Molecular Phylogenetics and Evolution 68 (2013) 269–281

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Misleading phylogenetic inferences based on single-exemplar sampling in the turtle genus *Pseudemys*

Phillip Q. Spinks[a,b,*], Robert C. Thomson[c], Gregory B. Pauly[d], Catherine E. Newman[e], Genevieve Mount[a,b], H. Bradley Shaffer[a,b]

[a] *Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA*
[b] *La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095, USA*
[c] *Department of Biology, University of Hawaii at Manoa, Honolulu, HI 96822, USA*
[d] *Section of Herpetology, Natural History Museum of Los Angeles County, 900 Exposition Blvd., Los Angeles, CA 90007, USA*
[e] *Department of Biological Sciences, Museum of Natural Science, Louisiana State University, 119 Foster Hall, Baton Rouge, LA 70803, USA*

ARTICLE INFO

Article history:
Received 24 October 2012
Revised 5 March 2013
Accepted 25 March 2013
Available online 12 April 2013

Keywords:
Molecular systematics
Mitonuclear discord
Multiple exemplars
Single exemplar
Pseudemys
Emydidae

ABSTRACT

Reconstructing species trees for clades containing weakly delimited or incorrectly identified taxa is one of the most serious challenges facing systematists because building phylogenetic trees is generally predicated on correctly identifying species membership for the terminals in an analysis. A common practice, particularly in large-scale phylogenetic analyses, is to use single-exemplar sampling under the implicit assumption that the resulting phylogenetic trees will be poorly supported if the sampled taxa are not good species. We examine this fundamental assumption in the North American turtle genus *Pseudemys*, a group of common, widely distributed freshwater turtles whose species boundaries and phylogenetic relationships have challenged systematists for over half a century. We sequenced 10 nuclear and three mitochondrial genes from the nine currently recognized species and subspecies of *Pseudemys* using geographically-widespread sampling of each taxon, and analyzed the resulting 86-individual data set using population-genetic and phylogenetic methods. We found little or no evidence supporting the division of *Pseudemys* into its currently recognized species/subspecies. Rather, our data strongly suggest that the group has been oversplit and contains fewer species than currently recognized. Even so, when we conducted 100 replicated, single-exemplar phylogenetic analyses of these same nine taxa, most Bayesian trees were well resolved, had high posterior probabilities, and yet returned completely conflicting topologies. These analyses suggest that phylogenetic analyses based on single-exemplar sampling may recover trees that depend on the individuals that are sampled, rather than the underlying species tree that systematists assume they are estimating. Our results clearly indicate that final resolution of *Pseudemys* will require an integrated analysis of morphology and historical biogeographic data coupled with extensive geographic sampling and large amounts of molecular data, and we do not recommend taxonomic changes based on our analyses. If our 100-tree resampling experiments generalize to other taxa, they suggest that single-exemplar phylogenies should be interpreted with caution, particularly for groups where species are shallowly diverged or inadequately delimited.

1. Introduction

Molecular phylogenetics encompasses a wide range of evolutionary problems, from recovering the deepest nodes in the Tree of Life to delimiting recently derived species, and methodological progress has moved forward at both ends of this spectrum. However, taxa that fall between relatively well-differentiated phylogenetic lineages and potentially subdivided populations often remain problematic because of the stochastic nature of gene-tree coalescence, potential introgression, and low information content of molecular sequences (Degnan and Rosenberg, 2006; Hudson and Coyne, 2002; Maddison and Knowles, 2006; Moore, 1995). These problems pose a challenge for species delimitation and downstream species-tree reconstruction because species-delimitation methods often require the use of a fully resolved input species phylogeny (e.g. Knowles and Carstens, 2007; Yang and Rannala, 2010), while species-tree reconstruction models assume little or no horizontal gene flow and often require that individuals be assigned to species *a priori* (Heled and Drummond, 2010; Kubatko et al., 2009; Liu and Pearl, 2007). Groups that contain many described but weakly differentiated (and potentially interbreeding) species, and

---

* Corresponding author. Address: Department of Ecology and Evolutionary Biology, University of California, 621 Charles E. Young Dr. South, Los Angeles, CA 90095-1606, USA.

E-mail address: pqspinks@ucla.edu (P.Q. Spinks).

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.03.031

those that exhibit relatively little phylogenetic structure can be particularly problematic. In some cases, the most desirable solution would be an integrated approach that resolves both the species boundaries and the species phylogeny, but these methods are parameter-rich and require relatively informative input data for reliable inferences. Thus the joint resolution of phylogeny and species boundaries might not be feasible for recently diverged rapid radiations or other groups characterized by extremely low levels of genetic variation (Carstens and Dewey, 2010; O'Meara, 2010; Polihronakis, 2010; Weisrock et al., 2012).

Although clear solutions to these taxonomic and phylogenetic challenges need further development, one way forward is to hypothesize that currently-recognized species are real and then use single or multiple-exemplar sampling to estimate species trees. A critical question when following this approach is, what is the relationship between the accuracy of the hypothesized species lineages and the resulting species tree? Although seldom made explicit, two underlying assumptions often characterize this approach in the phylogenetics literature. First, the reconstruction of single-exemplar species phylogenies implies that the contained species are distinct lineages, although this is seldom tested. Second, if the named species are not distinct lineages, then both single and multiple-exemplar phylogenies should return poorly resolved trees, owing to the shifting and uncertain placement of problematic taxa or individuals within the final collection of trees. To our knowledge this relationship between correctly delimited species and resulting species trees has never been formally explored with simulated or real data. If these assumptions are correct, and taxonomic inflation (Isaac et al., 2004) or oversplitting (Dayrat, 2005) has led to the naming of indistinct lineages, then we can make two predictions about resulting species trees. It seems reasonably clear that analyses based on multiple exemplars per species should recover poorly resolved, paraphyletic species lineages. More controversially, single-exemplar trees should have low bootstrap or posterior probabilities for nodes involving these lineages. The converse logic should also hold: if phylogenies generated from single or multiple exemplars per species are well resolved, it implies that the taxa under study are themselves well-resolved lineages. One goal of the current study is to test these predictions empirically.

Turtles present many examples of taxonomically problematic groups that exhibit hybridization and introgression, incomplete coalescence, and low information content resulting from an overall reduced rate of molecular evolution in the group (Shaffer et al., 2013), ms). Taken together, these problems have produced several taxonomic controversies throughout the turtle tree of life. Well-known examples include the Mediterranean Spur-thighed Tortoise (*Testudo graeca*) complex, Australian and New Guinean members of the genus *Emydura* and the North American cooters in the genus *Pseudemys*. Each of these radiations is widespread, morphologically variable, generally common, and well studied. Each also contains one or more taxa listed as an endangered species, making it even more critical that species are accurately delimited. The *T. graeca* complex is considered to be a single, or as many as 10 species (Parham et al., 2006; Turkozan et al., 2010), while *Emydura* has been considered to comprise four species or up to seven species including eight contained subspecies (Georges and Thomson, 2010). However, within chelonians, *Pseudemys* may be the most extreme example of a taxonomically confusing group. *Pseudemys* is a group of freshwater turtles (family Emydidae, subfamily Deirochelyinae) distributed throughout the southeastern US/northern Mexico, from New Mexico, Texas and adjacent Mexico east to Florida and north to Massachusetts (Conant and Collins, 1991, Fig. 1). For turtles, which comprise only 331 living species (TTWG, 2012), *Pseudemys* is a relatively large group (7–9 recognized species), but species boundaries among most named entities are uncertain and have been the subject of extensive and conflicting revision (Carr, 1952; Carr and Crenshaw, 1957; Fahey, 1980; Jackson, 1995; Seidel, 1994; TTWG, 2012; Ward and Jackson, 2008).

Pseudemys has long been recognized as something of a taxonomic quagmire, and has been the subject of a confusing litany of taxonomic arrangements (Fahey, 1980). For example, Leary et al. (2008, pp. 019.1) wrote:

The Alabama red-bellied turtle (*P. alabamensis*) was considered to be an invalid taxon and was designated as a "mutant of *P. floridana mobilensis*" (=*P. concinna mobilensis*) (Carr, 1938), or a variant of "*P. floridana suwanniensis*" (=*P. c. suwanniensis*) (Carr, 1952). It was also included within what is now *P. nelsoni* (De Sola, 1935), or considered a subspecies of *P. rubriventris* (Steneger, 1938; Wermuth and Mertens, 1961, 1977).

The taxonomic history of other *Pseudemys* species is similarly convoluted (Carr, 1952; Fahey, 1980; Jackson, 1995; Seidel, 1994; TTWG, 2012; Ward and Jackson, 2008), suggesting that *Pseudemys* constitutes a reasonable case study to explore the relationship between species delimitation and species-tree reconstruction in groups where both have been difficult to determine.

In this paper, we use a large, multilocus dataset to analyze both species delimitation and phylogenetic relationships across *Pseudemys*. Clarification of species boundaries and phylogeny of this clade is important for at least three reasons: *Pseudemys* constitutes an abundant part of the aquatic vertebrate fauna of the southeastern US; the currently recognized taxonomy makes *Pseudemys* a major contributor (nine taxa) to the identification of the Gulf Coast region of the US as the area of greatest chelonian species richness on earth (Buhlmann et al., 2009); and, it contains an endangered species, *P. alabamensis*, under the US Endangered Species Act (ESA). As an initial working hypothesis, we follow the TTWG (2012) and recognize nine taxa within *Pseudemys* including seven species of which one has three subspecies. As detailed in the annotations to earlier versions of the Turtle Taxonomy Working Group, both the taxonomy and content of several species complexes has remained controversial, rendering it difficult to identify a single taxonomy for *Pseudemys*. Many recent authors recognize three species, *P. alabamensis*, *P. nelsoni*, *P. rubriventris* that are assigned to the redbelly, or "*rubriventris*" group. The remaining six generally recognized taxa include *P. gorzugi*, *P. peninsularis*, *P. texana*, *P. concinna concinna*, *P. concinna floridana*, and *P. concinna suwanniensis,* which are collectively assigned to the river cooter, or "*concinna*" group (Seidel, 1994). Most current authors agree on the recognition of *P. gorzugi*, *P. texana*, and *P. concinna* as taxa; the most active debate currently centers on *P. peninsularis* (distinct species vs. subspecies of *P. floridana*), *P. floridana* (distinct species or subspecies of *P. concinna*), and *P. suwanniensis* (distinct species or subspecies of *P. concinna*). Although each of these species was initially recognized based on color pattern and morphological features, many of these characters show marked overlap among different hypothesized species (Carr, 1952; Carr and Crenshaw, 1957; Fahey, 1980; Seidel, 1994), leading to the unsettled taxonomy for the group. We follow the configuration of seven species and three additional subspecies of *P. concinna* (*concinna*, *floridana*, and *suwanniensis*) in this study.

Despite the many taxonomic revisions of *Pseudemys* (Carr and Crenshaw, 1957; Jackson, 1995; Seidel, 1994), only six phylogenetic analyses have been completed for the genus, and only two have incorporated multiple individuals/species. Stephens and Wiens (2003, 2008, 2009), and Wiens et al. (2010) generated single-exemplar species phylogenies from mitochondrial DNA (mtDNA), nuclear DNA (nuDNA) and morphological characters for the turtle family Emydidae (which includes *Pseudemys*), while Jackson et al. (2012) and Seidel (1994) focused on *Pseudemys*, generating phylogenies from multiple individuals/species (mtDNA, and morphometric plus one protein electrophoretic character,