Molecular Phylogenetics and Evolution 127 (2018) 356–366

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

The effect of missing data on coalescent species delimitation and a taxonomic revision of whipsnakes (Colubridae: *Masticophis*)

Kyle A. O'Connell[a,b,*], Eric N. Smith[a]

[a] Department of Biology, The University of Texas at Arlington, Arlington, TX 76019, USA
[b] Department of Vertebrate Zoology; National Museum of Natural History, Washington, DC 20007, USA

ARTICLE INFO

Keywords:
Coalescent species delimitation
Coluber
Integrative taxonomy
North America
Masticophis, Whipsnakes

ABSTRACT

A stable alpha taxonomy is essential to understanding evolutionary processes and achieving effective conservation aims. Taxonomy depends on the identification of independently evolving lineages, and the delimitation of these lineages based on multiple lines of evidence. Coalescent species delimitation within an integrative framework has increased the rigor of the delimitation process. Here we use genome-wide SNP data and coalescent species delimitation to explore lineage relationships within several North American whipsnake species, test the species status of several lineages, and test the effect of missing data on species delimitation. We find support for the elevation of several previously recognized subspecies to full species status, and formally elevate two species. This study demonstrates the power of molecular data and model-based delimitation methods to identify evolutionary relationships, and finds that missing data have little impact on the outcome of delimitation analyses.

1. Introduction

The field of species delimitation has received increased attention in recent years (Sites and Marshall, 2003). Since the foundational work of de Queiroz (2007), the definition of the general lineage species concept has decoupled species conceptualization from species delimitation. As such, various lines of evidence can be used to assess lineage independence, but the status of the species is not dependent on any one type of evidence (de Queiroz, 2007). In the pre-molecular era, species delimitation primarily depended on morphological data, although ecological, distributional, or other types of data were used support a species' status (Padial et al., 2010, and Sites and Marshal, 2004). In the case of allopatrically distributed species, reproductive isolation is demonstrable, but in species with overlapping ranges, researchers traditionally relied on morphological differences as a proxy for reproductive isolation (Fujita et al., 2012). However, morphological or ecological variation may not accurately represent the evolutionary history of a species (Ruane et al., 2014). The advent of molecular data revolutionized taxonomy and species delimitation, but a dependence on a small number of loci often misled inferences of phylogeny due to incomplete lineage sorting and hybridization (Knowles and Carstens, 2007, and Streicher et al., 2016). Fortunately, genomic data, and the subsequent increase in available loci, have helped to mitigate many of these shortcomings by estimating species trees more accurately, and by allowing for robust testing of species hypotheses (Liu et al., 2015, Leaché et al., 2014, and Faircloth et al., 2012).

Species delimitation methods attempt to accurately quantify independently evolving lineages (Knowles and Carstens, 2007, Petit and Excoffier, 2009, and Sites and Marshall, 2003). The species delimitation process is comprised of two steps: lineage identification and hypothesis testing (Carstens et al., 2013). Lineage identification relies on a variety of methods, including morphological or ecological variation, disjunct geographic distributions, or molecular phylogenies (Wiens, 2007). However, lineages identified by one or more of these methods may not reflect the accurate evolutionary history of lineages, creating the need to test hypotheses regarding species composition and relationships (Fontaneto et al., 2015). Several recent techniques leverage coalescent theory to test species delimitation hypotheses (Fujita et al., 2012, and Pante et al., 2015). Bayes Factor Delimitation (with genomic data; BFD*) is one method for testing hypotheses of species relationships that utilizes genome-wide SNP data (Leaché et al., 2014). This method is advantageous to other coalescent species delimitation methods because it does not require a guide tree, but rather directly estimates the species tree from biallelic markers, and can calculate a marginal likelihood estimate (MLE) for each species model (Leaché et al., 2014). Nonetheless, recent criticism of this method suggests that it may 'over-split' populations instead of species, and that integrative taxonomic approaches should be used to balance these shortcomings (Sukumaran

* Corresponding author at: Department of Vertebrate Zoology, National Museum of Natural History, Washington, DC 20007, USA.
E-mail address: kyleaoconnell22@gmail.com (K.A. O'Connell).

https://doi.org/10.1016/j.ympev.2018.03.018
Received 31 May 2017; Received in revised form 10 January 2018; Accepted 14 March 2018
Available online 20 March 2018
1055-7903/

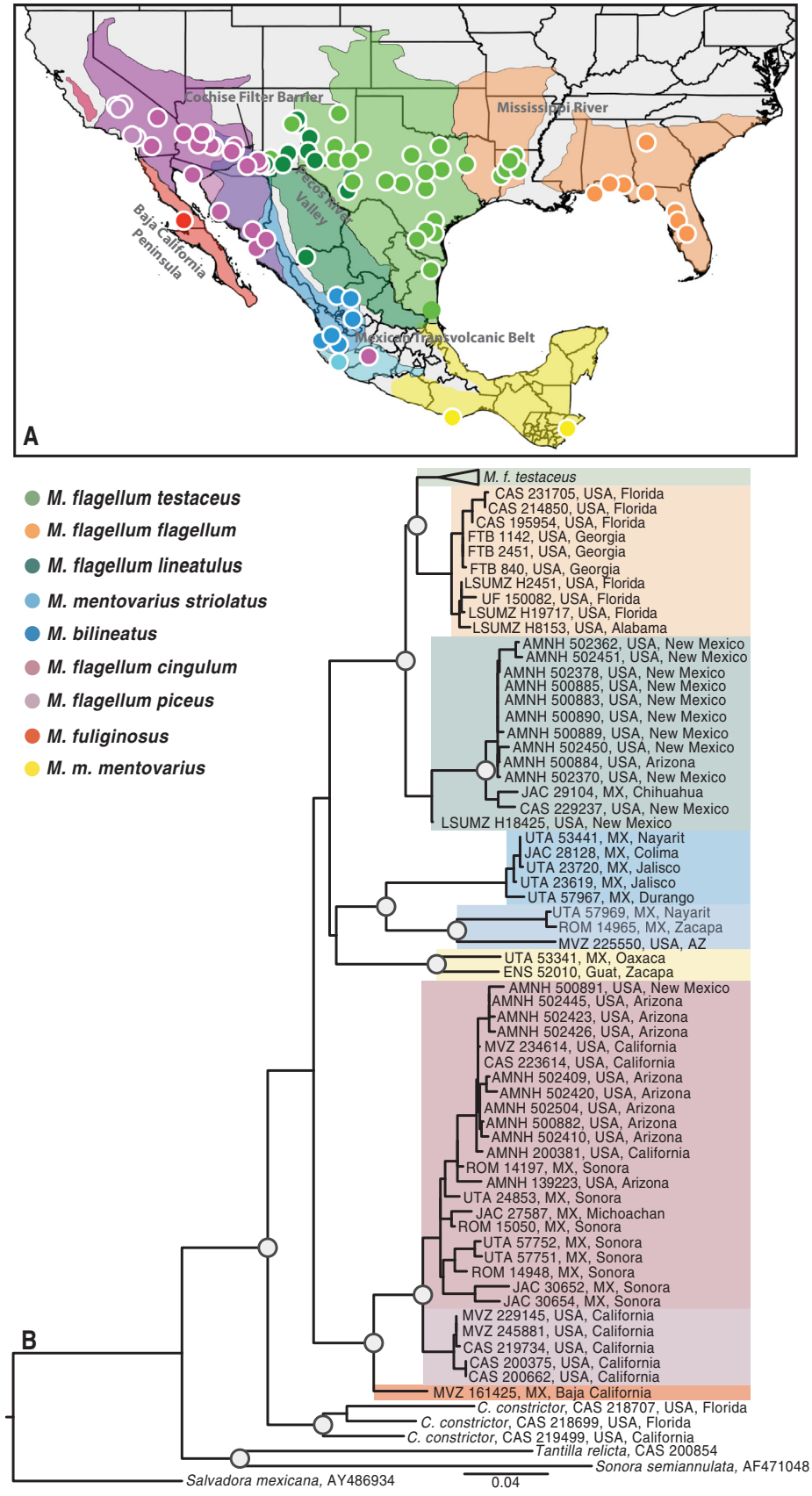

Fig. 1. (A) Map showing the approximate distributional ranges of each subspecies investigated in this study as described by Wilson (1970) and Stebbins (2003). Circles represent sampling localities for mitochondrial data. (B) Maximum likelihood phylogeny including several species of whipsnakes. Grey circles show nodes with at least 70% bootstrap support. Colors on each clade correspond to the colors used in the range map. We collapsed the clade pertaining to *M. flagellum testaceus* to save space.

and Knowles, 2017).

Missing data in restriction-site-associated datasets is a common problem due to both allelic dropout and sequencing stochasticity between samples (Arnold et al., 2013; Eaton et al., 2017). Missing data was previously thought to negatively bias estimates of population genetic parameters and phylogenetic estimation (Huang and Knowles, 2014, Hodel et al., 2017). Yet, recent work suggests that the inclusion of more loci may outweigh biases associated with missing data (Huang and Knowles, 2014). Recent studies have recovered more resolved phylogenies and more accurate population genetic parameters due to the inclusion of more loci (Eaton et al., 2017,Hodel et al, 2017, Tripp et al., 2017). Nonetheless, the effect of missing data on model-based coalescent species delimitation remains uninvestigated.

Here, we utilize a species delimitation framework with molecular data to explore species relationships within North American whipsnakes, and to test the species status of several lineages. Whipsnakes are a widespread clade of slender colubrid snakes distributed across North America into Colombia and Venezuela (Johnson, 1977, and Wilson, 1970, Fig. 1A). Despite recent taxonomic uncertainty about the monophyly of *Masticophis* in relation to *Coluber constrictor* (Nagy et al. 2004; Crother, 2012), we retain the use of *Masticophis* following Myers et al. (2017, Burbrink and Myers, 2015). This study focuses on the systematics of three species of whipsnakes: *M. flagellum, M. bilineatus* and *M. mentovarius*. *Masticophis flagellum* (Shaw, 1802) is a large bodied snake distributed across North America, and across several diverse ecoregions (Roze 1953, Johnson, 1977, Conant & Collins, 1998; and Uetz & Hošek, 2017). Color pattern and scale count variation is dramatic in this species group across geographic space (Wilson, 1970), leading to the recognition of six subspecies for *M. flagellum* (Fig. 1A): (1) *M. f. flagellum* (Type locality, Carolina and Virginia, USA), (2) *M. f. testaceus* (Type locality, Pueblo County, Colorado, USA), (3) *M. f. lineatulus* (Type locality, Chihuahua, MX), (4) *M. f. cingulum*, (Type locality, Moctezuma, Sonora, MX), (5) *M. f. piceus* (Type locality, Graham County, Arizona, USA), and (6) *M. f. ruddocki* (Type locality, Kern County, California, USA). *Masticophis fuliginosus* was classified as a seventh subspecies of *M. flagellum* from Baja California, Mexico, until Grismer (1994) elevated it to evolutionary species status. In this study, we refer to *M. fuliginosus* as part of the *M. flagellum* group, but retain its classification as a separate species.

The second species we investigate, *M. mentovarius,* has experienced a turbulent taxonomic history. This species currently encompasses five recognized subspecies, including *M. m. mentovarius* (Type locality, Chisec, GT), *M. m. centralis* (Type locality, Guajira, CO), *M. m. suborbitalis* (Type locality, Caracas, VE), *M. m. striolatus* (Type locality, Colima, MX), and *M. m. variolosus* (Type locality, Maria Magdalena Island, MX, Uetz and Hošek, 2017).

Finally, *Masticophis bilineatus* inhabits a much smaller geographic range than both *M. flagellum* and *M. mentovarius;* it is primarily restricted to the Sonoran Desert. Originally, *M. bilineatus* was divided into two subspecies. The first was *M. b. slevini* from the Isla San Esteban in the Gulf of California, Sonora, MX (Lowe and Norris, 1955), but was elevated to species status by Grismer (1999). The second subspecies was *M. b. lineolatus* (Hensley, 1950), which has since been synonymized with *M. bilineatus*. The ranges of *M. bilineatus* and several *M. flagellum* lineages overlap in and around the Cochise Filter Barrier (CFB: the boundary between the Sonoran and Chihuahuan Deserts), and Stebbins (2003), expressed uncertainty of subspecies limits of *M. flagellum* in this region, where the ranges of *M. f. testaceus*, *M. f. piceus*, *M. f. cingulum*, and *M. bilineatus* all overlap (Fig. 1A).

Here we explore the number of lineages within *M. flagellum*, *M. mentovarius*, *and M. bilineatus* using mitochondrial and genomic data, and conduct species delimitation using BFD* under different missing data regimes to test the effect of missing data on species delimitation. We also collect morphological data for each lineage from the literature. We use these data to address three questions: (1) Do subspecies of *M. flagellum* represent independently evolving lineages? (2) Where should *M. m. striolatus* be placed in the whipsnake phylogeny? (3) What is the effect of missing data on coalescent species delimitation?

2. Materials and methods

2.1. Mitochondrial sequencing and phylogenetic analyses

We utilized 65 mitochondrial sequences of the cytochrome B oxidase gene from O'Connell et al. (2017), available on GenBank (KT713652-KT713738), as well as 46 additional cytochrome B sequences downloaded from Genbank. Sequences used in this study included *M. f. flagellum* (n = 10), *M. f. testaceus* (n = 44), *M. f. lineatulus* (n = 13), *M. f. cingulum* (n = 22), *M. f. piceus* (n = 5), *M. fuliginosus* (n = 1), *M. bilineatus* (n = 3), *M. m. striolatus* (n = 5), and *M. m. mentovarius* (n = 2), *Coluber constrictor* (n = 3), *Salvadora mexicana* (n = 1), *Tantilla relicta* (n = 1), and *Sonora semiannulata* (n = 1). All sequences were aligned with the Geneious Aligner under default settings (Kearse et al., 2012). Uncorrected average pairwise distances were calculated between lineages in Mega v7 (Kumar et al., 2016). We selected the most probable model of nucleotide evolution for Likelihood analyses using Bayesian information criteria implemented in PartitionFinder (Lanfear et al., 2012), partitioning by codon position. We estimated a maximum likelihood phylogeny using raxmlGUI v1.3 with 1000 rapid bootstrap iterations (Silvestro and Michalak, 2012) and visualized the final phylogeny in FigTree v1.4.3 (Rambaut, 2017). Nodes with bootstrap values $\geq$70 were considered strongly supported.

2.2. Genomic sequence generation and computational analysis

We utilized double-digest restriction associated DNA sequencing (ddRADseq) data for 30 individuals from O'Connell et al. (2017) to evaluate relationships between mitochondrial lineages using nuclear data. Subspecific classifications in this study were based on location of field collection. Our sampling included *M. f. flagellum* (n = 6), *M. f. testaceus* (n = 6), *M. f. lineatulus* (n = 4), M. *fuliginosus* (n = 1), *M. bilineatus* (n = 1), *M. f. cingulum* (n = 3), *M. m. striolatus* (n = 3), and *M. m. mentovarius* (n = 6).

We processed ddRADseq data using the STACKS v1.12 pipeline (Catchen et al., 2013). We followed the recommended workflow which implemented the following scripts and programs: (i) process_radtags, which filtered out reads below 90% quality score threshold, (ii) ustacks, which set a maximum distance of 4 between 'stacks', (iii) cstacks, which creates a catalogue of all loci within all individuals (-n flag; setting of 0) (iv) sstacks, which searches the stacks created in ustacks against the catalogue from cstacks, and (v) populations, which genotypes each individual according to the matched loci from sstacks. After running populations, we used custom python scripts (available at https://github.com/dportik/Stacks_pipeline) to filter out invariant loci, and loci with more than two haplotypes. We produced several SNP datasets that differed in the species included as well as the percent missing data. Dataset A included one to three individuals from all species in our study. We limited the number of individuals to maximize taxonomic diversity while minimizing allelic dropout. Thus, dataset A included 14 individuals and 365 loci, including three *M. f. cingulum*, *M. m. striolatus*, *M. m. mentovarius*, and one *M. fuliginosus*, *M. f. flagellum*, *M. f. testaceus*, *M. f. lineatulus*, and *M. bilineatus*. We allowed up to 30% missing data per locus. Next we created species-specific datasets that used two different missing data thresholds to test species limits within *M. flagellum* (datasets B-C), and within *M. mentovarius* and *M. bilineatus* (datasets D-E). Dataset B included four *M. f. lineatulus*, six *M. f. flagellum*, six *M. f. testaceus*, three *M. f. cingulum*, one *M. fuliginosus*, and six *M. m. mentovarius*. We allowed up to 50% missing loci, resulting in 2079 loci. We also created dataset C with the same individuals but only allowed up to 20% missing loci resulting in 325 loci. Dataset D included three *M. m. mentovarius*, three *M. m. striolatus*, and one *M. f. flagellum*, *M. f. lineatulus*, *M. f. testaceus*, and *M. f. bilineatus*. We allowed up to 50% missing

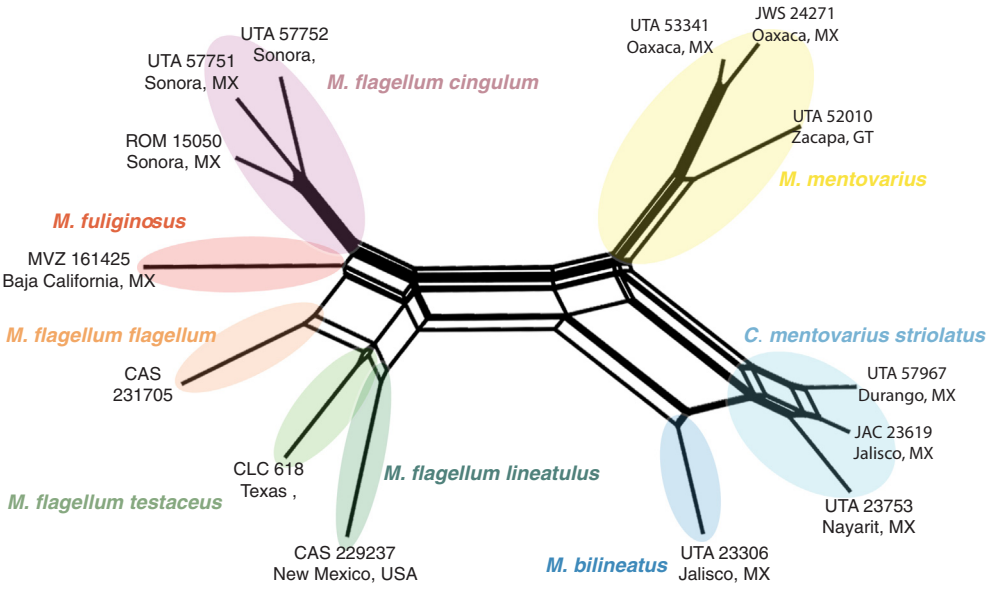

Fig. 2. Neighbor network generated using SPLITSTREE from 356 SNPs and 15 individuals representing each lineage for which we had nuclear sampling. The colors correspond to the range map in Fig. 1A.

loci, resulting in 1464 loci. Finally, dataset E included the same 10 individuals, but allowed up to 20% missing loci resulting in 216 loci.

2.3. Investigating species relationships within Masticophis using phylogenetic networks

We investigated phylogenetic relationships between all study species using dataset A in SPLITSTREE v4.13.1 (Huson and Bryant, 2006; Fig. 2). SPLITSTREE uses a distance-based method to estimate an unrooted phylogenetic network, rather than estimating a strict phylogeny. We used the neighbor-net algorithm under default settings, and visualized the network using EqualAngle distances.

2.4. Coalescent species delimitation and species tree estimation

Our mitochondrial and nuclear phylogenetic analyses agreed with O'Connell et al. (2017) and Myers et al. (2017) to suggest that several historically recognized subspecies represented evolutionary lineages. To test if these lineages should be elevated to species status, we conducted Bayes Factor Delimitation with genomic data following Leaché et al. (2014; BFD*) and Grummer et al. (2013; BFD). Bayes Factor delimitation utilizes the SNAPP (Bryant et al., 2012) plugin of the BEAST2 platform (Kühnert et al., 2014) to calculate a MLE for alternative models using path sampling. One advantage of this method for SNP data is that it can accommodate missing data between individuals (among species), although the impact of these missing data on the analysis is untested. The method also allows for varying numbers of individuals per species to accommodate unequal sampling between lineages. Using a Bayes Factor (BF), we compared and ranked models to select the best-supported species hypothesis. We calculated the BF by subtracting the absolute value of the MLE of the model representing the current taxonomic classification of each dataset from each alternative model. Following Kass and Raftery (1995) we considered a BF over 10 to provide strong support for a model. We subsequently ranked all models and chose the model with the highest BF (Table 1). In addition to testing lineage limits, we wanted to test the effect of missing data on species delimitation. Thus, we conducted four sets of analyses, two for each species group using the ≤50% and ≤20% missing loci datasets (Table 1; Fig. 3). We assigned individuals to lineages based on the genomic clustering analyses of O'Connell et al. (2017) and the mitochondrial analysis from this study. Our first two analyses tested species limits within *M. flagellum* using datasets B and C (Fig. 3). We tested the following models: (1) current taxonomy, all *M. flagellum* were lumped, and split from *M. fuliginosus* and *M. m. mentovarius*, (2) lumped all *M. flagellum* with *M. fuliginosus*, split *M. mentovarius*, (3) split *M. f. cingulum*, *M. fuliginosus* and *M. m. mentovarius,* lumped *M. f. flagellum*, *M. f. testaceus*, and *M. f. lineatulus*, (4) lumped *M. f. cingulum* with *M. fuliginosus*, lumped all other *M. flagellum*, split *M. m. mentovarius*, (5) lumped *M. f. testaceus* with *M. f. lineatulus*, split all other lineages, (6) lumped *M. f. flagellum* with *M. f. testaceus*, split all other lineages, (7) split all lineages, (8) split all lineages, but mixed *M. f. flagellum*, *M. f. testaceus*, *M. f. lineatulus*, and *M. f. cingulum* randomly. Our second set of analyses utilized datasets D and E and tested the following models (Fig. 3): (1) current taxonomy, where we split *M. flagellum*, *M. m. mentovarius*, *M. m. striolatus*, and *M. bilineatus*, (2) lumped *M. m. striolatus* and *M. bilineatus*, (3) lumped *M. m. striolatus* with *M. m. mentovarius*, (4) lumped *M. m. striolatus* with *M. flagellum*, (5) split all lineages but mixed them randomly. We allowed BEAUti to estimate the mutation rate, and confirmed that both U and V were approximately equal to one. We assigned a Gamma distribution to our Lambda prior, with an Alpha of 1 and a Beta of 77. On our Snap prior we assigned an Alpha of 1, a Beta of 100, and a Lambda of 77. We performed 48 path sampling steps, with 100,000 MCMC generations, and 10,000 burnin generations. We calculated the BF by subtracting the absolute value of the MLE of the all models from the current taxonomic classification (model 1).

We estimated the species tree for each dataset using SNAPP v1.0. We assigned species identities based on the best supported model from our BFD* analyses. We utilized the same parameters as above, but we ran the analyses for 10 000 000 MCMC generations, sampling every 1000 generations. We visualized the complete tree sets in DENSITREE v1.0 (Bouckaert, 2010), and removed the first 10% of trees as burn-in.

2.5. Morphological data collection

We collected ventral and subcaudal counts from the literature and from three museum specimens. We recorded count data for the *M. flagellum* group (n = 1452) from Wilson (1970), for *M. m. striolatus* (n = 91), *M. m. variolosus* (n = 39), and *M. m. mentovarius* (n = 92) from Zweifel (1960) and Johnson (1977), and for *M. bilineatus* (n = 4) from Hensley (1950). We summarized counts for each recognized subspecies based on current distributions, except in *M. f. flagellum*, where we classified all *M. flagellum* west of the Mississippi River as *M. f.*

Table 1
Bayes Factor delimitation results are shown for each analysis. The number of species represents the number of species included in each analysis after lumping or splitting lineages. The number of loci represents the number of loci shared between all species in each analysis.

Model	Species	Loci	MLE	BF	Rank
Analysis 1, *Coluber flagellum cingulum* & *C. fuliginosus*, dataset B, 50% missing data					
1. Current taxonomy	3	832	7992.089	–	5
2. lump *C. flagellum* with *C. fuliginosus*	2	1409	13807.758	−5815.669	8
3. *C. f. cingulum* independent	4	770	13807.7376	−5815.6486	7
4. lump *C. f. cingulum* and *C. fuliginosus*	3	1239	10577.2167	−2585.1277	6
5. *C. f. flagellum* independent	5	737	5446.333	2545.756	2
6. *C. f. lineatulus* independent	5	734	5744.23	2247.859	3
7. split all subspecies	6	698	4957.0188	3035.0702	1
8. split all, mix all *C. flagellum*	6	810	7768.476	223.613	4
Analysis 2, *Coluber flagellum cingulum* & *C. fuliginosus*, dataset C, 20% missing data					
1. Current taxonomy	3	268	2648.984	–	6
2. lump *C. flagellum* with *C. fuliginosus*	2	321	3302.128	−653.144	8
3. *C. f. cingulum* independent	4	264	2298.313	350.671	4
4. lump *C. f. cingulum* and *C. fuliginosus*	3	312	2783.404	−134.42	7
5. *C. f. flagellum* independent	5	264	2081.4077	567.5763	2
6. *C. f. lineatulus* independent	5	264	2186.11	462.874	3
7. split all subspecies	6	264	2003.415	645.569	1
8. split all, mix all *C. flagellum*	6	268	2640.044	8.94	5
Analysis 3, *Coluber mentovarius striolatus,* dataset D, 50% missing data					
1. Current taxonomy, all split	4	366	1883.26	–	1
2. Lump *C. m striolatus* with *C. bilineatus*	3	912	4468.79	−2585.53	4
3. Lump *C. m. striolatus* with *C. m. mentovarius*	3	456	2649.06	−765.8	2
4. Lump *C. m. striolatus* with *C. flagellum*	3	548	3999.75	−2116.49	3
5. Split all lineages but mix randomly	4	1151	7994.76	−6111.5	5
Analysis 4, *Coluber mentovarius striolatus*, dataset E, 80% missing data					
1. Current taxonomy, all split	4	159	879.786	–	1
2. Lump *C. m striolatus* with *C. bilineatus*	3	216	1192.552	−312.766	4
3. Lump *C. m. striolatus* with *C. m. mentovarius*	3	159	1049.436	−169.65	2
4. Lump *C. m. striolatus* with *C. flagellum*	3	159	1090.767	−210.981	3
5. Split all lineages but mix randomly	4	159	1713.235	−833.449	5

testaceus based on O'Connell et al. (2017). We also counted ventral and subcaudal scales for three individuals that we sequenced, including one male and one female *M. m. striolatus*, and one male *M. bilineatus*. The *M. bilineatus* was from central Mexico, on the southern end of the range from the *M. bilineatus* measured by Hensley (1950).

3. Results

3.1. Phylogenetic analyses revealed mito-nuclear discordance and suggested unrecognized species diversity

We used mitochondrial data to explore the number of lineages within whipsnakes (Fig. 1). Our mitochondrial analyses suggested that *M. flagellum* is composed of an eastern and western radiation, but that this species may be paraphyletic with respect to *M. bilineatus* and *M. mentovarius*. We found that *M. f. testaceus*, *M. f. flagellum*, and *M. f. lineatulus* (eastern *M. flagellum*) formed a monophyletic group with *M. bilineatus*, *M. m. striolatus*, and *M. m. mentovarius*. Within eastern *M. flagellum*, we found that clades did not strictly adhere to traditional subspecies range boundaries (Fig. 1A; S1). Specifically, *M. f. flagellum* was traditionally thought to extend west of the Mississippi River into the east Texas pine forests, but we recovered a clear distinction between *M. flagellum* to the east and west of the Mississippi River. Likewise, we recovered two clades composed of samples pertaining to *M. f. lineatulus*. Both clades are restricted to the Chihuahua Desert. While the division between *M. f. lineatulus* and *M. f. testaceus* is clearly defined by the Pecos River Valley, where the Great Plains transition into the Chihuahua Desert, we sampled several individuals west of the Pecos River Valley with the *M. f. testaceus* haplotype. We recovered strong support (bootstrap ≥ 70) for relationships between eastern *M. flagellum*, namely, a sister relationship between *M. f. testaceus* and *M. f. flagellum*, and the inclusion of *M. f. lineatulus* to form a monophyletic group (Fig. 1B). We recovered *M. m. mentovarius*, *M. bilineatus*, and *M. m. striolatus* as sister to the eastern *M. flagellum*. However, this node did not receive high support, nor did the node between *M. m. mentovarius* and *M. bilineatus* and *M. m. striolatus*. We did recover strong support for the sister relationship between *M. m. striolatus* and *M. bilineatus*, and for the node between eastern and western *M. m. mentovarius.*

We recovered three clades pertaining to western *M. flagellum* representing *M. f. cingulum*, *M. f. piceus*, and *M. fuliginosus.* We recovered one monophyletic clade comprised of all individuals pertaining to *M. f. cingulum* extending as far south as Michoacán, Mexico. This *M. f. cingulum* clade also included individuals from Arizona, New Mexico, and California that were traditionally classified as *M. f. piceus*, with no notable divergence between individuals from each subspecies (Fig. 1). The *M. f. cingulum* clade was sister to samples from the San Joaquin Valley in Californian currently recognized as *M. f. piceus*. These two clades were sister to *M. fuliginosus* from Baja California, MX.

Interspecific genetic divergences ranged from 3.0% between *M. f. flagellum* and *M. f. testaceus,* to 12.8% between *M. m. mentovarius* and *M. f. piceus* (Table 2). We made several inferences from the distance matrix in Table 2. First, the eastern *M. flagellum* are closely related. Second, the western *M. flagellum* are more closely related to each other than to any other taxa in the matrix. Third, *M. bilineatus* and *M. m. striolatus* are distantly related to each other, and to all other whipsnakes.

Our distance-based neighbor network analysis using genomic SNP data (dataset A) recovered similar relationships within species as the mtDNA results, but very different relationships between species (Fig. 2). We recovered two genetic groups, one group including all individuals pertaining to *M. flagellum* and the other pertaining to *M. bilineatus* and *M. mentovarius*. Within *M. flagellum*, *M. fuliginosus* clustered closely to *M. f. cingulum* (*M. f. piceus* was not included). These two species were related to a cluster that included *M. f. flagellum*, *M. f. testaceus*, and *M. f. lineatulus*. *Masticophis m. striolatus*, and *M. bilineatus* were more closely related to each other than to any other species, and clustered more closely to *M. m. mentovarius* than to *M. flagellum*.

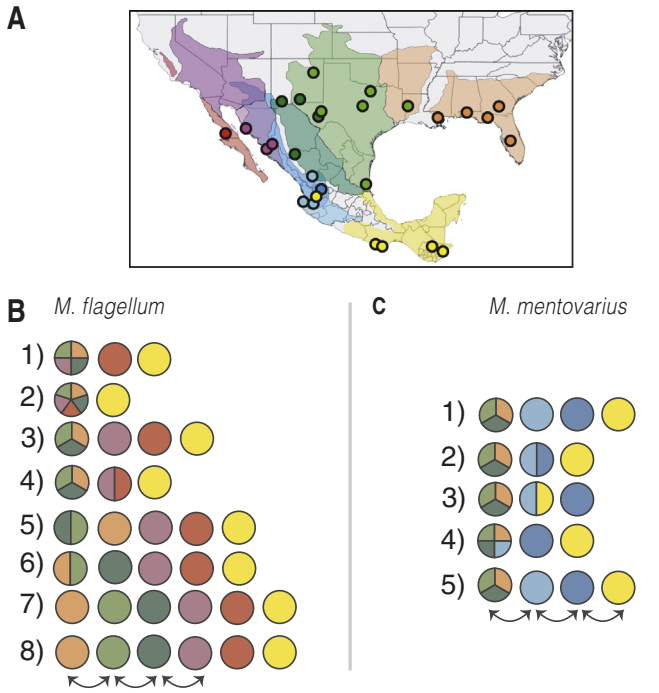

Fig. 3. Cartoon representing the different species relationship models we tested using Bayes Factor Delimitation. Pie charts represent models where lineages were lumped in the model. (A) Map showing the distribution of whipsnake lineages. Circles represent samples with nuclear data used in the species delimitation analysis. (B) Delimitation models for the *M. flagellum* species group. Orange = *M. f. flagellum*, light green = *M. f. testaceus*, dark green = *M. f. lineatulus*, purple = *M. f. cingulum*, red = *M. fuliginosus*, yellow = *M. m. mentovarius*. (C) Delimitation models testing *M. m. striolatus*. The three *M. flagellum* lineages were lumped in each model. Light blue = *M. m. striolatus*, dark blue = *M. bilineatus*, yellow = *M. m. mentovarius*.

Table 2
Mean between-group divergences generated from uncorrected *p* distances among Cytochrome b haplogroups using Mega 7.

	1	2	3	4	5	6	7	8	9
1. *M. f. flagellum*									
2. *M. f. testaceus*	3								
3. *M. f. lineatulus*	5.5	6							
4. *M. f. cingulum*	9.6	10.3	9.3						
5. *M. f. piceus*	9.6	9.8	9.8	4.6					
6. *M. fuliginosus*	9	9.9	9.1	6.6	6.5				
7. *M. m. striolatus*	9.3	9.8	9.2	12	11.3	10.8			
8. *M. m. mentovarius*	10.1	9.8	10.1	12.1	12.8	10.3	12		
9. *M. bilineatus* AZ	8.4	9.4	8.1	12	11.6	11.2	9.4	11.5	
10. *M. bilineatus* MX	9.7	8.9	8.4	11.4	11.8	11.5	10.3	10.6	6.4

3.2. Species delimitation supports the elevation of several recognized subspecies to species status

We tested eight species delimitation models to identify the number of lineages in *M. flagellum* (Table 1; Fig. 3). With both missing data thresholds, we recovered a general pattern where more split models were better supported than those that lumped discrete lineages. With dataset B (≤50%), our best-supported model involved splitting all possible lineages (model 7, BF = 3035.07), followed by classifying *M. f. flagellum* as an independent lineage (model 5, BF = 2545.76), and then by classifying *M. f. lineatulus* as an independent lineage (model 6, BF = 2247.86). Interestingly, the fourth best-supported model involved splitting all lineages, but mixing all *M. flagellum* subspecies randomly. With dataset C (≤20%), our best-supported model split all lineages (model 7, BF = 645.57), while the second and third best classified *M. f. flagellum* (model 5, BF = 576.58) and *M. f. lineatulus* (model 6, BF = 462.87) as independent species. Between both analyses, the ranking of models that were lumped and mixed varied, but the top three models remained consistent. The dataset with higher levels of missing data, but more loci (dataset B), recovered lower ML estimates, and subsequently higher BF values than dataset C, which had less missing data, but fewer loci.

We tested five species delimitation models to test the placement of *M. m. striolatus* using datasets D (≤50%) and E (≤20%). Results for dataset E are shown in brackets (Table 1). The best supported model involved splitting all lineages (model 1). The other four models were ranked as follows: lumping *M. m. striolatus* with *M. m. mentovarius* (model 3, BF = −765.8 [−169.65]), lumping *M. m. striolatus* with *M. flagellum* (model 4, −2116.49 [−210.98]), lumping *M. m. striolatus* with *M. bilineatus* (model 2, BF = −2585.53 [−312.77]), and mixing all lineages randomly (model 5, BF = −611.5 [−813.45]). In summary, we found that BFD* supported the elevation of several lineages currently recognized as subspecies to species status. However, we note that this method may have a bias towards splitting lineages rather than lumping them (Sukumaran & Knowles, 2017).

We estimated the species trees for the two best-supported models in each species group using datasets B-E. We found that nodes received higher support with datasets B and D, where < 50% of loci where missing, but where more loci were available (Fig. 4; S2). Our analyses with datasets C and E had less missing data, but received lower nodal

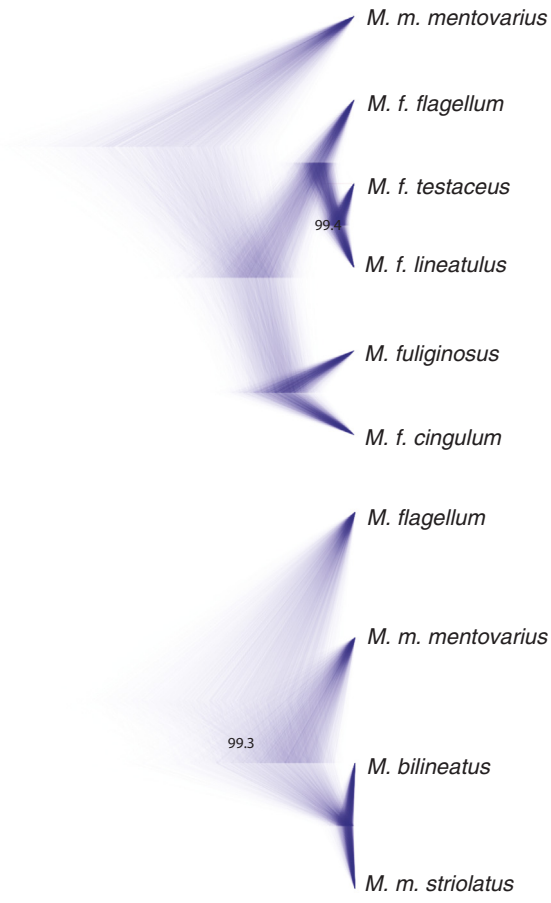

Fig. 4. Species trees generated using SNAPP based on the best-supported models from our Bayes Factor delimitation analysis shown in Fig. 3 for datasets B and D (≤50% missing loci). Support values are labeled for each node that is not fully supported.

support (Table 1). Our species trees of the *M. flagellum* group revealed several differences with the mitochondrial data. First, we found that the nuclear data supported the monophyly of *M. flagellum* relative to *M. mentovarius*. We found two sister clades within *M. flagellum*. In one clade we recovered a sister relationship between *M. f. cingulum* and *M. fuliginosus*. In the other clade we recovered a strongly supported sister relationship between *M. f. testaceus* and *M. f. lineatulus*, which were sister to *M. f. flagellum*. This also differs from the mtDNA analysis, which recovered a sister relationship between *M. f. flagellum* and *M. f. testaceus*. All nodes had greater than 99% support. We recovered the same topology between datasets B and C, but dataset C only received 91% support for the sister relationship between *M. f. testaceus* and *M. f. lineatulus*, compared with over 99% support.

Our species tree analyses for *M. mentovarius* and *M. bilineatus* recovered the same topology for dataset D and E, but with large differences in support values at deeper nodes (Fig. 4; S2). In dataset D, we recovered a well resolved phylogeny with *M. m. striolatus* closely related to *M. bilineatus*, and these two lineages were sister to *M. m. mentovarius* with 99.27% support. In dataset E, we recovered the same topology, but the sister relationship between *M. m. mentovarius* and *M. m. striolatus/M. bilineatus* was only supported by 84.50% support.

3.3. Morphological variation corresponds to discrete lineages

We collected ventral and subcaudal scale counts for each lineage investigated in this study (Table 3). We found little variation in the mean values of subcaudal counts between species, indicating that this character does not effectively differentiate whipsnake lineages. Ventral scale counts have historically been the primary character used to define subspecies in *Masticophis* (Wilson, 1970). While we observed variation in ventral scale counts between lineages, we also found substantial overlap in the ranges of counts between geographically adjacent lineages. We found that *M. bilineatus* had the highest mean number of ventral scales for males with a count of 203.8 (198–205.25; data lacking for females). *Masticophis f. flagellum* also had a high ventral scale count, with a mean in males of 202.7 (201–203.7), and in females of 200.5 (196–203). This contrasts with the lowest number of ventral scales in *M. m. striolatus*, which had a mean count in males of 187 (176–195), and in females of 186.5 (166–202). It should also be noted that while *M. f. flagellum* had a high ventral count, *M. f. testaceus* had a much lower count, with a difference between the means of 10.3 scales in males and 8.2 scales in females. This corresponded to the discontinuity observed at the Mississippi River between these two lineages.

4. Discussion

We use genetic and genomic data to explore lineage diversity within whipsnakes, conduct species delimitation with genomic data to test species delimitation models for several lineages, and test the effect of missing data on coalescent species delimitation. We found that species diversity within whipsnakes is currently underdescribed. Namely, we found that *M. flagellum* is composed of eastern and western clades divided by the Cochise Filter Barrier. Within the eastern clade, we found support for three lineages corresponding to *M. f. flagellum*, *M. f. testaceus*, and *M. f. lineatulus*. In the western group we found support for two lineages, corresponding to *M. f. cingulum*, and *M. fuliginosus*, although *M. f. piceus* was not sampled in the nuclear data. Within *M. mentovarius* we found that *M. m. striolatus* is most closely related to *M. bilineatus*, rather than *M. m. mentovarius*. Our species delimitation analyses supported the elevation of *M. f. cingulum* to evolutionary species, which we elevate to *M. piceus* (Appendix A). We also found support for the elevation of *M. m. striolatus* to full species status, which we elevate to *M. lineatus* (Appendix A). Finally, we found that datasets with more loci, despite higher levels of missing data, provided stronger support for species delimitation models.

4.1. Species delimitation in the face of conflicting data

4.1.1. Morphological and molecular discordance

Whipsnake taxonomy has traditionally relied on morphological data, namely, dorsal color and pattern, supralabrial scale counts, and ventral scale counts (Wilson, 1970, Johnson, 1977). Generally, the combination of these characters accurately delimited lineages, but for a few exceptions. Within *M. flagellum*, the boundary of the lineages classified as *M. f. flagellum* and *M. f. testaceus* was misdiagnosed based on color pattern, despite a marked discontinuity in ventral scale counts at the Mississippi River (Table 3, Wilson, 1970). Additionally, the boundary between the western lineages of *M. flagellum*, specifically, *M. f. cingulum* and *M. f. piceus*, was traditionally defined near the border of the United States and Mexico, based on color pattern variation (Wilson, 1970). However, this study, and that of Myers et al. (2017), found that these two subspecies only encompass one genetic lineage. Finally, morphological data misled past researchers regarding the placement of *M. m. striolatus*. Using several morphological characters, including ventral and supralabial scale counts, color pattern, and head to body width-ratios, this lineage was classified as an independent species, as a subspecies of *M. flagellum*, and finally as a subspecies of *M. mentovarius* (Smith, 1941, 1943, Bogert and Oliver, 1945, Zweifel and Norris, 1955, Johnson, 1977). However, molecular data places this lineage as most closely related to *M. bilineatus*, a result unforeseen by purely

Table 3
Morphological data is summarized for each species. The mean for each taxon is shown for males and females for ventral and subcaudal scale counts as collected from the literature and our own specimen counts. Species are sorted by male ventral count in descending order. In parentheses is the range, followed by the sample size. We collected data for *Masticophis flagellum* from Wilson (1970), for *M. bilineatus* from Hensley (1950), for *M. m. striolatus* and *M. m. variolosus* from Zweifel (1960) and Johnson (1977), and for *M. m. mentovarius* from Johnson (1977).

	Ventral		Subcaudal	
	Male	Female	Male	Female
M. bilineatus	205.25 (203.00–204.00; 4)	–	–	–
M. f. flagellum	202.70 (201.00–203.70; 114)	200.50 (196.00–203.00; 117)	112.81 (108.00–116.00; 41)	109.26 (106.60–113.60; 50)
M. f. cingulum	195.30 (193.80–197.20; 174)	195.10 (185.00–205.00; 45)	108.10 (101.20–112.20; 91)	104.50 (99.80–106.50; 40)
M. m. variolosus	194.85 (190.00–204.00; 33)	194.50 (190.00–197.00; 6)	125.40 (119.00–132.00; 6)	115.70 (113.00–120.00; 6)
M. f. ruddocki	193.40 (193.40–193.40; 71)	194.00 (192.80–196.70; 100)	107.30 (107.30–107.30; 6)	108.00 (104.20–115.00; 50)
M. f. lineatulus	193.30 (191.10–197.00; 62)	193.30 (193.30–193.30; 7)	105.90 (104.30–108.10; 41)	102.00 (98.00–104.50; 31)
M. fuliginosus	193.30 (186.00–199.20; 82)	192.90 (187.30–198.00; 58)	117.86 (109.50–123.00; 32)	114.70 (108.80–119.20; 33)
M. f. testaceus	192.40 (188.00–196.00; 470)	192.85 (190.10–196.50; 73)	108.59 (105.50–115.10; 184)	103.30 (99.20–107.60; 181)
M. f. piceus	191.90 (189.10–195.30; 71)	192.60 (192.60–189.90; 54)	110.40 (105.30–115.30; 42)	112.10 (104.00–123.00; 45)
M. m. mentovarius	191.30 (181.00–203.00; 47)	192.30 (106.00–113.60; 384)	111.90 (102.00–120.00; 47)	–
M. m. striolatus	187 (176.00–195.00; 47)	186.50 (166.00–202.00; 46)	118.50 (111.00–123.00; 31)	113.60 (107.00–121.00; 29)

morphological analyses.

4.1.2. Mito-nuclear discordance

Our phylogenetic analyses demonstrated two instances of mito-nuclear discordance. First, we found that *M. bilineatus*, *M. m. striolatus*, and *M. m. mentovarius* rendered *M. flagellum* paraphyletic in the mitochondrial analysis, a finding supported by O'Connell et al. (2017, Fig. 1). Yet, we recovered a monophyletic *M. flagellum* (including *M. fuliginosus*) in our species tree analyses, a result also found by Myers et al. (2017). Second, in the mitochondrial data, we found support for a sister relationship between *M. f. flagellum* and *M. f. testaceus*, but nuclear data supported a sister relationship between *M. f. testaceus* and *M. f. lineatulus*, a pattern also recovered by Myers et al. (2017). This suggests that mitochondrial data was sufficient to identify lineages within whipsnakes, but not suitable for the estimation of interspecific relationships.

4.2. Missing data and species delimitation

Allelic dropout has been discussed extensively in the literature (Arnold et al., 2013). At deeper divergences, mutations in the digestion cut site lead to a reduction in homologous loci shared between species. This can lead to large amounts of missing data in more divergent taxa, which can present a challenge when conducting analyses at the phylogenetic level (Rubin et al., 2012, Cariou et al., 2013, Huang and Knowles, 2014, Streicher et al., 2014, Collins and Hrbek, 2015, Leaché et al., 2015, Eaton et al., 2017). As a result, our analyses that included more divergent taxa (*M. bilineatus* and *M. m. striolatus*) resulted in fewer loci. Another challenge with ddRADseq is that it is less effective with lower quality DNA samples which may not digest well, or may fail size selection (Suchan et al., 2015). We hypothesize that DNA degradation due to the collection on roads of dead specimens reduced the number of available loci in several samples, especially with *M. bilineatus* and *M. m. striolatus*.

Much of the discussion regarding the effects of missing SNP data have focused on likelihood analyses of concatenated datasets (Eaton et al., 2017), but few studies have examined the effects of missing data when using SNAPP, which does not accommodate missing data between assigned species. We found that the inclusion of more loci, even at the expense of very high amounts of missing data, led to higher BF and better resolved species trees than datasets with less missing data but fewer loci. This is because less stringent filtering retains lineage-specific loci, which may help coalescent methods better delimit lineages (Huang and Knowles, 2014). Thus, we advocate that SNP based analyses should focus on maximizing total loci and lineage-specific (highly variable) loci, although the filtering regime will be different for each study.

4.3. Whipsnake taxonomy

We make several taxonomic recommendations for whipsnakes. First, we recommend leaving *M. f. testaceus, M. f. flagellum,* and *M. f. lineatulus* as *M. flagellum*. We emphasize that although the evolutionary distinctiveness of each of these lineages is clearly defined and supported by BFD*, the three lineages form a monophyletic group. Mitochondrial divergence between *M. f. testaceus* and *M. f. flagellum* is low (3.0%), and the divergence between *M. f. lineatulus* is on average 5.8% from the other two subspecies. In addition, O'Connell et al. (2017) found support for gene flow between *M. f. testaceus* and *M. f. lineatulus*. Finally, the mito-nuclear discordance in this group suggests possible mitochondrial introgression between the three lineages (Fig. 1; 4).

Second, we recommend synonymizing *M. f. cingulum* and *M. f. piceus*, and elevating both lineages to species status as *M. piceus* (Appendix A). We tentatively group individuals referred to as *M. f. ruddocki* into this species. Our mitochondrial data placed *M. f. cingulum* and *M. f. piceus* as sister clades, and our nuclear data strongly supported the delimitation of *M. f. cingulum* from other *M. flagellum*. The mitochondrial data places *M. f. piceus* as only 4.6% divergent from *M. f. cingulum*, and 6.6% divergent from *M. fuliginosus*. This is less divergent than *M. f. lineatulus* from the other eastern lineages. Morphological variation also corroborates splitting the eastern and western *M. flagellum* lineages into two species. Ventral scale counts vary less between western lineages than they do between eastern lineages (Table 3), yet the degree of color polymorphisms in the western lineages is higher than in the east, leading to the recognition of the different western subspecies (Wilson, 1970). For this reason, Wilson (1970) hypothesized a two-lineage scenario with western *M. flagellum* diversifying from the ancestor of *M. f. piceus*, and eastern *M. flagellum* from *M. f. testaceus*. We emphasize however that in the absence of nuclear data from *M. f. ruddocki*, we cannot rule out a scenario where *M. f. piceus*, *M. f. cingulum,* and *M. f. ruddocki* each represent independent species, rather than populations. Based on the amount of mito-nuclear discordance observed between the different *M. flagellum* lineages, it may be that these three lineages form a monophyletic group, or, they could be paraphyletic with respect to *M. fuliginosus*. Additional taxonomic sampling would help to more fully resolve these relationships.

Third, we retain the evolutionary species status of *M. fuliginosus*. We found that this species was 6.3% divergent from *M. f. cingulum* and 5.9% divergent from *M. f. piceus*. Our species delimitation analysis confirmed its distinctiveness from *M. f. cingulum* (Table 1). Based on the variation of ventral scale counts from Wilson (1970), this species varies from north to south in scale counts, and has a range of 186–199 ventral scales in males, and 187–198 in females. This is a much wider range than we found in other whipsnakes from the literature. Additionally, this species exhibits a dark and light color morph on the Baja California Peninsula (Wilson, 1970). This may represent polymorphism within the species, or may represent the presence of a second lineage on the peninsula.

Fourth, we recommend the elevation of *M. m. striolatus* to *M. lineatus*. We describe this species in Appendix A. The taxonomic history of this species is complex. Whipsnakes ranging from southern Sonora to southern Jalisco were originally described as *Bascanion lineatus* by Bocourt (1890). Ortenburger (1923) then placed this species within *Masticophis*. Mertens (1934) proposed *striolatus* as a substitute for *lineatus* because it was a secondary homonym of *Masticophis lineatus* (originally described as *Lygophis lineatus* by Linnaeus), and made it a subspecies of *M. mentovarius*. Smith (1941) classified *M. striolatus* as a subspecies of *M. flagellum*, but Bogert and Oliver (1945) and Zweifel and Norris (1955) provided morphological evidence to differentiate *M. f. cingulum* and *M. striolatus*. Zweifel (1960) classified this species as *Masticophis lineatus,* feeling it was unnecessary to suppress secondary homonyms. However, Johnson (1977) considered this species to be conspecific with *M. mentovarius*, and described it as the subspecies *M. m. striolatus*.

If we were to suppress secondary homonyms, we would classify this species as *M. striolatus*; we chose instead to follow Smith and Tayor (1945), and Zweifel (1960), and retain *M. lineatus*. In addition, Smith (1943) and Smith and Taylor (1945) classified the *M. striolatus* from the Tres Marias Islands as a distinct subspecies, which they named *M. m. variolosus*. Zweifel (1960) classified *M. variolosus* as an island population of *M. striolatus*, and used *Masticophis lineatus* to encompass all whipsnakes from these two subspecies. At present, both *M. m. striolatus* from mainland western Mexico, and *M. m. variolosus* from the Tres Marias Islands are recognized subspecies. Lacking sampling of *M. m. variolosus* from the Tres Marias Islands, we can not recommend taxonomic changes for *M. m. variolosus*. However, we hypothesize that this population is closely related to what we describe below as *M. lineatus*. The complex taxonomic history described in this section exemplifies the difficulty of placing this species within the whipsnake phylogeny. While morphological data places this species with *M. mentovarius* or *M. flagellum*, molecular data placed it as sister to *M. bilineatus* (Figs. 1, 2, 4). Yet, *M. lineatus* and *M. bilineatus* were 9.6% divergent in mitochondrial DNA, suggesting deep divergence (Table 2).

While our study has helped to clarify the taxonomy of several lineages within whipsnakes, we encourage further investigation to answer some remaining questions. Specifically, would nuclear data support the synonymy of *M. f. piceus* and *M. f. ruddocki* with *M. f. cingulum*? Second, would more extensive sampling support the species status of *M. fuliginosus*, or should it be lumped with the other western lineages? Third, does hybridization occur at putative contact points between lineages? We found evidence of mito-nuclear discordance, perhaps indicative of mitochondrial introgression between *M. f. testaceus, C f. flagellum*, and *M. f. lineatulus*. However, would more comprehensive nuclear sampling show evidence for introgression between lineages that overlap at the Cochise Filter Barrier, including *M. bilineatus*, *M. f. lineatulus*, *M. f. piceus* and *M. f. cingulum*? Finally, does hybridization occur between overlapping lineages in western Mexico, including *M. f. cingulum*, *M. f. lineatulus*, *M. m. striolatus*, or *M. bilineatus*?

5. Conclusions

We demonstrate the power of using genetic data to explore lineage composition, and the use of genomic data to test models of species relationships to resolve recalcitrant taxonomic classifications, exemplified by *M. m. striolatus*. Our phylogenetic analyses recovered support for several lineages within *M. flagellum*, all of which pertained to previously recognized subspecies. We support the elevation of *M. f. cingulum* to evolutionary species status, which we describe with *M. f. piceus* and *M. f. ruddocki* as *M. piceus* (Appendix A). We found that *M. m. striolatus* was most closely related to *M. bilineatus*, and based on coalescent species delimitation, elevate this subspecies to full species status as *M. lineatus* (Appendix A). We encourage further genomic sampling of western whipsnake lineages to further understand their phylogeny, and to investigate potential admixture at putative contact zones.

Acknowledgments

We sincerely thank Ed Myers and Frank Burbrink for the use of their extensive sequencing data of western US coachwhips. We also TJ Firneno, D. Rivera, E. Wostl, DM Portik, and MK Fujita for their feedback on this manuscript. This work was supported by a Theodore Roosevelt Grant from the American Museum of Natural History.

Appendix A

We revise the taxonomy of *M. flagellum* and *M. mentovarius*. We made several taxonomic recommendations in Section 4.3, including the elevation of *M. f. cingulum* and *M. m. striolatus* to full species status. We emphasize that these species designations are well-supported hypotheses, but additional nuclear sampling from *M. f. piceus* and *M. f. ruddocki* would confirm these hypotheses, or reveal additional undocumented diversity. In addition to molecular data, we emphasize that whipsnake lineages are generally diagnosable by ecoregion, geographic distribution, ventral color, ventral scale count, and supralabial scale count.

Masticophis piceus **(**Cope 1892**)**

Type locality: Fort Grant, Graham County, Arizona

Holotype. USNM 7891

Range

This species is distributed across the range of *M. f. cingulum*, *M. f. piceus*, and *M. f. ruddocki* (Fig. 1A). This range includes the San Joaquin Valley in California, western Nevada, southern Utah, western Arizona, and in Mexico, northern Baja California, Sonora, and Sinaloa. We also recovered on individual from Michoacán that belonged to this species; thus this species may extend as far south as Michoacán. More sampling is needed in western Mexico to confirm the southern limit of this species.

Diagnosis

This species exhibits exceptional variation in ventral color and ventral scale numbers. However, four general color morphs are present within this wide-ranging species. First, what was considered *M. f. ruddocki* in the San Joaquin Valley, California, is a light-bodied snake lacking dark dorsal neck banding (Battstrom and Warren, 1953). Second, what was considered *M. f. piceus* exhibits some isolated areas of snakes with all black dorsal scales, but primarily has red dorsal coloration, often with dark neck bands (Klauber, 1942). Finally, *M. f. cingulum* was characterized by Lowe & Woodin (1954) as having dark red-brown ground color on upper surfaces, broken by complete transverse, narrow, light-colored pink cross-bands, which break the ground color into large, dark, longitudinally oblong sections; the cross bands are doubled (or paired) posteriorly; a single outstanding light band cross the nape. Thus, across most of its range, this snake exhibits a reddish dorsal color, often dark or light-red dorsal banding, usually close to the neck, with all black or very light populations isolated within its range. This species can be differentiated from *M. lineatus* (see below) by mean ventral counts. In *M. piceus*, ventral counts average 193.15 (189.1–197.2) in males, and 195.1 (185–205) in females, compared with in *M. lineatus*, 184.5 (176–195) in males, and 185.7 (166–202) in females. This species can best be differentiated from *M. f. lineatulus* by the lighter brown or yellow dorsal color with the presence of a dark lateral stripe on each dorsal scale in *M. f. lineatulus*. *Masticophis flagellum* usually lacks the horizontal banding present in *M. piceus*. This species can be readily differentiated for *M. m. mentovarius* in the southern end of its range by a supralabial count of 8–8, instead of the 7–7 present in *M. m. mentovarius*. *Masticophis piceus* can also be diagnosed by its geographic range (primarily Sonoran and Mojave deserts), mitochondrial DNA, (Genbank Accession numbers (AY486928, KT713629, KT713648, KT713650–KT713652, KT713674, KT713676, KX835758–KX835764, KX835772, KX835773, KX835775, KX835780–KX835787, KY007698) and by nuclear DNA (NCBI Sequence Read Archive SRS1047243, SRS1047245, SRS1047292, SRS1047293).

Masticophis lineatus **(Boucourt, 1890)**

Type locality: Mexico.

Type specimen. Three syntypes: MNHP 1519 and 1520 from Izucar (though to be Izucar de Matamoros, Puebla, Mexico), and MNHP 1648 from Colima, Mexico. Johnson (1977) designated 1519 and 1520 as intergrades between *M. m. striolatus* and *M. m. mentovarius*, and designated 1648 as the lectotype. Smith and Taylor (1945) designated the type locality as Presidio de Mazatlan, Sinaloa, Mexico. We retain the lectotype classification of Johnson (1977).

Range

This species has been recorded from southern Sonora southward through Sinaloa, Nayarit, Jalisco and Colima. It also extends into Durango, Zacatecas, Aguascalientes, northwestern Michoacán. The species has also been reported from Guerrero, Morelos, Puebla, and as far south as Oaxaca.

Diagnosis

According to the description of Johnson (1977), this species is characterized by lacking prominent mottling on the sides of the head, chin, and anterior ventrals. The supralabials are 8–8, with two entering the orbit. The juvenile pattern is banded. This species has a mean ventral scale count of 184.50 (176.00–195.00) for males, and 185.75 (166.00–202.00) for females. This differs substantially from *M. m. variolosus* on the Tres Maria Islands with a mean count of 194.85 (190.00–204.00) for males and 194.50 (190.00–197.00) for females. This also differs from *M. m. mentovarius* with 191.30 (181.00–203.00) in males and 195.10 (185.00–205.00) in females, *M. piceus* with 195.30

(193.80–197.20) in males, and 194.00 (192.80–196.70) in females, and *M. bilineatus* with 205.25 (203.00–204.00) in males (female data lacking; Table 3). This species can also be diagnosed from *M. piceus* by a wider head and heavier body, with a head width/length ratio of 0.528–0.536 compared with 0.429–0.489 in *M. piceus*. *Masticophis lineatus* is easily diagnosed from *M. bilineatus* by the lack of dorsal striping. This species can be diagnosed from *M. m. mentovarius* by supralabial count of 8–8, in contrast to the 7–7 supralabials usually found in *M. m. mentovarius*. This species can also be diagnosed using mitochondrial DNA (Genbank Accessions KT713692 , KT713693, KT713694 , KT713695) and nuclear DNA (NCBI Short Read Archive SRS1047296, SRS1047267, SRS1047268, SRS1047265).

References

Arnold, B., Corbett Detig, R.B., Hartl, D., Bomblies, K., 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol. Ecol. 22, 3179–3190. http://dx.doi.org/10.1111/mec.12276.

Battstrom, B.H., Warren, J.W., 1953. A new subspecies of racer, *Masticophis flagellum* from the San Joaquin Valley of California. Herpetologica 9, 177–179.

Bogert, C.M., Oliver, J.A., 1945. A preliminary analysis of the herpetofauna of Sonora. B. Am. Mus. Nat. Hist. 83, 297–426.

Bouckaert, R.R., 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics 26, 1372–1373.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932. http://dx.doi.org/10.1093/molbev/mss086.

Burbrink, F.T., Myers, E.A., 2015. Both traits and phylogenetic history influence community structure in snakes over steep environmental gradients. Ecography 38, 1036–1048.

Cariou, M., Duret, L., Charlat, S., 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecol. Evol. 3, 846–852. http://dx.doi.org/10.1002/ece3.512.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140. http://dx.doi.org/10.1111/mec.12354.

Collins, R.A., Hrbek, T., 2015. An in silico comparison of reduced-representation and sequence-capture protocols for phylogenomics. bioRxiv 032565.

Conant, R., Collins, J.T., 1998. Reptiles and Amphibians: Eastern/Central North America. Houghton Mifflin Co, Boston.

Cope, E.D., 1892. A critical review of the characters and variation of the snakes of North America. ProM. U.S. Nat. Mus. 14, 589–694.

Crother, B.I., (ed.), 2012. Standard Common and Current Scientific Names for North American Amphibians, Turtles, Reptiles, and Crocodilians, Seventh Edition. SSAR Herpetol. CirM. 39, 1–92.

De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886.

Eaton, D.A., Spriggs, E.L., Park, B., Donoghue, M.J., 2017. Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst. Biol. syw092. http://dx.doi.org/10.1093/sysbio/syw092.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726. http://dx.doi.org/10.1093/sysbio/sys004.

Fontaneto, D., Flot, J.-F., Tang, C.Q., 2015. Guidelines for DNA taxonomy, with a focus on the meiofauna. Marine Biodiv. 45, 433–451.

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, 480–488.

Grismer, L.L., 1994. The origin and evolution of the peninsular herpetofauna of Baja California. Mexico. Herpetol. Nat. Hist. 2, 51–106.

Grimer, L.L., 1999. Checklist of amphibians and reptiles on islands in the Gulf of California. Bull Sern. Calif. Acad. Sci. 98, 45–56.

Grummer, J.A., Bryson Jr, R.W., Reeder, T.W., 2013. Species delimitation using Bayes factors: simulations and application to the Sceloporus scalaris species group (Squamata: Phrynosomatidae). Syst. Bio. 63, 119–133.

Hensley, M.M., 1950. Results of a herpetological reconnaissance in extreme South-Western Arizona and adjacent Sonora, with a description of a new subspecies of the sonoran whipsnake Masticophis bilineatus. Trans. Kansas Acad. Sci. 53, 270–288.

Hodel, R.G., Chen, S., Payton, A.C., McDaniel, S.F., Soltis, P., Soltis, D.E., 2017. Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. Sci. Rep. 7, 17598.

Huang, H., Knowles, L.L., 2014. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Syst. Biol. 65, 357–365.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267. http://dx.doi.org/10.1093/molbev/msj030.

Johnson, J.D., 1977. The taxonomy AND Distribution of the neotropical whipsnake Masticophis mentovarius (Reptilia, Serpentes, Colubridae). J. Herpetol. 11, 287–309.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Amer. Stat. Assoc. 90, 773–795.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.

Klauber, L.M., 1942. The status of the black whipsnake. Copeia 1942, 88–97.

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

Kühnert, D., Vaughan, T., Wu, C.-H., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLOS Comput. Biol. 10, e1003537–e1003546. http://dx.doi.org/10.1371/journal.pcbi.1003537.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecul. Biol. Evol. 33, 1870–1874.

Lanfear, R., Calcott, B., Ho, S.Y., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29, 1695–1701.

Leaché, A.D., Banbury, B.L., Felsenstein, J., de Oca, A.N.-M., Stamatakis, A., 2015. Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. Syst. Biol. 64, 1032–1047.

Leaché, A.D., Fujita, M.K., Bouckaert, R.R., Minin, V.N., 2014. Species delimitation using genome-wide SNP data. Syst. Biol. 63, 534–542.

Liu, L., Xi, Z., Wu, S., Davis, C.C., Edwards, S.V., 2015. Estimating phylogenetic trees from genome-scale data. Ann. NY Acad. Sci. 1360, 36–53.

Lowe, C.H., Norris, K.S., 1955. Analysis of the herpetofauna of Baja California, Mexico. III. New and revived reptilian subspecies of Isla de San Esteban, Gulf of California, Sonora, Mexico, with notes on other satellite island of Isla Tiburon. Herpetologica 11, 89–96.

Lowe, C.H., Woodin, W.H., 1954. A new Racer (Genus Masticophis) from Arizona and Sonora, Mexico. ProM. Biol. SoM. W.A. 67, 247–250.

Mertens, R., 1934. Die Insel-Reptilian, ihre Ausbreitenung, Variation und Artbildung, Zooligcal, Stuttgart 84, 1–209.

Myers, E.A., Burgoon, J.L., Ray, J.M., Martinez-Gomez, J.E., Matias-Ferrer, N., Mulcahy, D.G., Burbrink, F.T., 2017. Coalescent species tree inference of masticophis and masticophis. Copeia 105, 640–648.

Nagy, Z.T., Lawson, R., Joger, U., Wink, M., 2004. Molecular systematics of racers, whipsnakes, and relatives (Reptilia: Colubridae) using mitochondrial and nuclear markers. J. Zool. Syst. Evol. Res. 42, 223–233.

O'Connell, K.A., Streicher, J.S., Smith, E.N., Fujita, M.K., 2017. Geographical features are the predominant driver of molecular diversification in widely distributed North American Whipsnakes. Mol. Ecol. 26, 5729–5751.

Ortenburger, A.I., 1923. A note on the genera *Masticophis* and *Masticophis* and a description of a new species of Masticophis. OcM. Papers. Mus. Zool. Univ. Michigan 139, 1–14.

Padial, J.M., Miralles, A., De la Riva, I., Vences, M., 2010. The integrative future of taxonomy. Front. Zool 7, 16. http://dx.doi.org/10.1186/1742-9994-7-16.

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S.C., Boisselier, M.-C., Samadi, S., 2015. Use of RAD sequencing for delimiting species. Heredity 114, 450–459.

Petit, R.J., Excoffier, L., 2009. Gene flow and species delimitation. Trends Ecol. Evol. 24, 386–393.

Rambaut, A., 2017. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 12/15/16.

Carstens, B.C., Pelletier, T.A., Reid, N.M., Satler, J.D., 2013. How to fail at species delimitation. Mol. Ecol. 22, 4369–4383.

Roze, J.A., 1953. The *Rassenkreis Masticophis* (*Masticophis*) *mentovarius* (Dumeril, Bibron et Dumeril), 1854. Herpetologica 9, 113–120.

Ruane, S., Bryson, R., Pyron, A., Burbrink, F., 2014. Coalescent species delimitation in milksnakes (genus Lampropeltis) and impacts on phylogenetic comparative analyses. Syst. Biol. 63, 231–250. http://dx.doi.org/10.1093/sysbio/syt099.

Rubin, B.E., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLOS ONE 7, e33394.

Shaw, G., 1802. General Zoology, or Systematic Natural History. Vol.3, part 2. G. Kearsley, Thomas Davison, London, 313–615.

Silvestro, D., Michalak, I., 2012. raxmlGUI: a graphical front-end for RAxML. Organis. Divers. Evol. 12, 335–337.

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

Sites, J.W., Marshall, J.C., 2003. Delimiting species: a Renaissance issue in systematic biology. Trends Ecol. Evol. 18, 462–470.

Smith, H.M., 1941. Notes on Mexican snakes of the genus *Masticophis*. J.W.A. Acad. Sci. 31, 388–398.

Smith, H.M., 1943. Summary of the collections of snakes and crocodilians made in Mexico under the Walter Rathbone Bacon Traveling Scholarship. Ibid 93, 393–504.

Smith, H.M., Taylor, E.H., 1945. An annotated checklist and key to the snakes of Mexico. Bull. U.S. Nat. Hist. Mus. 187, 1–239.

Stebbins, R.C., 2003. Western Reptiles and Amphibians. Houghton Mifflin Co, New York.

Streicher, J.W., Devitt, T.J., Goldberg, C.S., Malone, J.H., Blackmon, H., Fujita, M.F., 2014. Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. Mol. Ecol. 23, 3273–3291. http://dx.doi.org/10.1111/mec.12814.

Streicher, J.W., McEntee, J.P., Drzich, L.C., Card, D.C., Schield, D.R., Smart, U., Parkinson, C.L., Jezkova, T., Smith, E.N., Castoe, T.A., 2016. Genetic surfing, not allopatric divergence, explains spatial sorting of mitochondrial haplotypes in venomous coralsnakes. Evolution 70, 1435–1449.

Suchan, T., Pitteloud, C., Gerasimova, N., Kostikova, A., Arrigo, N., Pajkovic, M., Ronikier, M., 2015. Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on museum collection specimens. PLOS ONE. http://dx.doi.org/10.1101/025551.

Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species. PNAS 114, 1607–1612.
Tripp, E.A., Tsai, Y.H.E., Zhuang, Y., Dexter, K.G., 2017. RADseq dataset with 90% missing data fully resolves recent radiation of Petalidium (Acanthaceae) in the ultra-arid deserts of Namibia. Ecol. Evol 7, 7920–7936.
Uetz, P., Hošek, J., 2017. The Reptile Database. http://www.reptile-database.org, accessed Jan 15, 2017.
Wiens, J.J., 2007. Species delimitation: new approaches for discovering diversity. Syst. Biol. 56, 875–878.
Wilson, L.D., 1970. The coachwhip snake, *Masticophis flagellum* (Shaw): taxonomy and distribution. Tulane Stud. Zool. Bot. 16, 31–99.
Zweifel, R.G., 1960. Herpetology of the Tres Maria Islands. Results of the Puritan-American Museum of Natural History expedition to Western Mexico. Bull. Amer. Mus. Nat. Hist. 119, 77–128.
Zweifel, R.G., Norris, K.S., 1955. Contribution to the herpetology of Sonora, Mexico: descriptions of new subspecies of snakes (*Micruroides euryxanthus* and *Lampropeltis getulus*) and miscellaneous collecting notes. Amer. Midl. Nat. 54, 230–248.

Glossary

Bayes Factor Delimitation (with genomic data): A coalescent-based species delimitation method that allows for the ranking of species delimitation models using Bayes Factors
Coluber: Monotypic genus of snakes in the New World known as racers
General lineage concept: a species concept that defines a species as an independently evolving lineage
Integrative taxonomy: The use of multiple data-types to delimit species
Masticophis: Genus of snakes in the new world known as whipsnakes
Species tree: a phylogenetic tree representing relationships between lineages (species), rather than between genes (gene tree)