Molecular Phylogenetics and Evolution 70 (2014) 130–136

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Assessing species boundaries and the phylogenetic position of the rare Szechwan ratsnake, *Euprepiophis perlaceus* (Serpentes: Colubridae), using coalescent-based methods

Xin Chen [a,b,*,1], Ke Jiang [c,1], Peng Guo [d], Song Huang [e], Dingqi Rao [c], Li Ding [f], Hirohiko Takeuchi [g], Jing Che [c], Ya-ping Zhang [c], Edward A. Myers [a,b], Frank T. Burbrink [a,b]

[a] Department of Biology, The College of Staten Island, The City University of New York, 2800 Victory Blvd., Staten Island, NY 10314, United States
[b] Department of Biology, The Graduate School and University Center, The City University of New York, 365 Fifth Ave., NY, NY 10016, United States
[c] State Key Laboratory of Genetic Resources and Evolution, and Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, PR China
[d] College of Life Science and Food Engineering, Yibin University, Yibin 644007, Sichuan, PR China
[e] College of Life and Environment Sciences, Huangshan University, Huangshan 245021, Anhui, PR China
[f] Department of Herpetology, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, Sichuan, PR China
[g] Department of Zoology, Graduate School of Science, Kyoto University, Sakyo, Kyoto 606-8502, Japan

ARTICLE INFO

Article history:
Received 24 May 2013
Revised 3 September 2013
Accepted 7 September 2013
Available online 21 September 2013

Keywords:
Species delimitation
Species tree
Coalescence
Gene tree
Euprepiophis perlaceus

ABSTRACT

Delimiting species and clarifying phylogenetic relationships are the main goals of systematics. For species with questionable taxonomic status, species delimitation approaches using multi-species coalescent models with multiple loci are recommended if morphological data are unavailable or unhelpful. Moreover, these methods will also reduce subjectivity based on genetic distance or requirement of monophyletic genetic lineages. We determine the validity and phylogenetic position of a rare and long controversial species of Chinese reptile, the Szechwan ratsnake (*Euprepiophis perlaceus*), using multi-locus data from multiple individuals and coalescent-based approaches. Species were first delimited using Bayesian Phylogenetics & Phylogeography (BP&P), Brownie and Bayes Factor model comparison approaches, while relationships among species were estimated using species tree inference in *BEAST. Results indicate that *Euprepiophis perlaceus* is a distinct species sister to *Euprepiophis mandarinus*. Despite gene tree discrepancy, the coalescent model-based approaches used here demonstrate the taxonomic validity and the phylogenetic position of *Euprepiophis perlaceus*. These approaches objectively test the validity of questionable species diagnoses based on morphological characters and determine their phylogenetic position.

1. Introduction

Delimiting species and clarifying the phylogenetic relationship among taxa are crucial objectives in systematics, both of which are important for enumerating and understanding the origins of biodiversity and implementing conservation management strategies (Myers et al., 2013; Fujita et al., 2012; O'Meara, 2010; Sites and Marshall, 2004; Wilson, 2003). For taxon delimitation, a species concept and the criteria to define the species boundaries are necessary. However, lineage separation and divergence are temporally extended processes. Researchers have until recently relied on multiple species concepts based on criteria to designate species at different stages of the speciation continuum. This has impaired consensus on a unified species definition and taxonomic stability (de Queiroz, 2007). Traditionally, the use of morphological traits has dominated species delimitation and taxonomic inquiry. However, relying solely on these data may underestimate diversity by failing to detect cryptic taxa where morphological variation among species is slight or nonexistent. Dependency on these characters may ultimately mislead phylogenetic inference due to convergent evolution of states under similar natural selection pressures (Bickford et al., 2007; Edwards et al., 2012; Paris et al., 1989; Yang and Rannala, 2010). As the ease of generating molecular data has increased dramatically, genetic markers have been widely used in systematic studies. For example, genetic barcoding has been used as a method to assign unknown samples into existing species using a single-locus (Hebert et al., 2003; Tautz et al., 2003) and gene tree methods are often used to infer phylogeny (Felsenstein,

---

* Corresponding author at: Department of Biology, The College of Staten Island, The City University of New York, 2800 Victory Blvd., Staten Island, NY 10314, United States. Fax: +1 718 982 3852.
E-mail address: xchen1@gc.cuny.edu (X. Chen).

[1] These authors contributed equally to this work.

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.09.003

2004; Hillis and Moritz, 1996). Nevertheless, methods based on criteria such as genetic distance or reciprocal monophyly to identify species boundaries are also subjective (Hey, 2009) and fail to account for incomplete lineage sorting, introgression and gene duplication causing gene-tree incongruence with the species tree. Genetic barcoding may also fail to assign individuals into the corresponding species (Edwards, 2009; Hickerson et al., 2006; O'Meara, 2010).

According to some authors, integrating multiple sources of data (e.g. molecular, morphological, ecological, and behavioral) for taxonomy is expected to reduce subjectivity in delimiting taxa (Dayrat, 2005; Leaché et al., 2009; Padial et al., 2010; Schlick-Steiner et al., 2010; Yeates et al., 2011). However, in situations where cryptic species with indistinct morphological and ecological properties are present, integrative methods will not yield a clear result since different types of data may yield discordant conclusions (Fujita et al., 2012; Wiens and Penkrot, 2002). Coalescent models, which describe the evolutionary processes of populations by tracing alleles back to their most recent common ancestor in a mathematical and probabilistic framework (Kingman, 2000; Wakeley, 2008) have been recommended for species delimitation and species tree inference (Fujita et al., 2012; O'Meara, 2010). Gene trees of different loci should sort randomly when assuming a panmictic population without selection, migration or linkage. A speciation event will yield gene trees that are more similar among loci than expected under the neutral hypothesis (O'Meara, 2010). Similarly, coalescent model based methods use multi-locus markers to estimate important parameters for inferring a species tree given lineage sorting (e.g. population size θ and divergence time τ, Fujita et al., 2012). In this way, the discordance between species trees and gene trees can be avoided and still provide an objective estimation of speciation events and phylogenies (Camargo et al., 2012; Edwards, 2009).

We use multiple loci and multi-species coalescent-based approaches to test the validity of the rare and ambiguous species of Chinese ratsnake, the Szechwan ratsnake (*Euprepiophis perlaceus*), as well as clarify its phylogenetic relationship to other ratsnakes. *Euprepiophis perlaceus* has a narrow distribution in mountainous regions with an elevation range from 1650 m to 2500 m in western Sichuan, China (Fig. 1). Preferred habitats are rocky areas in humid broad-leaf forests and shrubby grassland (Hu et al., 2002). Stejneger (1929) originally described the species as *Elaphe perlacea* using morphological characters from just one male specimen. Since this

Fig. 1. The distribution of *Euprepiophis perlaceus* (red), *Euprepiophis mandarinus* (green), and *Euprepiophis conspicillatus* (orange, Schulz, 1996; Zhao and Adler, 1993). Samples of *E. perlaceus* (red squares), *E. mandarinus* (blue circles), and *E. conspicillatus* (orange triangles) used in this study are indicated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

description, only three additional specimens were collected in the 1980s, however tissue samples were not preserved for molecular data (Zhao, 1990). Thus, the taxonomic certainty of this species has been called into doubt. Schulz (1989) suggested assigning this species as a subspecies of *Elaphe mandarina*. Zhao (1990) argued that *Elaphe perlacea* should be considered as a valid species based on a combination of unique scale characters and distinctive color patterns that differ from *Elaphe mandarina*. The last treatment of this species was by Schulz (1996), who followed Zhao (1990) and listed *Elaphe perlacea* as a full species in his monograph of the genus *Elaphe*. More recently, molecular data have been used to evaluate phylogenetic relationships within the genus *Elaphe*. Based on a mitochondrial gene tree, *Elaphe mandarina* and *Elaphe conspicillata* were placed in the genus *Euprepiophis*, as was *Elaphe perlacea* because of its close association with these two species, yet the validity and phylogenetic placement of *Euprepiophis perlaceus* remains untested because of the lack of molecular data for this taxon (Utiger, 2002). In 2010, three specimens of *Euprepiophis perlaceus* were collected and tissue samples preserved, permitting us to determine species status and investigate the taxonomic affinities of this species.

The distribution of *Euprepiophis mandarinus* and *Euprepiophis pearlaceus* overlap in mainland China (Fig. 1) and share many morphological similarities (Schulz, 1996), whereas *Euprepiophis conspicillatus* is only distributed on the islands of Japan and is morphologically distinct from the other two species in this genus. Thus, it has been hypothesized that *Euprepiophis perlaceus* is a valid species and is sister to *Euprepiophis mandarinus* (Schulz, 1996; Zhao, 1990), yet these descriptions rely primarily on labile color and scutellation characters (Burbrink, 2001; Burbrink et al., 2000). To test this hypothesis, multiple loci were sequenced from several individuals of all three species in the genus *Euprepiophis* as well as five other closely related species to serve as outgroups (Burbrink and Lawson, 2007). With these data, we delimit species using coalescent model based approaches and provide a comparison of methods from Bayesian Inference (Yang and Rannala, 2010), a nonparametric heuristic search (O'Meara, 2010) and the relatively new Bayes Factor (BF) model comparison framework (Carstens and Dewey, 2010; Grummer et al., in press; Knowles and Carstens, 2007). Species tree methods are used to estimate phylogeny to avoid gene tree/species tree conflicts.

2. Materials and methods

2.1. Tissue samples and DNA sequencing

We collected 3 tissue samples of *Euprepiophis perlaceus*, 10 samples of *Euprepiophis mandarinus*, 2 samples of *Euprepiophis conspicillatus*(Fig. 1) and multiple individuals from 5 close related taxa to serve as outgroups; these include *Elaphe carinata*, *Orthriophis taeniurus*, *Oreocryptophis porphyraceus*, *Rhadinophis frenatus* and *Ptyas korros* and have been chosen based on pervious phylogenetic analyses (Burbrink and Lawson, 2007; Utiger, 2002). Although *Euprepiophis mandarinus* has a wide distribution in East Asia, the availability of genetic tissue samples is limited. Therefore, we collected samples that covered the range of its distribution (the approximate localities of the 10 samples are labeled in Fig. 1). Voucher numbers of all samples used for this study are listed in Supplementary data (Appendix Table S1). DNA extractions were conducted using QIAGEN DNeasy Kits and four gene regions were amplified and sequenced including one mitochondrial protein-coding gene cytochrome *b* (cyt-b, 1102 bp) and three nonprotein-coding nuclear fragments: SPTBN1 (774 bp, Matthee et al., 2001), Vimentin Intron 4 (722 bp, Zehner and Paterson, 1983) and Vimentin Intron 5 (587 bp, Zehner and Paterson, 1983) following the protocols listed in Pyron and Burbrink (2009). The primers used are listed in Supplementary data (Appendix Table S2). GenBank accession numbers of genes are given in Supplementary data (Appendix Table S1). Sequences were edited manually in SEQUENCHER 4.2 (Genecodes Corp.), were aligned in Geneious Pro 4.7.5 (Biomatters Ltd.) using the MUSCLE algorithm with default parameters (Edgar, 2004) and checked by eye for ambiguous alignments. An open reading frame in the final alignment was maintained for the protein coding gene cyt-b. The completeness of taxa for cyt-b, SPTBN1, Vimentin Introns 5 is 100%; while only one *Euprepiophis mandarinus* (M7, GP73) did not sequence for Vimentin Intron 4.

2.2. Phylogenetic inference

The appropriate substitution models for each gene were determined in the program jModelTest 2.1.2 based on Bayesian Information Criterion (BIC) values (Darriba et al., 2012; Guindon and Gascuel, 2003). The most probable pair of alleles for each nuclear gene was determined using the program PHASE v2.1.1 with 90% as the threshold for genotype certainty (Stephens and Donnelly, 2003; Stephens et al., 2001). Species trees were constructed with phased sequence data set using *BEAST, which estimates the species tree and all gene trees in one Bayesian MCMC analysis simultaneously, in the software package BEAST v. 1.7.4 (Drummond et al., 2012; Heled and Drummond, 2010). Eight terminal taxa were defined including all three putative species within the genus *Euprepiophis* and the five outgroup taxa. The substitution, clock and tree models were unlinked among genes and an uncorrelated lognormal rate distribution was used under a birth-death process. Cytochrome *b*, a protein-coding gene, was partitioned by codon position. We ran two *BEAST analyses each with 600 million generations and discarded the first 100 million generations as burn-in after checking the status of parameters in Tracer 1.5 (Rambaut and Drummond, 2007) to ensure the stationary of each run. The XML code for species tree estimation is available in Dryad: doi:http://dx.doi.org/10.5061/dryad.053m3. To compare the estimated phylogenies using species tree methods and gene tree methods, we also inferred the phylogeny with a concatenated gene matrix, using both Bayesian Inference (BI) and Maximum Likelihood (ML) approaches. We generated a BI concatenated gene tree in the program MrBayes 3.2.1 (Ronquist et al., 2012) with a model partitioned by genes and codon position in the protein-coding gene. Two runs were performed each with 25 million generations; each analysis was run with one cold and three heated chains. The first 25% of generations were discarded as burn-in after checking the status in Tracer. A ML tree was constructed and assessed for nodal support with the rapid-boot strapping algorithm using 1000 nonparametric bootstraps in the program RaxML v 7.2.8 (Stamatakis, 2006b; Stamatakis et al., 2008) by partitioning genes and codon positions under the GTRCAT model (Stamatakis, 2006a).

2.3. Species delimitation

Multiple coalescent-based species delimitation methods have been developed under different statistical frameworks. To validate the species *Euprepiophis perlaceus*, we performed species delimitation analyses using three approaches. First, we used Bayesian Phylogenetics & Phylogeography (BP&P). This program calculates the posterior probabilities of models where species numbers differ using reversible-jump Markov chain Monte Carlo (rjMCMC, Yang and Rannala, 2010) and assumes no migration post speciation. Each model contains three parameters, θ (N_e [effective population size] $\times$ μ [mutation rate]), τ_a (the time at which the species originated), τ_d (the time in which one species split into two descendent species). This method requires a fully resolved phylogeny as a guide tree because the rjMCMC algorithm can only evaluate the status of a node on the tree by either splitting (1) or collapsing (0) the node without branch swapping (Leaché and Fujita, 2010). We performed species delimitation analyses in the program BP&P v2.1b (Yang and Rannala, 2010) using the phased sequence data of each locus in two different tests. One test included just *Euprepiophis mandarinus* and *Euprepiophis perlaceus* in order to assess whether the BP&P supports the splitting of the two species or alternatively if the program would merge both taxa into one species. The second test included the addition of *Euprepiophis conspicillatus* sequences into our dataset, where we used the topology from the *BEAST species tree as the BP&P guide tree. We performed multiple runs using algorithm 0 with fine-tune parameters 2, 5, 10, 20 for ε and algorithm 1 with fine-tune parameters 1, 1.5, 2 for α and 0.5, 1, 2 for m to confirm the stability of the rjMCMC. We set up three different prior combinations of ancestral population size (θ) and root age (τ_0) using a gamma distribution $\Gamma(1,10)$, $\Gamma(2,1000)$, $\Gamma(2,2000)$ as suggested by (Leaché and Fujita, 2010; Yang and Rannala, 2010) to test the effect of these prior distributions on the model posterior probability. For each test, BP&P was run for 1,000,000 generations with starting seed equals to -1 and the first 20,000 generations were discarded as burn-in. The fine-tuning variables were adjusted to keep the acceptance proportions for MCMC moves within the interval of (0.15, 0.7). To test the robustness of these analyses, we repeated each test by randomly assigning individuals into different lineages.

Next, we performed analyses using a nonparametric delimitation approach, which assumes that for a speciation event, the corresponding nodes on gene trees will be more consistent with each other than the divergences within species. This algorithm has been implemented in the program Brownie v2.1 (O'Meara, 2010; O'Meara et al., 2006). The best species tree and species delimitation model was estimated simultaneously from input gene trees of the different loci where the species tree topology is not constrained a priori. Similar to BP&P, gene flow is assumed not to occur among species. Gene trees for the four loci were generated in the software package BEAST. For nuclear genes, we randomly picked one allele from each individual. We partitioned substitution models and linked clock models and trees. Substitution models, clock models and tree priors were setup as described in Section 2.2. Two analyses were performed and each analysis was run for 300 million generations with the first 50 million generations removed as burn-in after checking the stationary of parameters in Tracer. For the heuristic search parameters in Brownie, the number of random starting species trees (NReps) was set to 100, all possible taxon reassignments on leaf splits were explored (Subsample = 1), and the minimum number of samples per species (MinSamp) was set to 2. Tests were repeated multiple times to ensure the optimal delimited species tree.

Additionally, we attempted to delimit species by comparing two species models assuming *Euprepiophis perlaceus* and *Euprepiophis mandarinus* either as two independent species (Model 1) or as one species (Model 2) following methods similar to that of Carstens and Dewey (2010) using a Bayes Factor (BF) approach for model selection. Unlike the other model comparison approaches (likelihood ratio test [LRT], Akaike Information Criterion [AIC], and Bayesian Information Criterion [BIC], etc.), BF calculates the ratio of the marginal likelihood of two models, which has the advantage of taking into account priors used in Bayesian analyses (Xie et al., 2011). The marginal likelihood values of these two competing models (Model 1, Model 2) were estimated using recently developed techniques including path sampling (PS, Lartillot and Philippe, 2006) and stepping-stone sampling (SS, Xie et al., 2011), which have previously demonstrated to have better performance than the harmonic mean estimator (HME, Newton and Raftery, 1994) by not overestimating the true marginal likelihood (Baele et al., 2012a,b; Fan et al., 2011; Grummer et al., in press; Xie et al., 2011). Calculations of PS and SS were performed in BEAST v 1.7.4. First, we ran species tree analyses in *BEAST based on Model 1 and Model 2 respectively, using the same phased dataset for phylogenetic inference including all three lineages of *Euprepiophis* and five outgroup species. Model 1 assigns eight taxa with *Euprepiophis perlaceus* and *Euprepiophis mandarinus* as two independent species; Model 2 assigns seven taxa after merging *Euprepiophis perlaceus* and *Euprepiophis mandarinus* as one species. All parameters were setup as described in the Section 2.2. This allows for the direct comparison of the two models considering both the topology and the branch lengths of species trees. After generating XML files from BEAUti v1.7.4 (Drummond et al., 2012), we added additional code in order to collect samples along the path between the prior and the posterior (available at http://beast.bio.ed.ac.uk/Model_selection; Baele et al., 2012a,b). The first 100 million generations of species tree results were discarded as burn-in; samples collected during two MCMC runs for PS and SS were combined using LogCombiner v 1.7.4 in the BEAST package (Drummond et al., 2012). Marginal likelihood values of the two models were calculated using PS and SS; all XML files used in running the MCMC and to calculate the marginal likelihood are available on Dryad: doi:http://dx.doi.org/10.5061/dryad.053m3.

3. Results

3.1. Species delimitation

Species delimitation analyses in the program BP&P using either the two taxon dataset (*Euprepiophis perlaceus, Euprepiophis mandarinus*) or the three taxon dataset (*Euprepiophis perlaceus, Euprepiophis mandarinus, Euprepiophis conspicillatus*) support the node separating *Euprepiophis perlaceus* and *Euprepiophis mandarinus* with a posterior probability (Pp) = 1.0,where all ESS parameters are above 5000 (Appendix Table S3). Additionally, tests where samples were randomly assignedinto each lineage support collapsing these randomized lineages into one (Pp > 0.98), verifying the robustness of these analyses. The results from the nonparametric method also support *Euprepiophis perlaceus* as a separate species, sister to *Euprepiophis mandarinus* with the highest score 6.4. Bayes Factor model selection results show that PS and SS marginal likelihood estimators strongly prefer Model 1 ($2\ln \mathrm{BF} > 10$, Kass and Raftery, 1995), which supports *Euprepiophis perlaceus* as a valid species (Table 1).

3.2. Phylogenetic inference

The most appropriate substitution models for Cytochrome *b* (cyt-b), SPTBN1, Vimentin Introns 4, Vimentin Introns 5 were HKY + Γ, HKY + Γ, GTR + Γ, HKY + Γ, respectively. The relationships among genus *Euprepiophis* and the other outgroup genera here are consistent with previous studies (Burbrink and Lawson, 2007; Utiger, 2002). The species tree estimated in *BEAST supports *Euprepio-*

Table 1
The species delimitation results using Bayes Factor (BF) model selection approaches. Model 1 supports *Euprepiophis perlaceus* and *Euprepiophis mandarinus* as two independent species and Model 2 supports *Euprepiophis perlaceus* and *Euprepiophis mandarinus* as a single species. The log likelihood transformed marginal likelihood values of two models are listed in the table with the preferred model labeled with "*". PS = path sampling; SS = stepping-stone sampling.

	PS	SS
Marginal likelihood estimator		
Model 1	-11830.53^*	-11830.39^*
Model 2	-11849.47	-11849.00
$2\ln \mathrm{BF}$	37.88	37.22

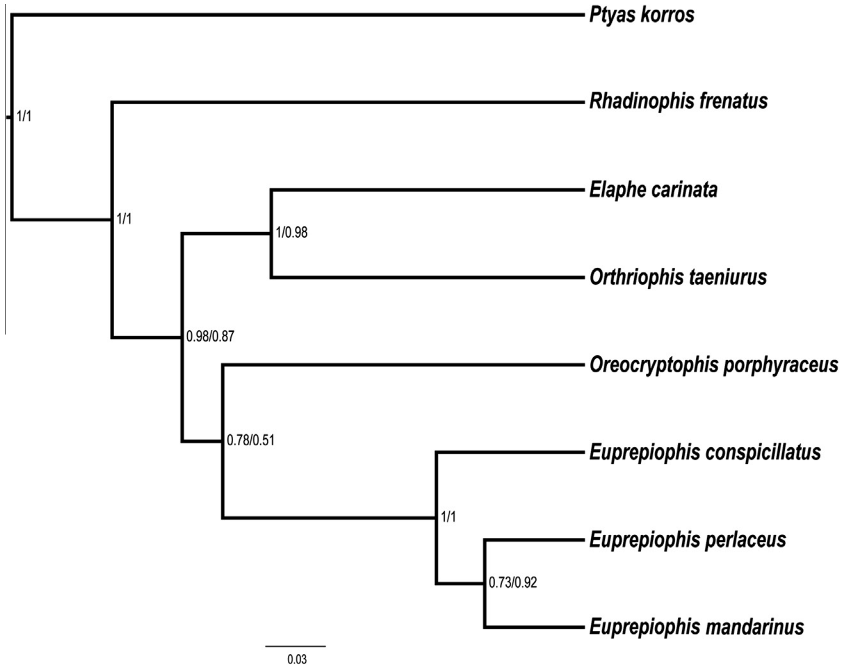

Fig. 2. Species tree estimated using *BEAST. The numbers on nodes are posterior probability support values. Posterior probabilities on the left side of the slash were generated using phased nuclear data and cyt-b; posterior probabilities on the right side of the slash were generated using only phased nuclear genes.

phis perlaceus as sister to *Euprepiophis mandarinus* (Fig. 2). However, concatenated gene trees are incongruent with respect to supporting a sister relationship between *Euprepiophis perlaceus* and *Euprepiophis conspicillatus* (Appendix Fig. S1a). This phylogenetic relationship is also shown in the cyt-b gene tree (Pp = 1, Appendix Fig. S1b) but not in the three nuclear gene trees. Thus, to quantitatively examine the impact of each locus to the concatenated gene tree, we compared the similarity of the concatenated gene tree with the gene tree of each locus by calculating Robinson–Foulds (RF) distance (Steel and Penny, 1993) using package Phangorn (Schliep, 2011) in the statistical program R (R Development Core Team, 2012). These results indicate that the concatenated gene tree is most similar to the cyt-b gene tree (RF distance = 16; Appendix Fig. S1), suggesting that the topology of concatenated gene trees is dominated by signal from cyt-b. To test the influence of the mitochondrial/nuclear gene conflict, we repeated the species tree analyses using only the phased sequence data of the three nuclear loci. The resulting species tree with only nuclear data has the same topology supporting the relationship of *Euprepiophis perlaceus* and *Euprepiophis mandarinus* as sister lineages with higher Pp (0.92, Fig. 2).

4. Discussion

4.1. Coalescent species delimitation of the Szechwan ratsnake

We have determined that the rare Szechwan ratsnake, *Euprepiophis perlaceus*, which was originally described based on morphological traits of one specimen, is indeed a distinct species. Morphological differences are slight between *Euprepiophis perlaceus* and *Euprepiophis mandarinus*, yet they are congruent with our results using genetic data and coalescent analyses. While all evidence suggests *Euprepiophis perlaceus* is a distinct species, very little information is known about the conservation status of this rare and narrowly distributed taxon (Schulz, 1996). Similar cases are not uncommon and assessing the species boundaries of rare taxa is critical for properly evaluating regional biodiversity (Lim et al., 2011).

With respect to delimiting species, such as in the case of Chinese ratsnakes of the genus *Euprepiophis*, multi-species coalescent methods overcome the potential subjectivity stemming from the use of morphological traits, incorrect assumptions based on strict reciprocal monophyly of gene trees, and increase the objectivity of delimiting species using molecular data. Of the three species-delimitation approaches used here, the BP&P method calculates the posterior probability (Pp) of each node either supporting lineage splitting or lineage collapsing. Since incorrect phylogenetic relationships in the guide tree will mislead the results supporting lineage splitting with high Pp, a fully resolved and correct guide tree is necessary (Leaché and Fujita, 2010). Thus, the other two tests using Brownie and BF model comparison, without constraining the tree topology a priori were also used and yield similar results to BP&P. For the BF model comparison approach, among multiple marginal likelihood estimators used for calculating the likelihood values including harmonic mean estimator (HME), the stabilized/smoothed harmonic mean estimator (sHME, Redelings and Suchard, 2005), path sampling (PS), and stepping-stone sampling (SS), HME and sHME (an extension of HME) have been demonstrated to perform poorly in choosing the correct model (Baele et al., 2012a,b; Fan et al., 2011; Xie et al., 2011). Specifically, a simulation study comparing the power of HME, sHME, PS and SS for the species model selection in *BEAST shows that only PS and SS estimators are sensitive in identifying the correct species model with both over splitting or lumping of lineages (Grummer et al., in press). Thus, to ensure the accuracy of marginal likelihood estimations, we chose to use results from the PS and SS estimators to calculate BF. In the case with the ratsnakes here, all of our results suggest that these species are distinct and that all of these methods, even with very different assumptions, yield the same answer with high support.

4.2. Gene-tree and species-tree conflicts on phylogenetic inference

Unlinked genetic loci have independent evolutionary histories, and thus, gene trees based on different loci from different regions of the genome may not generate congruent topologies or represent the real species phylogeny (Tajima, 1983). The conflicts between gene trees and species trees could be caused by incomplete lineage sorting, hybridization, gene duplication or branch length heterogeneity (Edwards, 2009; McCracken and Sorenson, 2005; McGuire et al., 2007). Therefore, results from species-level phylogenies that have used only a single locus or a concatenated gene matrix should be interpreted cautiously (Wahlberg et al., 2009). Species tree methods and concatenated gene tree methods used here to infer the phylogeny show inconsistent topologies where species trees support the previous hypothesis based on morphology and biogeographic proximity that the Chinese taxa, *Euprepiophis mandarinus* and *Euprepiophis perlaceus* are sister taxa (Fig. 2); while concatenated gene trees support *Euprepiophis perlaceus* as sister to the Japanese *Euprepiophis conspicillatus* (Fig. 1, Appendix Fig. S1). Comparing gene trees of the four genetic markers with the concatenated tree, the Cytochrome *b* gene tree topology is most similar and supports *Euprepiopihs perlaceus* as sister to *Euprepiopihs conspicillatus* with high Pp (1.0). This suggests that the concatenated gene tree may be dominated by site patterns from the mitochondrial gene (Appendix Fig. S1a and b). The SPTBN1 and Vimentin Intron 4 gene trees show the individuals of three *Euprepiophis* lineages are unsorted (Appendix Fig. S1c) and only the Vimentin Intron 5 gene tree gives the same topology as that of the species tree (Appendix Fig. S1d). Thus, our results provide empirical support for the hypothesis that inferring phylogeny using a concatenated gene matrix or a single locus has a high probability of being different from the species tree (Edwards, 2009; Maddison, 1997). Coalescent-based species tree methods avoid gene tree conflicts and integrate the information of gene trees to estimate the phylogeny representing the evolution of species or populations. The species tree generated with four loci provides low Pp support for the node uniting *Euprepiophis perlaceus* and *Euprepiophis mandarinus* (Fig. 2). Considering the mitochondrial/nuclear gene conflict, we repeated the species tree analysis with the three nuclear loci and obtained a species tree with the same topology and higher Pp (0.92) supporting *Euprepiophis perlaceus* and *Euprepiophis mandarinus* as sister taxa (Fig. 2). This result is also consistent with the previous hypothesis based on the information from the geographic distribution and morphological similarities of these taxa (Schulz, 1996; Zhao, 1990). The mitochondrial gene tree, which shows an unconventional topology, yet with high support, indicates that mtDNA may be sorted incorrectly relative to the species tree and thus responsible for the alternative placement of taxa in *Euprepiophis*. Currently, with the development of new sequencing techniques, large phylogenomic data sets are readily available. Based on numerous loci across the genome, reconstructed phylogenies are expected to avoid the ambiguity of a single gene tree and therefore be robust to reflect phylogenetic relationship of different groups.

5. Conclusion

In this study, consistent results from three coalescent-based species delimitation methods confirm the validity of rare and controversial species, *Euprepiophis perlaceus*. Despite gene tree discrepancies, the reconstructed phylogeny using species tree methods shows *Euprepiophis perlaceus* and *Euprepiophis mandarinus* as sister species, which is also supported by geographic and morphological evidence. We provide more empirical support that coalescent-based methods are indeed valuable for delimiting rare species. We therefore recommend these approaches, which avoid ambiguities of morphological traits and subjectively defined criteria, to test the validity of questionable diagnoses that are based on ambiguous morphological characters analyzed from few specimens.

Acknowledgments

This research was supported, in part, by the National Science Foundation Grants CNS-0958379 and CNS-0855217 and the City University of New York High Performance Computing Center. We thank all of the researchers whose time and effort went in to producing sequence data used in this project. We thank Jun Zhang to help collecting *Euprepiophis perlaceus*. We thank Hui Zhao who provided the photo of *Euprepiophis mandarinus*. We thank Jared A. Grummer, Julianne R. Goldenberg for providing suggestions about using the BF method for species delimitation. Field work for collecting *Euprepiophis mandarinus*, *Euprepiophis perlaceus*, *Elaphe carinata*, *Orthriophis taeniurus*, *Ptyas korros*, *Oreocryptophis porphyraceus* tissue samples was partially supported by grants from the National Natural Science Foundation of China (NSFC30870290, 31071891) to Song Huang, from the National Natural Science Foundation of China (NSFC 31071892, NSFC31372152) and the Department of education of Sichuan Province (13TD0027) to Peng Guo and from the Chinese Academy of Sciences (KSCX2-EW-Z-2; KSCX2-EW-Q-9) and the Bureau of Science and Technology of Yunnan, China (2010CI045) to Jing Che. Additional financial support was provided by L. Clampitt, S. Harris, and D. Rosenberg.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev. 2013.09.003.

References

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M.A., Alekseyenko, A.V., 2012a. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29 (9), 2157–2167.

Baele, G., Li, W.L.S., Drummond, A.J., Suchard, M.A., Lemey, P., 2012b. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. Mol. Biol. Evol. 30 (2), 239–243.

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K., Das, I., 2007. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22 (3), 148–155.

Burbrink, F.T., 2001. Systematics of the eastern ratsnake complex (*Elaphe obsoleta*). Herpetol. Monogr. 15, 1–53.

Burbrink, F.T., Lawson, R., 2007. How and when did old world ratsnakes disperse into the new world? Mol. Phylogenet. Evol. 43 (1), 173–189.

Burbrink, F.T., Lawson, R., Slowinski, J.B., 2000. Mitochondrial DNA phylogeography of the polytypic North American ratsnake (*Elaphe obsoleta*): a critique of the subspecies concept. Evolution 54 (6), 2107–2118.

Camargo, A., Morando, M., Avila, L.J., Sites, J.W., 2012. Species delimitation with ABC and other coalescent-based methods: a test of accuracy with simulations and an empirical example with lizards of the *Liolaemus darwinii* complex (Squamata: Liolaemidae). Evolution 66 (9), 2834–2849.

Carstens, B.C., Dewey, T.A., 2010. Species delimitation using a combined coalescent and information-theoretic approach: an example from North American *Myotis* bats. Syst. Biol. 59 (4), 400–414.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModeltest 2: more models, new heuristics and parallel computing. Nat. Meth. 9 (8), 772.

Dayrat, B., 2005. Towards integrative taxonomy. Biol. J. Linn. Soc. 85 (3), 407–415.

de Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56 (6), 879–886.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29 (8), 1969–1973.

Edgar, R.C., 2004. Muscle: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792–1797.

Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63 (1), 1–19.

Edwards, S., Vanhooydonck, B., Herrel, A., Measey, G.J., Tolley, K.A., 2012. Convergent evolution associated with habitat decouples phenotype from phylogeny in a clade of lizards. PLoS One 7 (12), e51636.

Fan, Y., Wu, R., Chen, M.-H., Kuo, L., Lewis, P.O., 2011. Choosing among partition models in Bayesian phylogenetics. Mol. Biol. Evol. 28 (1), 523–532.

Felsenstein, J., 2004. Inferring Phylogenies. Sinauer Associates Inc..

Fujita, M.K., Leaché, A.D., Burbrink, F.T., McGuire, J.A., Moritz, C., 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27 (9), 480–488.

Grummer, J.A., Rob, W., Bryson, J., Reeder, T.W., 2013. Species delimitation using Bayes Factors: Simulations and application to the *Sceloporus scalaris* species group (Squamata: Phrynosomatidae). Syst. Biol. (in press).

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52 (5), 696–704.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. London, B: Biol. Sci. 270 (1512), 313–321.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27 (3), 570–580.

Hey, J., 2009. On the arbitrary identification of real species. In: Butlin, R.K., Bridle, J.R., Schluter, D. (Eds.), Speciation and Patterns of Diversity. Cambridge University Press, Cambridge, UK, pp. 15–28.

Hickerson, M.J., Meyer, C.P., Moritz, C., 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. Syst. Biol. 55 (5), 729–739.

Hillis, D.M., Moritz, C., 1996. Molecular Systematics, second ed. Sinauer Associates Inc..

Hu, J., Li, Y., Li, C., Hu, J., Wang, H., 2002. The present status of transverse-banded racer. Sichuan J. Zool. 3 (21), 184–185.

Kass, R.E., Raftery, A.E., 1995. Bayes Factors. J. Am. Statist. Assoc. 90 (430), 773–795.

Kingman, J.F.C., 2000. Origins of the coalescent: 1974–1982. Genetics 156 (4), 1461–1463.

Knowles, L.L., Carstens, B.C., 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56 (6), 887–895.

Lartillot, N., Philippe, H., 2006. Computing Bayes Factors using thermodynamic integration. Syst. Biol. 55 (2), 195–207.

Leaché, A.D., Fujita, M.K., 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proc. R. Soc. London, B: Biol. Sci. 277 (1697), 3071–3077.

Leaché, A.D., Koo, M.S., Spencer, C.L., Papenfuss, T.J., Fisher, R.N., McGuire, J.A., 2009. Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). Proc. R. Soc. London, B: Biol. Sci. 106 (30), 12418–12423.

Lim, G.S., Balke, M., Meier, R., 2011. Determining species boundaries in a world full of rarity: singletons, species delimitation methods. Syst. Biol. 61 (1), 165–169.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46 (3), 523–536.

Matthee, C.A., Burzlaff, J.D., Taylor, J.F., Davis, S.K., 2001. Mining the mammalian genome for artiodactyl systematics. Syst. Biol. 50 (3), 367–390.

McCracken, K.G., Sorenson, M.D., 2005. Is homoplasy or lineage sorting the source of incongruent mtDNA and nuclear gene trees in the stiff-tailed ducks (*Nomonyx-oxyura*)? Syst. Biol. 54 (1), 35–55.

McGuire, J.A., Linkem, C.W., Koo, M.S., Hutchison, D.W., Lappin, A.K., Orange, D.I., Lemos-Espinal, J., Riddle, B.R., Jaeger, J.R., 2007. Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. Evolution 61 (12), 2879–2897.

Myers E. A., Rodríguez-Robles J. A., DeNardo D. F., Staub R. E., Stropoli A., Ruane S., and Burbrink F. T., Multilocus phylogeographic assessment of the California Mountain Kingsnake (Lampropeltis zonata) suggests alternative patterns of diversification for the California Floristic Province, Mol. Eco. 2013, n/a–n/a. doi:10.1111/mec.12478.

Newton, M.A., Raftery, A.E., 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. J. Roy. Stat. Soc. B 56, 3–48.

O'Meara, B.C., 2010. New heuristic methods for joint species delimitation and species tree inference. Syst. Biol. 59 (1), 59–73.

O'Meara, B.C., Ané, C., Sanderson, M.J., Wainwright, P.C., 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60 (5), 922–933.

Padial, J., Miralles, A., De la Riva, I., Vences, M., 2010. The integrative future of taxonomy. Front. Zool. 7 (1), 16.

Paris, C.A., Wagner, F.S., Wagner Jr., W.H., 1989. Cryptic species, species delimitation, and taxonomic practice in the homosporous ferns. Am. Fern J. 79 (2), 46–54.

Pyron, R.A., Burbrink, F.T., 2009. Neogene diversification and taxonomic stability in the snake tribe Lampropeltini (Serpentes: Colubridae). Mol. Phylogenet. Evol. 52 (2), 524–529.

R Development CoreTeam, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.4. <http://beast.Bio.Ed.Ac.Uk/tracer>.

Redelings, B.D., Suchard, M.A., 2005. Joint Bayesian estimation of alignment and phylogeny. Syst. Biol. 54 (3), 401–418.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61 (3), 539–542.

Schlick-Steiner, B.C., Steiner, F.M., Seifert, B., Stauffer, C., Christian, E., Crozier, R.H., 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. Annu. Rev. Entomol. 55 (1), 421–438.

Schliep, K.P., 2011. Phangorn: phylogenetic analysis in R. Bioinformatics 27 (4), 592–593.

Schulz, K.-D., 1989. Die hinterasiatischen kletternattern der gattung *Elaphe*. Teil xvi: *Elaphe perlacea*, Stejneger, 1929. Sauria 11 (2), 15–16.

Schulz, K.-D., 1996. A Monograph of the Colubrid Snakes of the Genus *Elaphe* Fitzinger. Koeltz Scientific Books.

Sites, J.W., Marshall, J.C., 2004. Operational criteria for delimiting species. Annu. Rev. Ecol. Evol. Syst. 35, 199–227.

Stamatakis, A., 2006a. Phylogenetic models of rate heterogeneity: a high performance computing perspective. In: Proceedings of the 20th IEEE/ACM International Parallel and Distributed Processing Symposium (IPDPS 2006). Rhodos, Greece.

Stamatakis, A., 2006b. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22 (21), 2688–2690.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57 (5), 758–771.

Steel, M.A., Penny, D., 1993. Distributions of tree comparison metrics—some new results. Syst. Biol. 42 (2), 126–141.

Stejneger, L., 1929. A snake from China. Proc. Biol. Soc. Wash. 42, 129–130.

Stephens, M., Donnelly, P., 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73 (5), 1162–1169.

Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68 (4), 978–989.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105 (2), 437–460.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., Vogler, A.P., 2003. A plea for DNA taxonomy. Trends Ecol. Evol. 18 (2), 70–74.

Utiger, 2002. Molecular systematics and phylogeny of Old and New World ratsnakes, *Elaphe* auct., and related genera (Reptilia, Squamata, Colubridae). Russ. J. Herpetol. 9 (2), 105–124.

Wahlberg, N., Weingartner, E., Warren, A.D., Nylin, S., 2009. Timing major conflict between mitochondrial and nuclear genes in species relationships of Polygonia butterflies (Nymphalidae: Nymphalini). BMC Evol. Biol. 9, 92.

Wakeley, J., 2008. Coalescent Theory: An Introduction. Roberts & Company Publishers.

Wiens, J.J., Penkrot, T.A., 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Syst. Biol. 51 (1), 69–91.

Wilson, E.O., 2003. The encyclopedia of life. Trends Ecol. Evol. 18 (2), 77–80.

Xie, W., Lewis, P.O., Fan, Y., Kuo, L., Chen, M.-H., 2011. Improving marginal likelihood estimation for bayesian phylogenetic model selection. Syst. Biol. 60 (2), 150–160.

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. Proc. Natl. Acad. Sci. USA 107, 9264–9269.

Yeates, D.K., Seago, A., Nelson, L., Cameron, S.L., Joseph, L.E.O., Trueman, J.W.H., 2011. Integrative taxonomy, or iterative taxonomy? Syst. Entomol. 36 (2), 209–217.

Zehner, Z.E., Paterson, B.M., 1983. Characterization of the chicken Vimentin gene: single copy gene producing multiple mRNAs. Proc. Natl. Acad. Sci. USA 80 (4), 911–915.

Zhao, E., 1990. The validity of *Elaphe perlacea*, a rare endemic snake from Sichuan province, China. Asiatic Herpetol. Res. 3, 101–103.

Zhao, E., Adler, K., Herpetology of China. 1993, SSAR, Oxford, Ohio.