Amphibia-Reptilia 33 (2012): 150-154

Phylogenetic position of *Pelusios williamsi* and a critique of current GenBank procedures (Reptilia: Testudines: Pelomedusidae)

Uwe Fritz[1,*], Mario Vargas-Ramírez[1], Pavel Široký[2]

Abstract. We re-examine the phylogenetic position of *Pelusios williamsi* by merging new sequences with an earlier published data set of all *Pelusios* species, except the possibly extinct *P. seychellensis*, and the nine previously identified lineages of the closely allied genus *Pelomedusa* (2054 bp mtDNA, 2025 bp nDNA). Furthermore, we include new sequences of *Pelusios broadleyi*, *P. castanoides*, *P. gabonensis* and *P. marani*. Individual and combined analyses of the mitochondrial and nuclear data sets indicate that *P. williamsi* is sister to *P. castanoides*, as predicted by morphology. This provides evidence for the misidentification of GenBank sequences allegedly representing *P. williamsi*. Such mislabelled GenBank sequences contribute to continued confusion, because only the original submitter can revise their identification; an impractical procedure impeding the rectification of obvious mistakes. We recommend implementing another option for revising taxonomic identifications, paralleling the century-old best practice of natural history museums for new determinations of specimens. Within *P. broadleyi*, *P. gabonensis* and *P. marani*, there is only shallow genetic divergence, while some phylogeographic structuring is present in the wide-ranging species *P. castaneus* and *P. castanoides*.

Keywords: Africa, Kenya, mtDNA, nDNA, phylogeny, systematics.

African hinged terrapins (genus *Pelusios*) are with 17-18 currently recognized species one of the most speciose genera of turtles and tortoises (Ernst, Altenburg and Barbour, 2000; Fritz and Havaš, 2007; Rhodin et al., 2010; Fritz et al., 2011). However, in a recent molecular study using mitochondrial and nuclear DNA sequences of all *Pelusios* species except the possibly extinct *P. seychellensis*, we have identified many taxonomic issues. Among others, GenBank sequences of *P. williamsi* were phylogenetically embedded within sequences of *P. castaneus* (Fritz et al., 2011). According to its morphology, *P. williamsi* should be closely allied to *P. castanoides* and not to *P. castaneus* (Bour, 1986). In contrast to all other studied species, *P. williamsi* was only represented by GenBank sequences in our previous study. Therefore, we have concluded that the phylogenetic position of *P. williamsi* needs to be re-investigated using fresh material. In the present paper, we add newly generated sequences of two *P. williamsi* to the data set of Fritz et al. (2011) and examine their phylogenetic placement. Furthermore, we include in our analyses sequences of new samples of *P. broadleyi*, *P. gabonensis* and *P. marani*. The latter two species were represented by only one individual in our previous study. *Pelusios broadleyi* is a badly known species, in which we found little variation before (Fritz et al., 2011). Moreover, we supplement our previous data with sequences from a Kenyan specimen of *P. castanoides*, so that this species is now represented by terrapins from Kenya, South Africa, Madagascar and the Seychelles. Like in our earlier paper, we include in our analyses all nine lineages of the only other pelomedusid genus *Pelomedusa*, as identified by Vargas-Ramírez et al. (2010).

Using blood samples of two *Pelusios broadleyi*, one *P. castanoides*, two *P. gabonensis*, four *P. marani* and two *P. williamsi*, the same mitochondrial (12S rRNA, cyt *b*, ND4 plus adjacent DNA, in part coding for tRNAs) and nuclear genes (C-mos, R35, Rag2) as in Fritz et al. (2011) were sequenced. The *P. broadleyi*, *P. castanoides* and *P. williamsi* were directly imported from Kenya by the international

1 - Museum of Zoology, Senckenberg Dresden, A. B. Meyer Building, D-01109 Dresden, Germany
2 - Department of Biology and Wildlife Diseases, Faculty of Veterinary Hygiene and Ecology, University of Veterinary and Pharmaceutical Sciences, Palackého 1-3, CZ-612 42 Brno, Czech Republic
*Corresponding author; e-mail: uwe.fritz@senckenberg.de

 DOI:10.1163/156853812X627204

pet trade, while the other terrapins were wild-caught and have precise locality data (table 1). Laboratory procedures were described in Fritz et al. (2011). For PCR and sequencing, the following primers were applied: L1091 + H1478 (12S rRNA), L-ND4 + H-Leu (ND4 plus flanking DNA), CB1 + L (cyt *b*), G136 + G137 (C-mos), R35Ex1 + R35Ex2 (R35), and F2-1 + R2-1 (Rag2). However, for sequencing the ND4 gene of challenging samples, the newly designed primer pair Pelusios_ND4_Seq_F (TAAATATAGCCCTYCCMCC) and Pelusios_ND4_Seq_R (AGTAGAGYGCTGAYATTA) was used. Other primer sequences were given in Fritz et al. (2011). Due to small sample size or degraded DNA, not all genes could be produced for each sample (table 1).

The new sequences were aligned in BIOEDIT 7.0.5.2 (Hall, 1999) with the *Pelusios* and *Pelomedusa* sequences of Fritz et al. (2011; see there for GenBank accession numbers). Phylogenetic relationships were inferred for three data sets, viz. (i) the concatenated mitochondrial sequences, corresponding to a total length of 2054 bp, (ii) the concatenated nuclear sequences, 2025 bp, and (iii) a supermatrix of 4079 bp length, consisting of the merged nuclear and mitochondrial data sets. The nuclear data set and the supermatrix contained only 10 new sequences, because one sample (*Pelusios broadleyi*) did not work for the nuclear genes. For phylogenetic analyses the original data set of Fritz et al. (2011) was pruned, so that for each species maximally two of the previously sequenced individuals were retained, except for *P. castanoides*. This species is thought to be closely allied to *P. williamsi* (Bour, 1986), which is why all *P. castanoides* sequences of Fritz et al. (2011) were kept. *Podocnemis expansa* served as outgroup (see Fritz et al., 2011 for GenBank accession numbers). Phylogeny was inferred using the Maximum Likelihood (ML) approach as implemented in RAxML 7.2.6 (Stamatakis, 2006) via the graphical user interface raxmlGUI 0.93 (Silvestro and Michalak, 2011). The partition scheme was the same as in Fritz et al. (2011); the default GTR + G model was applied across all partitions. Five independent ML searches were performed with the fast bootstrap algorithm to explore the robustness of the branching patterns by comparing the best-scored trees. Subsequently, 1000 thorough bootstrap replicates were calculated and plotted against the tree with the highest likelihood value. In addition, for each data set Bayesian analyses (BA) were run using MrBAYES 3.1.2 (Ronquist and Huelsenbeck, 2003). The best-fit model of sequence evolution was selected using the Akaike Information Criterion (AIC) as implemented in MODELTEST 3.7 (Posada and Crandall, 1998) and incorporated into a single tree search (mixed model approach). The models for the respective partitions were: 12S rRNA – GTR + I + G; cyt *b* – TVM + I + G; ND4 – TVM + I + G; tRNA-His – HKY + G; non-annotated partition (see Fritz et al., 2011) – HKY + G; tRNA-Leu – K81; C-mos – K81 + G; R35 – HKY + I; and Rag2 – HKY. Two simultaneous analyses with four Markov chains were run for 10 000 000 generations, with every 100th generation being saved. Posterior probabilities were obtained from the 50% majority rule consensus trees. For each independent run, the variation of likelihood scores was examined by plotting $-\ln L$ scores against the number of generations, and the burn-in was set to sample only the plateau of the most likely trees.

Tree topologies and their support values were in close agreement with the RAxML and BA trees of our previous paper (Fritz et al., 2011); minor differences occurred only with respect to weakly supported branching patterns. There was little intraspecific variation within *Pelusios broadleyi*, *P. gabonensis*, *P. marani* and *P. williamsi*. In all trees, the new sequences of *P. williamsi* were with high support sister to the East African species *P. castanoides*, while GenBank sequences identified with *P. williamsi* were placed among West African *P. castaneus*. In the generally weakly resolved nuclear trees (not shown), the sister group relationship of the new *P. williamsi* sequences and *P. castanoides* received a bootstrap support value of 88 (ML) and a posterior probability of 1.0 (BA). In the mitochondrial trees (not shown) and the trees based on the combined data set (fig. 1), the sister group relationship was maximally supported under both methods. Compared to the other mentioned species, intraspecific variation was more pronounced within *P. castaneus* and *P. castanoides*. In the nuclear trees, the Kenyan sample of *P. castanoides* was slightly distinct from the others. In the mitochondrial trees and the trees using the combined data set, the Kenyan and South African terrapins had a basal position within *P. castanoides* and the Seychellois and Malagasy terrapins were sister groups.

Our new sequence data of *P. williamsi* support, in agreement with morphology (Bour, 1986), that this species is closely allied to the East African *P. castanoides*. This provides clear evidence that GenBank sequences labelled as *P. williamsi* (12S rRNA: accession number U81324, cyt *b*: U81347, R35: AY339629) are misidentified. These sequences cluster in phylogenetic analyses with high support with the West African species *P. castaneus* (Fritz et al., 2011; this study). Such mislabelled GenBank sequences contribute to continued confusion, because many authors use BLAST searches and GenBank sequences as standard tools for

Table 1. Studied *Pelusios* samples and their GenBank accession numbers. Sample numbers are MTD T numbers (Museum of Zoology Dresden, Tissue Collection). The ND4 fragment contained also adjacent mtDNA, in part coding for tRNAs.

Sample	Species	Provenance	GenBank accession numbers					
			12S rRNA	cyt *b*	ND4	C-mos	R35	Rag2
7290	*Pelusios broadleyi*	Kenya	JQ352029	JQ352037	JQ352048	JQ352059	JQ352068	JQ352078
7291	*Pelusios broadleyi*	Kenya	–	JQ352038	JQ352049	–	–	–
7288	*Pelusios castanoides*	Kenya	JQ352030	JQ352039	JQ352050	JQ352060	JQ352069	JQ352079
7283	*Pelusios gabonensis*	Democratic Republic of the Congo: Gbadolite village	–	JQ352040	JQ352051	JQ352061	JQ352070	JQ352080
7284	*Pelusios gabonensis*	Democratic Republic of the Congo: Gbadolite village	JQ352031	JQ352041	JQ352052	JQ352062	JQ352071	JQ352081
7293	*Pelusios marani*	Gabon: Mourimatsengui, 5 km from Yombi	JQ352032	JQ352042	JQ352053	JQ352063	JQ352072	JQ352082
7294	*Pelusios marani*	Gabon: Mourimatsengui, 5 km from Yombi	JQ352033	JQ352043	JQ352054	JQ352064	JQ352073	JQ352083
7295	*Pelusios marani*	Gabon: Mourimatsengui, 5 km from Yombi	JQ352034	JQ352044	JQ352055	–	JQ352074	JQ352084
7296	*Pelusios marani*	Gabon: Mourimatsengui, 5 km from Yombi	–	JQ352045	JQ352056	JQ352065	JQ352075	JQ352085
7287	*Pelusios williamsi*	Kenya	JQ352035	JQ352046	JQ352057	JQ352066	JQ352076	JQ352086
7289	*Pelusios williamsi*	Kenya	JQ352036	JQ352047	JQ352058	JQ352067	JQ352077	JQ352087

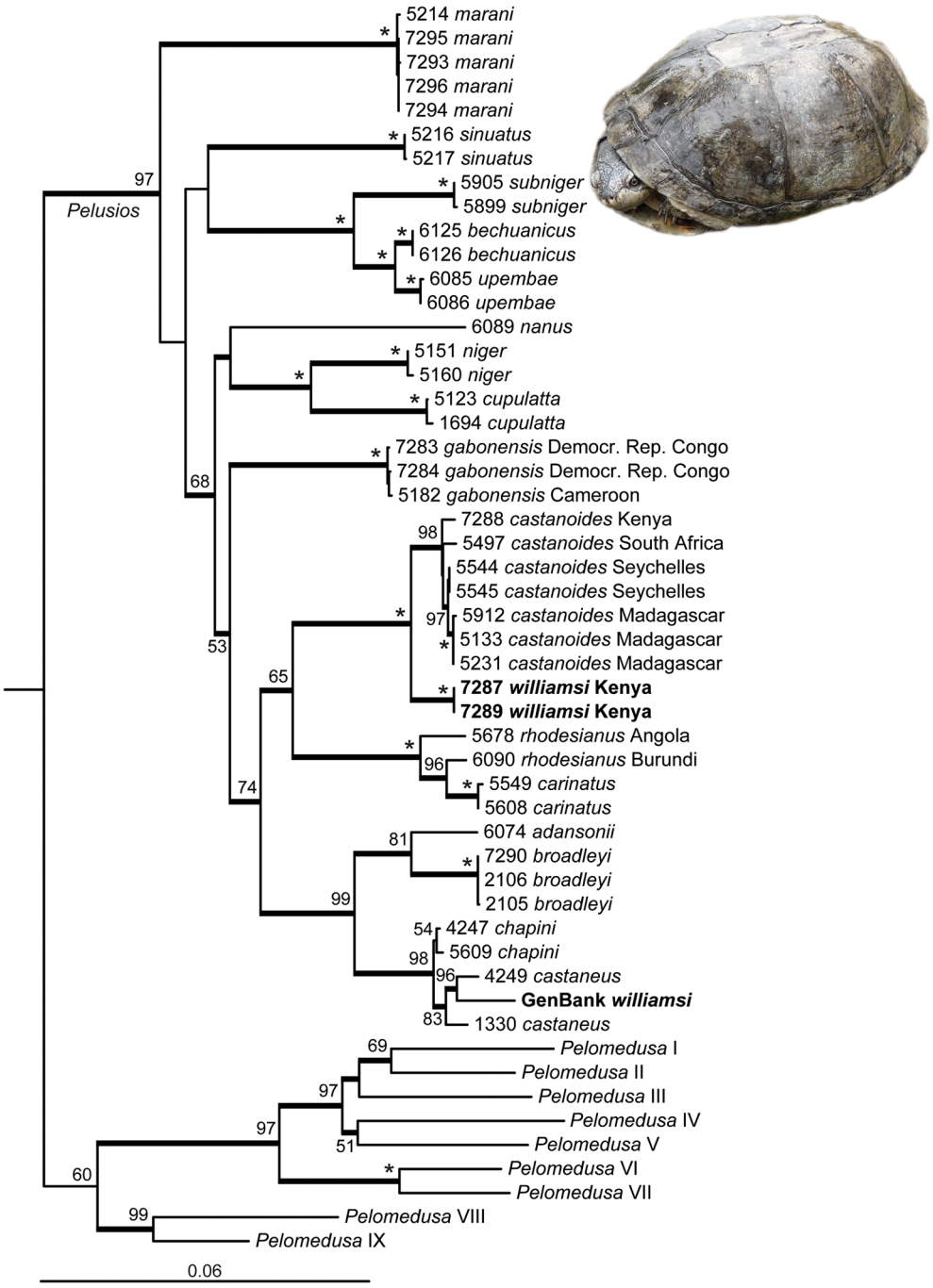

Figure 1. Phylogeny of *Pelusios* species and the nine lineages of *Pelomedusa* as inferred by Maximum Likelihood analysis, based on an alignment of 2054 bp of mitochondrial and 2025 bp of nuclear DNA. New sequences of *Pelusios williamsi* and GenBank sequences allegedly representing the same species in boldface. Sample codes preceding taxon names refer to table 1 or Fritz et al. (2011: table S1). Numbers along branches are thorough bootstrap values > 50; asterisks indicate maximum support. Support values are not shown for some terminal clades with short branches. Branches in bold are supported by posterior probabilities $\geqslant 0.99$ in Bayesian analyses. Outgroup (*Podocnemis expansa*) removed for clarity. Inset: female *Pelusios williamsi* from Kenya (photo by H. Prokop). This figure is published in colour in the online version.

the identification of DNA sequences. We learnt from correspondence with GenBank that their current regulations do not allow the correction of any taxonomic misidentification, unless it is done by the original submitter – an impractical and overly complicated procedure, and for many obvious reasons with uncertain outcome. Among others, it necessitates identifying and convincing the original submitter to admit a mistake.

Taxonomic misidentifications occur frequently, not only with respect to GenBank sequences, and every scientist working in collections of natural history museums knows that it is good practice there that the original identifiers, the original labels, are kept with the specimens. However, when there is new evidence, taxonomic identifications are revised by adding another label with the new determination and the reviser's name. We strongly recommend that GenBank and its allied data bases adopt a similar procedure.

Acknowledgements. Hynek Prokop donated samples of captive terrapins. Most laboratory work was done by Anja Rauh. Ylenia Chiari and an anonymous reviewer provided helpful comments.

References

Bour, R. (1986): Note sur *Pelusios adansonii* (Schweigger, 1812) et sur une nouvelle espèce affine du Kenya (Chelonii, Pelomedusidae). Stud. Geol. Salmant., Vol. Esp. **2** (Studia Palaeocheloniologica II): 23-54.

Ernst, C.H., Altenburg, R.G.M., Barbour, R.W. (2000): Turtles of the World. World Biodiversity Database, CD-ROM Series, Windows, Version 1.2. Amsterdam, Biodiversity Center of ETI.

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

Fritz, U., Branch, W.R., Hofmeyr, M.D., Maran, J., Prokop, H., Schleicher, A., Široký, P., Stuckas, H., Vargas-Ramírez, M., Vences, M., Hundsdörfer, A.K. (2011): Molecular phylogeny of African hinged and helmeted terrapins (Testudines: Pelomedusidae: *Pelusios* and *Pelomedusa*). Zool. Scr. **40**: 115-125.

Hall, T.A. (1999): BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. **41**: 95-98.

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

Rhodin, A.G.J., van Dijk, P.P., Iverson, J.B., Shaffer, H.B. (2010): Turtles of the world, 2010 update: annotated checklist of taxonomy, synonymy, distribution, and conservation status. Chelon. Res. Monogr. **5**: 000.85-000.164.

Ronquist, F., Huelsenbeck, J.P. (2003): MrBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**: 1572-1574.

Silvestro, D., Michalak, I. (2011): raxmlGUI: a graphical front-end for RAxML. Org. Divers. Evol.: DOI:10.1007/s13127-011-0056-0.

Stamatakis, A. (2006): RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**: 2688-2690.

Vargas-Ramírez, M., Vences, M., Branch, W.R., Daniels, S.R., Glaw, F., Hofmeyr, M.D., Kuchling, G., Maran, J., Papenfuss, T.J., Široký, P., Vieites, D.R., Fritz, U. (2010): Deep genealogical lineages in the widely distributed African helmeted terrapin: evidence from mitochondrial and nuclear DNA (Testudines: Pelomedusidae: *Pelomedusa subrufa*). Mol. Phylogenet. Evol. **56**: 428-440.

Received: November 15, 2011. Accepted: January 18, 2012.