Molecular Phylogenetics and Evolution 92 (2015) 108–117

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Concordant species delimitation from multiple independent evidence: A case study with the *Pachytriton brevipes* complex (Caudata: Salamandridae) ☆

Yunke Wu [a,*], Robert W. Murphy [b]

[a] Museum of Comparative Zoology and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[b] Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, 100 Queen's Park, Toronto, Ontario M5S 2C6, Canada

ARTICLE INFO

Article history:
Received 15 October 2014
Revised 16 June 2015
Accepted 17 June 2015
Available online 25 June 2015

Keywords:
Amphibians
Statistical species delimitation
Bayesian methods
Ecological niche modeling
Vicariance
Southeastern China

ABSTRACT

Mitochondrial DNA (mtDNA) sequence data are widely used to delimit species. However, owing to its strict maternal inheritance in most species, mtDNA tracks female dispersion and dispersal only. The accuracy of mtDNA-derived species delimitation is often not explicitly tested using other independent evidence, such as nuclear DNA (nDNA) data, morphological data, or ecological data. Because species are independent evolutionary lineages that can form testable hypotheses, we present a multi-evidence case study on species delimitation that combines statistical approaches with spatially explicit ecological analysis. Montane salamanders of the *Pachytriton brevipes* complex (Salamandridae) from southeastern China exhibit conservative morphology and variable color patterning that impede species diagnosis. Recent studies proposed splitting *P. brevipes* into four species based on deep mtDNA divergence but also found discordance between mtDNA and nDNA trees. In this study, we test evolutionary independence of these hypothesized species lineages using two coalescent-based Bayesian methods (Bayes factor and BP&P). Despite significant conflict between mtDNA gene tree and the species phylogeny, the results reinforce the inference of at least four species in the complex as opposed to the one species recognized for over 130 years. Correlative ecological niche modeling and statistical analysis of environmental data indicate that suitable habitats for each species are isolated by incompatible intervening lowland regions, so the likelihood of gene flow among species is very low, which means species lineages should maintain their evolutionary independence. We demonstrate that concordance among independent evidence confirms species status, which forms the basis for accurate assessment of regional biodiversity.

1. Introduction

Species are the fundamental units of comparative biology. They are the basis for communicating and classifying biodiversity and for performing ecological assessments. Species concepts are controversial, yet we cannot communicate without referring to species. Delimitation of species often depends on the underlying species concept and its particular operational criteria (Mayden, 1997). For example, the phylogenetic species concept requires monophyly, but hybridization violates the concept's operational basis, even in *Homo sapiens* (Vernot and Akey, 2014). Species also are independently evolving lineages (Simpson, 1961; Wiley, 1978). In the evolutionary species concept, emphasis shifts from the search for diagnostic characters among lineages to tests of explicit hypotheses of lineage independence (Wiley, 1978; Pons et al., 2006; Yang and Rannala, 2010; Fujita et al., 2012).

Divergence of mitochondrial DNA (mtDNA) has been widely used in the discovery of cryptic geographic species (e.g., Vieites et al., 2009) and in species delimitation (e.g., Pons et al., 2006; Monaghan et al., 2009). For example, during 2011 to August 2013, there were 133 new species (spp.) described among Asian amphibians (AmphibiaWeb, 2014). Seventy-two of these species—more than 50%—were recognized based on large mtDNA distance from known species or mitochondrial gene-tree monophyly. This trend is not declining: 2011, 38/65 spp.; 2012, 17/36 spp.; and 2013, 17/32 spp. However, because the mitochondrial genome evolves as a single locus, deep coalescence and/or mitochondrial introgression via hybridization may hamper identification of the most appropriate number of independently evolving lineages (Edwards, 2009; Near and Keck, 2013). In other words, divergent mtDNA lineages may not equal evolutionary species. For example,

☆ This paper was edited by the Associate Editor A. Larson.

* Corresponding author at: Department of Ecology and Evolutionary Biology, Cornell University, Corson Hall, Ithaca, NY 14853, USA.
E-mail address: yw578@cornell.edu (Y. Wu).

http://dx.doi.org/10.1016/j.ympev.2015.06.010
1055-7903/

a Gulf of California spiny-tailed iguana, *Ctenosaura nolascensis*, contains two highly divergent mtDNA haplotype clades that do not form a monophyletic group with respect to other species (Davy et al., 2011).

Earlier studies rarely approach taxonomy from the basis of explicit hypothesis testing, despite species being hypotheses of genetic cohesiveness. Yet, recent developments offer statistical methods for species delimitation. These methods provide a direct, repeatable and objective means of testing hypotheses of evolutionary independence by using probabilistic models that do not rely on monophyly of haplotypes within species lineages (Fujita et al., 2012). Independence is critical because species hybridize. Delimitations involve nested or non-nested models in which likelihoods or posterior probabilities evaluate whether the mtDNA, nuclear DNA (nDNA) or combined data reject the hypothesis that populations studied form a single species lineage (Pons et al., 2006; Yang and Rannala, 2010; Grummer et al., 2014). Most methods allow for gene tree discordance owing to, for example, deep coalescence (e.g., incomplete lineage sorting), which they address by superimposing the coalescent process on the phylogeny. Such methods are robust to relatively low levels of gene flow (Eckert and Carstens, 2008; Zhang et al., 2011). Analyses assume that the best model gives the most appropriate classification of species boundaries, and that the results are testable using independent, non-genetic data sources. For example, ecological data can test the hypothesis that a species' ecology drives and will maintain lineage independence (e.g., Rissler and Apodaca, 2007).

Pachytriton brevipes (Salamandridae) is an aquatic, mostly nocturnal salamander (commonly called the spotted stout newt) that inhabits small montane streams in densely covered, "sky-island" forests from southeastern China (Zhao and Hu, 1984; Fei et al., 2006). After over 130 years of taxonomic stability, starting in 2011 *P. brevipes* was split into four species based largely on deep mtDNA divergences and reciprocal monophyly (Nishikawa et al., 2011; Wu et al., 2012). Because species lineages diagnosed by mtDNA data were not recovered by nDNA data, Wu et al. (2013) called them a species complex. Their conservative morphology and variable color patterning limits the number of diagnostic characters (Wu et al., 2010). These attributes make the species complex an ideal system for exploring species delimitation. Herein, we use the *P. brevipes* complex as an example to test whether the mtDNA-derived taxonomy depicts species diversity more reliably than alternative delimitation hypotheses. Based on multiple independent data (mtDNA, nDNA, and spatially explicit environmental data), we test for the null hypothesis of conspecificity using two coalescent-based Bayesian methods, correlative ecological niche modeling (ENM), and spatial statistical analysis. Topological constraints and Bayes factors further test if the mtDNA gene tree and the species tree conflict significantly. Our study provides an example of concordant species delimitation from independent evidence.

2. Material and methods

2.1. Data collection

We sampled 178 salamanders from 31 localities throughout the range of the *Pachytriton brevipes* complex, including *P. brevipes*, *P. feii*, *P. granulosus* and *P. xanthospilos* and their respective type localities (Fig. 1; Supplementary Materials, Table S1). This extended the sampling of the species complex in Wu et al. (2013) by 4.5 fold. Nishikawa et al. (2012) described *P. changi* based on two specimens of unknown origin obtained from the Japanese pet market. They considered *P. changi* and *P. xanthospilos* to be conspecific despite moderate genetic and morphometric differences (Nishikawa et al., 2013). We use the latter name in the absence of a type locality for the former; the absence of locality and nDNA data for *P. changi* (impossible to collect specimens) has precluded its use in most subsequent analyses. *Pachytriton inexpectatus*, *P. moi* and *P. archospotus* are used as outgroup taxa.

Total genomic DNA was extracted from ethanol-preserved liver or muscle tissue using a DNeasy Blood and Tissue Kit (QIAGEN Corp.). Two nDNA fragments covered the 3′ end of *RAG-1* (~1200 base pairs, bp) along with a non-coding region of the gene that encodes tyrosinase (*NCRT*, 600 bp). Two mitochondrial DNA (mtDNA) genes, *ND2* and *cytb* and their flanking tRNAs (each with ~1200 bp) were amplified and sequenced (Wu et al., 2010, 2013). DNA sequences were aligned manually using Se-Al 2.0 (Rambaut, 1995). Uncorrected *p*-distances were calculated in MEGA 5 (Tamura et al., 2011). Heterozygotes were evaluated using PHASE 2.1.1 (Stephens and Donnelly, 2003). When they were inferred with a probability of <0.9, we applied IUPAC ambiguity codes. PHASE was run for 100 iterations with a thinning interval of 1 after 100 burn-in iterations. Sequences were deposited in GenBank under accession numbers KT152285–KT152632.

2.2. MtDNA gene tree and nuclear networks

We inferred the mtDNA gene tree under the maximum likelihood (ML) criterion in RAxML-HPC 7.3.1 (Stamatakis, 2006) through the CIPRES Science Gateway (Miller et al., 2010). The mitochondrial genes were concatenated and partitioned by codon position and tRNAs (Wu et al., 2012). The GTR + Γ + I substitution model was selected for each partition based on the Akaike information criterion obtained from jModelTest 2 (Darriba et al., 2012). Nodal support was estimated using 1000 bootstrap pseudoreplicates. We also inferred the mtDNA gene tree and nodal support based on the 50% majority-rule consensus tree generated by MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) under the same partitioning strategy. We constructed statistical parsimony nuclear gene networks for *RAG-1* and *NCRT* using TCS 1.21 (Clement et al., 2000). Gaps were treated as the fifth state. We estimated the maximum number of differences among haplotypes as a result of single substitutions and 95% statistical confidence. Reticulations in the network were resolved as per Posada and Crandall (2001).

2.3. Statistical species delimitation

Two coalescent-based delimitation methods compared competing models. MtDNA and phased nDNA sequences were included in both analyses. First, Bayes factors were used to evaluate competing models, which were formulated based on the mtDNA gene tree, nDNA networks, and geographic distributions of species (Grummer et al., 2014). We employed models that nested within the current taxonomy of four species, which was used as the starting delimitation for hypothesis testing. For each model, individuals were assigned to species. The species-tree analyses used the *BEAST option in BEAST 1.7.4 (Drummond and Rambaut, 2007; Heled and Drummond, 2010). Marginal model likelihood used in Bayesian model comparisons (Bergsten et al., 2013) was calculated via both path-sampling (PS; Baele et al., 2012) and stepping-stone (SS; Xie et al., 2011) methods in BEAST. Substitution models for mtDNA and nDNA were determined using jModelTest 2. All analyses employed an uncorrelated lognormal relaxed molecular clock for each locus with the clock rate of mtDNA fixed to 1.0. The species-tree prior assumed a Yule process. The population-size model was assumed to be piecewise linear and to have a constant root. An inverse gamma prior ($\alpha = 3$, $\beta = 0.03$) was used to model the population sizes (θ) with a mean of 0.015. The Markov Chain Monte Carlo (MCMC) was performed for 50 million generations while sampling every 5000 generations. Chain convergence was