JOURNAL OF ZOOLOGICAL SYSTEMATICS AND EVOLUTIONARY RESEARCH

Accepted on 18 January 2010
J Zool Syst Evol Res doi: 10.1111/j.1439-0469.2010.00565.x

[1]*Museum of Zoology (Museum für Tierkunde), Senckenberg Dresden, Germany;* [2]*Evolutionary Genomics Group, Department of Botany and Zoology, University of Stellenbosch, South Africa;* [3]*Chelonian Biodiversity and Conservation, Department of Biodiversity and Conservation Biology, University of the Western Cape, Bellville, South Africa;* [4]*c/Monte 43 Bajos, Hospitalet (Barcelona), Spain;* [5]*Fundación AndígenA, Mérida, Venezuela;* [6]*Department of Biology and Wildlife Diseases, Faculty of Veterinary Hygiene and Ecology, University of Veterinary and Pharmaceutical Sciences, Brno, Czech Republic*

Mitochondrial phylogeography and subspecies of the wide-ranging sub-Saharan leopard tortoise *Stigmochelys pardalis* (Testudines: Testudinidae) – a case study for the pitfalls of pseudogenes and GenBank sequences

Uwe Fritz[1], Savel R. Daniels[2], Margaretha D. Hofmeyr[3], Juan González[4], César L. Barrio-Amorós[5], Pavel Široký[6], Anna. K. Hundsdörfer[1] and Heiko Stuckas[1]

Abstract

The leopard tortoise (*Stigmochelys pardalis*) is the most widely distributed sub-Saharan tortoise species, with a range extending from the Horn of Africa all over eastern Africa to the Republic of South Africa, Namibia and southernmost Angola. Using 1938 bp of mitochondrial DNA (cyt *b* gene, partial ND4 gene plus adjacent tRNA genes) from a nearly range-wide sampling, we examined its phylogeographic structure and compared our findings with previously published GenBank sequences. We identified seven major clades that are largely parapatrically distributed. A few records of distinct haplotypes at the same locality or in close proximity could be the result of translocation of tortoises by man. The greatest diversity occurs in the south of the species' range, with five out of the seven clades. Testing for isolation-by-distance suggests that the observed phylogeographic structure is the result of restricted geographical gene flow and not of historical vicariance. This is in sharp contrast to wide-ranging thermophilic reptiles from the western Palaearctic, whose phylogeographic structure was significantly shaped by Pleistocene range interruptions, but also by earlier dispersal and vicariant events. Most cyt *b* sequences of *S. pardalis* from GenBank turned out to be nuclear pseudogenes, or to be of chimerical origin from such pseudogenes and authentic mitochondrial sequences, which argues for caution regarding uncritical usage of GenBank sequences. The recent revalidation of the two subspecies of *S. pardalis* was based on such a chimerical sequence that was erroneously identified with the subspecies *S. p. babcocki*. Furthermore, according to our data, the distribution of mitochondrial clades does match neither the traditional subspecies ranges nor the pronounced geographical size variation of leopard tortoises. We conclude that there is no rationale for recognizing subspecies within *S. pardalis*.

Key words: *Stigmochelys* – Africa – isolation-by-distance – numts – phylogeography – pseudogenes – subspecies

Introduction

While mitochondrial phylogeographies of wide-ranging Palaearctic and Nearctic reptiles have yielded significant insights into their Pleistocene and Holocene range dynamics (reviews in Avise et al. 1998; Soltis et al. 2006; Joger et al. 2007), such data are unavailable for any widely distributed Afrotropical species. In sub-Saharan Africa, and especially in its south, land tortoises (family Testudinidae) constitute an important part of the local fauna (Branch 1998; Boycott and Bourquin 2000). With 18 currently recognized species in seven genera (Branch 2007; Fritz and Havaš 2007), sub-Saharan Africa harbours the world's greatest diversity of tortoises. The leopard tortoise, *Stigmochelys pardalis* (Bell, 1828), is the most widely distributed species, with a range extending from the Horn of Africa all over eastern Africa to the Republic of South Africa, Namibia and southernmost Angola (Fig. 1; Iverson 1992; Ernst et al. 2000; Fritz and Havaš 2007). *Stigmochelys pardalis* is well known for considerable geographical variation in body size, shell shape and colouration (Fig. 2; Table S1); these differences have been attributed either to subspecific variation (Loveridge 1935; Loveridge and Williams 1957) or the influence of latitudinal and environmental impact (Lambert 1995; Lambert et al. 1998; Hailey and Lambert 2002). For decades, two subspecies have been recognized, *Stigmochelys pardalis pardalis* (Bell, 1828), being confined to the southwest of the range, while *S. p. babcocki* (Loveridge, 1935) was thought to occupy the remainder (Fig. 1; Loveridge and Williams 1957; Wermuth and Mertens 1977; Ernst and Barbour 1989). Since the putatively diagnostic characters of shell shape and colouration had little value (Greig and Burdett 1976), the usage of subspecies was later abandoned (Boycott and Bourquin 2000; Ernst et al. 2000). But a recent molecular phylogeny of most land tortoise species found a significant difference between mitochondrial cytochrome *b* sequences of two individuals of *S. pardalis*, allegedly representing each subspecies, leading to their resurrection (Le et al. 2006). Yet, the leopard tortoises studied by Le et al. (2006) were of unknown geographical provenance (M. Le, personal communication), so that their taxonomic assignment should be treated with caution, and a reassessment using individuals of known geographical origin has to be undertaken.

Here, we present a nearly range-wide phylogeography of the leopard tortoise based on two mitochondrial DNA fragments (ND4, cyt *b*; total: 1938 bp) that have been shown to be highly informative for phylogeographic purposes in

Corresponding author: Uwe Fritz (uwe.fritz@senckenberg.de)
Contributing authors: Savel R. Daniels (srd@sun.ac.za), Margaretha D. Hofmeyr (mdhofmeyr@uwc.ac.za), Juan González (eurorep@arrakis.es), César L. Barrio-Amorós (cesarba@yahoo.com), Pavel Široký (sirokyp@vfu.cz), Anna K. Hundsdörfer (anna.hundsdoerfer@senckenberg.de) and Heiko Stuckas (heiko.stuckas@senckenberg.de)

Fig. 1. Distribution range of *Stigmochelys pardalis* and sampling sites. Light shading corresponds to putative range of *S. p. pardalis*; dark shading, *S. p. babcocki*. Symbols of sampling sites indicate distinct haplotype clusters and clades (Figs 4 and 5). Large symbols, imprecise localities; star, collection site of individual yielding GenBank sequence DQ080041 (Awdal Region, Somalia). Site numbers correspond to Table 1

chelonians (Lenk et al. 1999; Fritz et al. 2005, 2006a,b, 2007, 2008, 2009a,b; Daniels et al. 2007, 2010; Rosenbaum et al. 2007; Amato et al. 2008). By doing so, we specifically address the following questions: (1) Does phylogeographic structuring exist in *S. pardalis*, the most widely distributed sub-Saharan tortoise species? (2) If so, is vicariance or isolation-by-distance responsible for the observed differentiation? (3) Do phylogeographic breaks match the putative subspecies borders and, consequently, (4) do the subspecies constitute evolutionarily significant units? (5) How do previously published GenBank sequences fit into our phylogeographic scenario?

Materials and Methods

Sampling, DNA extraction, PCR and sequencing

Blood, muscle tissue or saliva samples of wild or captive known-locality leopard tortoises were taken, preserved in absolute ethanol

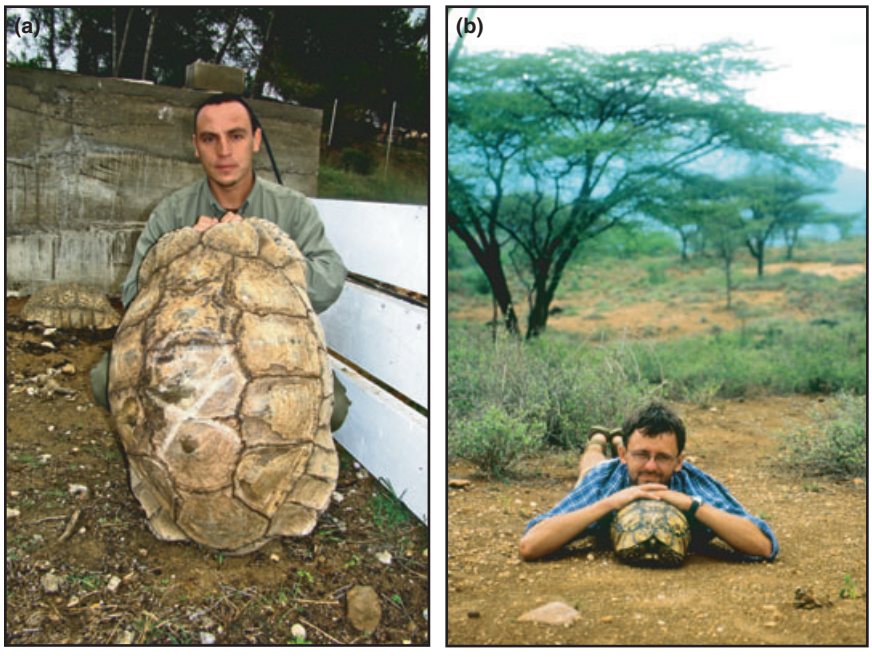

Fig. 2. Size differences in adult *Stigmochelys pardalis* from (a) Ethiopia, vicinity of Lake Metehara and (b) Kenya, South Horr. (Photographs taken by J. González and P. Nečas, respectively)

doi: 10.1111/j.1439-0469.2010.00565.x

(EtOH) and stored at −20°C until processing. Four samples were from tortoises of unknown provenance. In total, 43 samples representing 20 different collection sites were studied (Table 1).

Total DNA was extracted using standard protocols. To amplify the mitochondrial cytochrome *b* (cyt *b*) gene as well as the partial NADH dehydrogenase subunit 4 (ND4) gene plus adjacent tRNA genes, an array of newly developed and previously published primers was employed (Table S2). While PCR and sequencing of the ND4 fragment could be performed routinely, problems arose during processing the cyt *b* gene. Initially, only chelonian standard primers were applied, but these yielded contradictory results (see Results and Discussion). Therefore, the new primer pair CytB-F-pard/CytB-R-pard was designed to specifically target the conserved regions of the cyt *b* flanking $tRNA^{Gln}$ and $tRNA^{Thr}$ genes of the complete mitochondrial genomes of all testudinid species available from GenBank (*Malacochersus tornieri*, accession number DQ080042; *Manouria emys*, DQ080040; *Manouria impressa*, EF661586; *Indotestudo elongata*, DQ080043; *Indotestudo forstenii*, DQ080044; *Stigmochelys pardalis*, DQ080041; *Testudo graeca*, DQ080049; *Testudo horsfieldii*, DQ080045; *Testudo kleinmanni*, DQ080048; *Testudo marginata*, DQ080047), with an optimal fit to the *S. pardalis* sequence DQ080041 (Parham et al. 2006). Ironically, just this *S. pardalis* sequence subsequently turned out to be of chimerical origin (see Results).

PCR was performed using 1 unit *Taq* polymerase (Bioron, Ludwigshafen, Germany) with the following conditions: 35–40 cycles with denaturation at 94°C (5 min in the first cycle, then 45 s each), annealing for 45 s at fragment-specific temperature (Table S2) and extension at 72°C for 90 s and in the final cycle for 10 min. PCR products were purified using the ExoSAP-IT enzymatic cleanup (USB Europe GmbH, Staufen, Germany; 1 : 20 dilution; modified protocol: 30 min at 37°C, 15 min at 80°C).

PCR products were sequenced on an ABI 3130 genetic analyser (Applied Biosystems, Foster City, USA) using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and the primers indicated in Table S2. Cycle sequencing reaction products were purified by precipitation under the following conditions: 1 volume PCR product (10 μl), 0.1 volume 3 M NaAc (1 μl, pH 4.6) and 2.5 volumes EtOH (100%; 27.5 μl). For accession numbers of sequences produced in the present study, see Table 1.

Sequence analyses

All available cyt *b* and ND4 sequences of *Stigmochelys pardalis* were downloaded from GenBank (Table 2) and aligned with sequences produced in the present study using BIOEDIT 7.0.5.2 (Hall 1999). Alignments were further inspected in MEGA 4.0.2 (Tamura et al. 2007). In addition, cyt *b* sequences of about two-thirds of all extant testudinid species were downloaded from GenBank (for accession numbers, see Le et al. 2006; Fritz and Bininda-Emonds 2007) and compared with our data. Such a data set is not available for the mtDNA fragment comprising the partial ND4 gene.

Phylogenetic analyses were run for several alignments. One data set included all distinct cyt *b* haplotypes of *S. pardalis* identified in this study and from GenBank plus the cyt *b* sequences of all tortoise species constituting together with *S. pardalis* an African-Malagasy-Indian Ocean clade. Sequences of *Testudo horsfieldii* and *Manouria impressa* were added as distantly related taxa; *M. impressa* was used for rooting the tree (for accession numbers, see Fig. 3). The other data sets included only authentic mitochondrial haplotypes of *S. pardalis* produced in this study (see Results); sequences of *Psammobates tentorius* (ND4: AY673506, cyt *b*: DQ497318) and *T. horsfieldii* (DQ080045) served as outgroups. The genus *Psammobates* represents the sister group to *Stigmochelys*, while *T. horsfieldii* is a distantly related tortoise species (Fritz and Bininda-Emonds 2007). Heterogeneity of the two partitions (ND4, cyt *b* fragments) was assessed with the incongruence length difference test (Farris et al. 1995) as implemented in PAUP*4.0b10 (Swofford 2002; settings add = cl nreps = 100000). The test suggested significantly different phylogenetic signals (p = 0.001), which is why we computed trees for the distinct haplotypes of each partition alone as well as for a data set of concatenated haplotypes, acknowledging that the mitochondrial genome represents one and the same locus.

For each data set, Maximum Parsimony (MP) and Maximum Likelihood (ML) trees were calculated using the heuristic search option in PAUP* (commands pset gaps = new and hs add = cl for MP); parsimony statistics are summarized in Table S3. The best evolutionary model was established using MODELTEST 3.06 (Posada and Crandall 1998; Table S4). Bootstrap support was obtained using the additional setting nreps = 1000 (MP) with PAUP* as well as for ML with nreps = 1000 in the program GARLI 0.95 (Zwickl 2006). Each data set was also analyzed using Bayesian analysis (BA) as implemented in MRBAYES 3.1.2 (Ronquist and Huelsenbeck 2003; settings: ngen = 1000000 nchains = 4 nrun = 2 sample = 500 temp = 0.1; default priors; for the concatenated partitions using mixed-model approach). After discarding the first non-plateau trees applying an adequate burn-in, a majority rule consensus tree with posterior probabilities was computed.

Since infraspecific gene genealogies are often incompletely reflected by bifurcating trees because of the persistence of ancestral haplotypes and reticulate relationships (Posada and Crandall 2001), we also calculated haplotype networks using TCS 1.21 (Clement et al. 2000) for the combined data set. This software is based on statistical parsimony and connects haplotypes via a minimal number of mutational steps and allows for alternative pathways. A further advantage of such network analyses is that information about the age of haplotypes may be obtained. Interiorly located haplotypes, having more than one mutational connection, are thought to be ancestral to, and older than, tip haplotypes (Posada and Crandall 2001). TCS determines also the outgroup probability of each haplotype, which is, according to coalescent theory, also correlated with haplotype age (Donnelly and Tavaré 1986; Castelloe and Templeton 1994).

Furthermore, for the protein-coding parts of the mtDNA fragments of *S. pardalis*, the following molecular indices were computed using the software DNASP 5.00.07 (Rozas et al. 2003): haplotype diversity (H_d), number of segregating sites (S), Jukes and Cantor corrected nucleotide diversity at all sites (π), Jukes and Cantor corrected nucleotide diversity at synonymous sites (π_s) and Jukes and Cantor corrected nucleotide diversity at non-synonymous sites (π_a).

To determine whether the observed sequence variation can be attributed to vicariance or a differentiation process caused by isolation-by-distance, the following partitions were examined: (1.1) all known-locality samples and (1.2) the samples of perhaps allochthonous tortoises (Oranjemund, Benfontein, Calitzdorp; see Discussion) excluded from this data set; (2.1) a subset of only the samples from South Africa and Namibia and (2.2) the samples of perhaps allochthonous tortoises (Oranjemund, Benfontein, Calitzdorp) excluded from this subset of southern samples; (3) another subset comprising only the samples from Ethiopia, Kenya, Tanzania, Uganda and Zambia. Pairwise genetic distances between individual samples were estimated in PAUP* 4.0b10 using combined sequences of the cyt *b* and ND4 fragments based on the HKY + I + G model as determined by MODELTEST (ML settings base frequency = 0.3347 0.2793 0.1200; Nst = 2; TRatio = 7.5566; Rates = gamma; Shape = 0.7047; Pinvar = 0.8242). Geographical distances were calculated based on co-ordinates (Table 1) as shortest line between two localities using a internet platform for latitude/longitude distance calculation (http://jan.ucc.nau.edu/~cvm/latlongdist.html) and subsequently log-transformed as suggested by Rousset (1997). Correlations between genetic and geographical distances were determined using Mantel tests as implemented in the R package APE (Paradis et al. 2004); p values were obtained using 10^4 simulations.

Results

Identification of authentic sequences, pseudogenes and chimerical sequences

Amplification and sequencing of the mtDNA fragment containing the partial ND4 gene produced unambiguous results. Accordingly, GenBank sequences of the ND4 gene are identical with or closely resemble our data (Table 2).

By contrast, problems occurred during the processing of the cyt *b* gene. Initially, the chelonian standard primers

doi: 10.1111/j.1439-0469.2010.00565.x

Table 1. Samples of *Stigmochelys pardalis* used in the present study and their mitochondrial and pseudogene haplotypes (see text). Site numbers refer to the map (Fig. 1). Samples are housed in the collections of the Department of Botany and Zoology, University of Stellenbosch, South Africa (sample codes starting with Sp) and the Museum of Zoology, Senckenberg Dresden, Germany (sample codes starting with MTD)

Locality	Sample	Mitochondrial haplotype	Accession numbers (cyt *b*)	Accession numbers (ND4)	Pseudogene haplotype	Accession number
1 – Ethiopia: E Addis Ababa; 8°51′N 40°05′E*	MTD T 5447 (S)	IIb	FN646114	FN646157	–	–
1 – Ethiopia: E Addis Ababa; 8°51′N 40°05′E*	MTD T 5448 (S)	IIb	FN646115	FN646158	–	–
2 – Kenya: Seredupi, between Isiolo and Marsabit; 1°07′N 37°36′E	MTD T 5676 (B)	IIa	FN646116	FN646159	PG7	FN650623
3 – Kenya: Samburu Hills: Morijo; 1°14′N 36°36′E	MTD T 5496 (B)	IIc	FN646117	FN646160	PG3	FN650624
4 – Uganda; approx. 0°23′N 32°00′E*	MTD T 1006 (T)	Ia	FN646118	FN646161	PG6	FN650625
5 – Tanzania: northern Serengeti: Klein's Camp; 1°54′S 35°02′E	Sp13 (B)	Ia	FN646119	FN646162	–	–
6 – Tanzania; approx. 7°00′N 36°00′E*	MTD T 674 (T)	Ia	FN646120	FN646163	–	–
7 – Zambia; approx. 14°00′S 29°00′E*	MTD T 5528 (B)	Ia	FN646121	FN646164	PG6	FN650626
8 – Namibia: Cimarron; 24°18′S 17°44′E	Sp18 (B)	VIb	FN646122	FN646165	PG5	FN650627
8 – Namibia: Cimarron; 24°18′S 17°44′E	Sp19 (B)	VId	FN646123	FN646166	PG5	FN650628
9 – Namibia: Congella; 25°39′S 16°57′E	Sp21 (B)	VIIa	FN646124	FN646167	PG8	FN650629
9 – Namibia: Congella; 25°39′S 16°57′E	Sp20 (B)	VIIb	FN646125	FN646168	PG1	FN650630
9 – Namibia: Congella; 25°39′S 16°57′E	Sp22 (B)	VIIc	FN646126	FN646169	PG2	FN650631
9 – Namibia: Congella; 25°39′S 16°57′E	Sp23 (B)	VIId	FN646127	FN646170	PG8	FN650632
10 – Namibia: Oranjemund; 28°33′S 16°24′E	Sp1 (B)	IIId	FN646128	FN646171	–	–
11 – South Africa: Springbok; 29°41′S 17°48′E	Sp4 (B)	IVa	FN646129	FN646172	–	–
12 – South Africa: Groveput; 29°47′S 22°28′E	Sp11 (B)	VIa	FN646130	FN646173	–	–
13 – South Africa: Benfontein; 28°48′S 24°49′E	Sp15 (B)	Vc	FN646131	FN646174	–	–
13 – South Africa: Benfontein; 28°48′S 24°49′E	Sp17 (B)	Vc	FN646132	FN646175	–	–
13 – South Africa: Benfontein; 28°48′S 24°49′E	Sp16 (B)	Vd	FN646133	FN646176	–	–
13 – South Africa: Benfontein; 28°48′S 24°49′E	Sp14 (B)	VIc	FN646134	FN646177	–	–
14 – South Africa: Calitzdorp; 33°31′S 21°42′E	Sp9 (T)	Vb	FN646135	FN646178	–	–
15 – South Africa: Prince Albert; 33°13′S 22°01′E	Sp6 (B)	IIId	FN646136	FN646179	PG4	FN650633
15 – South Africa: Prince Albert; 33°13′S 22°01′E	Sp7 (B)	IIId	FN646137	FN646180	–	–
16 – South Africa: Oudtshoorn; 33°35′S 22°13′E	Sp8 (B)	IIIb	FN646138	FN646181	–	–
17 – South Africa: Victoria West; 31°23′S 23°06′E	Sp30 (B)	IIIa	FN646139	FN646182	–	–
17 – South Africa: Victoria West; 31°23′S 23°06′E	Sp27 (B)	IIIc	FN646140	FN646183	–	–
17 – South Africa: Victoria West; 31°23′S 23°06′E	Sp29 (B)	IIId	FN646141	FN646184	PG8	FN650634
17 – South Africa: Victoria West; 31°23′S 23°06′E	Sp32 (B)	IIId	FN646142	FN646185	–	–
17 – South Africa: Victoria West; 31°23′S 23°06′E	Sp34 (B)	IIIe	FN646143	FN646186	–	–
17 – South Africa: Victoria West; 31°23′S 23°06′E	Sp35 (B)	IIIe	FN646144	FN646187	–	–
17 – South Africa: Victoria West; 31°23′S 23°06′E	Sp33 (B)	IIIf	FN646145	FN646188	–	–
17 – South Africa: Victoria West; 31°23′S 23°06′E	Sp31 (B)	IIIh	FN646146	FN646189	–	–
17 – South Africa: Victoria West; 31°23′S 23°06′E	Sp26 (B)	IIIi	FN646147	FN646190	–	–
17 – South Africa: Victoria West; 31°23′S 23°06′E	Sp28 (B)	IIIj	FN646148	FN646191	–	–
18 – South Africa: Willowmore; 33°17′S 23°29′E	Sp3 (B)	IIIg	FN646149	FN646192	–	–
19 – South Africa: Eastern Cape: Cranemere; 32°21′S 24°59′E	Sp36 (B)	IIIe	FN646150	FN646193	–	–
19 – South Africa: Eastern Cape: Cranemere; 32°21′S 24°59′E	Sp37 (B)	IIIf	FN646151	FN646194	PG8	FN650635
20 – South Africa: Alice; 32°47′S 26°49′E	Sp12 (B)	IIId	FN646152	FN646195	PG8	FN650636
No locality	MTD T 908 (T)	Ia	FN646153	FN646196	PG6	FN650637
No locality	MTD T 1248 (T)	Ia	FN646154	FN646197	PG6	FN650638
No locality	MTD T 2583 (T)	Ia	FN646155	FN646198	–	–
No locality	MTD D 42984 (T)	Va	FN646156	FN646199	–	–

*Approximate coordinates, corresponding to the centre of the region where the tortoise was collected.
B, Blood sample; S, Saliva sample; T, Tissue sample.

doi: 10.1111/j.1439-0469.2010.00565.x

Table 2. Previously published GenBank sequences of *Stigmochelys pardalis* and their allocation. 'Length' indicates number of sites aligned with our sequences. For haplotypes identified in the present study, the geographical origin is given in brackets (cf. Table 1); many of these haplotypes do not differ in their ND4 section. All GenBank sequences were produced with universal chelonian PCR primers

Accession numbers	Fragment	Length (bp)	Identity according to GenBank	Identity according to this study	References
AY673458	ND4	680	*Stigmochelys pardalis*	Haplotype IIa (Kenya)	Cunningham (2002)
AY673459	ND4	680	*Stigmochelys pardalis*	Haplotype IIa (Kenya)	Cunningham (2002)
AY673460	ND4	680	*Stigmochelys pardalis*	Resembles haplotype IIa (Kenya; distinct by 1 step)	Cunningham (2002)
AY673461	ND4	680	*Stigmochelys pardalis*	Haplotype IIa (Kenya)	Cunningham (2002)
AY673462	ND4	680	*Stigmochelys pardalis*	Resembles haplotype IIIc = IIId = IIIg = IIIh = IIIi = IIIj (Namibia, South Africa; distinct by 2 steps)	Cunningham (2002)
AY673533	ND4	680	*Stigmochelys pardalis pardalis*	Haplotype VIIa = VIIb = VIIc = VIId (Namibia)	Cunningham (2002)
AY673534	ND4	680	*Stigmochelys pardalis babcocki*	Haplotype Va = Vc = Vd (South Africa)	Cunningham (2002)
DQ080041*	ND4*	802	*Stigmochelys pardalis*	Haplotype IIb = IIc (Ethiopia, Kenya)	Parham et al. (2006)
AF020898	cyt *b*	382	*Stigmochelys pardalis*	Pseudogene (GenBank haplotype 1)	Caccone et al. (1999)
AF371238	cyt *b*	401	*Stigmochelys pardalis*	Pseudogene (GenBank haplotype 1)	Austin and Arnold (2001)
AY678363	cyt *b*	425	*Stigmochelys pardalis*	Pseudogene (GenBank haplotype 2)	Cunningham (2002)
AY678364	cyt *b*	425	*Stigmochelys pardalis*	Pseudogene (GenBank haplotype 1)	Cunningham (2002)
AY678365	cyt *b*	425	*Stigmochelys pardalis*	Pseudogene (GenBank haplotype 1)	Cunningham (2002)
AY678366	cyt *b*	425	*Stigmochelys pardalis*	Pseudogene (GenBank haplotype 3)	Cunningham (2002)
AY678367	cyt *b*	425	*Stigmochelys pardalis*	Pseudogene (GenBank haplotype 2)	Cunningham (2002)
DQ080041*	cyt *b**	1136	*Stigmochelys pardalis*	Chimerical sequence of mitochondrial gene and pseudogene	Parham et al. (2006)
DQ497301	cyt *b*	1136	*Stigmochelys pardalis pardalis*	Resembles haplotype Vc (South Africa; distinct by 1 step)	Le et al. (2006)
DQ497302	cyt *b*	1136	*Stigmochelys pardalis babcocki*	Chimerical sequence of mitochondrial gene and pseudogene	Le et al. (2006)
U81353	cyt *b*	892	*Stigmochelys pardalis*	*Astrochelys radiata*	Shaffer et al. (1997)

*Complete mitochondrial genome; originates from the only tortoise with known geographical origin (Awdal Region, Somalia).

CytbG/mt-E-Rev2 and mt-c2/mt-f-na (Table S2) were used to amplify the cyt *b* gene in two fragments overlapping by approximately 300 bp. However, when the resulting PCR products were sequenced, two mismatching DNA fragments were obtained in most cases. One fragment was very similar to GenBank sequence DQ497301 labelled as *S. p. pardalis* by Le et al. (2006), while the other rather resembled GenBank sequence DQ497302, labelled *S. p. babcocki* by the same authors. To explore this unexpected situation, the primer combination CytbG/mt-f-na was used to amplify the whole cyt *b* gene; amplification was successful in 20 samples. The PCR product was again sequenced in two fragments. The resulting contigs were now consistent and corresponded to ten distinct haplotypes, eight of which were 1133 bp long (16 samples; Table 1). Their first half again resembled DQ497302 and the majority of shorter cyt *b* GenBank sequences (Table 2); yet, their second half was highly distinct (Table S5). The other two haplotypes, corresponding to four samples, were 1136 bp long and resembled GenBank sequence DQ497301. When our new PCR primers CytB-F-pard/CytB-R-pard, specifically targeting the conserved tRNA genes flanking the cyt *b* gene of *S. pardalis* and other testudinids, were applied to the same 20 samples, only sequences of 1136 bp length were obtained, also from the 16 samples that yielded the highly distinctive 1133 bp sequences with the other PCR primers (Table 1).

These 16 sequences of 1133 bp length, resulting from the amplicon of the standard primers CytbG/mt-f-na, share with GenBank sequence DQ497302 (*S. p. babcocki*) a characteristic deletion of 3 bp at position 446–448 of our 1136-bp-long alignment. Such a deletion does not occur in the cyt *b* gene of any of the 33 other testudinid species and subspecies studied by Le et al. (2006) and Fritz and Bininda-Emonds (2007). Most of the other GenBank sequences of *S. pardalis* are shorter and do not include the region where the deletion is located, so it cannot be determined whether they contain it or not. Nevertheless, these short sequences (GenBank haplotypes 1-3) have many distinctive mutations in common with DQ497302 and our sequences (Tables 2 and S5).

Our new PCR primers CytB-F-pard/CytB-R-pard were then used also to amplify the whole cyt *b* gene of the remaining 23 samples. The 1136-bp-long sequences of all 43 samples corresponded to 25 distinct haplotypes that are similar to GenBank sequence DQ497301 (*S. p. pardalis*); as in this and in another GenBank sequence (DQ080041), the deletion at position 446–448 does not occur.

The presence of two distinct, cyt *b*-like sequences in the same sample is consistent with two alternative explanations: heteroplasmy or the existence of nuclear mitochondrial insertions (numts). Heteroplasmy implies the coexistence of two functional cytoplasmatic mitochondrial genomes. By contrast, numts behave as pseudogenes and evolve in a different manner compared to the cytoplasmatic mitochondrial genome, and therefore are expected also to produce a different phylogenetic signal. Furthermore, numts often show frame-shift mutations because of relaxed evolutionary constraints of the pseudogene (Zhang and Hewitt 1996; Bensasson et al. 2001). Acknowledging this, we tested the hypothesis that one of the two cyt *b* variants might represent a numt using the following approaches: (1) by analyzing the coding sequence for the presence of frame-shift mutations, (2) by comparing nucleotide diversity of the two cyt *b* variants and the protein-coding part of the mtDNA fragment containing the partial ND4 gene and (3) by phylogenetic analyses and comparisons.

doi: 10.1111/j.1439-0469.2010.00565.x

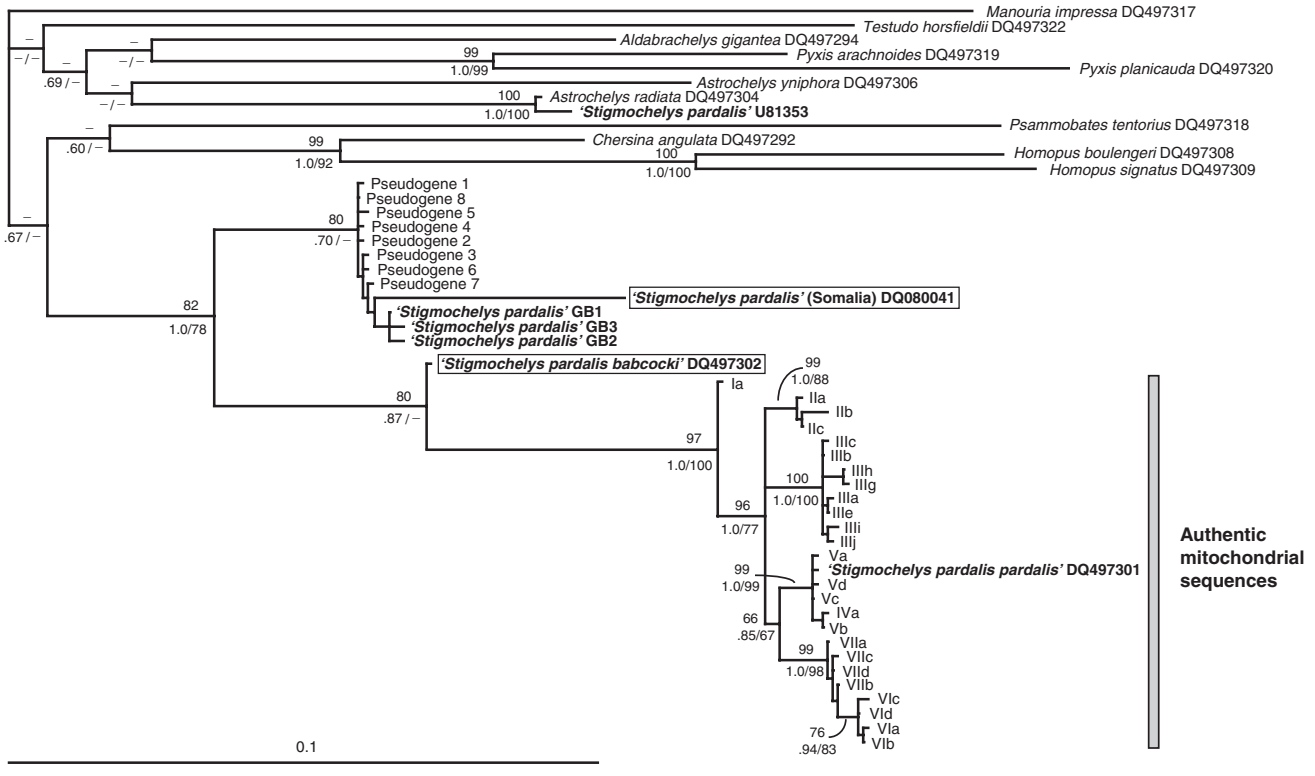

Fig. 3. Maximum Likelihood tree (-ln L = 6554.9015) for cyt *b* haplotypes of *Stigmochelys pardalis* (GenBank, this study) and tortoise species of the African-Malagasy-Indian Ocean clade based on a 1136-bp-long alignment. Nomenclature follows Fritz and Bininda-Emonds (2007). GenBank haplotypes of *S. pardalis* in bold; chimerical sequences in boxes. Codes following species names are GenBank accession numbers; for accession numbers of GenBank haplotypes (GB1-GB3), see Table 2. Numbers above nodes are ML bootstrap values; below nodes, Bayesian posterior probabilities/MP bootstrap values. Topology of the BA tree is identical with respect to the position of *S. pardalis* sequences. The 50% majority rule consensus tree of 5860 equally parsimonious trees (1145 steps; CI = 0.5013, RI = 0.7727) places, with bootstrap support below 50%, the short GenBank haplotypes GB1-GB3 together with the chimerical sequence DQ497302 in a distinct clade. This clade constitutes together with another clade of our authentic *S. pardalis* sequences the sister group to our pseudogene sequences plus DQ080041

When GenBank sequence DQ497302 and our 1133-bp-long sequences are translated into amino acids, their 3-bp-long deletion results in the absence of a leucine at amino acid position 150/151. In all 33 other testudinid taxa from the data sets of Le et al. (2006) and Fritz and Bininda-Emonds (2007), two leucines are present at positions 150/151, suggesting that this difference represents not intraspecific variation of functional mitochondrial lineages in *S. pardalis*, but rather the result of relaxed evolutionary constraints as expected for numts.

Further evidence for this hypothesis is provided by molecular diversity indices (Table 3). The ratio π_a/π_s of nucleotide diversity estimates derived from non-synonymous (π_a) and synonymous (π_s) sites is some 0.81 for our 1133-bp-long sequences, as expected for neutrally evolving genes. This is approximately the tenfold of the value (0.06) for the 1136-bp-long cyt *b* sequences obtained from the same 16 samples. In addition, the value for all 43 sequences of 1136 bp length and the values for the partial ND4 fragment are similar (all approximately 0.06), as expected for mitochondrial genes under purifying selection.

Nuclear copies accumulate fewer mutations because of the presence of repair mechanisms in the nucleus, which results in a slower evolution rate and less variation compared to functional mitochondrial genes (Brown et al. 1982; Zhang and Hewitt 1996). In fact, while haplotype diversity is similar in the two cyt *b* variants and ND4, nucleotide diversity is approximately 7.5 times smaller in the putative numt (Table 3). This is mirrored by different numbers of segregating sites, which is particularly obvious when the two cyt *b* variants of the same subsample are compared. A similar pattern of low evolution rate of numts compared to their mitochondrial paralogues has also been described, for instance, in primates (Brown et al. 1982; Zischler et al. 1995) and ants (Martins et al. 2007). The latter authors also found for pseudogenes significantly decreased nucleotide diversity and a significantly increased π_a/π_s ratio, similar to our data.

Based on this evidence, we conclude that the universal PCR primers CytbG/mt-f-na amplified in most cases a numt of the mitochondrial cyt *b* gene of *S. pardalis*, while our newly designed primers CytB-F-pard/CytB-R-pard consistently yielded authentic mitochondrial sequences. This hypothesis is confirmed by phylogenetic analyses of all distinct haplotypes of such putative numts and authentic cyt *b* sequences produced in this study (Fig. 3). The numt haplotypes constitute a deeply divergent clade that is sister to all other cyt *b* sequences of *S. pardalis*, as expected for such a 'molecular fossil' (Zischler et al. 1995; Bensasson et al. 2001). Only one homologous GenBank haplotype (DQ497301) is nested within our authentic haplotypes, while all other GenBank sequences are highly distinct. GenBank haplotypes 1-3 (Table 2) are easily identified by their position as numts. However, GenBank sequence DQ080041, a sequence without the characteristic 3-bp deletion, is also associated with this numt clade, while another

doi: 10.1111/j.1439-0469.2010.00565.x

Table 3. Molecular diversity indices of protein-coding mtDNA fragments produced with different PCR primers (see text for explanations). The subsample corresponds to sequences from the same 16 specimens (cf. Table 1)

Sample (*n*)	PCR primer	Resulting Fragment	bp	S	*h*	H_d	π*	π_a*	π_s*	π_a/π_s*
All (43)	CytB-F-pard/ CytB-R-pard	mtDNA: cyt *b*	1136	66	25	0.931	0.01470	0.00322	0.05050	0.064
Subsample (16)	CytB-F-pard/ CytB-R-pard	mtDNA: cyt *b*	1136	41	10	0.900	0.01403	0.00289	0.04944	0.059
Subsample (16)	CytbG/mt-f-na	nDNA: cyt *b*-like numt	1133	10	8	0.858	0.00187	0.00177	0.00220	0.805
All (43)	L-ND4/H-Leu	mtDNA: ND4	679	25	11	0.879	0.00928	0.00191	0.03070	0.062
Subsample (16)	L-ND4/H-Leu	mtDNA: ND4	679	18	6	0.858	0.00989	0.00204	0.03271	0.062

Translation table: mammalian mitochondrion; * Jukes Cantor corrected; *n*: number of nucleotide sequences analyzed; S: number of segregating sites; *h*: number of haplotypes; π: nucleotide diversity; H_d: haplotype diversity; π_a: nucleotide diversity at non-synonymous sites; π_s: nucleotide diversity at synonymous sites.

sequence in which this deletion occurs (DQ497302, *S. p. babcocki*) clusters under BA and ML basally to the authentic cyt *b* sequences of *S. pardalis*. When the alignment of these sequences is compared with authentic cyt *b* and pseudogene sequences produced in this study, it is obvious that DQ080041 and DQ497302 are of chimerical origin from concatenating mitochondrial and pseudogene fragments (Table S5), explaining their unexpected allocation in the trees. The chimerical origin of DQ080041, representing the complete mitochondrial genome of a leopard tortoise from Somalia, is further corroborated by the complete identity of its ND4 fragment with the homologous sequences of our two samples collected in close proximity in Ethiopia (Table 2). By contrast, the cyt *b* sequences of our two samples resemble only the 5′-part of DQ080041 (Table S5).

Finally, it may be noted that another distinct GenBank haplotype (U81353) has been erroneously assigned to *S. pardalis* (Shaffer et al. 1997). According to our phylogenetic analyses (Fig. 3), U81353 clearly represents the Malagasy tortoise species *Astrochelys radiata*. The identity of U81353 with *A. radiata* was confirmed by a BLAST search in GenBank. Ninety-nine per cent of U81353 are identical with another cyt *b* sequence of *A. radiata* (DQ497304, score 1754), and only 89% match the cyt *b* sequences of three other tortoise species (*Chersina angulata*, DQ497292, score 1294; *Pyxis arachnoides*, DQ497319, score 1318; *S. p. pardalis*, DQ497301, score 1308; *S. p. babcocki*, DQ497302, score 1326). When this misidentified sequence is excluded from further considerations, it is evident that most GenBank data represent not the mitochondrial cyt *b* gene of *S. pardalis*, but a paralogous pseudogene. Only one GenBank sequence (DQ497301) is of mitochondrial origin indeed and two GenBank sequences (DQ080041, DQ497302) are chimerical. In the following, only mitochondrial sequences produced in the present study are used for further phylogeographic analyses.

Phylogeography of the leopard tortoise

Our 43 samples of leopard tortoises yielded for the ND4 sequences 11 distinct haplotypes and for authentic mitochondrial cyt *b* sequences 25 distinct haplotypes. When the ND4 and cyt *b* sequences were concatenated for each individual, there resulted 27 distinct haplotypes. Such combined haplotypes differed for many individuals only in the cyt *b* section, while significantly less variation occurred in their ND4 sequences. Yet, two haplotypes (IIIb, IIIf) differed from others only in their ND4 section.

All phylogenetic analyses of the concatenated data set (cyt *b* + ND4) resulted in the same seven major clades (bearing consecutive Roman numerals in Fig. 4a). Clade I consists of only one haplotype (Ia) that has a wide geographical distribution (Uganda, Tanzania, Zambia). Clade II, comprising three individual haplotypes, was found in samples from Ethiopia and Kenya; the other five clades, III-VII with 1-10 haplotypes each, originate from the Republic of South Africa and Namibia. The branching pattern for these seven clades differs between the tree-building methods, however, and is not well-resolved. All methods agree in the sister group relationship of clades (IV + V) and (VI + VII). Furthermore, ML analysis suggests that a more inclusive clade ((IV + V) + (VI + VII)) is sister to (II + III) and that clade I represents the most basal branch. By contrast, BA results in a basal multifurcation of I, II, III and ((IV + V) + (VI + VII)). A similar basal polytomy is also suggested by the 50% majority rule consensus of 371 equally parsimonious trees (516 steps). Yet, under parsimony, clade I is basal to all others, as it is under the likelihood criterion, and the clades (IV + V) and (VI + VII) are not sister groups.

Both single-gene analyses yielded a much worse resolution. Since the ND4 fragment of many haplotypes is not or barely differentiated, only clades I, II and V were retrieved as distinct (Fig. 4b), in agreement with the analyses of the combined data set. The remaining haplotypes cluster in three clades that do not correspond to any entity identified by the analyses of the combined data set. Clade I is suggested as basal to a polytomy comprising all other haplotypes. Using the cyt *b* data set (Fig. 4c), clades I, II and III were returned as distinct, while clade V is now paraphyletic with respect to the only haplotype of clade IV (IVa), and haplotypes of clade VII are paraphyletic with respect to the embedded and well-supported clade VI. Clade I occurs now in a polytomy together with clades II, (IV + V) and (VI + VII).

Using the default 95% connection limit of TCS, three unconnected networks and one isolated haplotype were obtained for the concatenated cyt *b* + ND4 sequences. One network corresponds to the three haplotypes of Ethiopian and Kenyan tortoises (clade II), and another one to the 10 haplotypes of clade III from the Republic of South Africa and Namibia. The third network is the most diverse one, comprising the remaining haplotype clades IV, V, VI and VII from South Africa and Namibia. Haplotype Ia, found in leopard tortoises from Uganda, Tanzania, Zambia and three pet-trade tortoises, remained unconnected. Under the 90% threshold, haplotypes of these entities were connected by pathways of a minimum of 22–28 mutational steps (Fig. 5).

doi: 10.1111/j.1439-0469.2010.00565.x

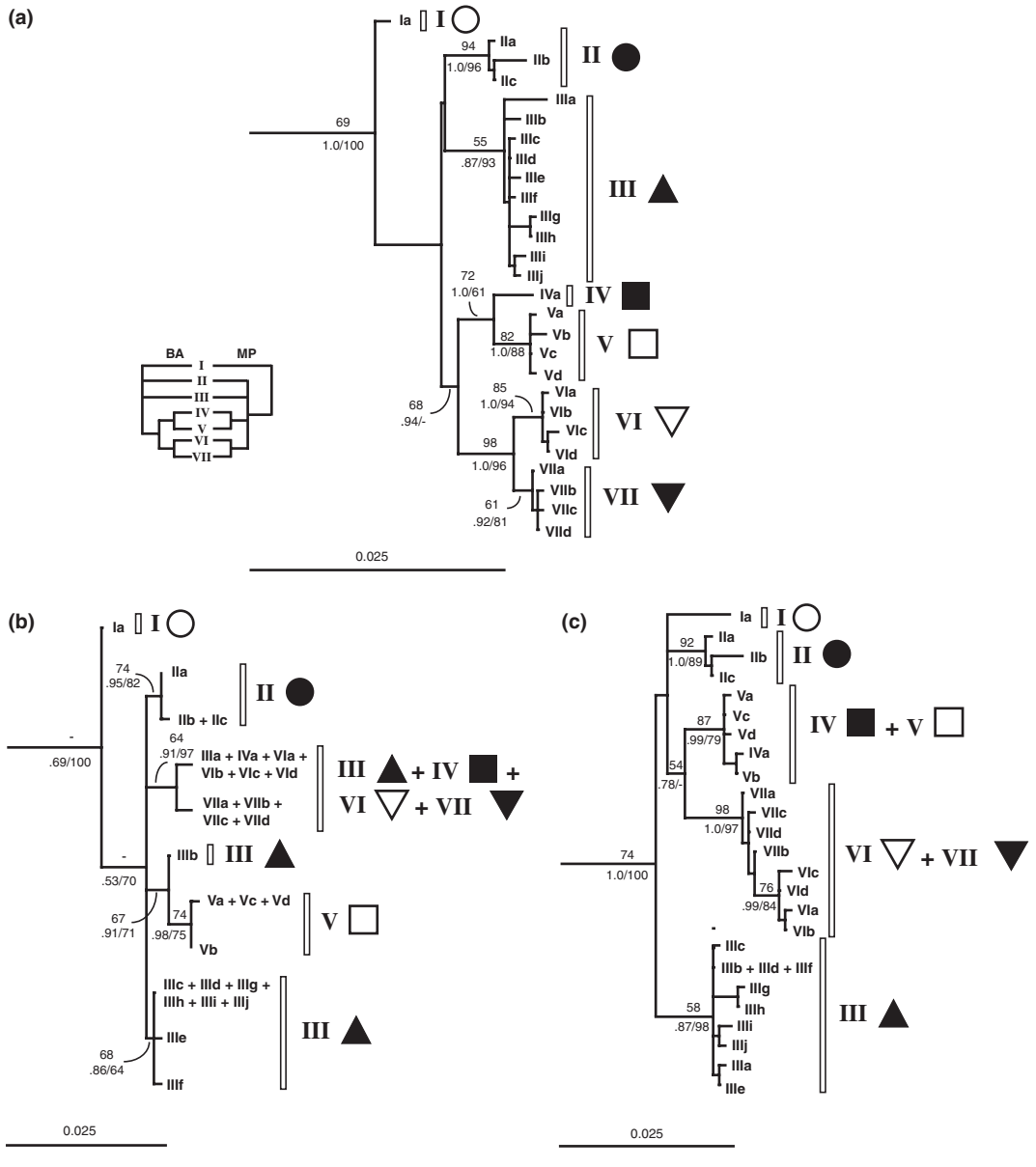

Fig. 4. (a) Maximum Likelihood tree (-ln L = 4972.9323) for haplotypes of *Stigmochelys pardalis* (combined data set of cyt *b* + ND4; 1938 bp, only sequences from this study). Roman numerals on the right indicate clades; letters, individual haplotypes. Clade symbols correspond to Fig. 1. Inset: alternative gross topologies as obtained under Bayesian analysis or Maximum Parsimony (50% consensus of 371 trees of 516 steps; CI = 0.8624; RI = 0.8436). (b) Maximum Likelihood tree (-ln L = 1858.9472) for the ND4 data partition (802 bp, only sequences from this study). The gross topologies of the BA tree and the 50% majority rule consensus tree of four equally parsimonious trees (182 steps; CI = 0.9176; RI = 0.7656) are identical. (c) Maximum Likelihood tree (-ln L = 2992.6225) for the cyt *b* data partition (1136 bp, only sequences from this study). The gross topologies of the BA tree and the 50% majority rule consensus tree of 1531 equally parsimonious trees (317 steps; CI = 0.877, RI = 0.88) are identical. Outgroups (*Psammobates tentorius*, *Testudo horsfieldii*) removed for clarity. For further explanation, see Fig. 3 and text

The greatest outgroup probability (0.2090) has the most frequent haplotype IIId, representing also the ancestral haplotype of the subnet comprising all haplotypes of clade III. Haplotypes of the South African and Namibian clades IV, V, VI and VII appear as clearly distinct clusters in the network analyses, whereas they were not or weakly differentiated in the single-gene phylogenetic analyses.

The Mantel tests revealed for all examined data sets a highly significant correlation between genetic and geographical distances ($p < 0.001$), suggesting that the observed sequence variation is more likely to be caused by isolation-by-distance processes than by historical vicariant events (cf. Irwin 2002). The inclusion or exclusion of perhaps allochthonous tortoises from the South African and Namibian localities Benfontein, Calitzdorp and Oranjemund (Fig. 1) had no impact on the significance level.

Discussion

Misleading GenBank sequences and numts

Our results provide evidence that most cyt *b* sequences of *Stigmochelys pardalis* from GenBank are numts or chimerical sequences, resulting from the usage of standard PCR primers. One GenBank sequence of *S. pardalis* (U81353) turned out to represent a completely different tortoise species, the Malagasy *Astrochelys radiata*. This argues for caution with respect to the

doi: 10.1111/j.1439-0469.2010.00565.x

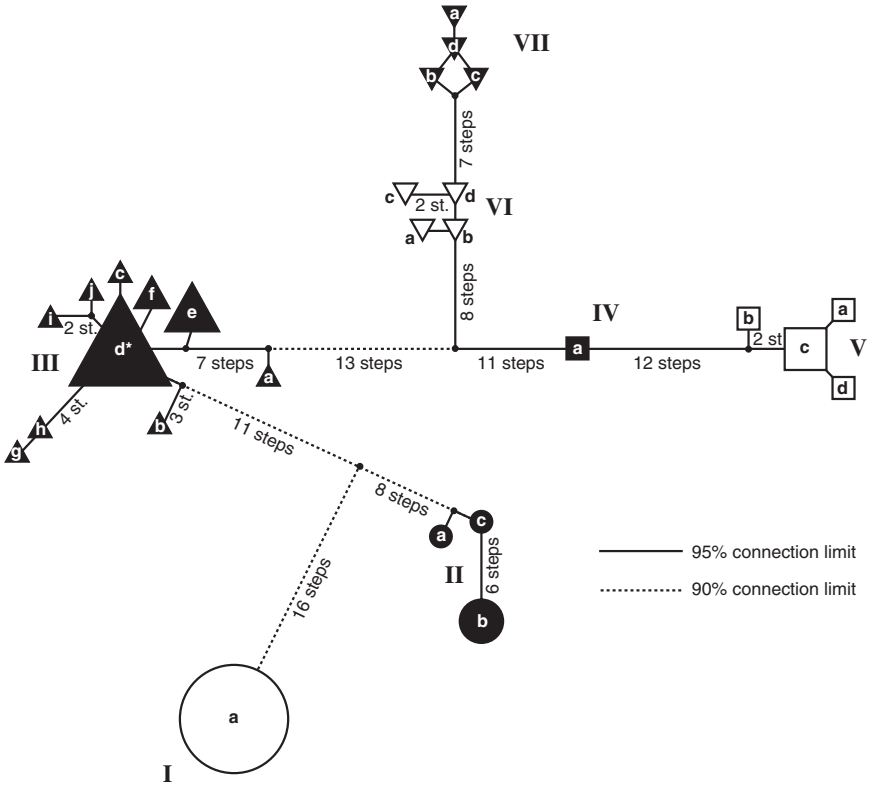

Fig. 5. Parsimony network for haplotypes (cyt *b* + ND4) of *Stigmochelys pardalis* (only sequences from this study). Haplotype symbols correspond to other figures; symbol size, approximate haplotype frequency; missing node haplotypes, small black circles. Connections using the 95% or 90% probability threshold indicated. Each line joining haplotypes represents one mutational step unless otherwise specified. The haplotype with the greatest outgroup probability (IIId) is asterisked

application of standard PCR primers and the uncritical usage of GenBank data. Several authors have previously highlighted the risks involved in using sequence data from GenBank or other public data bases (e.g., Forster 2003; Harris 2003; Vilgalys 2003; Seberg 2004; Meier et al. 2006), but this practice continues. For instance, recently an exhaustive study has been published that relied entirely on GenBank data for the reconstruction of a species-level phylogeny of chelonians (Thomson and Shaffer 2010).

Some authors describing the discovery of numts reported the occurrence of 'ghost bands' in their PCR products or ambiguities in sequencer chromatograms (e.g., Sorenson and Quinn 1998; Thalmann et al. 2004; Buhay 2009; for chelonians: Spinks and Shaffer 2007; Shi et al. 2008; see also the review by Bensasson et al. 2001), necessitating sophisticated laboratory procedures (long-range PCR and cloning; Bensasson et al. 2001; Thalmann et al. 2004; Buhay 2009) for obtaining clean sequences of the pseudogene and authentic mtDNA. By contrast, we encountered only one amplicon using standard primers. Such PCR products yielded single, clear sequences that, however, differed in part significantly from sequences amplified from the same samples using other PCR primers. A similar finding has been described by Sorenson and Quinn (1998) for the wandering whistling-duck (*Dendrocygna arcuata*), underlining that the numt and its mitochondrial counterpart are not necessarily amplified together. In fact, we are convinced that such clean PCR products and numt sequences contributed to the confusion of the numt with an authentic mitochondrial sequence in previous studies using *S. pardalis* (Caccone et al. 1999; Austin and Arnold 2001; Cunningham 2002; Le et al. 2006; Parham et al. 2006). As Sorenson and Quinn (1998) pointed out, primer design is a crucial issue here: ''Because numts may evolve more slowly than mtDNA following transposition […], they diverge less from the ancestral sequences of related taxa. As a result, primers based on sequences from species other than the study taxa and so-called 'universal' primers may be particularly prone to amplification of numts.'' We could imagine that numts like the one identified in the present study have also been confounded with authentic mitochondrial sequences in other cases.

Sorenson and Quinn (1998), Spinks and Shaffer (2007) and Shi et al. (2008) concluded that numts are more frequently amplified from blood samples than from tissues in reptiles and birds because their erythrocytes are nucleated and depauperate in mtDNA. While at first glance this seems convincing, we wish to point out that we obtained clear PCR products and sequences of the numt not only from blood samples, but also from tissue samples (Table 1). Also Le et al. (2006) and Parham et al. (2006) used tissues for their investigations and obtained partial sequences of the pseudogene of *S. pardalis*. This suggests that primer match plays a key role and highlights our concerns about standard primers.

Subspecies and phylogeography of *Stigmochelys pardalis*

The recent revalidation of the subspecies *S. p. babcocki* was based on a chimerical cyt *b* sequence (DQ497302) that was erroneously identified as this subspecies (Le et al. 2006). This renders all recent considerations about the validity of the subspecies *S. p. babcocki* obsolete. Nevertheless, our results indicate a pronounced phylogeographic structure within *S. pardalis*, the most widely distributed sub-Saharan tortoise species. Although our locality sampling was somewhat patchy, our data clearly suggest that the greatest diversity occurs in the southernmost part of the species' range (Fig. 1). Southern Africa is a hotspot of tortoise diversity and harbours 14 of the 18 currently recognized sub-Saharan African species (Branch

doi: 10.1111/j.1439-0469.2010.00565.x

2007; Fritz and Havaš 2007), a situation echoed by the occurrence of five out of the seven mitochondrial clades of *S. pardalis* there. Haplotypes of the five southern clades are largely parapatrically distributed. Yet, a few records of distinct haplotypes at the same locality or in close proximity (localities Oranjemund, Benfontein and Calitzdorp in Fig. 1) could be the result of translocation of tortoises by man. Leopard tortoises are favourite pets in South Africa and Namibia, but people often release them when they get big and start causing damage to their gardens. Furthermore, people sometimes pick up tortoises that cross the road when they travel along highways and release them elsewhere (M. D. Hofmeyr, personal observation). But the inclusion or exclusion of the perhaps allochthonous tortoises does not alter the results of our Mantel tests, clearly suggesting that the observed geographical variation is the result of isolation-by-distance and not of historical vicariant events. This is in sharp contrast to wide-ranging western Palaearctic reptiles, in particular to chelonians and other thermophilic species, whose phylogeographic structure was significantly shaped by Pleistocene range interruptions, but also by earlier dispersal and vicariant events (Lenk et al. 1999; Fritz et al. 2006a,b, 2007, 2008, 2009a,b; Ursenbacher et al. 2006a,b, 2008; Böhme et al. 2007; Joger et al. 2007; Sommer et al. 2007, 2009; Guicking et al. 2008, 2009). Nonetheless, the differences in the mitochondrial diversity of leopard tortoises from the south and north of their ranges, with a distinctly higher number of clades and haplotypes in the south, might indicate that the species originated in southern Africa and that the northerly regions were colonized later. This is corroborated by the southern African distribution of *Psammobates*, the sister group of *S. pardalis* (Le et al. 2006; Fritz and Bininda-Emonds 2007), suggesting that their last common ancestor occurred here.

Mitochondrial differentiation in *S. pardalis* is not paralleled by obvious morphological variation, although it is well known that considerable size differences exist in different regions. Leopard tortoises from the far south and the far north are much larger than tortoises from the centre of the range (Lambert 1995; Lambert et al. 1998; Hailey and Lambert 2002; see also Fig. 2). Yet, the distribution of large or small tortoises does not match our clades. For instance, the gigantean Ethiopian leopard tortoises and populations of small-bodied tortoises from Kenya (Fig. 2; Table S1) occur in the same clade (Fig. 4: clade II). Such a marked disparity between morphology and mitochondrial divergence has been described for other tortoise species as well (*Testudo* spp.: Fritz et al. 2005, 2007, 2009a,b; Attum et al. 2007; Široký and Fritz 2007; *Indotestudo forstenii*: Ives et al. 2008; *Homopus signatus*: Daniels et al. 2010), which suggests considerable phenotypic plasticity in testudinids. In western Palaearctic *Testudo graeca* and *T. marginata*, populations of large-bodied or small-bodied tortoises do not differ genetically; and in these species, adult size is positively correlated with factors such as food availability and environmental humidity (Fritz et al. 2005, 2007). Similar body size and environmental correlations, coupled with a high survival rate in northern populations, have also been suggested for leopard tortoises (Lambert 1995; Lambert et al. 1998; Hailey and Lambert 2002).

Although phylogeographic differentiation exists in *S. pardalis*, the distributions of the mitochondrial clades and the traditional subspecies ranges do not match (Fig. 1). Furthermore, the revalidation of the two subspecies of *S. pardalis* (Le et al. 2006) was based on a chimerical cyt *b* sequence misidentified as representing *S. p. babcocki*, while all other mitochondrial and nuclear genomic markers studied by Le et al. (2006) were not or only barely distinct. Prior to the study of Le et al. (2006), the subspecies had been synonymized (Boycott and Bourquin 2000; Ernst et al. 2000) because the purported morphological differences had little diagnostic value (Greig and Burdett 1976). Hence, we conclude it is best to return to the situation before *S. p. babcocki* was revalidated: The usage of subspecies within *S. pardalis* should be abandoned.

Acknowledgements

We thank conservation agencies in the Western (CapeNature), Northern and Eastern Cape Provinces, South Africa, and in Namibia for permits to collect biological material, private landowners for allowing sampling on their properties, and Theunis Hofmeyr, Fanie Avenant, Alfred Schleicher, Marius Burger, Thomas Leuteritz, Victor Loehr, Johan Marais and Krystal Tolley for assistance with sampling or for providing samples. MDH and SRD are indebted to the South African National Research Foundation and the Universities of the Western Cape and Stellenbosch for financial support. Sisay Taye Endalew, the Ethiopian Wildlife Conservation Department, and the Dirección Territorial y Provincial de Comercio Barcelona enabled the work with Ethiopian tortoises. Research in Kenya was facilitated by the Biota East Africa Project; we are deeply indebted to Jörn Köhler for generous help and to Richard Bagine (Kenyan Wildlife Service) for assistance and issuing the necessary permits. Thanks go to David Modrý for his photograph together with a Kenyan leopard tortoise; he and Martin Kamler helped during field work in Kenya.

Zusammenfassung

Mitochondriale Phylogeographie und Unterarten der weitverbreiteten subsaharischen Pantherschildkröte Stigmochelys pardalis *(Testudines: Testudinidae) – ein Beispiel für die Fallstricke von Pseudogenen und GenBank-Sequenzen*

Die Pantherschildkröte (*Stigmochelys pardalis*) ist mit einem Areal, das sich vom Horn von Afrika über Ostafrika bis nach Südafrika, Namibia und ins südliche Angola erstreckt, die am weitesten verbreitete subsaharische Landschildkrötenart. Anhand mitochondrialer DNA-Sequenzen (1938 bp; Cytochrom *b*-Gen, partielles ND4-Gen plus angrenzende tRNA-Gene) von Exemplaren aus fast dem gesamten Verbreitungsgebiet wurde die phylogeographische Struktur der Art untersucht. Die in dieser Arbeit gewonnenen Daten wurden zudem mit vorher veröffentlichten GenBank-Sequenzen verglichen. Sieben größere Kladen wurden identifiziert, die fast komplett parapatrisch verbreitet sind; die wenigen Nachweise von unterschiedlichen Haplotypen am selben Fundort oder von unmittelbar benachbarten Fundorten könnten auf Schildkröten zurückzuführen sein, die vom Menschen verschleppt wurden. Im Süden des Verbreitungsgebietes, wo fünf der sieben Kladen vorkommen, ist die Diversität am größten. Isolation-by-Distance-Tests deuten darauf hin, dass die phylogeographische Struktur der Pantherschildkröte auf eingeschränkten geographischen Genfluss und nicht auf historische Vikarianzereignisse zurückzuführen ist. Dies stellt einen klaren Unterschied zu weitverbreiteten wärmeliebenden Reptilienarten aus der Westpaläarktis dar, deren phylogeographische Struktur ganz entscheidend durch pleistozäne Arealzerschneidungen, aber auch durch ältere Ausbreitungs- und Vikarianzereignisse beeinflusst wurde. Die meisten *S. pardalis* zugeschriebenen Cytochrom *b*-Sequenzen aus der GenBank stellten sich als Pseudogene oder chimärische Sequenzen aus authentischen mitochondrialen und Pseudogen-Fragmenten heraus. Dies unterstreicht, dass GenBank-Sequenzen nicht unkritisch verwendet werden dürfen. Auch die kürzlich erfolgte Revalidierung der beiden Unterarten von *S. pardalis* ist auf eine chimärische Sequenz zurückzuführen, die irrtümlich der Unterart *S. p. babcocki* zugeschrieben worden war. Wie die Daten der vorliegenden Arbeit zeigen, deckt sich die Verbreitung der mitochondrialen Kladen weder mit den traditionell anerkannten

doi: 10.1111/j.1439-0469.2010.00565.x

Unterarten von *S. pardalis*, noch mit den beträchtlichen geographischen Größenunterschieden bei der Pantherschildkröte, weswegen es keinen Grund gibt, bei dieser Spezies Unterarten anzuerkennen.

References

Amato ML, Brooks RJ, Fu J (2008) A phylogeographic analysis of populations of the wood turtle (*Glyptemys insculpta*) throughout its range. Mol Ecol **17:**570–581.

Attum O, Baha el Din S, Carranza S, Earley R, Arnold EN, Kingsbury B (2007) An evaluation of the taxonomic validity of *Testudo werneri*. Amphib-Reptil **28:**393–401.

Austin JJ, Arnold EN (2001) Ancient mitochondrial DNA and morphology elucidate an extinct island radiation of Indian Ocean giant tortoises (*Cylindraspis*). Proc R Soc Lond B **268:**2515–2523.

Avise JC, Walker D, Johns GC (1998) Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc R Soc London B **265:**1707–1712.

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol **16:**314–321.

Böhme MU, Fritz U, Kotenko T, Džukić G, Ljubisavljević K, Tzankov N, Berendonk TU (2007) Phylogeography and cryptic variation within the *Lacerta viridis* complex. Zool Scr **36:**119–131.

Boycott RC, Bourquin O (2000) The Southern African Tortoise Book. A Guide to Southern African Tortoises, Terrapins and Turtles. Privately Printed, KwaZulu-Natal.

Branch B (1998) Field Guide to Snakes and Other Reptiles of Southern Africa. Third Revised Edition. Ralph Curtis Books Publishing, Sanibel Island, Florida.

Branch WR (2007) A new species of tortoise of the genus *Homopus* (Chelonia: Testudinidae) from southern Namibia. Afr J Herpetol **56:**1–21.

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol **18:**225–239.

Buhay JE (2009) ''COI-like'' sequences are becoming problematic in molecular systematic and DNA barcoding studies. J Crustacean Biol **29:**96–110.

Caccone A, Amato G, Gratry OC, Behler J, Powell JR (1999) A molecular phylogeny of four endangered Madagascar tortoises based on mtDNA sequences. Mol Phylogenet Evol **12:**1–9.

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. Mol Phylogenet Evol **3:**102–113.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol **9:**1657–1660.

Cunningham J (2002) A Molecular Perspective on the Family Testudinidae Batsch, 1788. PhD dissertation. University of Cape Town, Cape Town, South Africa.

Daniels SR, Hofmeyr MD, Henen BT, Crandall KA (2007) Living with the genetic signature of Miocene induced change: evidence from the phylogeographic structure of the endemic angulate tortoise *Chersina angulata*. Mol Phylogenet Evol **45:**915–926.

Daniels SR, Hofmeyr MD, Henen BT, Baard EHW (2010) Systematics and phylogeography of a threatened tortoise, the speckled padloper. Anim Conserv **12:** DOI:10.1111/j.1469-1795.2009.00323.x.

Donnelly P, Tavaré S (1986) The ages of alleles and a coalescent. Adv Appl Probab **18:**1–19.

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

Ernst CH, Altenburg RGM, Barbour RW (2000) Turtles of the World. World Biodiversity Database, CD-ROM Series, Windows Version 1.2. Biodiversity Center of ETI, Amsterdam.

Farris J, Källersjö M, Kluge AG, Bult C (1995) Testing significance of incongruence. Cladistics **10:**315–319.

Forster P (2003) To err is human. Ann Hum Genet **67:**2–4.

Fritz U, Bininda-Emonds ORP (2007) When genes meet nomenclature: tortoise phylogeny and the shifting generic concepts of *Testudo* and *Geochelone*. Zoology **110:**298–307.

Fritz U, Havaš P (2007) Checklist of chelonians of the world. Vertebr Zool **57:**149–368.

Fritz U, Široký P, Kami H, Wink M (2005) Environmentally caused dwarfism or a valid species—Is *Testudo weissingeri* Bour, 1996 a distinct evolutionary lineage? New evidence from mitochondrial and nuclear genomic markers. Mol Phylogenet Evol **37:**389–401.

Fritz U, Auer M, Bertolero A, Cheylan M, Fattizzo T, Hundsdörfer AK, Martín Sampayo M, Pretus JL, Široký P, Wink M (2006a) A rangewide phylogeography of Hermann's tortoise, *Testudo hermanni* (Reptilia: Testudines: Testudinidae): implications for taxonomy. Zool Scr **35:**531–543.

Fritz U, Barata M, Busack SD, Fritzsch G, Castilho R (2006b) Impact of mountain chains, sea straits and peripheral populations on genetic and taxonomic structure of a freshwater turtle, *Mauremys leprosa*. Zool Scr **35:**97–108.

Fritz U, Hundsdörfer AK, Široký P, Auer M, Kami H, Lehmann J, Mazanaeva LF, Türkozan O, Wink M (2007) Phenotypic plasticity leads to incongruence between morphology-based taxonomy and genetic differentiation in western Palaearctic tortoises (*Testudo graeca* complex; Testudines, Testudinidae). Amphib-Reptil **28:**97–121.

Fritz U, Ayaz D, Buschbom J, Kami HG, Mazanaeva LF, Aloufi AA, Auer M, Rifai L, Šilić T, Hundsdörfer AK (2008) Go east: phylogeographies of *Mauremys caspica* and *M. rivulata*—Discordance of morphology, mitochondrial and nuclear genomic markers and rare hybridization. J Evol Biol **21:**527–540.

Fritz U, Auer M, Chirikova MA, Duysebayeva TN, Eremchenko VK, Kami HG, Kashkarov RD, Masroor R, Moodley Y, Pindrani A, Široký P, Hundsdörfer AK (2009a) Mitochondrial diversity of the widespread Central Asian steppe tortoise (*Testudo horsfieldii* Gray, 1844): implications for taxonomy and relocation of confiscated tortoises. Amphib-Reptil **30:**245–257.

Fritz U, Harris DJ, Fahd S, Rouag R, Graciá Martínez E, Giménez Casalduero A, Široký P, Kalboussi M, Jdeidi TB, Hundsdörfer AK (2009b) Mitochondrial phylogeography of *Testudo graeca* in the Western Mediterranean: old complex divergence in North Africa and recent arrival in Europe. Amphib-Reptil **30:**63–80.

Greig JC, Burdett PD (1976) Patterns in the distribution of southern African terrestrial tortoises (Cryptodira: Testudinidae). Zool Afr **11:**249–273.

Guicking D, Joger U, Wink M (2008) Molecular phylogeography of the viperine snake *Natrix maura* (Serpentes: Colubridae): evidence for strong intraspecific differentiation. Org Divers Evol **8:**130–145.

Guicking D, Joger U, Wink M (2009) Cryptic diversity in a Eurasian water snake (*Natrix tessellata*, Serpentes: Colubridae): evidence from mitochondrial sequence data and nuclear ISSR-PCR fingerprinting. Org Divers Evol **9:**201–214.

Hailey A, Lambert MRK (2002) Comparative growth patterns in Afrotropical giant tortoises. Trop Zool **15:**121–139.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser **41:**95–98.

Harris DJ (2003) Can you bank on GenBank? Trends Ecol Evol **18:**317–319.

Irwin DE (2002) Phylogeographic breaks without geographic barriers to gene flow. Evolution **56:**2383–2394.

Iverson JB (1992) A Revised Checklist with Distribution Maps of the Turtles of the World. Privately Printed, Richmond, Indiana.

Ives I, Spinks PQ, Shaffer HB (2008) Morphological and genetic variation in the endangered Sulawesi tortoise *Indotestudo forstenii*: evidence of distinct lineages? Conserv Genet **9:**709–713.

Joger U, Fritz U, Guicking D, Kalyabina-Hauf S, Nagy ZT, Wink M (2007) Phylogeography of western Palaearctic reptiles—Spatial and temporal speciation patterns. Zool Anz **246:**293–313.

Lambert MRK (1995) On geographical size variation, growth, and sexual dimorphism of the leopard tortoise, *Geochelone pardalis*, in Somaliland. Chelonian Conserv Biol **1:**269–278.

Lambert MRK, Campbell KLI, Kabigumila JD (1998) On growth and morphometrics of leopard tortoises, *Geochelone pardalis*, in Serengeti National Park, Tanzania, with observations on effects of bushfires and latitudinal variation in populations of eastern Africa. Chelonian Conserv Biol **3:**46–57.

Le M, Raxworthy CJ, McCord WP, Mertz L (2006) A molecular phylogeny of tortoises (Testudines: Testudinidae) based on mitochondrial and nuclear genes. Mol Phylogenet Evol **40:**517–531.

doi: 10.1111/j.1439-0469.2010.00565.x

Lenk P, Fritz U, Joger U, Wink M (1999) Mitochondrial phylogeography of the European pond turtle, *Emys orbicularis* (Linnaeus 1758). Mol Ecol **8:**1911–1922.

Loveridge A (1935) Scientific results of an expedition to rain forest regions in eastern Africa. I. New Reptiles and Amphibians from East Africa. Bull Mus Comp Zool **79:**1–19.

Loveridge A, Williams EE (1957) Revision of the African tortoises and turtles of the suborder Cryptodira. Bull Mus Comp Zool **115:**163–557.

Martins J, Solomon SE, Mikheyev AS, Mueller UG, Oritz A, Bacci M (2007) Nuclear mitochondrial-like sequences in ants: evidence from *Atta cephalotes* (Formicidae: Attini). Insect Mol Biol **16:**777–784.

Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst Biol **55:**715–728.

Paradis E, Claude J, Strimmer K (2004) APE: analyses of Phylogenetics and Evolution in R language. Bioinformatics **20:**289–290.

Parham JF, Macey JR, Papenfuss TJ, Feldman CR, Türkozan O, Polymeni R, Boore J (2006) The phylogeny of Mediterranean tortoises and their close relatives based on complete mitochondrial genome sequences from museum specimens. Mol Phylogenet Evol **38:**50–64.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol **16:**37–45.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:**1572–1574.

Rosenbaum PA, Robertson JM, Zamudio KR (2007) Unexpectedly low genetic divergences among populations of the threatened bog turtle (*Glyptemys muhlenbergii*). Conserv Genet **8:**331–342.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics **145:**1219–1228.

Rozas J, Sánchez-del Barrio JC, Messeguer X, Rozas R (2003) DNASP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics **19:**2496–2497.

Seberg O (2004) The future of systematics: assembling the tree of life. Systematist **23:**2–8.

Shaffer HB, Meylan P, McKnight ML (1997) Tests of turtle phylogeny: molecular, morphological and paleontological approaches. Syst Biol **46:**235–268.

Shi H, Fong JJ, Parham JF, Pang J, Wang J, Hong M, Zang Y-P (2008) Mitochondrial variation of the ''eyed'' turtles (*Sacalia*) based on known-locality and trade specimens. Mol Phylogenet Evol **49:**1025–1029.

Široký P, Fritz U (2007) Is *Testudo werneri* a distinct species? Biologia **62:**228–231.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. Mol Ecol **15:**4261–4293.

Sommer RS, Persson A, Wieseke N, Fritz U (2007) Holocene recolonization and extinction of the pond turtle, *Emys orbicularis* (L., 1758), in Europe. Quat Sci Rev **26:**3099–3107.

Sommer RS, Lindqvist C, Persson A, Bringsøe H, Rhodin AGJ, Schneeweiss N, Široký P, Bachmann L, Fritz U (2009) Unexpected early extinction of the European pond turtle (*Emys orbicularis*) in Sweden and climatic impact on its Holocene range. Mol Ecol **18:**1252–1262.

Sorenson MD, Quinn TW (1998) Numts: a challenge for avian systematics and population biology. Auk **115:**214–221.

Spinks PQ, Shaffer HB (2007) Conservation phylogenetics of the Asian box turtles (Geoemydidae, *Cuora*): mitochondrial introgression, numts, and inferences from multiple nuclear loci. Conserv Genet **8:**641–657.

Swofford DL (2002) PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4, Sinauer Associates, Sunderland, Massachusetts.

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

Thalmann O, Hebler J, Poinar HN, Pääbo S, Vigilant L (2004) Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. Mol Ecol **13:**321–335.

Thomson RC, Shaffer HB (2010) Sparse supermatrices for phylogenetic inference: taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. Syst Biol **59:**42–58.

Ursenbacher S, Carlsson M, Helfer V, Tegelstrom H, Fumagalli L (2006a) Phylogeography and Pleistocene refugia of the adder (*Vipera berus*) as inferred from mitochondrial DNA sequence data. Mol Ecol **15:**3425–3437.

Ursenbacher S, Conelli A, Golay P, Monney JC, Zuffi MAL, Thiery G, Durand T, Fumagalli L (2006b) Phylogeography of the asp viper (*Vipera aspis*) inferred from mitochondrial DNA sequence data: evidence for multiple Mediterranean refugial areas. Mol Phylogenet Evol **38:**546–552.

Ursenbacher S, Schweiger S, Tomovic L, Crnobrnja-Isailovic J, Fumagalli L, Mayer W (2008) Molecular phylogeography of the nose-horned viper (*Vipera ammodytes*, Linnaeus (1758)): evidence for high genetic diversity and multiple refugia in the Balkan peninsula. Mol Phylogenet Evol **46:**1116–1128.

Vilgalys R (2003) Taxonomic misidentification in public DNA databases. New Phytol **160:**4–5.

Wermuth H, Mertens R (1977) Testudines, Crocodylia, Rhynchocephalia. Das Tierreich **100:**I–XXVII + 1–174.

Zhang D-X, Hewitt GM (1996) Nuclear integrations: challenges for mitochondrial DNA markers. Trends Ecol Evol **11:**247–251.

Zischler H, Geisert H, von Haeseler A, Pääbo S (1995) A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. Nature **378:**489–492.

Zwickl DJ (2006) Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion. PhD dissertation. University of Texas, Austin, Texas. Genetic Algorithm for Rapid Likelihood Inference software available at: http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html.

Supporting Information

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

Table S1. Straight-line carapace lengths in *Stigmochelys pardalis* from different parts of the range.

Table S2. Primer pairs used for PCR and sequencing mitochondrial DNA and their annealing temperatures. New primers were designed using the software OLIGO EXPLORER (http://molbiol-tools.ca/molecular_biology_freeware.htm). Primers also used for sequencing asterisked.

Table S3. Parsimony statistics for the three data sets analyzed.

Table S4. Best-fit substitution models (AIC) and their parameters, established using MODELTEST 3.06.

Table S5. Polymorphic sites (130 of 1136 bp) of authentic mitochondrial, chimerical and pseudogene sequences of *Stigmochelys pardalis*. For pseudogene haplotypes identified in the present study, see also Table 1. Authentic mitochondrial sequences highlighted with yellow; pseudogene sequences, green; deletion characteristic for the pseudogenes, red. GenBank haplotypes 1–3 are represented by short sequences that terminate prior to the deletion.

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

doi: 10.1111/j.1439-0469.2010.00565.x