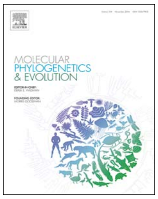
Molecular Phylogenetics and Evolution 120 (2018) 1–15

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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



Resolving taxonomic turbulence and uncovering cryptic diversity in the musk turtles (*Sternotherus*) using robust demographic modeling

Peter A. Scott[a,c,*], Travis C. Glenn[b], Leslie J. Rissler[a,d]

[a] *Department of Biological Sciences, University of Alabama, Tuscaloosa, AL 35487, USA*
[b] *Department of Environmental Health Science and Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA*
[c] *Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA*[1]
[d] *Directorate for Biological Sciences, National Science Foundation, 4201 Wilson Blvd. Arlington, VA 22230, USA*[1]

ARTICLE INFO

Keywords:
Sternotherus
Species delimitation
Demographic models
Cryptic species
Species tree conflict

ABSTRACT

Accurate and consistent delimitation of species and their relationships provides a necessary framework for comparative studies, understanding evolutionary relationships, and informing conservation management. Despite the ever-increasing availability of genomic data, evolutionary dynamics can still render some relationships exceedingly difficult to resolve, including underlying speciation events that are rapid, recent, or confounded by post-speciation introgression. Here we present an empirical study of musk turtles (*Sternotherus*), which illustrates approaches to resolve difficult nodes in the Tree of Life that robust species-tree methods fail to resolve. We sequence 4430 RAD-loci from 205 individuals. Independent coalescent-based analyses, corroborated with morphology and geography, strongly support the recognition of cryptic species within *Sternotherus*, but with conflicting or weak support for some intraspecific relationships. To resolve species-tree conflict, we use a likelihood-based approach to test support for alternative demographic models behind alternative speciation scenarios and argue that demographic model testing has an important role for resolving systematic relationships in recent, rapid radiations. Species-tree and demographic modeling strongly support the elevation of two nominal subspecies in *Sternotherus* to species and the recognition of a previously cryptic species (*S. intermedius* sp. nov.) described within. The evolutionary and taxonomic history of *Sternotherus* is discussed in the context of these new species and novel and well-supported systematic hypotheses.

1. Introduction

Accurate species delimitation and inference of species relationships are the two main goals of systematics as they provide the framework for comparative biology and understanding the evolution of life. Our ability to sequence 10–1000 s of loci (Edwards, 2009; McCormack et al., 2013) and model the evolutionary relationships of multiple non-model species using sophisticated multi-species coalescent models (Kingman, 1982; Maddison, 1997) has provided increased resolution to even the most difficult phylogenetic hypotheses (Rokas et al., 2003; Edwards, 2009). Computational methodologies (e.g. Liu and Pearl, 2007; Liu, 2008; Liu et al., 2010; Heled and Drummond, 2010; Bryant et al., 2012; Mirarab et al., 2014a) have greatly advanced our ability to account for the causes of incongruence between gene trees and species trees (e.g., incomplete lineage sorting [ILS], horizontal gene transfer, and paralogy; Maddison, 1997; Edwards, 2009). However, many methods focus primarily on modeling conflicting support for the true species tree due to ILS, and fail to account for other demographic sources of gene tree/species tree discordance (e.g. *BEAST [Heled and Drummond, 2010]; BEST [Liu and Pearl, 2007; Liu, 2008]; MP-EST [Liu et al., 2010]; ASTRAL [Mirarab et al., 2014a]).

Some of the most challenging nodes to resolve in the Tree of Life are the so-called "bushes" (Rokas and Carroll, 2006), which are thought to be the result of high amounts of ILS at nodes with rapid or near-simultaneous lineage diversification (Patel et al., 2013). There are numerous empirical examples (reviewed in Whitfield and Lockhart, 2007) where multiple, relatively ancient short branches have yielded poorly resolved species trees despite the use of large amounts of genomic data (i.e. in mammals: Song et al., 2012; Tarver et al., 2016; or Aves: Hackett et al., 2008; Jarvis et al., 2014; Prum et al., 2015). Additionally, one of the most common methods of multi-locus phylogenetic inference with genomic data (concatenation of multiple loci to a super matrix) can often lead to highly supported but incorrect inference of the species tree - a situation where the bulk of individual gene trees, and thus node

---

[*] Corresponding author at: Department of Ecology and Evolutionary Biology, 610 Charles E. Young Dr. East, Terasaki Life Science Building, #4129, Los Angeles, CA 90095-7246, USA.
E-mail address: pete.a.scott@gmail.com (P.A. Scott).

[1] Current address.

https://doi.org/10.1016/j.ympev.2017.11.008
Received 18 August 2017; Received in revised form 9 November 2017; Accepted 16 November 2017

support in the concatenated phylogeny, fall into the "anomaly zone" (Degnan and Rosenberg, 2006; Kubatko and Degnan, 2007), which may be especially prevalent when multiple phylogenetic splits are separated by short internodes.

Rapid diversification or recent speciation can still pose challenges for resolving some nodes when ILS is explicitly modeled. For example, in deep radiations there can be confounding genetic signal (caused by ILS and then compounded by a lack of time for informative mutations to accrue in short internal evolutionary branches) and homoplastic or novel mutations whose signals drive gene-tree conflict (Whitfield and Lockhart, 2007; Patel et al., 2013; Suh et al., 2015; Linkem et al., 2016). Furthermore, these issues may be subsequently magnified in more recent speciation events, where contemporary or ongoing hybridization and introgression may obfuscate underlying phylogenetic signal, compress branch lengths, and lead to poor estimation of demographic parameters (Eckert and Carstens, 2008; Nielsen and Wakeley, 2001; Leaché et al., 2014), and thus the species tree. For example, despite high morphological diversity, resolving relationships within the rapid radiation of Hawaiian mints has been problematic due to exceedingly low intraspecific genomic variation (Lindquist and Albert, 2002; Roy et al., 2015; Welch et al., 2016), even when sequenced on the "genomic" scale. Additionally, in the well-known and very recent radiations of Rift Valley (African) cichlids, complex patterns of vicariance and secondary contact, introgression, and rapidly changing demographic histories have made both species identification and resolving relationships in this radiation exceedingly difficult (Brawand et al., 2014; Malinsky et al., 2015, 2017; Nater et al., 2015).

Inference of species trees and species limits while accounting for ILS fall into two major categories: (1) those that simultaneously infer gene or SNP genealogies while inferring the underlying species tree (simultaneous methods, e.g.: *BEAST [Heled and Drummond, 2010]; SNAPP [Bryant et al., 2012]; SVDquartets [Chifman and Kubatko, 2014, 2015]), and (2) those that infer the species tree from individual gene trees assumed to be inferred without error (summary methods, e.g.: ASTRAL [Mirarab et al., 2014a]; NJst [Liu and Yu, 2011]; STEM [Kubatko et al., 2009]; ASTRID [Vachaspati and Warnow, 2015]). In both categories, the most favored methods are computationally stable (i.e. with enough data they are mathematically guaranteed to converge on the true underlying species tree), as long as model assumptions are met. However, data collected from empirical systems, especially those that have undergone rapid diversification, will never satisfy all assumptions of currently available species-tree methods. Specifically, the assumption that individual gene trees are inferred without error, as necessitated by summary methods, can often not be met when using the most common short-read genomic methods (e.g. RADseq [where each locus may have less informative sites than termini], Mirarab et al., 2014b; Harvey et al., 2016). Additionally, all computational methodologies ignore some biologically relevant demographic parameters (i.e. hybridization between species or population structure within species) and/or impose necessary approximations (i.e. constant N_e), which can influence topology, node support, and/or branch length (Eckert and Carstens, 2008; Carstens and Dewey, 2010; Leaché et al., 2014; Solís-Lemus et al., 2016). For example, in a simulation study, Leaché et al. (2014), found that even in situations where species-tree topologies were robust to gene flow, introgression could cause drastic overestimation of population size and underestimation of branch lengths. Lastly, all common species-tree methods (e.g. BEST, *BEST, STEM, ASTRAL) assume lineage bifurcation, even when the empirical data are poorly modeled by the multi-species coalescent (Huson and Bryant, 2006; Reid et al., 2013; Brown, 2014). Empirically, Cui et al. (2013) and Solís-Lemus and Ané (2016), found strong support for novel relationships among Mesoamerican *Xiphophorus* fishes after accounting for reticulate evolution when inferring species trees. Thus, despite major advancements in species-tree methods, our inability to account for all aspects of demographic history of each lineage, while inferring a species tree, may cause incorrect and weakly supported inferences, especially when dealing with recent or rapid diversifications or systems with expected gene flow.

Although we know that few evolutionary relationships are likely to conform to the simplifying assumptions underlying species-tree methods, and that failing to account for demography can further bias our inferences, we often have a limited understanding of the demographic processes underlying diversification for most species. To address this uncertainty, researchers have begun incorporating models that can explicitly estimate migration rates and shifting population sizes when inferring demographic parameters underlying the species tree (e.g. the isolation with migration [IM] model; Nielsen and Wakeley, 2001; Hey and Nielsen, 2004, 2007; Hey, 2010), or to verify species delimitation results (Hey and Pinho, 2012). However, these models (e.g. IMa2, Hey, 2010; migrate-n, Beerli and Palczewski, 2010) rely on *post hoc* inference of population demography and migration once species delimitation and tree inference have been completed, while assuming the true underlying topology has not been influenced by the gene flow researchers may be investigating. Such modeling is not limited solely to inferring histories, but can also be used to test alternative and complex patterns of evolutionary history by simulating genetic data under alternative demographic hypotheses, and applying a model-testing procedure to discriminate alternative species trees while better accounting for underlying demographic histories of species of interest (Steele and Storfer, 2006; Hickerson et al., 2010; Csilléry et al., 2010; Eaton and Ree, 2013; Excoffier et al., 2013). Incorporating explicit demographic model-testing into species delimitation and species-tree inference provides an important accompaniment to current systematic methods that are unable to account for complex demographic histories in recent radiations.

Turtles provide many opportunities to test methods for resolving the most complex evolutionary nodes in the Tree of Life. Many deep and shallow relationships within turtles have been difficult to resolve (Barley et al., 2010) due to exceedingly low rates of molecular evolution (Spinks et al., 2013; Shaffer et al., 2013), homoplasy of morphological characters (Le et al., 2013), long overlapping generations, rampant hybridization (Spinks et al., 2012), and ontogenetic and environmentally mediated (e.g. wearing down of diagnostic shell characteristics) variability in diagnostic morphological characters. The genus *Sternotherus*, currently comprising four species, is no exception to the pervasive pattern of difficult-to-resolve turtle relationships. The genus' distribution spans one of the world's greatest turtle (Mittermeier et al., 2015) and global biodiversity (North American Coastal Plain; Myers et al., 2000; Noss et al., 2015) hotspots and contains at least one species (*S. depressus*) that is critically endangered. Thus, a thorough understanding of evolutionary relationships within *Sternotherus* is critical for comparative biological studies and conservation management of these threatened taxa.

Sternotherus has a long history of taxonomic uncertainty with nearly every possible systematic relationship for species in this genus having been proposed (Fig. 1; Tinkle, 1958; Zug, 1966; Seidel et al., 1981; Walker et al., 1988; Iverson, 1991, 1998; Iverson et al., 2007, 2013; Spinks et al., 2014). One species, *S. minor*, is thought to contain two geographically cohesive and morphologically distinct subspecies, *S. m. minor* and *S. m. peltifer*, united by a large hybrid zone (Fig. 2; Tinkle, 1958; Iverson, 1977a; Seidel et al., 1981; Ernst et al., 1988). Additionally, *S. depressus* has been synonymized to a subspecies of *S. minor* or elevated to species status by different authorities (Tinkle, 1958; Mount, 1975; Iverson, 1977b; Walker et al., 1988) because it may hybridize with *S. m. peltifer* (Walker et al., 1988). Lastly, the subspecies *S. m. peltifer* was once considered a subspecies of *S. carinatus* (e.g. Carr, 1952). This taxonomic confusion has led to a poor understanding of the geographic boundaries of described taxa and putative hybrid zones, which limits effective conservation efforts (e.g. to recognize putative *S. depressus* × *S. m. peltifer* hybrids; Mount, 1975; Ernst et al., 1983, 1988).

Hypotheses for taxonomic relationships in *Sternotherus* have