*, respectively. In accordance with previous studies, *C. bourreti* represented the sister clade to *C. galbinifrons* and these two clades together were sister to *C. picturata*. Haplotype networks revealed pronounced mitochondrial divergences between the taxa. STRUCTURE and PCA analyses using 12 microsatellite loci also confirmed three distinct clusters that are in agreement with the recognized species. Only a few specimens with admixed ancestry (hybrids) or mismatched mitochondrial identity were revealed, suggesting extremely limited gene flow among the three species. However, this pattern could also reflect the separate captive management of the individual taxa and an underrepresentation of geographic contact zones in our sampling.

Key words. Testudines, Geoemydidae, biogeography, conservation, genetics, endangered species, microsatellites, phylogenetics, Principal Component Analyses, Southeast Asia, STRUCTURE.

Introduction

Asian box turtles of the genus *Cuora* Gray, 1856 (Geoemydidae) inhabit the tropical forests of Southeast and East Asia. Decades of unsustainable harvest for human consumption, traditional medicine, and international trade along with habitat destruction and other anthropogenic pressures have driven most *Cuora* species to the brink of extinction (Fiebig & Lehr 2000, Lau & Shi 2000, Lehr 1997, Parham et al. 2001, TCC 2018, van Dijk 2000). Originally, *Cuora galbinifrons* Bourret, 1939 was regarded as a monotypic species which was only known from a handful of individuals (Bourret 1939, Li 1958, Petzold 1963, 1965). Specimens from Hainan were originally described by Li (1958) as a new subspecies of *C. flavomarginata* (Gray, 1863), before this subspecies was assigned to *C. galbinifrons* and either regarded as the distinct subspecies *C. g. hainanensis* (Li, 1958) (e.g., Ernst & Barbour 1989, Iverson 1986, Zhao 1986) or synonymized with *C. galbinifrons* (e.g., Iverson & McCord 1992, Lehr et al.

1998, Obst & Reimann 1994, Zong & Pan 1989). Around 1980, an increasing number of specimens started to appear in the international pet trade (Buskirk 1988, Pauler 1980) as the corollary of a non-sustainable harvest for the Chinese food market (Lehr 1996, 1997), before local populations collapsed and international protection measures were implemented (Li et al. 2020, McCormack & Stuart 2020, McCormack et al. 2020). As a by-product of this overexploitation, several new subspecies of *C. galbinifrons* were described based on material from the pet trade. Two of them are now regarded as full species (Stuart & Parham 2004, TTWG 2021), *C. bourreti* Obst & Reimann, 1994 and *C. picturata* Lehr et al., 1998, whereas *C. g. serrata* Iverson & McCord, 1992 is invalid because this taxon was based on species hybrids (Parham et al. 2001).

Using mtDNA sequences (ND4 and COI) of 26 individuals (9 *C. galbinifrons*, 5 "*C. g. serrata*", 7 *C. bourreti*, and 5 *C. picturata*) Stuart & Parham (2004) demonstrated the *C. galbinifrons* complex to contain three major monophyletic clades corresponding to *C. galbinifrons*, *C. bourreti*, and *C. picturata*. The authors identified all three as independent evolutionary lineages that should be treated as full species. Later Spinks et al. (2012) confirmed these results when examining five representatives of the *C. galbinifrons* complex (1 *C. bourreti*, 2 *C. galbinifrons*, 2 *C. picturata*) in the frame of a genus-level phylogeny. Based on phylogenetic analyses of the mitochondrial ND1 gene and 16 nuclear loci, they found all three species to be monophyletic and highly divergent from one another. In both studies, *C. bourreti* represented the sister taxon of *C. galbinifrons*, and these two taxa together were sister to *C. picturata*. However, nuclear networks yielded evidence for gene flow and/or incomplete lineage sorting, particularly between *C. bourreti* and *C. galbinifrons* (Spinks et al. 2012). In contrast, a recently published global phylogeny of turtles using 15 nuclear loci (Thomson et al. 2021) suggested *C. bourreti* as sister taxon of a clade comprised of *C. galbinifrons* and *C. picturata*, but the tree presented in the supporting information matched the topology reported in other studies, suggesting misidentification. In another study, 24 box turtles (10 *C. galbinifrons*, 9 *C. bourreti*, 5 *C. picturata*) from a Chinese turtle breeding farm were examined (Liu et al. 2019). These authors compared morphological features, as well as sequences for the complete mitochondrial genome, the mitochondrial COI gene, and one nuclear marker (Rag-1). Their results were inconclusive. Nevertheless, they concluded that *C. picturata* represents a distinct species whereas subspecies status should be assigned to *C. galbinifrons* and *C. bourreti*. Using osteological evidence, conspecificity between *C. bourreti* and *C. galbinifrons* had already been suggested earlier (Fritz et al. 2006), based on intermediate character states in putative hybrid individuals.

Presently, the *C. galbinifrons* complex is generally accepted to contain three species (TTWG 2021). The Indochinese box turtle (*C. galbinifrons*) is native to northern Vietnam, adjacent northeastern Laos, the Guangxi province of China, and Hainan Island (China). Bourret's box turtle (*C. bourreti*) is restricted to central Vietnam and adjacent eastern Laos (Stuart et al. 2011). The Southern Vietnamese box turtle (*C. picturata*) is confined to southern central Vietnam (Ly et al. 2011, TTWG 2021).

All three species still face intense collection pressure with an estimated population loss of over 90% during the last 60 years (McCormack et al. 2020, McCormack & Stuart 2020, Li et al. 2020). These taxa are classified as Critically Endangered by the International Union for Conservation of Nature since 2002 (Li et al. 2020) and have been elevated to Appendix I by the Convention on International Trade in Endangered Species in 2022 (CITES 2023). *Cuora galbinifrons*, *C. bourreti*, and *C. picturata* rank among the 25 most endangered chelonian species in the world (TCC 2018).

Despite the obvious need for effective conservation planning and management for these three species, there is still uncertainty and ongoing discussions regarding their validity. The paucity of field-collected samples for genetic investigations and the frequent hybridization reported between recognized *Cuora* taxa contribute to the confusion regarding phylogenetic relationships and species boundaries within the *C. galbinifrons* complex (Spinks et al. 2012). Legislative restrictions (e.g., the 'Convention on Biological Diversity' [CBD] and since 2014, the additional 'Nagoya Protocol on Access and Benefit-sharing') render obtaining sufficient numbers of genetic samples for research increasingly difficult (e.g., Neumann et al. 2018, Prathapan et al. 2018 and references therein). Hence, the taxonomy of the *C. galbinifrons* complex is still not sufficiently resolved, despite several previous efforts (Liu et al. 2019, Spinks et al. 2012, Stuart & Parham 2004, Thomson et al. 2021, see also Fritz et al. 2006).

In the present study, we re-investigate relationships and potential hybridization within the *C. galbinifrons* complex based on a comprehensive dataset of 394 individuals. For doing so, we combine sequences of two mtDNA fragments (COI and ND4 plus adjacent tRNA genes) with information from 12 nuclear microsatellite loci. We supplement our dataset with 36 previously published mtDNA sequences available from GenBank or provided directly by the authors (Liu et al. 2019) to allow comparison with previous studies.

Material and methods

Sampling

We compiled 394 genetic samples housed in the tissue collection of the Museum of Zoology, Senckenberg Dresden (MTD T), among them 136 morphologically identified *C. galbinifrons*, 200 *C. bourreti*, 49 *C. picturata*, and nine pet individuals allegedly from Hainan. Among these samples are a few for which the complete specimen is preserved in the herpetological collection (MTD D) of the same institute. For these specimens, we refer below to the voucher in the herpetological collection.

Due to the paucity of field-collected material, our samples were mainly obtained from captive assurance colonies

across the USA (123 samples) and Europe (234 samples, 68 of which with known origin) and supplemented with 36 samples from museum specimens (Fig. 1, Supplementary Table S1), all exported to the USA or Europe prior to the 12 October 2014 when the 'Nagoya Protocol on Access and Benefit-sharing' took effect. Our samples also include nine turtles from the international pet trade, supposedly from the putative intergradation zone of *C. galbinifrons* and *C. bourreti* on Hainan Island. Also, one field-collected sample from a turtle morphologically identified as *C. bourreti* from Quảng Bình Province, northern central Vietnam, was studied (MTD T 357). This site lies in the putative mainland hybrid zone of *C. galbinifrons* and *C. bourreti* (Fritz et al. 2002, Lehr et al. 1998). Tissue (scute clippings) or a small amount of blood (~100 µl) was collected by veterinarians or trained keepers from the jugular vein or the subcarapacial sinus using a 1 ml syringe and 22–25 g needle. Whole blood was placed in Lithium heparin microtainer tubes, while scale clippings were preserved in 90% ethanol. All samples were refrigerated until shipping to the molecular genetic laboratory of the Museum of Zoology, Senckenberg Dresden.

Laboratory procedures

Total genomic DNA of fresh tissue and blood samples was extracted using the innuPREP DNA Mini Kit and the innuPREP Blood DNA Mini Kit (Analytik Jena AG, Jena, Germany). Two mitochondrial gene fragments (ND4 including flanking tRNA genes and the barcoding gene COI), commonly used for phylogenetic and phylogeographic purposes in geoemydid turtles (e.g., Barth et al. 2004, Fritz et al. 2008, Spinks et al. 2004, Stuart & Parham 2004) were selected. Both mtDNA fragments were amplified using a PCR volume of 25 µl containing 1.25 units Taq polymerase (Bioron, Ludwigshafen, Germany) with PCR buffer 10× including $MgCl_2$, 0.2 mM of each dNTP (Thermo Scientific, St. Leon-Rot, Germany), 0.4 mM of each primer, 0.02 µg/µl bovine serum albumine (BSA; Thermo Fisher Scientific Inc., Waltham, MA, USA), ultrapure H_2O, and 10–40 ng of total genomic DNA. To amplify the ND4 fragment, the primer pair L-ND4 and H-Leu (Stuart & Parham 2004) was used, while for COI the primers L-turtCOIc and H-turtCOIc (Stuart & Parham 2004) were used. The thermocycling conditions were: initial denaturation at 94°C for 5 min, 35 cycles with denaturation at 94°C for 45 s, annealing for 45 s at 57°C and extension at 72°C for 80 s, with a final extension step at 72°C for 10 min. PCR products were purified using the ExoSAP-IT enzymatic clean-up (USB Europe GmbH, Staufen, Germany; 1:20 dilution, modified protocol: 30 min at 37°C; 15 min at 80°C) and sequenced on an ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and either the primers L-ND4int and H-Leuint for the ND4 fragment or L-turtCOIc, and H-turtCOIc for the COI fragment (Stuart & Parham 2004). Cycle sequencing reactions were run with an initial denaturation at 96°C for 60 s, and 25 cycles with denaturation at 96°C for 10 s, annealing at 50°C for 5 s, and elongation at 60°C for 76 s. For purification, 400 µl Sephadex (GE Healthcare, München, Germany; 1:20 dilution) per well in a Performa DTR V3 96-Well Short Plate was used (Edge Biosystems, Gaithersburg, MD, USA). COI sequences could be obtained for 345 turtles (fragment length 741 bp). Sequencing of ND4 (814 bp, including 152 bp of flanking tRNA genes) was successful for 328 samples (Supplementary Table S1). All sequences were checked using the original electropherograms and

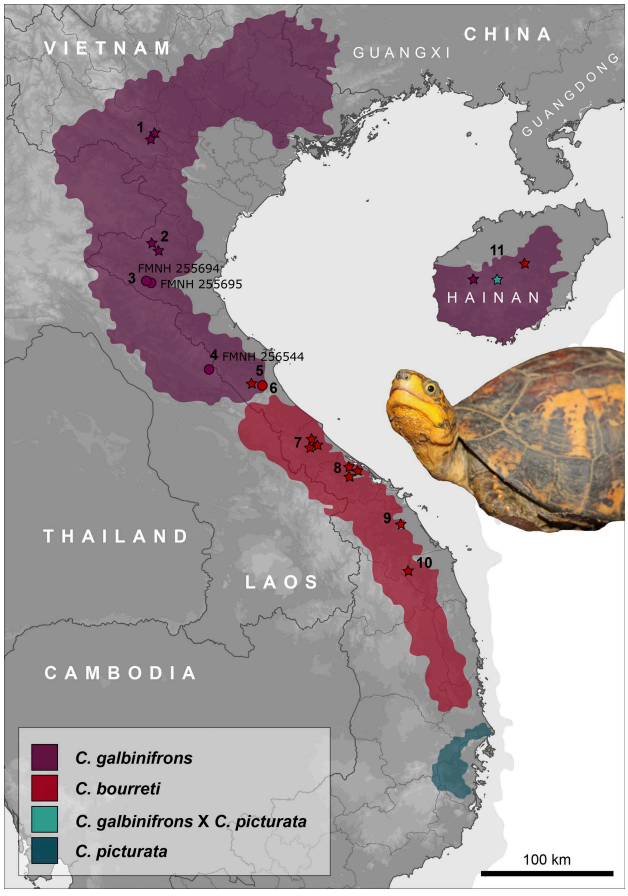

Figure 1. Distribution of the *Cuora galbinifrons* complex. Putative ranges according to TTWG (2021). Numbers refer to the following sample sites: 1: Sơn La Province, Vietnam; 2: Nghệ An Province, Vietnam; 3: Pu Mat Nature Reserve, Vietnam; 4: Khammouane Province, Lao People's Democratic Republic; 5: Quảng Bình Province, Vietnam; 6: Quảng Bình Province, Vietnam (field-collected sample); 7: Quảng Trị Province, Vietnam; 8: Thừa Thiên Huế Province, Vietnam; 9: Quảng Nam Province, Vietnam; 10: Quảng Ngãi Province, Vietnam; 11: Hainan, People's Republic of China. Except for the three FMNH samples, coloration is in accordance with microsatellite cluster assignment. Exposed landmasses during the Last Glacial Maximum are displayed in light grey. Uncertain and unreliable localities are shown as stars. Inset: *C. bourreti*, Turtle Conservation Centre (TCC), Cuc Phuong National Park, Vietnam.

subsequently manually aligned in BIOEDIT 7.0.5.2 (Hall 1999). European Nucleotide Archive (ENA) accession numbers and sampling sites are summarized in Supplementary Table S1.

In addition, 16 nuclear microsatellite loci previously used for taxon delimitation in the genus *Cuora* (Tiedemann et al. 2014) were amplified using multiplex PCRs (Supplementary Table S2). Each PCR was conducted in a final volume of 10 µl containing 0.5 units Taq polymerase (Bioron) together with the buffer 10x recommended by the supplier, 2.0 mM $MgCl_2$ (Bioron), 0.2 mM of each dNTP (Thermo Scientific), 0.02 µg/µl of BSA (Thermo Scientific), 20–40 ng of total DNA and a specific set of primers at a specific concentration as described in Supplementary Table S2. For thermocycling conditions, see Supplementary Table S3. PCR products were diluted with water in a ratio of 1:100. Fragment lengths were determined on an ABI 3730 Genetic Analyzer (Applied Biosystems) using the GeneScan–600 LIZ Size Standard (Applied Biosystems) and the software PEAK SCANNER 1.0 (Life Technologies, Carlsbad, CA). Twelve out of the 16 microsatellite loci were highly polymorphic and were used for further analyses (Supplementary Table S2).

Data analysis

Our mtDNA data were supplemented with 33 sequences available from GenBank (Supplementary Table S1). These include sequences of the only three field-collected *C. galbinifrons* samples available (FMNH 255694, FMNH 255695, FMNH 256544) from Stuart & Parham (2004). Liu et al. (2019) did not make their sequence data publicly available, but provided upon request 12 COI sequences and three mitochondrial genomes. An inspection and comparison of the COI sequences generated by Liu et al. (2019) revealed multiple alignment gaps suggestive of poor sequence quality. Since gaps may indicate sequencing errors or nuclear mitochondrial insertions (numts), we did not use these sequences for our calculations. However, we extracted COI and ND4 sequences from the three mitochondrial genomes provided by Liu et al. and incorporated these into our dataset (Supplementary Table S1).

For phylogenetic calculations, the mtDNA sequences were concatenated (421 sequences), resulting in an alignment of 1555 bp. The optimal partition scheme and the best-fitting nucleotide substitution model was inferred using PARTITIONFINDER 2 (Lanfear et al. 2012, 2016, Supplementary Table S4) and the Bayesian Information Criterion (BIC). Bayesian trees were calculated with MRBAYES 3.2.7 (Ronquist et al. 2012) and the implemented Markov chain Monte Carlo (MCMC) algorithm. Two independent runs (each with 4 chains) were performed with 10 million generations each, sampling every 500th generation, until the average standard deviation of split frequencies fell below 0.01. Results of the MCMC runs were summarized and the initial 25% of each run were discarded as burn-in.

Maximum Likelihood trees were calculated using RAxML 8.2.10 (Stamatakis 2014) using the default GTR + G model for all partitions. Five independent fast bootstrap searches were conducted to compute ML trees, starting from distinct randomized maximum parsimony trees. Subsequently, thorough bootstrap replicates were run until they converged with a cut-off of 1%. Convergences occurred after 400 replicates. TRACER 1.7.1 (Rambaut et al. 2018) was used to test for parameter convergence of both runs using the Effective Sample Sizes (ESS) of parameters prior to generating consensus trees. Homologous sequences for *Mauremys reevesii* (GenBank accession number AP019398) along with two *Cuora trifasciata* (GenBank accession numbers KF574821, NC022857) were used for rooting the phylogenetic trees.

Parsimony networks were constructed for both mtDNA fragments using TCS 1.2.3 (Clement et al. 2000), with gaps treated as fifth character state and a connection limit of 100 steps. Since missing data compromises network calculation (Joly et al. 2007), individuals represented by short sequences were excluded, resulting in 377 sequences for COI and 361 for ND4.

The 12 nuclear microsatellite loci were analyzed with an unsupervised Bayesian clustering approach as implemented in STRUCTURE 2.3.4 (Hubisz et al. 2009, Pritchard et al. 2000) using the admixture model and correlated allele frequencies. STRUCTURE searches the data set for partitions which are, as far as possible, in Hardy-Weinberg equilibrium and linkage equilibrium, and unsupervised analyses use only genetic data but no information about the collection sites or presumed taxonomic identity. For all STRUCTURE calculations, the upper bound was set arbitrarily to K = 10, and the most likely number of clusters (K) was determined using the ΔK method (Evanno et al. 2005) as implemented in STRUCTURE HARVESTER (Earl & vonHoldt 2012). Calculations were repeated 10 times for each K using a MCMC chain of 750,000 generations for each run, after a burn-in of 250,000 generations. Population structuring and individual admixture were visualized using DISTRUCT 1.1 (Rosenberg 2004).

To test the resolution power of STRUCTURE for inferring hybrid status and pure ancestry, simulations were run using HYBRIDLAB 1.0 (Nielsen et al. 2006). Twenty samples each of *C. galbinifrons*, *C. bourreti*, and *C. picturata* were selected as pure parental genotypes based on their pure microsatellite genotypes and concordant mitochondrial identities. Using these data, 20 genotypes of each hybrid class (F1, F2, and the two backcrosses) were modeled in HYBRIDLAB. Then, the obtained simulated hybrid data were subjected to analyses using STRUCTURE, together with the data of the 20 pure individuals of each species. According to our simulation analyses, a threshold of 93% (Supplementary Table S5) was used for identifying pure cluster membership, treating individuals with lower values as having mixed ancestries.

In addition, we examined microsatellite data of the inferred clusters by calculating population genetic diversity

and divergence indices. Numbers and sizes of microsatellite alleles were compared by a frequency table produced in CONVERT 1.31 (Glaubitz 2004). For inferring locus-specific observed (H_O) and expected heterozygosities (H_E) and for performing a locus-by-locus analysis of molecular variance (AMOVA, 10,000 permutations), ARLEQUIN 3.5.2.1 (Excoffier & Lischer 2010) was used. Locus-specific excess or deficiency of heterozygotes as expressed by the inbreeding coefficient F_{IS} (Weir & Cockerham 1984) was calculated with FSTAT 2.9.3.2 (Goudet 1995). The same software was also used for computing values for locus-specific allelic richness and testing statistical significance of F_{IS} for each locus and across all loci using randomizations and Bonferroni correction (Rice 1989). Two Principal Component Analyses (PCAs) were run on microsatellite genotypes to illustrate multidimensional relationships and to corroborate the STRUCTURE results without population-genetic presumptions following the recommendations by Puechmaille (2016). PCAs were calculated using the ADEGENET package (Jombart 2008) for R 3.6.3. The first PCA included only individuals identified as pure by STRUCTURE. The second PCA included all genotypes, i.e., from pure *C. galbinifrons*, *C. bourreti*, *C. picturata*, and from turtles with admixed ancestries.

Results

Our phylogenetic analyses of the concatenated mtDNA sequences revealed three major clades corresponding to *C. galbinifrons*, *C. bourreti*, and *C. picturata* (Fig. 2). Bayesian and ML approaches yielded almost identical and mostly well-supported topologies. In concordance with previous publications (Spinks et al. 2012, Stuart & Parham 2004), *C. bourreti* was retrieved as the sister clade of *C. galbinifrons* and both together were sister to *C. picturata*. The clades containing sequences of *C. galbinifrons* and *C. bourreti* showed some weak substructure.

In a few cases, the phenotypic identity conflicted with the phylogenetic placement, suggesting either misidentification or mitochondrial introgression. Sequence data extracted from one of the mt-genomes provided by Liu et al. (2019) and identified by them as *C. bourreti* clustered in the *C. galbinifrons* clade. Also, our sample MTD T 16819, morphologically identified as *C. bourreti*, was placed in the *C. galbinifrons* clade, in accordance with its microsatellite identity (Supplementary Table S1). MTD D 42945, morphologically identified as *C. bourreti*, corresponded mitochondrially to *C. picturata*. Another live turtle, according to its morphology and microsatellite identity a *C. galbinifrons* (MTD T 16738), mitochondrially matched *C. bourreti*, and two morphologically identified *C. picturata* (MTD D 42872, MTD T 18547) and a GenBank sequence (JF712890.1) of *C. picturata* clustered in the *C. bourreti* clade. The microsatellite genotypes of MTD D 42872 and MTD T 18547 support that these turtles are misidentified *C. bourreti* (Supplementary Table S1).

Eight samples from trade specimens putatively from Hainan Island were placed in the *C. galbinifrons* clade (voucher specimens MTD D 41957, 42575, 42868, 42875, 42895, 42896, 42944, and 42946). Another alleged Hainan turtle occurred, in agreement with its microsatellite genotype (Supplementary Table S1), in the *C. bourreti* clade (MTD D 42869). Our single wild-collected sample from Quảng Bình Province, northern central Vietnam (MTD T 357), was placed in the *C. bourreti* clade.

Haplotype networks showed three highly distinct haplotype clusters corresponding to the mtDNA clades in phylogenetic analyses. Each cluster was distinct by at least 16 mutational steps (Fig. 3).

For the microsatellite data, the ΔK method revealed K = 3 as the optimal number of clusters, with the three clusters matching the three currently recognized species and the distinct mtDNA clades (Fig. 4). The analyses revealed only seven specimens with less than 93% cluster membership probability, suggesting admixed ancestry. Three individuals with mixed ancestries from *C. galbinifrons* and *C. bourreti* were captive turtles from the Turtle Survival Center, USA (MTD T 16744, Q: 0.671, 0.318; MTD T 16763, Q: 0.820, 0.176) and the Allwetterzoo Münster, Germany (MTD T 18638, Q: 0.896, 0.100). All three had mitochondrial haplotypes of *C. galbinifrons* (Supplementary Table S1). A single individual from the collection of the Smithsonian National Zoological Park, USA (MTD T 16822, Q: 0.840, 0.147), was inferred to have a mixed ancestry between *C. bourreti* and *C. picturata*. In both cases, taxa are involved that theoretically could have natural contact and hybrid zones according to their distribution. Surprisingly, also three admixed turtles between the northernmost species *C. galbinifrons* and the completely allopatric southernmost species *C. picturata* were found (MTD D 42875, Q: 0.838, 0.151, trade specimen, allegedly from Hainan; MTD D 42643, Q: 0.860, 0.118, trade specimen, allegedly from China; MTD T 16840, Q: 0.845, 0.110, unknown origin, Buffalo Zoo). All three turtles yielded mitochondrial haplotypes of *C. galbinifrons* (Supplementary Table S1). Our single wild-collected sample from Quảng Bình Province, Vietnam (MTD T 357), was inferred as pure *C. bourreti*, without admixture. This is in agreement with its mitochondrial identity (Supplementary Table S1).

The two PCAs (including and excluding admixed individuals, respectively) using microsatellite genotypes confirmed three distinct clusters (Fig. 5, left and right) that agree with the distinct species, the three mtDNA clades and STRUCTURE clusters (Figs. 4, 5). In the PCA including admixed genotypes, the admixed individuals were either placed between the clusters or were retrieved within the *C. galbinifrons* cluster (Fig. 5, right).

To assess the genetic diversity within and the divergence among the three taxa using microsatellites, turtles of mixed ancestry were excluded from the dataset. The number of alleles per locus ranged from 6 to 39 (Supplementary Table S2). From a total of 216 alleles, 98 represented private alleles (Table 1, see also for further genetic

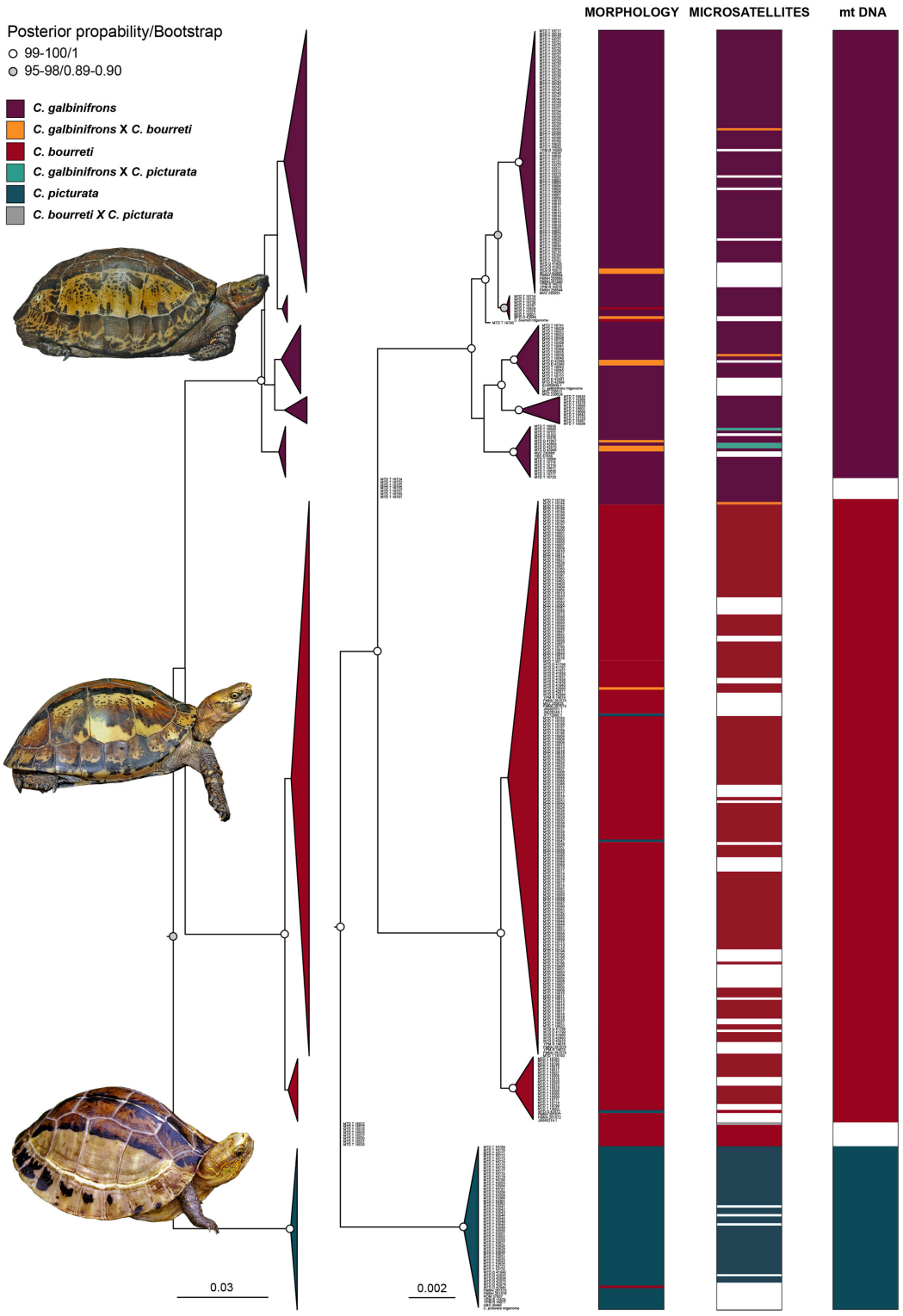

Figure 2. Bayesian consensus tree based on sequences for 415 *Cuora* samples using a concatenated mitochondrial alignment (1555 bp; 741 bp: COI, 814 bp: ND4 with flanking tRNA genes). Terminal clades of all taxa were collapsed, outgroups (*Mauremys reevesii*, *C. trifasciata*) and posterior probabilities of terminal clades not shown for clarity. Posterior probabilities exceeding 0.95 are displayed as color-coded nodes as explained in the inset. Bars display species identity for each sample according to phenotype, microsatellite analyses, and mtDNA (horizontal color sections; white sections represent missing data). Insets from top to bottom: *C. galbinifrons*, *C. bourreti*, and *C. picturata*. Photos by E. Lehr.

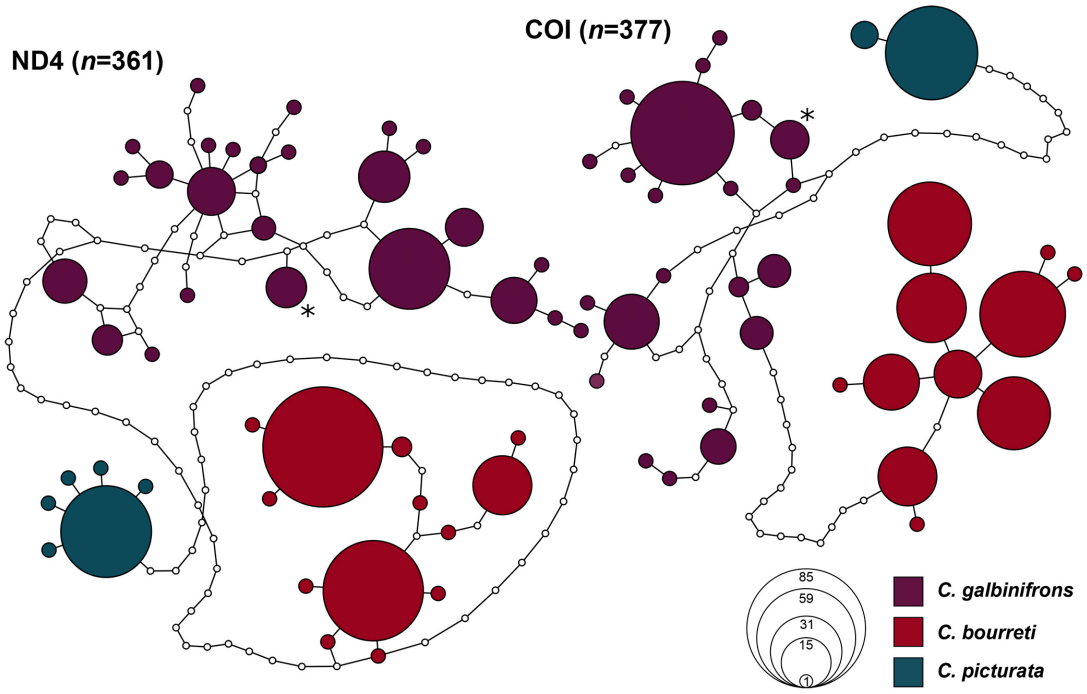

Figure 3. Parsimony networks for individual mtDNA fragments of *Cuora* samples. Symbol size corresponds to haplotype frequency; lines connecting haplotypes represent one mutation step. Small white circles are missing haplotypes. Colors of circles correspond to those of mitochondrial clades (Fig. 2). The haplotypes containing the respective mtDNA fragment of the misidentified *C. bourreti* mt-genome from Liu et al. (2019) are marked by asterisks.

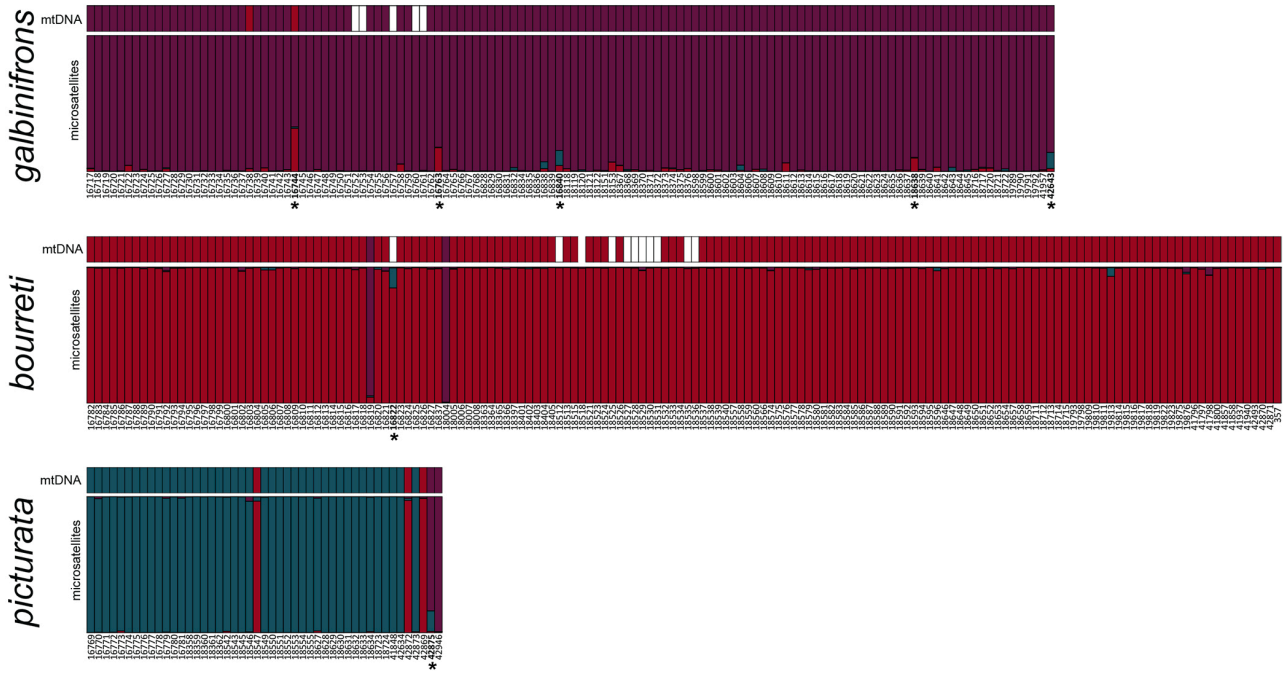

Figure 4. Genotypic structuring of 333 box turtles of the genus *Cuora* for K = 3 using 12 microsatellite loci. Shown is the STRUCTURE run with the best probability value. Within each cluster, an individual is represented by a vertical segment that reflects its ancestry. Mixed ancestries are indicated by differently colored sections corresponding to inferred genetic percentages of the respective cluster. Individuals with admixed ancestry are marked with an asterisk. For each sample, its identity according to mitochondrial lineage is shown above the STRUCTURE diagram (see also Supplementary Table S1).

Table 1. Genetic diversity indices and population genetic parameters of the pure representatives of the three *Cuora* species using 12 microsatellite loci: n = sample size; n_A = total number of alleles per cluster; $n_{\bar{A}}$ = mean number of alleles per locus; n_P = number of private alleles; AR, allelic richness; F_{IS}, inbreeding coefficient; * statistically significant; H_O, average observed heterozygosity; H_E, average expected heterozygosity; HWE, number of loci in Hardy-Weinberg equilibrium.

	n	n_A	$n_{\bar{A}}$	n_P	AR	F_{IS}	H_o	H_E	HWE
Cuora bourreti	158	165	13.5	52	10.7	0.250*	0.534	0.712	1
Cuora galbinifrons	126	148	12.33	37	9.81	0.194*	0.606	0.751	2
Cuora picturata	42	49	4.08	9	4.07	0.221*	0.254	0.326	5

Table 2. Pairwise fixation indices (F_{ST} values) for the STRUCTURE clusters (K = 3) of Indochinese box turtles. Turtles with mixed ancestries were excluded, resulting in a data set of 326 individuals. All F_{ST} values are significantly different from zero

Species	*C. bourreti*	*C. galbinifrons*	*C. picturata*
C. bourreti	–		
C. galbinifrons	0.189	–	
C. picturata	0.343	0.417	–

diversity indices and population genetic parameters). *Cuora bourreti* had the highest number of private alleles and the highest diversity indices. The highest F_{ST} value was found between *C. galbinifrons* and *C. picturata* (F_{ST} = 0.417, Table 2). The AMOVA indicated that 27.12% of the observed global variation occurred among pure representatives of the individual species and 72.88% within them.

Discussion

Analyses of two mtDNA fragments (together 1555 bp) and 12 microsatellite loci yielded for the *Cuora galbinifrons* complex largely concordant results supporting, in accordance with morphology, three distinct taxa matching the three currently recognized species. A few admixed individuals and turtles with mismatched mitochondrial identity indicate that hybridization does occasionally occur in captivity. Our only wild-collected sample from Quảng Bình Province, Vietnam, a putative natural hybrid zone of *C. galbinifrons* and *C. bourreti* (Fritz et al. 2002), was inferred as *C. bourreti* without admixture. This is in accordance with the morphological identification of the specimen. Based on shell shape and color pattern, Lehr et al. (1998) concluded that the Hainan population of what was then *C. galbinifrons* sensu lato is morphologically intermediate between *C. galbinifrons* and *C. bourreti* and suggested that this population represents a relic of a former intergradation zone that connected the two taxa when Hainan Is-

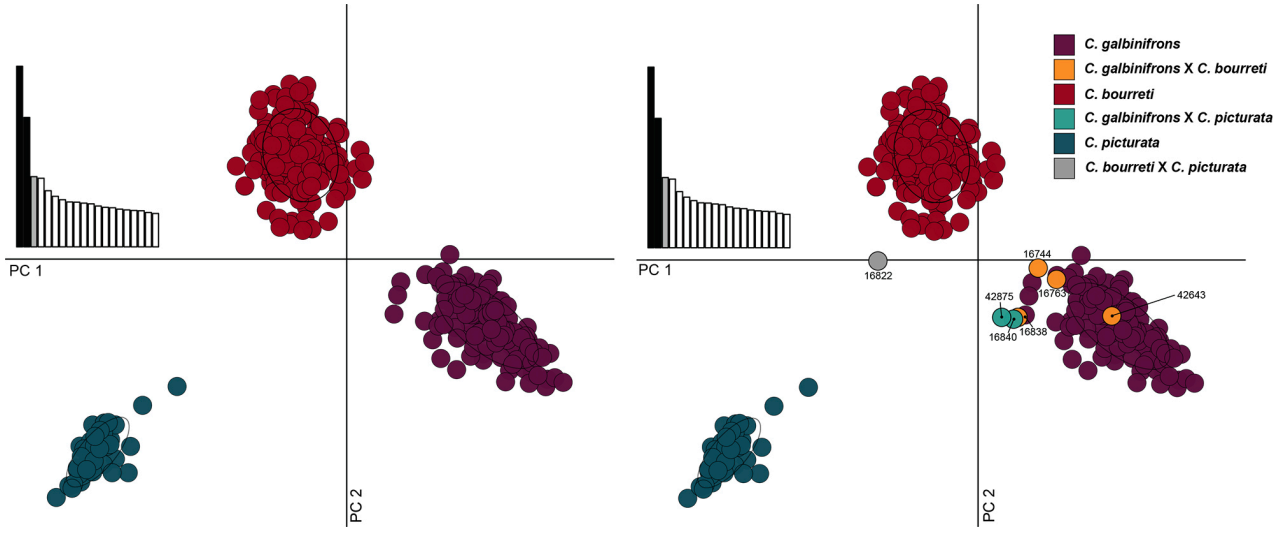

Figure 5. Principal Component Analysis (PCA) using microsatellite genotypes. Left: PC 1–2 for pure representatives of *C. bourreti*, *C. galbinifrons*, and *C. picturata*. The first, second, and third principal components explain 5.63%, 4.03%, and 2.19% of variation, respectively. Right: PC 1–2 for pure representatives of the three species plus admixed individuals. The first, second, and third principal components explain 5.54%, 3.97%, and 2.16% of the variation, respectively. Oval outlines represent 95% confidence intervals; non-overlapping lines denote significantly different clusters. Dots correspond to individual genotypes; colors according to STRUCTURE clusters.

land was part of the Asian mainland during Pleistocene low sea level stands. This hypothesis was further supported by osteological evidence from three skeletons from Hainan, with one specimen showing the character states of *C. bourreti* (MTD D 40849), the second of *C. galbinifrons* (MTD D 42895), and the third with an intermediate character state (MTD D 42896, Fritz et al. 2006). Mitochondrial DNA sequences of two of these specimens (MTD D 42895 and MTD D 42896) placed them in the *C. galbinifrons* clade, but microsatellites could not be amplified. Five trade specimens allegedly from Hainan were genotypically identified as pure *C. galbinifrons* (MTD D 41957, 42946) or *C. bourreti* (MTD D 42869), without any evidence of admixture. Another pet trade specimen allegedly from Hainan (MTD D 42875) had an admixed, but completely unexpected, ancestry and turned out as a hybrid between the northernmost species *C. galbinifrons* and the southernmost species *C. picturata*. Since these two taxa are completely allopatric, natural hybridization is impossible, suggesting that this turtle was captive-bred (e.g., in a turtle farm), where deliberate or accidental cross-breeding occurred. The unexpected absence of any signal of admixture between *C. galbinifrons* and *C. bourreti* among our few trade samples supposedly from Hainan Island does not necessarily provide evidence for the absence of hybridization. It could either be the result of faulty or intentionally fabricated localities, or caused by genetic backcrossing. Evidence of historical gene flow between distinct species is impossible to detect using microsatellites alone (Beebee & Rowe 2008, Sanz et al. 2009, Vamberger et al. 2017) because a few generations of backcrossing obscure genetic admixture. Thus, more research is needed including samples with reliable collection sites from the putative contact zones of *C. galbinifrons* and *C. bourreti* and additional marker systems (nuclear loci, SNPs, or whole genome sequencing) to either verify or refute the existence of intergradation or ancient hybridization, both on Hainan and in Vietnam.

With respect to the identified captive turtles with mixed ancestries (MTD T 16744, 16763 from the Turtle Survival Center, USA, MTD T 18638 from the Allwetterzoo Münster, Germany, and MTD T 16822 from the Smithsonian National Zoological Park, USA), we recommend that these individuals should be excluded from breeding to prevent further genetic pollution.

Acknowledgements

We thank E. Meier and C. Langner of the Internationales Zentrum für Schildkrötenschutz (IZS) at the Allwetterzoo Münster, F. Schmidt (now Zoo Basel) and A. Jacken and the Zoo Leipzig, M. Neff, S. Koeth, and B. Hughes and the TSA Turtle Survival Center, the Smithsonian National Zoo, the Buffalo Zoo, the Wildlife Conservation Society Bronx Zoo, the Fresno Chaffee Zoo, the Tennessee Aquarium, the Woodland Park Zoo, the Bristol Zoo, the Paignton Zoo, and the Chester Zoo as well as the following private keepers who supported this study by providing genetic samples: H. Becker, S. Braun, and S. Nickl. T. Ziegler provided our sole field collected sample along with valuable information on this specimen, for which we are very grateful. We thank E. Lehr for photographs to enhance our figures. FI was supported by the German Research Foundation (DFG IH 133/1-1) and the Maria Reiche Postdoctoral Fellowship of TU Dresden & the Programme for Women Professors III of the Federal Government and the States (Germany). We also thank the National Zoological Park (Smithsonian Institution) for general support in this project, and P. Joyner for her assistance during grant application. LA is grateful for financial support provided for the acquisition of genetic samples by the American Association of Zoo Keepers (AAZK) Research Grant.

References

Barth, D., D. Bernhard, G. Fritzsch & U. Fritz (2004): The freshwater turtle genus *Mauremys* (Testudines, Geoemydidae) – a textbook example of an east-west disjunction or a taxonomic misconcept? – Zoologica Scripta, **33**: 213–221.

Beebee, T. J. C. & G. Rowe (2008): An Introduction to Molecular Ecology. – Oxford University Press, Oxford.

Bourret, R. (1939): Notes herpétologiques sur l'Indochine française. XVllI. Reptiles et batraciens reçus au Laboratoire des Sciences Naturelles de l'Université au cours de l'année 1939. Descriptions de quatre espèces et d'une variété nouvelles. – Annexe au Bulletin Général de l'Instruction Publique, Hanoi, **1939**(4): 1–40.

Buskirk, J. R. (1988): The Indochinese box turtle. A perspective on captive management. – The Vivarium, Lakeside (CA), **1**(1): 22–25.

CITES (2023): Appendices I, II and III valid from 23 February 2023. – IUCN, Gland. https://cites.org/sites/default/files/eng/app/2023/E-Appendices-2023-02-23.pdf, accessed on April 04th 2023.

Clement, M. J., D. Posada & K. A. Crandall (2000): TCS: A computer program to estimate gene genealogies. – Molecular Ecology, **9**: 1657–1659.

Earl, D. A. & B. M. vonHoldt (2012): STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. – Conservation Genetic Resources, **4**: 359–361.

Ernst, C. H. & R. W. Barbour (1989): Turtles of the World. – Smithsonian Institution Press, Washington, DC.

Evanno, G., S. Regnaut & J. Goudet (2005): Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. – Molecular Ecology, **14**: 2611–2620.

Excoffier, L. & H. E. Lischer (2010): ARLEQUIN suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. – Molecular Ecology Resources, **10**: 564–567.

Fiebig, J. & E. Lehr (2000): Haltung und Erstnachzucht der Bourret-Scharnierschildkröte *Cuora galbinifrons bourreti* Obst & Reimann, 1994, mit Anmerkungen zum Bedrohungsstatus. – Salamandra, **36**: 147–156.

Fritz, U., M. Auer & A. Petzold (2006): Osteology in the *Cuora galbinifrons* complex suggests conspecifity of *C. bourreti* and *C. galbinifrons*, with notes on shell osteology and phalangeal formulae within the Geoemydidae. – Amphibia-Reptilia, **27**: 195–205.

Fritz, U., D. Guicking, M. Auer, R. S. Sommer, M. Wink & A. K. Hundsdörfer (2008): Diversity of the Southeast Asian leaf turtle genus *Cyclemys*: How many leaves on its tree of life? – Zoologica Scripta, **37**: 367–390.

Fritz, U., T. Ziegler, H. W. Herrmann & E. Lehr (2002): Intergradation between subspecies of *Cuora galbinifrons* Bourret, 1939 and *Pyxidea mouhotii* (Gray, 1862) in southern North Vietnam (Reptilia: Testudines: Geoemydidae). – Faunistische Abhandlungen, **23**: 59–74.

Glaubitz, J. C. (2004): CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. – Molecular Ecology Notes, **4**: 309–310.

Goudet, J. (1995): FSTAT (version 1.2): A computer program to calculate F-statistics. – Journal of Heredity, **86**: 485–486.

Hall, T. A. (1999): BIOEDIT: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. – Nucleic Acids Symposium Series, **41**: 95–98.

Hubisz, M. J., D. Falush, M. Stephens & J. K. Pritchard (2009): Inferring weak population structure with the assistance of sample group information. – Molecular Ecology Resources, **9**: 1322–1332.

Iverson, J. B. (1986): Checklist with distribution maps of the turtles of the world. – Privately printed, Richmond, IN.

Iverson, J. B. & W. P. McCord (1992): A new subspecies of *Cuora galbinifrons* (Testudines: Batagurinae) from Hainan Island, China. – Proceedings of the Biological Society of Washington, **105**: 433–439.

Joly, S., M. I. Stevens & B. J. van Vuuren (2007): Haplotype networks can be misleading in the presence of missing data. – Systematic Biology, **56**: 857–862.

Jombart, T. (2008): ADEGENET: A R package for the multivariate analysis of genetic markers. – Bioinformatics, **24**: 1403–1405.

Lanfear, R., B. Calcott, S. Y. Ho & S. Guindon (2012): PARTITIONFINDER: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. – Molecular Biology and Evolution, **29**: 1695–1701.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld & B. Calcott (2016): PARTITIONFINDER 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. – Molecular Biology and Evolution, **34**: 772–773.

Lau, M. & H. T. Shi (2000): Conservation and trade of terrestrial and freshwater turtles and tortoises in the People's Republic of China. – Chelonian Research Monographs, **2**: 30–38.

Lehr, E. (1996): Zur innerartlichen Variabilität der hinterindischen *Cuora galbinifrons* (Testudines: Emydidae; Bourret, 1939) untersucht mit biogeographischen, morphologischen und genetischen Methoden. – Diplomarbeit, Institut für Genetik (Fachbereich Biologie), Johannes-Gutenberg-Universität, Mainz.

Lehr, E. (1997): Untersuchungen zum Schildkrötenhandel in Vietnam zwischen 1993 und 1996. – Mitteilungen der Zoologischen Gesellschaft für Arten- und Populationsschutz, München, **13**: 12–16, 19.

Lehr, E., U. Fritz & F. J. Obst (1998): Die Unterarten von *Cuora galbiniforns* Bourret, 1939. – Zoologische Abhandlungen, **50**: 77–97.

Li, P., T. McCormack, D.-Q. Rao, H. Shi, B. L. Stuart & L. Wang (2020): *Cuora galbinifrons*. – The IUCN Red List of Threatened Species, **2020**: e.T97357437A123816666. – https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T97357437A123816666.en, accessed on March 14th 2023.

Li, Z.-X. (1958): [Research report on the reptiles of Hainan Island]. – Chinese Journal of Zoology, **2**: 234–239 [in Chinese].

Liu, X., W. Li, Z. Ye, Y. Zhu, X. Hong & X. Zhu (2019): Morphological characterization and phylogenetic relationships of Indochinese box turtles – the *Cuora galbinifrons* complex. – Ecology and Evolution, **9**: 13030–13042.

Ly, T., H. D. Hoang & B. L. Stuart (2011): Market turtle mystery solved in Vietnam. Biological Conservation, **144**, 1767–1771.

McCormack, T. & B. Stuart (2020): *Cuora bourreti*. – The IUCN Red List of Threatened Species, **2020**: e.T163447A123815632. – https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T163447A123815632.en, accessed on March 14th 2023.

McCormack, T., B. Stuart & T. Blanck (2020): *Cuora picturata*. – The IUCN Red List of Threatened Species, **2020**: e.T163463A123815952. – https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T163463A123815952.en, accessed on March 14th 2023.

Neumann, D., A. V. Borisenko, J. A. Coddington, C. L. Häuser, C. R. Butler, A. Casino, J. C. Vogel, G. Haszprunar & P. Giere (2018): Global biodiversity research tied up by juridical interpretations of access and benefit sharing. – Organisms Diversity & Evolution, **18**: 1–12.

Nielsen, E. E., L. A. Bach & P. Kotlicki (2006): HYBRIDLAB (version 1.0): A program for generating simulated hybrids from population samples. – Molecular Ecology Notes, **6**: 971–973.

Obst, F. J. & M. Reimann (1994): Bemerkenswerte Variabilität bei *Cuora galbinifrons* Bourret, 1939, mit Beschreibung einer neuen geographischen Unterart: *Cuora galbinifrons bourreti* subsp. nov. – Zoologische Abhandlungen, **48**: 125–138.

Parham, J. F., W. B. Simison, K. H. Kozak, C. R. Feldman & H. Shi (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.

Pauler, I. (1980): Die Schildkrötengattung *Cuora*. – Herpetofauna, **6**: 15–18.

Petzold, H.-G. (1963): Über einige Schildkröten aus Nord-Vietnam im Tierpark Berlin. – Senckenbergiana biologica, **44**: 1–20.

Petzold, H.-G. (1965): *Cuora galbinifrons* und andere südostasiatische Schildkröten im Tierpark Berlin. – Die Aquarien- und Terrarienzeitschrift (DATZ), **18**: 87–91, 119–121.

Prathapan, K. D., R. Pethiyagoda, K. S. Bawa, P. H. Raven, P. D. Rajan & 172 co-signatories from 35 countries (2018): When the cure kills – CBD limits biodiversity research. – Science, **360**: 1405–1406.

Pritchard, J. K., M. Stephens & P. Donnelly (2000): Inference of population structure using multilocus genotype data. – Genetics, **155**: 945–959.

Puechmaille, S. J. (2016): The program STRUCTURE does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. – Molecular Ecology Resources, **16**: 608–627.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele & M. A. Suchard (2018): Posterior summarization in Bayesian phylogenetics using TRACER 1.7. – Systematic Biology, **67**: 901.

Rice, W. R. (1989): Analyzing tables of statistical tests. – Evolution, **43**: 223–225.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, M. A. Suchard & J. P. Huelsenbeck (2012): MRBAYES 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. – Systematic Biology, **61**: 539–542.

Rosenberg, N. A. (2004): DISTRUCT: A program for the graphical display of population structure. – Molecular Ecology Notes, **4**: 137–138.

Sanz, N., R. M. Araguas, R. Fernández, M. Vera & J.-L. García-Marín (2009): Efficiency of markers and methods for detecting hybrids and introgression in stocked populations. – Conservation Genetics, **10**: 225–236.

Spinks, P. Q., H. B. Shaffer, J. B. Iverson & W. P. McCord (2004): Phylogenetic hypotheses for the turtle family Geoemydidae. – Molecular Phylogenetics and Evolution, **32**: 164–182.

Spinks, P. Q., R. C. Thomson, Y. Zhang, J. Che, Y. Wu & H. B. Shaffer (2012): Species boundaries and phylogenetic relationships in the critically endangered Asian box turtle genus *Cuora*. – Molecular Phylogenetics and Evolution, **63**: 656–667.

Stamatakis, A. (2014): RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. – Bioinformatics, **30**: 1312–1313.

Stuart, B. L., C. D. Hallam, S. Sayavong, C. Nanthavong, S. Sayaleng, O. Vongsa & W. G. Robichaud (2011): Two additions to the turtle fauna of Laos. – Chelonian Conservation and Biology, **10**: 113–116.

Stuart, B. L. & J. F. Parham (2004): Molecular phylogeny of the critically endangered Indochinese box turtle (*Cuora galbinifrons*). – Molecular Phylogenetics and Evolution, **31**: 164–177.

TCC [Turtle Conservation Coalition] (2018): Turtles in Trouble: The World's 25+ Most Endangered Tortoises and Freshwater Turtles – 2018. – IUCN SSC Tortoise and Freshwater Turtle Specialist Group, Turtle Conservancy, Turtle Survival Alliance, Turtle Conservation Fund, Chelonian Research Foundation, Conservation International, Wildlife Conservation Society, and Global Wildlife Conservation, Ojai, CA.

Thomson, R. C., P. Q. Spinks & H. B. Shaffer (2021): A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins. – Proceedings of the National Academy of Sciences of the U.S.A., **118**: e2012215118.

Tiedemann, R., A. R. Schneider, K. Havenstein, T. Blanck, E. Meier, M. Raffel, H. Zwartepoorte & M. Plath (2014): New microsatellite markers allow high-resolution taxon delimitation in critically endangered Asian box turtles, genus *Cuora*. – Salamandra, **50**: 139–146.

TTWG [Turtle Taxonomy Working Group] (2021): Turtles of the world. Annotated checklist and atlas of taxonomy, synonymy, distribution, and conservation status (9th edition). – Chelonian Research Monographs, **8**: 1–472.

Vamberger, M., H. Stuckas, M. Vargas-Ramírez, C. Kehlmaier, D. Ayaz, A. A. Aloufi, P. Lymberakis, P. Široký & U. Fritz (2017): Unexpected hybridization patterns in Near Eastern terrapins (*Mauremys caspica, M. rivulata*) indicate ancient gene flow across the Fertile Crescent. – Zoologica Scripta, **46**: 401–413.

van Dijk, P. P. (2000): The status of turtles in Asia. – Chelonian Research Monographs, **2**: 15–23.

Weir, B. S. & C. C. Cockerham (1984): Estimating F-statistics for the analysis of population structure. – Evolution, **38**: 1358–1370.

Zhao, E. (1986): A revised catalogue of Chinese tortoises. – Acta Herpetologica Sinica, **5**: 145–148.

Zong, Y. & L. Pan (1988): Studies on the genus *Cuora* of the Testudoformes. – Current Herpetology in East Asia, **1**: 198.

Supplementary data

The following data are available online:

Supplementary Table S1. Studied samples and GenBank sequences.

Supplementary Table S2. PCR number, microsatellite loci, fluorescent labels, primer concentration, number of alleles, and allele size ranges.

Supplementary Table S3. Thermocycling conditions for microsatellites.

Supplementary Table S4. Optimal partition scheme and the best-fitting evolutionary model.

Supplementary Table S5. Simulated data.