

Cryptic species delimitation in the southern Appalachian *Antrodiaetus unicolor* (*Araneae: Antrodiaetidae*) species complex using a 3RAD approach

Lacie Newton¹, James Starrett¹, Brent Hendrixson², Shahan Derkarabetian³, and Jason Bond⁴

¹University of California Davis

²Millsaps College

³Harvard University

⁴UC Davis

May 5, 2020

Abstract

Although species delimitation can be highly contentious, the development of reliable methods to accurately ascertain species boundaries is an imperative step in cataloguing and describing Earth's quickly disappearing biodiversity. Spider species delimitation remains largely based on morphological characters; however, many mygalomorph spider populations are morphologically indistinguishable from each other yet have considerable molecular divergence. The focus of our study, *Antrodiaetus unicolor* species complex which contains two sympatric species, exhibits this pattern of relative morphological stasis with considerable genetic divergence across its distribution. A past study using two molecular markers, COI and 28S, revealed that *A. unicolor* is paraphyletic with respect to *A. microunicolor*. To better investigate species boundaries in the complex, we implement the cohesion species concept and employ multiple lines of evidence for testing genetic exchangeability and ecological interchangeability. Our integrative approach includes extensively sampling homologous loci across the genome using a RADseq approach (3RAD), assessing population structure across their geographic range using multiple genetic clustering analyses that include STRUCTURE, PCA, and a recently developed unsupervised machine learning approach (Variational Autoencoder). We evaluate ecological similarity by using large-scale ecological data for niche-based distribution modeling. Based on our analyses, we conclude that this complex has at least one additional species as well as confirm species delimitations based on previous less comprehensive approaches. Our study demonstrates the efficacy of genomic-scale data for recognizing cryptic species, suggesting that species delimitation with one data type, whether one mitochondrial gene or morphology, may underestimate true species diversity in morphologically homogenous taxa with low vagility.

Introduction

Cataloguing and describing species is a crucial first step towards understanding Earth's biodiversity. Species are the foundation of biological questions including evolutionary processes, ecological systems, physiological mechanisms, and are key to formulating conservation efforts. However, determining species limits remains a contentious topic (Freudenstein et al., 2016; Hey, 2001; Wake, 2006), particularly for species delimitation in taxa prone to cryptic diversity. (i.e., morphologically indistinguishable species yet exhibit extensive molecular divergence; Bickford et al., 2007). Morphological homogeneity may be the result of recent divergence (i.e., not enough time has elapsed for morphological traits to evolve) or niche conservatism (i.e., two species inhabit ecologically similar niches but geographic separation led to genetic divergence due to lack of gene flow; Wiens & Graham, 2005). Cryptic species cannot be distinguished by morphology alone; thus, additional data types, such as molecular and ecological data, are necessary for robust species delimitation (Bickford

et al., 2007; Stockman & Bond, 2007). For example, recent studies using molecular data in addition to other types of data have unveiled cryptic species across many animal taxa, including birds (e.g., Battey & Klicka, 2017; Garg et al., 2016), arthropods (e.g., Bond et al., 2001; Bond & Sierwald, 2002; Bond & Stockman, 2008; Derkarabetian et al., 2011; Derkarabetian & Hedin, 2014; Garrick et al., 2018; Starrett et al., 2018; Starrett & Hedin, 2007; Zhang & Li, 2014), amphibians (e.g., Chan et al., 2017; Moritz et al., 2016; Ortega-Andrade et al., 2015; Ramírez-Reyes et al., 2017; Reynolds et al., 2012; Weisrock & Larson, 2006), cnidarians (e.g., Holland et al., 2004), and annelids (e.g., Barroso et al., 2010). These studies highlight that species delimitation utilizing only morphology potentially underestimates species diversity in taxa with relative morphological homogeneity.

Spiders in the infraorder Mygalomorphae (Bond et al. 2012; Opatova et al. 2019) present a perplexing situation for robust species delimitation when compared to other more diverse araneomorph spiders (as well as other taxa). Araneomorph spider species delimitation relies predominately on morphological criteria (e.g., distinctive differences in male genitalia, body size, etc.; Dupérré and Tapia 2015; Richardson 2016; also see Bond et al. in review for summary) whereas mygalomorph spiders are relatively morphologically homogeneous (Bond & Stockman, 2008; Hamilton et al., 2014; Harvey et al., 2018; Hendrixson & Bond, 2005a, 2005b; Huey et al., 2019; Opatova & Arnedo, 2014b) and subject to significant population genetic structuring at microgeographical scales (Bond et al., 2001; Hedin et al., 2015; Starrett et al., 2018). Mygalomorph life history traits that include limited dispersal abilities (with few exceptions; see Coyle 1983; 1985), relatively long life spans (10-30 years), habitat specialization, and site fidelity altogether make them ideal organisms for studying population divergence and phylogeography (Hamilton et al., 2014; Hedin et al., 2013; Hendrixson et al., 2013; Hendrixson & Bond, 2005a, 2005b; Opatova & Arnedo, 2014a; Satler et al., 2011; Starrett & Hedin, 2007; Stockman & Bond, 2007). Although morphology tends to under-split mygalomorph species, DNA barcoding and related approaches such as GMYC tend to over-split species owing to their population structuring at very fine geographical scales; it has been long recognized (Bond et al. 2001) that mygalomorph populations are highly structured over relatively short distances. Hamilton et al. (2014) posited that GMYC greatly overestimated species diversity in the tarantula genus, *Aphonopelma* (recognizing 114 species), whereas more integrative type approaches recovered fewer species (34). In addition to single locus approaches, multispecies coalescent methods using many loci are more inclined to detect population structure rather than speciation events (Sukumaran & Knowles, 2017). As a result, such attributes of mygalomorph morphological similarity and population structure typify a system requiring more integrative approaches when delimiting species; that is, it is important to evaluate molecular, geographical, ecological, and morphological lines of evidence as opposed to relying on only one data type, analysis, or conceptual approach (e.g., Bond and Stockman 2008; Hamilton et al. 2014; Hedin, Carlson, and Coyle 2015).

The focus of our study is the *Antrodiaetus unicolor* mygalomorph species complex. Like many related mygalomorph spider species, they have relatively long life spans (> 10 years), high site and microhabitat fidelity (e.g. mesic forests and stream banks), and limited dispersal capabilities (Coyle, 1971; Hendrixson & Bond, 2005a, 2005b). Sympatry within the species complex occurs where *Antrodiaetus unicolor* and *A. microunicolor* are co-distributed across part of their ranges in the eastern United States (Figure 1; Hendrixson and Bond 2005a; 2005b). *Antrodiaetus microunicolor* is found only in the southwestern region of the Appalachian Mountains, whereas *A. unicolor* is centered in the central and southern regions of the Appalachian Mountains with peripheral populations extending as far west as the Ozarks in Arkansas and east near the Atlantic Coast, with populations as far south as the Gulf Coast, and as far north to the central-eastern corner of Pennsylvania. In the mesic forests of the Appalachians these spiders are primarily found along creek banks or steep ravines, whereas in the peripheral populations they are isolated in small ravines in humid hardwood forests (Coyle, 1971; Hendrixson & Bond, 2005a). These spiders build subterranean burrows covered by a unique collapsible collar door where they sit and wait for prey; as a result, they rarely depart from their burrows unless disturbed, or to seek a mate in the case of mature males (Coyle, 1971; Hedin et al., 2019).

Antrodiaetus spiders show both morphological stasis and variation across their distribution (Coyle, 1971). Individuals from localities separated by hundreds of kilometers may be morphologically indistinguishable, yet spiders at the same location may exhibit significant disparities in size and coloration (Coyle, 1971;

Hendrixson & Bond, 2005a). Due to difficulty interpreting this intraspecific variation, Coyle (1971) was conservative when revising the genus by maintaining all populations of *A. unicolor* as one species. Later, Hendrixson and Bond (2005a) used morphological and behavioral data to distinguish two forms different in size, coloration, and setal patterns from the Coweeta Long Term Ecological Research site in southwestern North Carolina and described *A. microunicolor* as a new species. A subsequent molecular analysis using two genetic markers (mtDNA gene cytochrome oxidase I and nuclear ribosomal RNA gene 28S) to evaluate species boundaries between these two forms showed that *A. unicolor* is paraphyletic with respect to *A. microunicolor* (Hendrixson & Bond, 2005b), suggesting that these spiders possibly comprise a complex of multiple cryptic species.

With the advent of next-generation sequencing methods, it is now feasible to generate genomic-scale data for non-model organisms (Baird et al., 2008; Faircloth et al., 2012; Lemmon et al., 2012). These much larger, more comprehensive data sets provide a framework for more rigorous tests of species boundaries in systems where previously only a few targeted loci were available. RADseq (restriction-site associated DNA sequencing) approaches are one of the most widely used techniques for generating a reduced representation of the nuclear genome with extensively sampled homologous loci (Andrews et al., 2016; Baird et al., 2008). Various RADseq techniques have been utilized for addressing population genetic and phylogeographic studies (Andrews et al., 2016; Baird et al., 2008; Emerson et al., 2010), reconstructing phylogenetic relationships among both closely (e.g., Eaton et al. 2016) and distantly (e.g., Leaché et al. 2015) related species, and evaluating species boundaries and speciation processes (e.g., Battey & Klicka, 2017; De Jesús-Bonilla et al., 2019; Delgado-Machuca et al., 2019).

We generated RADseq data using 3RAD, a three-enzyme protocol that reduces DNA chimaera and adapter-dimer formation (Bayona-Vasquez et al., 2019), to investigate species boundaries and phylogenetic relationships within the *A. unicolor* species complex. Specifically, we employed the cohesion species concept (Templeton, 1989), which defines a species as a set of populations that derives from a single evolutionary lineage and meet the criteria of being genetically exchangeable and/or ecologically interchangeable. This integrative approach is particularly insightful when evaluating species limits in morphologically cryptic taxa with high population genetic structuring (e.g., Bond & Stockman, 2008; Hendrixson et al., 2013; Stockman & Bond, 2007). To evaluate cohesion species identity within the *A. unicolor* species complex, we assessed the amount of genetic population structure using clustering analyses to identify the number of evolutionary lineages (i.e., genetic exchangeability), and constructed niche-based distribution or ecological niche models (ENMs) for each lineage. ENMs were then compared to evaluate overlap/similarity to assess potential ecological interchangeability.

Methods

Genomic Library Prep

We sampled 157 individuals from 103 localities throughout the geographic distribution of the *Antrodiaetus unicolor* species complex (Supplementary Table S1; Figure 1). We extracted genomic DNA from leg tissue for each individual using the Qiagen DNeasy Blood and Tissue kit (Qiagen, Valencia, CA) following the manufacturer’s protocol. DNA was quantified using a Qubit 3.0 Fluorometer (Life Technologies, Inc., Grand Island) and checked for quality using an agarose gel.

RADseq libraries were prepared following the Adapterama III protocol (Bayona-Vasquez et al., 2019), which is a modified version of ddRAD (Peterson et al., 2012) that uses three enzymes for digesting genomic DNA (3RAD). This protocol alleviates the need for large quantities of DNA since the third enzyme prevents adapter dimers and DNA chimaeras from forming during the reaction (Bayona-Vasquez et al., 2019; Graham et al., 2015; Hoffberg et al., 2016). Based on previous studies that tested different enzyme combinations (Bayona-Vasquez et al., 2019; Burns et al., 2017), we chose EcoRI-HF, MspI, and ClaI as our cohort of enzymes for genomic DNA digestion.

For each sample, 100 ng of genomic DNA were digested for 2 hr at 37 degC in an 18 μ l solution consisting of 1X Cutsmart buffer, 20U EcoRI-HF, 20U MspI, 10U ClaI, 0.28 μ M each of forward and reverse adapters.

Immediately following digestion, the reaction was brought to 24 μ l total consisting 0.25x Ligase buffer, 100U T4 DNA Ligase, 0.75 μ M rATP, and incubated at 22 $^{\circ}$ C for 20 min and 37 $^{\circ}$ C for 10 min for 3 cycles and a final 20 min enzyme kill step at 80 $^{\circ}$ C. Libraries were pooled and cleaned with 1.5X volume Sera-Mag Speedbeads (Fisher Scientific, Pittsburgh, PA) and quantified using the Qubit.

PCR was used to attach the iTru5 primer. Six replicate 50 μ l reactions consisting of 10 μ l of pooled DNA, 1X Kapa HiFi Fidelity Buffer (Kapa Biosystems, Massachusetts, EUA), 0.3 μ M dNTP mix, 0.5 μ M iTru5-8N primer, and 1U KAPA HiFi Hotstart DNA polymerase with conditions at 95 $^{\circ}$ C for 2 min of initial denaturation, followed by 98 degC for 20 sec, 61 degC for 30 sec, and a 5.5 min 72 degC extension. The pools were combined and then 2X volume bead-cleaned. Twelve iTru7 primers were added using PCR with three reactions consisting of 10 μ l of pooled DNA, 1X Kapa HiFi Fidelity Buffer, 0.3 μ M dNTP mix, 0.5 μ M P5 primer, 0.125 μ M of four different iTru7 primers, and 1U KAPA HiFi Hotstart DNA polymerase with conditions at 95 $^{\circ}$ C for 2 min; 6 cycles at 98 degC for 20 sec, 61 degC for 15 sec, and 72 degC for 30 sec; and a final extension at 72 degC for 5 min. The products were pooled and 2X volume bead-cleaned and quantified.

Size selection was performed using Pippin Prep (Sage Science, Beverly, MA) to capture 570 bp +/- ~18% (470-670 bp range) fragments using a 1.5% agarose cassette. The last enrichment PCR for the final size-selected pool was run with 10 μ l of pooled DNA, 1X Kapa HiFi Fidelity Buffer, 0.3 μ M dNTP mix, 0.5 μ M each of P5 and P7 flowcell binding primers, and 1U KAPA HiFi Hotstart DNA polymerase with conditions at 95 $^{\circ}$ C for 2 min; 9 cycles at 98 degC for 20 sec, 61 degC for 15 sec, 72 degC for 45 sec; and lastly an extension at 72 degC for 5 min. Final libraries were sent to the Georgia Genome and Bioinformatics Core or UC Davis Genome Center for paired-end 150 bp mid-output sequencing with the Illumina NextSeq 550 platform.

Sequence Analysis

Short read processing was conducted using ipyrad v.0.7.1 (Eaton, 2014) on the Hopper HPC Cluster at Auburn University. The workflow for ipyrad involves the following steps: 1) demultiplexing raw reads; 2) quality filtering reads; 3) de novo clustering of data within samples using VSEARCH (Rognes et al., 2016) with clusters then aligned using MUSCLE (Edgar, 2004); 4) jointly estimating heterozygosity and error rate; 5) estimating consensus allele sequences from clustered reads; 6) de novo clustering of data and alignment across samples; 7) filtering/formatting output files for downstream analyses (see Table 1 for parameter settings used). Branched assemblies were created to assess different parameter settings (i.e. different clustering thresholds and minimum samples per locus) and differing number of individuals present in the matrix (i.e. all individuals and only *A. unicolor* B individuals), and the final data matrix including all individuals was chosen after comparing the efficacy of each built from the branched assemblies (see Results; Table 1).

Phylogenetic Analysis

We estimated the phylogeny for several datasets (see Table 1 for dataset descriptions) within a maximum likelihood framework implemented in RAxML v8.2.4 (Stamatakis, 2014). We used the GTR+G model of sequence evolution. For each analysis, 1000 bootstrap replicates were calculated using the rapid bootstrapping option implemented in RAxML. The phylogenies constructed from these datasets were viewed in Figtree v1.4.1 with midpoint rooting. The phylogenies were then compared to look for congruence among clades and bootstrap values, with the final tree having the largest amount of taxon coverage and high support values used for downstream analyses (All_M20 dataset with all 157 individuals and a minimum of 20% locus occupancy; Figure 2).

Bayesian inference of the All_M20 data set was conducted in ExaBayes v1.4.1 (Aberer et al., 2014). Two independent runs of 6×10^7 generations with four coupled chains each were run simultaneously starting with a random starting tree. Standard deviation of split frequencies was monitored (< 0.05), and the first 25% were discarded as burn-in. An extended majority rule consensus tree was generated using the program consense within ExaBayes (Supplemental Figure S1; Aberer, Kobert, and Stamatakis 2014).

A species tree was generated from a set of gene trees using the software ASTRAL-III (Supplemental Figure S2; Zhang et al. 2018). We used the MAGNET v0.1.5 pipeline (Bagley, 2019) to estimate a gene tree for each RAD locus within the All_M20 data set using RAxML v.8.2.4, with each RAxML run implementing the GTR+G model. These gene trees were used as input for ASTRAL-III. With uncertainty in the relationships among *A. unicolor* B, a well-supported clade in both maximum likelihood and Bayesian analyses but slightly differing sister relationships (see Results), additional species trees were generated using *A. unicolor* B-only matrices (UniB_M60 and UniB_M80; see Table 1) for aid in elucidating relationships among *A. unicolor* B.

Cohesion Species Delimitation

We employed the Bond and Stockman (2008) approach by utilizing two criteria to define cohesion species: genetic exchangeability and/or ecological interchangeability, which means that two or more populations exchanging genes and having similar ecological niche attributes are considered a cohesion species. To identify the number of evolutionary lineages, we used the RAxML topology (Figure 2) to establish a baseline evolutionary framework.

Genetic Exchangeability

Following the Bond and Stockman (2008) flowchart for delimiting cohesion species, we first established a “basal lineage” starting point, resulting in five lineages. Lineage name designations were based on clade labels from Hendrixson and Bond (2005b; see Figure 2): *A. microunicolor* and four lineages within *A. unicolor* (A, B1, B2, and B3). We assessed the potential for gene flow (i.e. genetic exchangeability) between sister lineages by evaluating the distributions of these lineages (Figure 1).

To help elucidate the potential for gene flow, several clustering analyses were performed. STRUCTURE v 2.3.4 (Pritchard et al., 2000) was run with several data matrices of unlinked SNPs for 1,000,000 generations and 100,000 burn-in using the admixture model featuring K values ranging from 1-10, with five replicate runs for each K value. The package pophelper (Francis, 2017) in R v.3.5.3 (R Core Team, 2019) was used to calculate the optimal K value via the Evanno method of calculating ΔK (Evanno, Regnaut, and Goudet 2005) as well as view the output (Figure 3a). In addition to STRUCTURE, we also used the R package adegenet (Jombart, 2008; Jombart & Ahmed, 2011) to perform a principal components analysis (dudi.pca function) and visualize the clustering of each lineage (Figure 3b). Also, an unsupervised machine learning approach, specifically a Variational Autoencoder (VAE), was implemented to provide another clustering algorithm (Figures 3c & 3d; for details see Derkarabetian et al. 2019). This approach, which relies on the inherent structure within the data for clustering individuals, alleviates the need for a priori hypotheses of species number, level of divergence, and any population parameters. VAEs, derived from Bayesian probability theory, provide relatively clear visualization for assessing cluster uncertainty due to the standard deviation around samples/clusters (Derkarabetian et al., 2019).

In cases of allopatric lineages, we considered the lineages isolated and genetic exchangeability was rejected. However, in cases of sympatric or parapatric lineages with uncertainty of whether a barrier to gene flow is present, we tentatively accept genetic exchangeability unless the clustering analyses indicate separate clusters. Because cohesion species can be rejected by either not being genetically exchangeable or ecologically interchangeable, these lineages were also tested for ecological interchangeability.

Ecological Interchangeability

We evaluated a proxy for ecological interchangeability by measuring the overlap between ENMs for the sister lineages being compared (generally following the approach of Stockman and Bond 2007) as well as taking into account previous morphological and behavioral criteria for distinguishing *A. microunicolor* (Hendrixson & Bond, 2005a). Current climate data from 1970-2000 for 19 bioclimatic variables at 30 arc-second resolution were downloaded for tiles 12 and 13 from WorldClim v.2 (<http://worldclim.org/version2>; Fick and Hijmans 2017). Climate data from the tiles were then merged into layers, cropped to the area of interest, and converted to a raster stack using the packages ‘raster’ (Hijmans, 2015) and ‘rgdal’ (Bivand et al., 2019) in R. The software ENMTools v1.4.4 (Warren et al., 2010) was used to estimate the amount of correlation

between the 19 bioclimatic variables, with significant correlation defined as $r > .90$ following (Jezkova et al., 2011). Six variables were removed and the 13 retained variables (BIO 2-9,12,13,15,16,18) were used for generating ENMs. Maxent v.3.4.1 (Phillips et al., 2006) was used to estimate ENMs with default settings for the lineages. The receiver-operating characteristic (ROC) plot's area under the curve (AUC) was used as a measure of model prediction accuracy, with values > 0.9 indicating optimal model performance as opposed to values < 0.7 indicating poor model performance (Swets, 1988). Occurrence records were based on specimens collected for this study and specimens from prior publications (Hendrixson & Bond, 2005a, 2005b) also included in this study.

To statistically compare the ENMs of each lineage, we conducted analyses of niche overlap, niche identity, and niche similarity (i.e. background) tests in ENMTools (Warren et al., 2008, 2010). Niche overlap for each sister lineage comparison was quantified using Schoener's D (Schoener, 1968), which ranges from 0 (no overlap between ENMs) to 1 (complete overlap of ENMs). To assess the significance of D , we employed both niche identity and niche similarity tests. For the niche identity test, 100 pseudoreplicates were used to construct a null distribution of niche overlap compared to the observed overlap value using a one-tailed test (Supplemental Figure S3). Warren, Glor, and Turelli (2008) highlighted that the niche identity test may be too strict and results in the null hypothesis often being rejected, so the more conservative niche similarity test was also conducted. The background regions for each lineage were generated based on minimum area polygons from occurrence points in ArcMap v10.7 (ESRI), and occurrence points of one lineage were tested against random points from the background region of the other lineage and vice versa. A hundred pseudoreplicates were used to construct a null distribution of niche similarity compared to the observed overlap value using a two-tailed test (Supplemental Figures S4-7). In order to reject ecological interchangeability with the niche similarity test, both of the background tests for a pair of lineages should have a D value that is more different (niche divergence) than expected by chance. Since some of our comparisons involved parapatric/sympatric lineages which results could potentially be affected by the defined background region, we conducted sensitivity tests by defining alternative background regions based on buffer polygons from occurrence points of differing distances (25 km, 50 km, 75 km, and 100 km) in ArcMap v10.7.

Results

3RAD Data

For the 157 specimens sampled throughout the distribution of the *A. unicolor* species complex, 234,116,852 raw reads were obtained ranging from 365,011 to 4,698,473 and averaging 1,491,190 reads per sample (Table 1). Table 1 outlines the number of loci and SNPs generated in each dataset with different `min_sample_locus` parameter values. Additional clustering thresholds (0.85, 0.88, 0.90) were tested as well, but there were minimal differences in the number of retained loci/SNPs across these thresholds; therefore, we chose the threshold value of 0.85 in ipyrad for generating our data matrices. When `min_sample_locus` values are lower for both the full dataset (All_samples) and the subset with *A. unicolor* B samples only (A.unicolorB_only), more loci and SNPs are retained compared to higher `min_sample_locus` values, which have a lower yield of loci and SNPs.

Phylogenetic Analysis

Maximum likelihood phylogenies produced from the All_M10, All_M20, All_M30, and All_M40 datasets all support three main clades: *A. microunicolor*, *A. unicolor* A, and *A. unicolor* B (Figure 2). However, low support values (< 70) for nodes in the All_M30 and All_M40 are more frequent when compared to All_M10 and All_M20 (see Discussion). Both All_M10 and All_M20 recovered a similar topology except for several sister relationships at the tips but with the majority of nodes having higher support values (> 70). The data matrix All_M20 was preferred over All_M10 for downstream analyses due to All_M20 having a higher taxon coverage than All_M10 while still maintaining well-supported relationships.

The Bayesian inference phylogeny produced from the All_M20 dataset was overall well-supported with posterior probabilities [?] 0.95 and very closely resembled the maximum likelihood All_M20 phylogeny with the exception of a few sister relationships within *A. unicolor* B (Supplemental Figure S1). Specifically, the clade

comprising MY2542, MY2541, MY2806, MY2807, ANTR74, and ANTR73 is sister to the rest of *A. unicolor* B in the Bayesian, whereas the clade is sister to all other samples within *A. unicolor* B3 in the maximum likelihood phylogeny. The node denoting the split between this clade and its sister group has low support for both All_M20 RAxML and Bayesian phylogenies; however, the All_M10 RAxML tree, which has the same sister group relationship as the RAxML All_M20, has moderate support. Therefore, the All_M20 maximum likelihood phylogeny was preferred over the Bayesian tree for downstream analyses.

The quartet-based species tree estimated in ASTRAL-III (Supplemental Figure S2) comprised 868,916,301 induced quartet trees from the gene trees, which represented 46.19% of all quartets present in the species tree. The low normalized quartet score indicates a very high level of gene tree discordance. The species tree yielded a different topology from both the Bayesian and maximum likelihood trees (Figure 2; Supplemental Figures S1 & S2). Two of the main clades *A. microunicolor* and *A. unicolor* A remain monophyletic; however, there is significant uncertainty in the relationships between individuals within *A. unicolor* B, which reflects a high level of gene tree discordance (see Discussion).

Genetic Exchangeability - Clustering Analyses

Three different genetic clustering analyses were used to evaluate the potential for gene flow (see Table 2 for summary). The STRUCTURE analysis with All_M20 resulted in an optimal K value of eight (Figure 3a), with the two monophyletic lineages *A. microunicolor* and *A. unicolor* A clustering separately; however, *A. unicolor* B formed six clusters that did not correspond to monophyletic groups. STRUCTURE analyses with UniB-only matrices produce two to three clusters with all but one data set recovering non-monophyletic clusters (Figure 4). UniB_M30 does recover the three monophyletic lineages within *A. unicolor* B, but some individuals have mixed cluster assignments (Figure 4). In contrast, both PCA and VAE clustering analyses (Figure 3b, 3c, & 3d) for All_M20 indicate three distinct clusters consistent with monophyletic lineages: *A. microunicolor*, *A. unicolor* A, and *A. unicolor* B. The VAE analysis with UniB_M30 shows some separation between *A. unicolor* B lineages; however, there is still some overlap between them when considering standard deviation around samples (Figure 4). The lineages within *A. unicolor* B were not recovered as discrete clusters in any of the three clustering analyses.

Ecological Interchangeability - Niche Identity/Similarity Analyses

Table 3 summarizes the ecological interchangeability analyses. Although *A. microunicolor* has morphological and behavioral differences indicating adaptive divergence from *A. unicolor* A, the other lineage comparisons do not possess any obvious morphological or known behavioral differences (Hendrixson & Bond, 2005a, 2005b). Niche identity was rejected for all four sister lineage comparisons. However, for the niche similarity test three of the comparisons were not significantly more different or similar than expected regardless of the background region threshold: 1) *A. microunicolor* and *A. unicolor* A; 2) *A. unicolor* B1 and *A. unicolor* B2; 3) *A. unicolor* B1 + *A. unicolor* B2 and *A. unicolor* B3. The fourth comparison between *A. microunicolor* + *A. unicolor* A and *A. unicolor* B had conflicting results with three of the four background regions (50km, 75km, and 100km) while the 25km background region indicated no niche divergence or niche conservatism. Comparing *A. microunicolor* + *A. unicolor* A occurrences to the background regions of *A. unicolor* B suggests niche divergence (i.e., significantly more different than expected), whereas comparing the occurrences of *A. unicolor* B to the background regions of *A. microunicolor* + *A. unicolor* A indicates niche conservatism (i.e., significantly more similar than expected).

Discussion

3RAD - Analytical considerations/caveats

First, when assembling 3RAD data it is essential to consider the effects various parameters will have on downstream analyses, especially the locus occupancy. Studies have shown that the amount of missing data can greatly affect resulting phylogenetic inferences (e.g., Crotti et al., 2019; Eaton et al., 2016; Huang & Knowles, 2016), with the trend toward a greater amount of missing data yielding a more robust phylogeny compared to a low amount of missing data. This result has been attributed to larger data sets comprising

more phylogenetic data/signal and informative sites being excluded when a greater percentage of taxon coverage is required for a locus to be retained in the final data matrix; therefore, quickly mutating sites become disproportionately omitted with increasing taxon coverage and exclude potentially variable and informative characters (Crotti et al., 2019).

We examined the effect of increasing the minimum sample per locus parameter for data sets containing all individuals as well as those with *A. unicolor* B samples only and found that increasing the number of samples required greatly lowered the number of SNPs retained in the final matrix (summarized in Table 1). Our phylogenetic inferences using these data sets (Figure 2; Supplementary Figures S1 & S2) reflect findings of previous studies (e.g., Crotti et al. 2019) showing disproportionate loss of informative SNPs resulting in phylogenies with ambiguous or unresolved evolutionary relationships. All_M30 (30% locus occupancy) and All_M40 (40% locus occupancy) matrices contain many nodes with low support (< 50) and polytomies present; however, All_M20 and All_M10 have very similar topologies and include a majority of highly-supported ([?] 95) nodes.

Not only did the amount of missing data affect our phylogenetic inferences, but it also influenced our STRUCTURE and VAE analyses (Figure 4). We explored the effects of missing data on these clustering analyses with our *A. unicolor* B only data sets (UniB_M30, UniB_M40, UniB_M50, UniB_M60, and UniB_M70). For STRUCTURE, the number of clusters decreased from $?K=3$ for UniB_M30 to $?K=2$ for UniB_M40, UniB_M50 and UniB_M60 while UniB_M70 had a peculiar increase back to $?K=3$. Although UniB_M30's clusters reflect population structure (albeit with some admixture) corresponding to the three lineages within *A. unicolor* B, the data sets with $?K=2$ cluster B2+B3 together with varying amounts of admixture. The three clusters inferred from the UniB_M70 dataset do not necessarily reflect the structure found in UniB_M30 because several individuals are not clustering with individuals from their lineage. This is most likely due to the greatly decreased number of informative sites to accurately detect any structure in the data.

VAE clustering across these data sets reflected the same general trend as STRUCTURE whereby increasing taxon coverage greatly decreases the amount of structure detectable in the data (Figure 4). UniB_M30 clusters had the greatest amount of structure detected for each lineage, though some overlap between them is still present. UniB_M70 has practically no noticeable structure in the data, which most likely results from the small number and mostly uniform loci retained in the data matrix. VAE, which uses the structure inherent in the data to train the model, is likely affected by the deficiency of informative sites.

Speciation and phylogeography

Like most mygalomorphs, the *Antrodiaetus unicolor* species complex has certain life history traits (i.e. long generation times, limited dispersal capabilities, morphological conservatism) that drastically influence population structure (Coyle, 1971). This in conjunction with extensive molecular divergence would most likely indicate that the populations have been isolated from gene flow for an extended period of time, which would increase speciation potential (Barraclough, 2019). While that seems to be the case when comparing *A. microunicolor*, *A. unicolor*A, and *A. unicolor*B, lineages within *A. unicolor*B remain ambiguous.

Specifically, the STRUCTURE analyses for both All_M20 and several UniB datasets recover conflicting clusters (Figures 3a & 4; also see 3RAD section above). These clusters generated for all datasets were incongruent with monophyletic groups, which might be reflective of the use of unlinked SNPs for STRUCTURE versus using all variation for phylogenetic inference. While VAE for UniB_M30 provides some evidence of structure separating these lineages, there is still slight overlap between them (Figure 4a). Derkarabetian et al. (2019) used STRUCTURE and VAE for species delimitation, which revealed clear VAE and STRUCTURE clusters agreeing with multiple lines of evidence (i.e., COI clades, DAPC, morphology) for species-level divergence within the harvestman *Metanonychus*, a group that is also known to have high population structure, conservative somatic morphology, with allopatry between species in the focal species group. Our study, however, includes a taxon with sympatry/parapatry across multiple lineages in addition to being morphologically homogeneous, with *A. microunicolor* being the exception. Given the more recent divergence in the *Antrodiaetus unicolor* species complex (~ 11.5 mya; Hendrixson & Bond, 2007) compared to *Metanonychus* (~ 25

mya; Derkarabetian et al., 2019), the underlying genetic patterns indicate speciation may not necessarily be complete and/or as clearly reflected in the VAE clusters.

Although the mixed cluster assignments of individuals suggest admixture between populations is occurring, it could also indicate incomplete lineage sorting (i.e., ancestral polymorphism) and/or recent range expansion (Avice, 2009). In addition, the species tree generated in ASTRAL (Supplemental Figure S2) showed considerable uncertainty for most of the relationships within *A. unicolor* B indicating an appreciable amount of gene tree discordance. There are several potential reasons for gene tree/species tree discordance (incomplete lineage sorting, reticulation, gene duplication, or horizontal gene transfer; see Maddison, 1997) but among these, incomplete lineage sorting (ILS) seems the most likely explanation due to the prevalence of ILS in taxa with long generation times, large effective population sizes, and/or low mutation rates (Degnan & Rosenberg, 2009). Additionally, mtDNA data in previous analyses hinted at the possibility of a recent range expansion (Hendrixson, unpublished data; see speciation scenarios below).

All spiders in this species complex appear to have similar habitat requirements (Coyle, 1971; Hendrixson & Bond, 2005a). However, morphology, behavior, and large-scale ecological (i.e., climatic) data offered some plausible insight into potential adaptive divergence. Morphological, geographical, and molecular data clearly demarcate *A. microunicolor* and *A. unicolor* A as distinct lineages, yet the niche similarity tests between these two lineages reveals no significantly divergent or conserved niches. Two of the background areas (75km and 100km) approach significant niche divergence when comparing *A. microunicolor* occurrence points to the background area of *A. unicolor* A but not vice versa. It is possible that niche divergence important for speciation in this complex occurs in other ecological variables not tested in our study; however, our clustering analyses show minimal gene flow between these lineages and suggests a speciation model in which reproductive isolation accumulated in allopatry without much ecological differentiation (i.e., niche divergence may not have been a large driver in the speciation process). With subsequent sympatry between *A. microunicolor* and *A. unicolor* B these lineages potentially maintained reproductive isolation through premating barriers, specifically character displacement with disparities in size and breeding periods (e.g., Bond and Sierwald 2002).

The other lineages, largely sympatric, seem morphologically and behaviorally similar, thus we use ENM ecological data. *A. microunicolor* + *A. unicolor* A and *A. unicolor* B niche similarity analyses yielded conflicting results. The comparison of *A. microunicolor* + *A. unicolor* A occurrences to the background area of *A. unicolor* B would suggest niche divergence (i.e., significantly more different than expected) whereas the comparison of occurrences of *A. unicolor* B to the background area of *A. microunicolor* + *A. unicolor* A suggests niche conservatism (i.e., significantly more similar than expected). For niche divergence, one possible explanation could be that *A. unicolor* B is much more widespread than both *A. microunicolor* and *A. unicolor* A and, therefore, has more potential for heterogeneous environmental conditions in the habitat available to them (e.g., more tolerance for lower elevation compared to *A. microunicolor* and *A. unicolor* A). Additionally, niche conservatism could be the result of *A. unicolor* B's distribution mostly overlapping *A. microunicolor* and *A. unicolor* A's distributions, suggesting that *A. unicolor* B can not only easily inhabit the same environmental conditions but also likely prefers similar habitats as *A. microunicolor* and *A. unicolor* A.

The niche similarity analyses involving lineages within *A. unicolor* B yielded no obvious results for either niche divergence or niche conservatism. The niche similarity analyses for *A. unicolor* B1 and *A. unicolor* B2 show that their niches are no more different or similar than expected for any of the background areas tested, although the 100km background area hinted at niche conservatism (not significantly though). The Little Tennessee River may be a potential geographic barrier between these two lineages facilitating reproductive isolation, but not enough time has occurred for the accumulation of neutral divergence indicative of allopatric speciation. Numerous taxa in the southern Appalachians exhibit similar phylogeographical patterns, identifying either the Asheville or Little Tennessee River Basins as potential barriers to gene flow: *Hypochilus* araneomorph spiders (Keith & Hedin, 2012), harvestmen *Sabacon cavicolens* (Hedin & McCormack, 2017) and *Fumontana deprehendor* (Thomas & Hedin, 2008), *Dasycerus carolinensis* beetles (Caterino & Langton-

Myers, 2019), *Narceus millipedes* (Walker et al., 2009), and plethodontid salamanders (Kozak & Wiens, 2010; Weisrock & Larson, 2006). Consequently, these lineages could prospectively be considered distinct species after sufficient time and divergence in allopatry. Additionally, the niche similarity analyses for *A. unicolor* B1 + *A. unicolor* B2 to *A. unicolor* B3, while not consistently significant for all background areas, does show signs of niche divergence. For the 100km and 75km background areas *A. unicolor* B3 occurrences were significantly different from the background of *A. unicolor* B1 and *A. unicolor* B2 but not quite significantly different for the reciprocal comparison. Also, the 50km background area hints at niche divergence for both comparisons. One explanation could be that the ancestor of *A. unicolor* B lineages had undergone range expansion where *A. unicolor* B3 later became geographically isolated from *A. unicolor* B1 and *A. unicolor* B2 for a significant period of time generating reproductive isolation, with subsequent secondary contact potentially driving niche divergence and, therefore, reinforcing reproductive isolation in sympatry.

One caveat for using large-scale ecological data like these is the potential for not having the resolution needed for detecting microhabitat differences, which are often important for divergent selection and driving speciation (Massatti & Knowles, 2014; Soberon, 2007). Mygalomorph spiders like *Antrodiaetus* are known for habitat specialization at small scales, such as preferring north-facing slopes, shaded ravines, and particular soil types (Hedin et al., 2015; Starrett et al., 2018). Therefore, it is possible that we could be overlooking significant microhabitat differences that potentially drive niche divergence, with speciation following. Another potential issue could be how we defined the background areas. Considering that these spiders have low dispersal capabilities, it may be that the regions with higher amounts of distance incorrectly show niche divergence due to incorporating potentially uninhabitable environments and, therefore, misinterpreting possible speciation mechanisms (Warren et al., 2008).

Species limits of the A. unicolor species complex

Species delimitation of the *Antrodiaetus unicolor* species complex has long been challenging due to both morphological stasis and variation across their distribution (Coyle, 1971; Hendrixson & Bond, 2005a). As already discussed, *A. microunicolor* is unambiguously recognized as a species based on morphological, behavioral, and genetic divergence from *A. unicolor*. However, the now paraphyletic assemblage of *A. unicolor* lineages lacks distinctive features that could be used to distinguish them. Because no nomenclatural changes have been made to elevate these lineages to species status, the species-level diversity remains underestimated in this complex.

This study employed genome-wide SNPs for several clustering analyses and niche-based distribution modeling to evaluate genetic and ecological exchangeability, which further elucidated potential species boundaries. Clustering analyses with all individuals present (All_M20) as well as morphology and behavior coincide with the three species hypothesis (*A. microunicolor*, *A. unicolor* A, and *A. unicolor* B). Alternatively, the niche identity analyses and STRUCTURE/VAE analyses for *A. unicolor* B individuals only (UniB_M30) support the five species hypothesis (additional splitting of *A. unicolor* B into B1, B2, and B3). Although it is possible that these lineages were once allopatric and diverged before recent secondary contact in the Southern Appalachians, the tentative interpretation of potential incipient speciation events within *A. unicolor* B and less conservative niche identity test results remain too ambiguous for confident species delineation within the clade. Our integrative approach utilizing morphological, behavioral, and a substantial amount of molecular characters with ecological niche modeling provided ample evidence of an additional cohesion species within the complex for a total of three (*A. microunicolor*, *A. unicolor* A, and *A. unicolor* B), which potentially warrants distinguishing these two genetically and ecologically different lineages as separate species. In that case, one lineage would be designated as the true name-bearing *Antrodiaetus unicolor* and one with a new name. Considering that Hendrixson and Bond (2005a) designated a neotype for *Antrodiaetus unicolor* from Desoto State Park in northeastern Alabama and specimens included in this study from the surrounding areas cluster together in the lineage *A. unicolor* B, this lineage should be considered the true name-bearing *Antrodiaetus unicolor*. To our knowledge, none of the existing available names (junior synonyms of *A. unicolor* sensu lato) would be attributed to any of the *A. unicolor* A individuals based on geography. As such a new name will need to be proposed for *A. unicolor* A.

Overall, our study demonstrates the efficacy of genomic-scale data for recognizing cryptic species, suggesting that species delimitation with one data type, whether one mitochondrial gene or morphology, underestimates species diversity in taxa with low vagility and relative morphological stasis. Incorporating multiple lines of evidence (i.e., morphological, behavioral, geographical, and ecological diversity) in addition to genomic-scale data underscores the robustness of integrative species delimitation approaches across all organismal diversity despite differences in biological or ecological characteristics. We were able to resolve species-level paraphyly within the *A. unicolor* complex by delimiting an additional putative species despite morphological homogeneity. In addition, our study highlights another instance of cryptic speciation in the southern Appalachians, with phylogeographical patterns contributing to our understanding of the processes generating biodiversity in this rich, geologically and environmentally complex region.

Acknowledgements

Research funding and other support for this study was provided by the National Science Foundation and Evert and Marion Schlinger Foundation. We thank Vera Opatova for her assistance and camaraderie in the field.

Data Accessibility

All data matrices, STRUCTURE-formatted files and VAE one-hot encoded files can be found here: [will have dryad repository]. All scripts used in this study other than the VAE script (found here:<https://github.com/sokrypton/sp-deli>) can be found here:https://github.com/lgnewton/A.unicolor_sp-delim

References

- [dataset] Newton, Lacie et al. (2020), Cryptic species delimitation in the southern Appalachian *Antrodiaetus unicolor* (Araneae: Antrodiaetidae) species complex using a 3RAD approach, UC Davis, Dataset, <https://doi.org/10.25338/B8Z61M>
- Aberer, A. J., Kobert, K., & Stamatakis, A. (2014). ExaBayes: Massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution*, 31 (10), 2553–2556. <https://doi.org/10.1093/molbev/msu236>
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17 (2), 81–92. <https://doi.org/10.1038/nrg.2015.28>
- Avise, J. C. (2009). Phylogeography: Retrospect and prospect. *Journal of Biogeography*, 36 (1), 3–15. <https://doi.org/10.1111/j.1365-2699.2008.02032.x>
- Bagley, J. C. (2019). *MAGNET v0.1.5* [GitHub Package]. <http://github.com/justincbagley/MAGNET>
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3 (10), e3376. <https://doi.org/10.1371/journal.pone.0003376>
- Barracough, T. (2019). *The Evolutionary Biology of Species*. Oxford University Press.
- Barroso, R., Klautau, M., Sole-Cava, A. M., & Paiva, P. C. (2010). Eurythoe complanata (Polychaeta: Amphinomidae), the ‘cosmopolitan’ fireworm, consists of at least three cryptic species. *Marine Biology*, 157 (1), 69–80. <https://doi.org/10.1007/s00227-009-1296-9>
- Batthey, C. J., & Klicka, J. (2017). Cryptic speciation and gene flow in a migratory songbird species complex: Insights from the red-eyed vireo (*Vireo olivaceus*). *Molecular Phylogenetics and Evolution*, 113, 67–75. <https://doi.org/10.1016/j.ympev.2017.05.006>
- Bayona-Vasquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., Bentley, K. E., Finger, J. W., Louha, S., Troendle, N., Diaz-Jaimes, P., Mauricio, R., & Faircloth, B. C. (2019).

Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *BioRxiv* .
<https://doi.org/10.1101/205799>

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* , 22 (3), 148–155. <https://doi.org/10.1016/j.tree.2006.11.004>

Bivand, R., Keitt, T., & Rowlingson, B. (2019). *rgdal: Bindings for the “geospatial” data abstraction library* [R package version 1.4-4]. . <https://CRAN.R-project.org/package=rgdal>

Bond, J. E., Hedin, M. C., Ramirez, M. G., & Opell, B. D. (2001). Deep molecular divergence in the absence of morphological and ecological change in the Californian coastal dune endemic trapdoor spider *Aptostichus simus* . *Molecular Ecology* , 10 (4), 899–910. <https://doi.org/10.1046/j.1365-294X.2001.01233.x>

Bond, J. E., Hendrixson, B. E., Hamilton, C. A., & Hedin, M. (2012). A reconsideration of the classification of the spider Infraorder Mygalomorphae (Arachnida: Araneae) based on three nuclear genes and morphology. *PLoS ONE* , 7 (6), e38753. <https://doi.org/10.1371/journal.pone.0038753>

Bond, J. E., & Sierwald, P. (2002). Cryptic speciation in the *Anadenobolus excisus* millipede species complex on the island of Jamaica. *Evolution* , 56 (6), 1123–1135.

Bond, J. E., & Stockman, A. K. (2008). An integrative method for delimiting cohesion species: Finding the population-species interface in a group of Californian trapdoor spiders with extreme genetic divergence and geographic structuring. *Systematic Biology* , 57 (4), 628–646. <https://doi.org/10.1080/10635150802302443>

Burns, M., Starrett, J., Derkarabetian, S., Richart, C. H., Cabrero, A., & Hedin, M. (2017). Comparative performance of double-digest RAD sequencing across divergent arachnid lineages. *Molecular Ecology Resources* , 17 (3), 418–430. <https://doi.org/10.1111/1755-0998.12575>

Caterino, M. S., & Langton-Myers, S. S. (2019). Intraspecific diversity and phylogeography in Southern Appalachian *Dasycerus carolinensis* (Coleoptera: Staphylinidae: Dasycerinae). *Insect Systematics and Diversity* , 3 (6). <https://doi.org/10.1093/isd/ixz022>

Chan, K. O., Alexander, A. M., Grismer, L. L., Su, Y.-C., Grismer, J. L., Quah, E. S. H., & Brown, R. M. (2017). Species delimitation with gene flow: A methodological comparison and population genomics approach to elucidate cryptic species boundaries in Malaysian Torrent Frogs. *Molecular Ecology* , 26 (20), 5435–5450. <https://doi.org/10.1111/mec.14296>

Coyle, F. A. (1971). Systematics and natural history of the mygalomorph spider genus *Antrodiaetus* and related genera (Araneae: Antrodiaetidae). *Bull. Mus. Comp. Zool* , 141 , 269–402.

Coyle, F. A. (1983). Aerial dispersal by Mygalomorph spiderlings (Araneae, Mygalomorphae). *Journal of Arachnology* , 11 (2), 283–286.

Coyle, F. A. (1985). Ballooning behavior of *Ummidia* spiderlings. (Araneae: Ctenizidae). *Journal of Arachnology* , 13 , 137–138.

Crotti, M., Barratt, C. D., Loader, S. P., Gower, D. J., & Streicher, J. W. (2019). Causes and analytical impacts of missing data in RADseq phylogenetics: Insights from an African frog (*Afrivalus*). *Zoologica Scripta* , 48 (2), 157–167. <https://doi.org/10.1111/zsc.12335>

De Jesus-Bonilla, V. S., Meza-Lazaro, R. N., & Zaldivar-Riveron, A. (2019). 3RAD-based systematics of the transitional Nearctic-Neotropical lubber grasshopper genus *Taeniopoda* (Orthoptera: Romaleidae). *Molecular Phylogenetics and Evolution* , 137 , 64–75. <https://doi.org/10.1016/j.ympev.2019.04.019>

Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* , 24 (6), 332–340. <https://doi.org/10.1016/j.tree.2009.01.009>

- Delgado-Machuca, N., Meza-Lazaro, R. N., Romero-Napoles, J., Sarmiento-Monroy, C. E., Amarillo-Suarez, A. R., Bayona-Vasquez, N. J., & Alejandro, Z. (2019). Genetic structure, species limits and evolution of the parasitoid wasp genus *Stenocorse* (Braconidae: Doryctinae) based on nuclear 3RAD and mitochondrial data. *Systematic Entomology* . <https://onlinelibrary.wiley.com/doi/abs/10.1111/syen.12373>
- Derkarabetian, S., Castillo, S., Koo, P. K., Ovchinnikov, S., & Hedin, M. (2019). A demonstration of unsupervised machine learning in species delimitation. *Molecular Phylogenetics and Evolution* , *139* , 106562. <https://doi.org/10.1016/j.ympev.2019.106562>
- Derkarabetian, S., & Hedin, M. (2014). Integrative Taxonomy and Species Delimitation in Harvestmen: A Revision of the Western North American Genus *Sclerobunus* (Opiliones: Laniatores: Travunioidea). *PLoS ONE* , *9* (8), e104982. <https://doi.org/10.1371/journal.pone.0104982>
- Derkarabetian, S., Ledford, J., & Hedin, M. (2011). Genetic diversification without obvious genitalic morphological divergence in harvestmen (Opiliones, Laniatores, *Sclerobunus robustus*) from montane sky islands of western North America. *Molecular Phylogenetics and Evolution* , *61* (3), 844–853. <https://doi.org/10.1016/j.ympev.2011.08.004>
- Duperre, N., & Tapia, E. (2015). Discovery of the first telemid spider (Araneae, Telemidae) from South America, and the first member of the family bearing a stridulatory organ. *Zootaxa* , *4020* (1), 191. <https://doi.org/10.11646/zootaxa.4020.1.9>
- Eaton, D. A. R. (2014). PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* , *30* (13), 1844–1849. <https://doi.org/10.1093/bioinformatics/btu121>
- Eaton, D. A. R., Spriggs, E. L., Park, B., & Donoghue, M. J. (2016). Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology* , syw092. <https://doi.org/10.1093/sysbio/syw092>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* , *32* (5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences* , *107* (37), 16196–16200. <https://doi.org/10.1073/pnas.1006538107>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology* , *14* (8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- 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. *Systematic Biology* , *61* (5), 717–726. <https://doi.org/10.1093/sysbio/sys004>
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas: new climate surfaces for global land areas. *International Journal of Climatology* , *37* (12), 4302–4315. <https://doi.org/10.1002/joc.5086>
- Francis, R. M. (2017). pophelper: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources* , *17* (1), 27–32. <https://doi.org/10.1111/1755-0998.12509>
- Freudenstein, J. V., Broe, M. B., Folk, R. A., & Sinn, B. T. (2016). Biodiversity and the species concept—Lineages are not enough. *Systematic Biology* , syw098. <https://doi.org/10.1093/sysbio/syw098>
- Garg, K. M., Tizard, R., Ng, N. S. R., Cros, E., Dejtardol, A., Chattopadhyay, B., Pwint, N., Packert, M., & Rheindt, F. E. (2016). Genome-wide data help identify an avian species-level lineage that is morphologically and vocally cryptic. *Molecular Phylogenetics and Evolution* , *102* , 97–103. <https://doi.org/10.1016/j.ympev.2016.05.028>

- Garrick, R. C., Newton, K. E., & Worthington, R. J. (2018). Cryptic diversity in the southern Appalachian Mountains: Genetic data reveal that the red centipede, *Scolopocryptops sexspinosus*, is a species complex. *Journal of Insect Conservation* , 22 (5–6), 799–805. <https://doi.org/10.1007/s10841-018-0107-3>
- Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., Manzon, R. G., Martino, J. A., Pierson, T., Rogers, S. M., Wilson, J. Y., & Somers, C. M. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources* , 15 (6), 1304–1315. <https://doi.org/10.1111/1755-0998.12404>
- Hamilton, C. A., Hendrixson, B. E., Brewer, M. S., & Bond, J. E. (2014). An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: A case study of the North American tarantula genus *Aphonopelma* (Araneae, Mygalomorphae, Theraphosidae). *Molecular Phylogenetics and Evolution* , 71 , 79–93. <https://doi.org/10.1016/j.ympev.2013.11.007>
- Harvey, M. S., Hillyer, M. J., Main, B. Y., Moulds, T. A., Raven, R. J., Rix, M. G., Vink, C. J., & Huey, J. A. (2018). Phylogenetic relationships of the Australasian open-holed trapdoor spiders (Araneae: Mygalomorphae: Nemesiidae: Anaminae): multi-locus molecular analyses resolve the generic classification of a highly diverse fauna. *Zoological Journal of the Linnean Society* , 184 (2), 407–452. <https://doi.org/10.1093/zoolinnean/zlx111>
- Hedin, M., Carlson, D., & Coyle, F. A. (2015). Sky island diversification meets the multispecies coalescent—Divergence in the spruce-fir moss spider (*Microhexura montivaga*, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. *Molecular Ecology* , 24 (13), 3467–3484. <https://doi.org/10.1111/mec.13248>
- Hedin, M., Derkarabetian, S., Alfaro, A., Ramirez, M. J., & Bond, J. E. (2019). Phylogenomic analysis and revised classification of atypoid mygalomorph spiders (Araneae, Mygalomorphae), with notes on arachnid ultraconserved element loci. *PeerJ* , 7 , e6864. <https://doi.org/10.7717/peerj.6864>
- Hedin, M., & McCormack, M. (2017). Biogeographical evidence for common vicariance and rare dispersal in a southern Appalachian harvestman (Sabaconidae, *Sabacon cavicolens*). *Journal of Biogeography* , 44 (7), 1665–1678. <https://doi.org/10.1111/jbi.12973>
- Hedin, M., Starrett, J., & Hayashi, C. (2013). Crossing the uncrossable: Novel trans-valley biogeographic patterns revealed in the genetic history of low-dispersal mygalomorph spiders (Antrodiaetidae, *Antrodiaetus*) from California. *Molecular Ecology* , 22 (2), 508–526. <https://doi.org/10.1111/mec.12130>
- Hendrixson, B. E., & Bond, J. E. (2005a). Two sympatric species of *Antrodiaetus* from southwestern North Carolina (Araneae, Mygalomorphae, Antrodiaetidae). *Zootaxa* , 872 (1), 1. <https://doi.org/10.11646/zootaxa.872.1.1>
- Hendrixson, B. E., & Bond, J. E. (2005b). Testing species boundaries in the *Antrodiaetus unicolor* complex (Araneae: Mygalomorphae: Antrodiaetidae): “Paraphyly” and cryptic diversity. *Molecular Phylogenetics and Evolution* , 36 (2), 405–416. <https://doi.org/10.1016/j.ympev.2005.01.021>
- Hendrixson, B. E., & Bond, J. E. (2007). Molecular phylogeny and biogeography of an ancient Holarctic lineage of mygalomorph spiders (Araneae: Antrodiaetidae: *Antrodiaetus*). *Molecular Phylogenetics and Evolution* , 42 (3), 738–755. <https://doi.org/10.1016/j.ympev.2006.09.010>
- Hendrixson, B. E., DeRussy, B. M., Hamilton, C. A., & Bond, J. E. (2013). An exploration of species boundaries in turret-building tarantulas of the Mojave Desert (Araneae, Mygalomorphae, Theraphosidae, *Aphonopelma*). *Molecular Phylogenetics and Evolution* , 66 (1), 327–340. <https://doi.org/10.1016/j.ympev.2012.10.004>
- Hey, J. (2001). The mind of the species problem. *Trends in Ecology & Evolution* , 16 (7), 326–329.
- Hijmans, R. J. (2015). *raster: Geographic data analysis and modeling* [R package version 2.5-2]. <https://cran.r-project.org/package=raster>

- Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A., Faircloth, B. C., Mauricio, R., & Glenn, T. C. (2016). RADcap: Sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources* , 16 (5), 1264–1278. <https://doi.org/10.1111/1755-0998.12566>
- Holland, B. S., Dawson, M. N., Crow, G. L., & Hofmann, D. K. (2004). Global phylogeography of Cassiopea (Scyphozoa: Rhizostomeae): molecular evidence for cryptic species and multiple invasions of the Hawaiian Islands. *Marine Biology* , 145 (6), 1119–1128. <https://doi.org/10.1007/s00227-004-1409-4>
- Huang, H., & Knowles, L. L. (2016). Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. *Systematic Biology* , 65 (3), 357–365. <https://doi.org/10.1093/sysbio/syu046>
- Huey, J. A., Hillyer, M. J., & Harvey, M. S. (2019). Phylogenetic relationships and biogeographic history of the Australian trapdoor spider genus Conothele (Araneae: Mygalomorphae: Halonoprocridae): diversification into arid habitats in an otherwise tropical radiation. *Invertebrate Systematics* . <https://doi.org/10.1071/IS18078>
- Jezkova, T., Olah-Hemmings, V., & Riddle, B. R. (2011). Niche shifting in response to warming climate after the last glacial maximum: Inference from genetic data and niche assessments in the chisel-toothed kangaroo rat (*Dipodomys microps*). *Global Change Biology* , 17 (11), 3486–3502. <https://doi.org/10.1111/j.1365-2486.2011.02508.x>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* , 24 (11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* , 27 (21), 3070–3071. <https://doi.org/10.1093/bioinformatics/btr521>
- Keith, R., & Hedin, M. (2012). Extreme mitochondrial population subdivision in southern Appalachian paleoendemic spiders (Araneae: Hypochilidae: *Hypochilus*), with implications for species delimitation. *Journal of Arachnology* , 40 (2), 167–181.
- Kozak, K. H., & Wiens, J. J. (2010). Niche conservatism drives elevational diversity patterns in Appalachian salamanders. *The American Naturalist* , 176 (1), 40–54. <https://doi.org/10.1086/653031>
- Leache, A. D., Chavez, A. S., Jones, L. N., Grummer, J. A., Gottscho, A. D., & Linkem, C. W. (2015). Phylogenomics of Phrynosomatid lizards: Conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution* , 7 (3), 706–719. <https://doi.org/10.1093/gbe/evv026>
- Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* , 61 (5), 727–744. <https://doi.org/10.1093/sysbio/sys049>
- Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology* , 46 (3), 523–536.
- Massatti, R., & Knowles, L. L. (2014). Microhabitat differences impact phylogeographic concordance of codistributed species: Genomic evidence in montane sedges (*Carex*) from the Rocky Mountains. *Evolution* , 68 (10), 2833–2846. <https://doi.org/10.1111/evo.12491>
- Moritz, C., Fujita, M. K., Rosauer, D., Agudo, R., Bourke, G., Doughty, P., Palmer, R., Pepper, M., Potter, S., Pratt, R., Scott, M., Tonione, M., & Donnellan, S. (2016). Multilocus phylogeography reveals nested endemism in a gecko across the monsoonal tropics of Australia. *Molecular Ecology* , 25 (6), 1354–1366. <https://doi.org/10.1111/mec.13511>
- Opatova, V., & Arnedo, M. A. (2014a). From Gondwana to Europe: Inferring the origins of Mediterranean Macrothele spiders (Araneae : Hexathelidae) and the limits of the family Hexathelidae. *Invertebrate Systematics* , 28 (4), 361. <https://doi.org/10.1071/IS14004>
- Opatova, V., & Arnedo, M. A. (2014b). Spiders on a Hot Volcanic Roof: Colonisation Pathways and

Phylogeography of the Canary Islands Endemic Trap-Door Spider *Titanidiops canariensis* (Araneae, Idiopidae). *PLoS ONE* , 9 (12), e115078. <https://doi.org/10.1371/journal.pone.0115078>

Opatova, V., Hamilton, C. A., Hedin, M., De Oca, L. M., Kral, J., & Bond, J. E. (2019). Phylogenetic systematics and evolution of the spider infraorder Mygalomorphae using genomic scale data. *Systematic Biology* , *syz064* . <https://doi.org/10.1093/sysbio/syz064>

Ortega-Andrade, H. M., Rojas-Soto, O. R., Valencia, J. H., Espinosa de los Monteros, A., Morrone, J. J., Ron, S. R., & Cannatella, D. C. (2015). Insights from Integrative Systematics Reveal Cryptic Diversity in Pristimantis Frogs (Anura: Craugastoridae) from the Upper Amazon Basin. *PLOS ONE* , 10 (11), e0143392. <https://doi.org/10.1371/journal.pone.0143392>

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* , 7 (5), e37135. <https://doi.org/10.1371/journal.pone.0037135>

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* , 190 (3–4), 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure Using multilocus genotype data. *Genetics* , 155 (2), 949–959.

R Core Team. (2019). *R: A language and environment for statistical computing* . R Foundation for Statistical Computing. <https://www.r-project.org/>

Ramirez-Reyes, T., Pinero, D., Flores-Villela, O., & Vazquez-Dominguez, E. (2017). Molecular systematics, species delimitation and diversification patterns of the *Phyllodactylus lanei* complex (Gekkota: Phyllodactylidae) in Mexico. *Molecular Phylogenetics and Evolution* , 115 , 82–94. <https://doi.org/10.1016/j.ympev.2017.07.008>

Reynolds, R. G., Niemiller, M. L., & Fitzpatrick, B. M. (2012). Genetic analysis of an endemic archipelagic lizard reveals sympatric cryptic lineages and taxonomic discordance. *Conservation Genetics* , 13 (4), 953–963. <https://doi.org/10.1007/s10592-012-0344-z>

Richardson, B. J. (2016). New genera, new species and redescriptions of Australian jumping spiders (Araneae: Salticidae). *Zootaxa* , 4114 (5), 501. <https://doi.org/10.11646/zootaxa.4114.5.1>

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahe, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ* , 4 , e2584. <https://doi.org/10.7717/peerj.2584>

Satler, J. D., Starrett, J., Hayashi, C. Y., & Hedin, M. (2011). Inferring species trees from gene trees in a radiation of California trapdoor spiders (Araneae, Antrodiaetidae, *Aliatypus*). *PLoS ONE* , 6 (9), e25355. <https://doi.org/10.1371/journal.pone.0025355>

Schoener, T. W. (1968). The *Anolis* lizards of Bimini: Resource partitioning in a complex fauna. *Ecology* , 49 (4), 704–726. <https://doi.org/10.2307/1935534>

Soberon, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters* , 10 (12), 1115–1123. <https://doi.org/10.1111/j.1461-0248.2007.01107.x>

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* , 30 (9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>

Starrett, J., Hayashi, C. Y., Derkarabetian, S., & Hedin, M. (2018). Cryptic elevational zonation in trapdoor spiders (Araneae, Antrodiaetidae, *Aliatypus janus* complex) from the California southern Sierra Nevada. *Molecular Phylogenetics and Evolution* , 118 , 403–413. <https://doi.org/10.1016/j.ympev.2017.09.003>

Starrett, J., & Hedin, M. (2007). Multilocus genealogies reveal multiple cryptic species and biogeographical complexity in the California turret spider *Antrodiaetus riversi* (Mygalomorphae, Antrodiaetidae): Cryptic di-

versifocation in Californian *A. riversi*. *Molecular Ecology*, 16 (3), 583–604. <https://doi.org/10.1111/j.1365-294X.2006.03164.x>

Stockman, A. K., & Bond, J. E. (2007). Delimiting cohesion species: Extreme population structuring and the role of ecological interchangeability. *Molecular Ecology*, 16 (16), 3374–3392. <https://doi.org/10.1111/j.1365-294X.2007.03389.x>

Sukumaran, J., & Knowles, L. L. (2017). Multispecies coalescent delimits structure, not species. *Proceedings of the National Academy of Sciences*, 114 (7), 1607–1612. <https://doi.org/10.1073/pnas.1607921114>

Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240 (4857), 1285–1293. <https://doi.org/10.1126/science.3287615>

Templeton, A. R. (1989). The meaning of species and speciation: A genetic perspective. In *Speciation and its Consequences* (pp. 3–27). Sinauer.

Thomas, S. M., & Hedin, M. (2008). Multigenic phylogeographic divergence in the paleoendemic southern Appalachian opilionid *Fumontana deprehendor* Shear (Opiliones, Laniatores, Triaenonychidae). *Molecular Phylogenetics and Evolution*, 46 (2), 645–658. <https://doi.org/10.1016/j.ympev.2007.10.013>

Wake, D. B. (2006). Problems with species: Patterns and processes of species formation in salamanders. *Annals of the Missouri Botanical Garden*, 93 (1), 8–23.

Walker, M. J., Stockman, A. K., Marek, P. E., & Bond, J. E. (2009). Pleistocene glacial refugia across the Appalachian Mountains and coastal plain in the millipede genus *Narceus*: Evidence from population genetic, phylogeographic, and paleoclimatic data. *BMC Evolutionary Biology*, 9 (1), 25. <https://doi.org/10.1186/1471-2148-9-25>

Warren, D. L., Glor, R. E., & Turelli, M. (2008). Environmental niche equivalency versus conservatism: Quantitative approaches to niche evolution. *Evolution*, 62 (11), 2868–2883. <https://doi.org/10.1111/j.1558-5646.2008.00482.x>

Warren, D. L., Glor, R. E., & Turelli, M. (2010). ENMTools: A toolbox for comparative studies of environmental niche models. *Ecography*. <https://doi.org/10.1111/j.1600-0587.2009.06142.x>

Weisrock, D. W., & Larson, A. (2006). Testing hypotheses of speciation in the *Plethodon jordani* species complex with allozymes and mitochondrial DNA sequences: Diagnosing plethodontid salamander species. *Biological Journal of the Linnean Society*, 89 (1), 25–51. <https://doi.org/10.1111/j.1095-8312.2006.00655.x>

Wiens, J. J., & Graham, C. H. (2005). Niche Conservatism: Integrating Evolution, Ecology, and Conservation Biology. *Annual Review of Ecology, Evolution, and Systematics*, 36 (1), 519–539. <https://doi.org/10.1146/annurev.ecolsys.36.102803.095431>

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19 (S6). <https://doi.org/10.1186/s12859-018-2129-y>

Zhang, Y., & Li, S. (2014). A spider species complex revealed high cryptic diversity in South China caves. *Molecular Phylogenetics and Evolution*, 79, 353–358. <https://doi.org/10.1016/j.ympev.2014.05.017>

Tables

Table 1. Summary of the data obtained in matrices with an 85% clustering threshold. Var, number of variable sites; Pis, number of parsimony-informative sites.

Matrix Individuals	Matrix Names	Locus occupancy percentage	Minimum sample per locus parameter	Total # of loci
All individuals	All_M10	10%	16	9561

Matrix Individuals	Matrix Names	Locus occupancy percentage	Minimum sample per locus parameter	Total # of loci
A. unicolorB-only	All_M20	20%	33	1719
	All_M30	30%	49	574
	All_M40	40%	66	290
	All_M50	50%	82	189
	All_M60	60%	99	145
	All_M70	70%	115	100
	All_M80	80%	132	56
	UniB_M10	10%	12	15651
	UniB_M20	20%	24	5384
	UniB_M30	30%	36	2408
	UniB_M40	40%	48	1277
	UniB_M50	50%	60	781
	UniB_M60	60%	72	512
	UniB_M70	70%	84	325
	UniB_M80	80%	96	202
	UniB_M90	90%	108	41

Table 2. Summary of the genetic exchangeability (GE) criterion for cohesion species delimitation between populations of the *A. unicolor* species complex.

Lineage Comparison	Geographic Barrier	STRUCTURE	PCA	VAE
A. unicolorA to <i>A. microunicolor</i>	Allopatric	Separate clusters	Separate clusters	Separate clusters
A. unicolorA+A. <i>microunicolor</i> to A. unicolorB	None/parapatric	Separate clusters	Separate clusters	Separate clusters
A. unicolorB1 to A. unicolorB2	None/parapatric	admixture	Cluster overlap	Cluster overlap
A. unicolorB1+A. unicolorB2 to A. unicolorB3	None/parapatric	admixture	Cluster overlap	Cluster overlap

Table 3. Summary of the ecological interchangeability (EI) criterion for cohesion species delimitation between populations of the *A. unicolor* species complex. We evaluated a proxy for ecological interchangeability by measuring the overlap between ENMs for the sister lineages comparisons as well as taking into account previous morphological and behavioral criteria for distinguishing *A. microunicolor*. N_a and N_b values are the number of occurrence records for the first and second lineages used in a comparison, respectively. Niche identity refers to the "niche equivalency test", and niche similarity refers to the "niche similarity (i.e. background) test" proposed by Warren et al. (2008) (see Methods for details). ^a ENMs are no more similar or different than expected by chance; ^b ENMs are more similar than expected by chance for one comparison, yet ENMs are more different than expected by chance for the reciprocal comparison; see Discussion.

Lineage Comparison	Morphology/Behavior Differences	N_a , N_b	Niche C
A. unicolorA to <i>A. microunicolor</i>	Size, setal characters, non-overlapping mating season	6, 20	0.2964,
A. unicolorA+A. <i>microunicolor</i> to A. unicolorB	No obvious differences	26, 82	0.5899,
A. unicolorB1 to A. unicolorB2	No obvious differences	29, 37	0.2041,
A. unicolorB1+A. unicolorB2 to A. unicolorB3	No obvious differences	66, 16	0.7566,

Figures

Figure 1. Geographic distributions of each *A. unicolor* lineage as defined in the Methods. See the legend denoting color and symbol for each lineage.

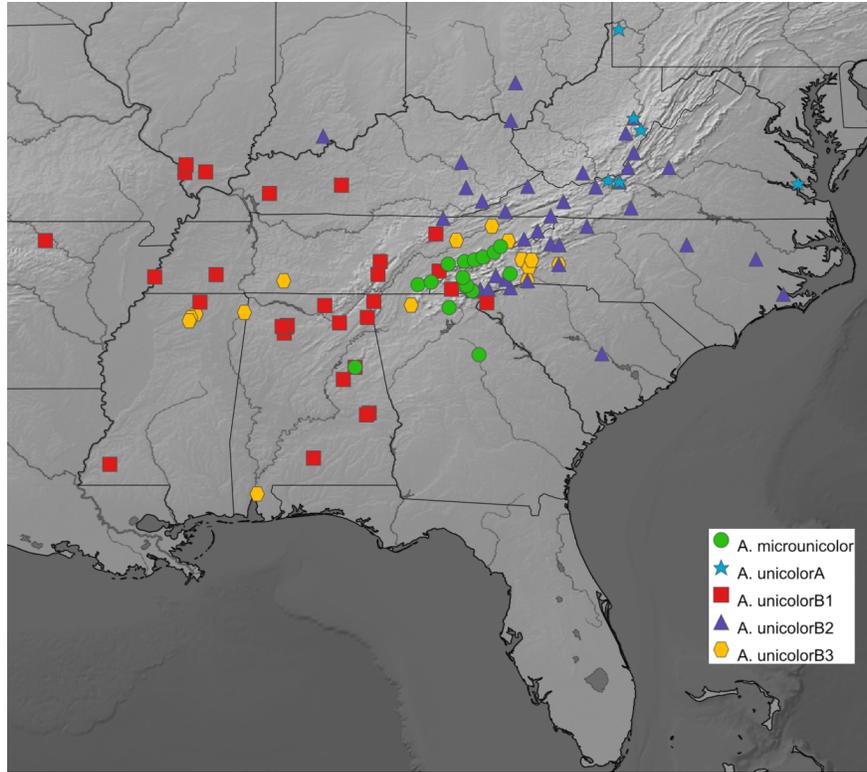


Figure 2. All_M20 RAxML phylogeny with midpoint rooting. Nodes with support < 90 are denoted with black dots. Each lineage has been labeled and highlighted with color scheme from Figure 1. Bottom left: male (top spider) and female (bottom spider) *Antrodiaetus unicolor* specimens.

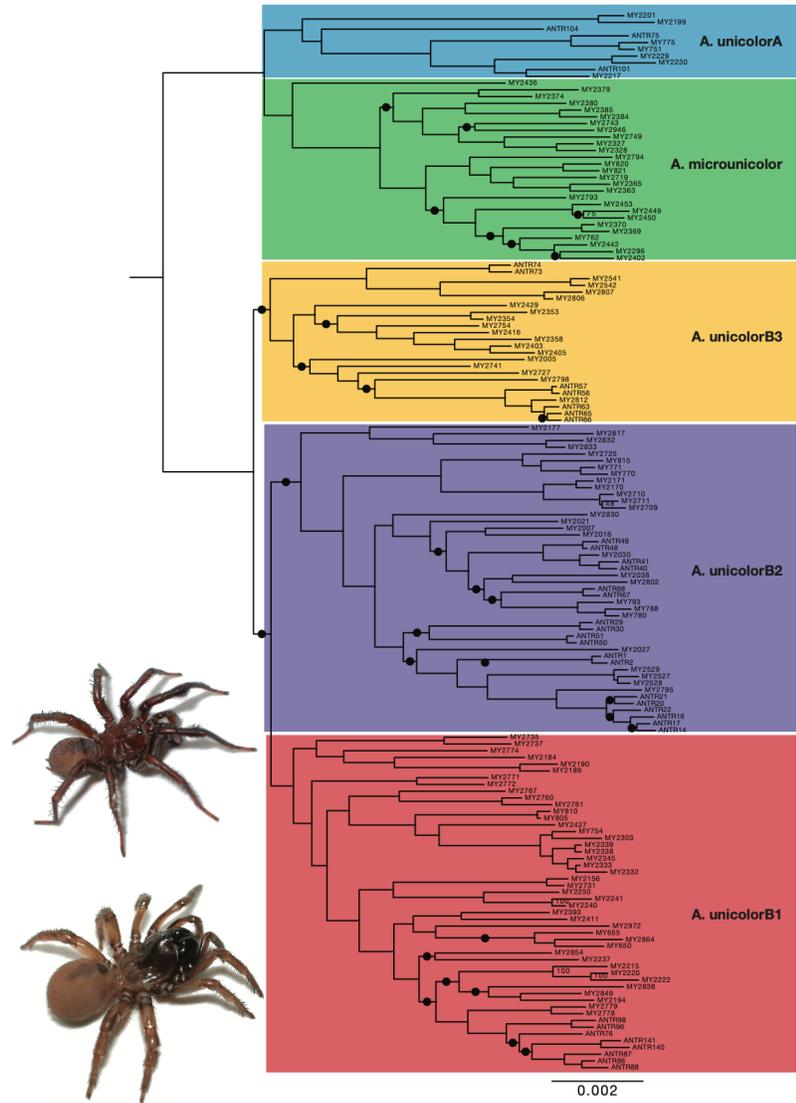


Figure 3. Cluster analysis plots of the All_M20 dataset with the same color scheme for each lineage as previous figures unless otherwise stated. a) STRUCTURE plot with an optimal K value of 8 ($?K=8$), showing a large amount of admixture. *A. microunicolor* and *A. unicolorA* retain their color from previous figures while *A. unicolorB* does not due to additional clusters (see key for lineage coloration). b) PCA plot, showing significant cluster overlap of *A. unicolorB* lineages. c) VAE plot, displaying no clear distinctions among *A. unicolorB* lineages and d) VAE results with encoded mean (μ) and standard deviation (σ) for each sample.

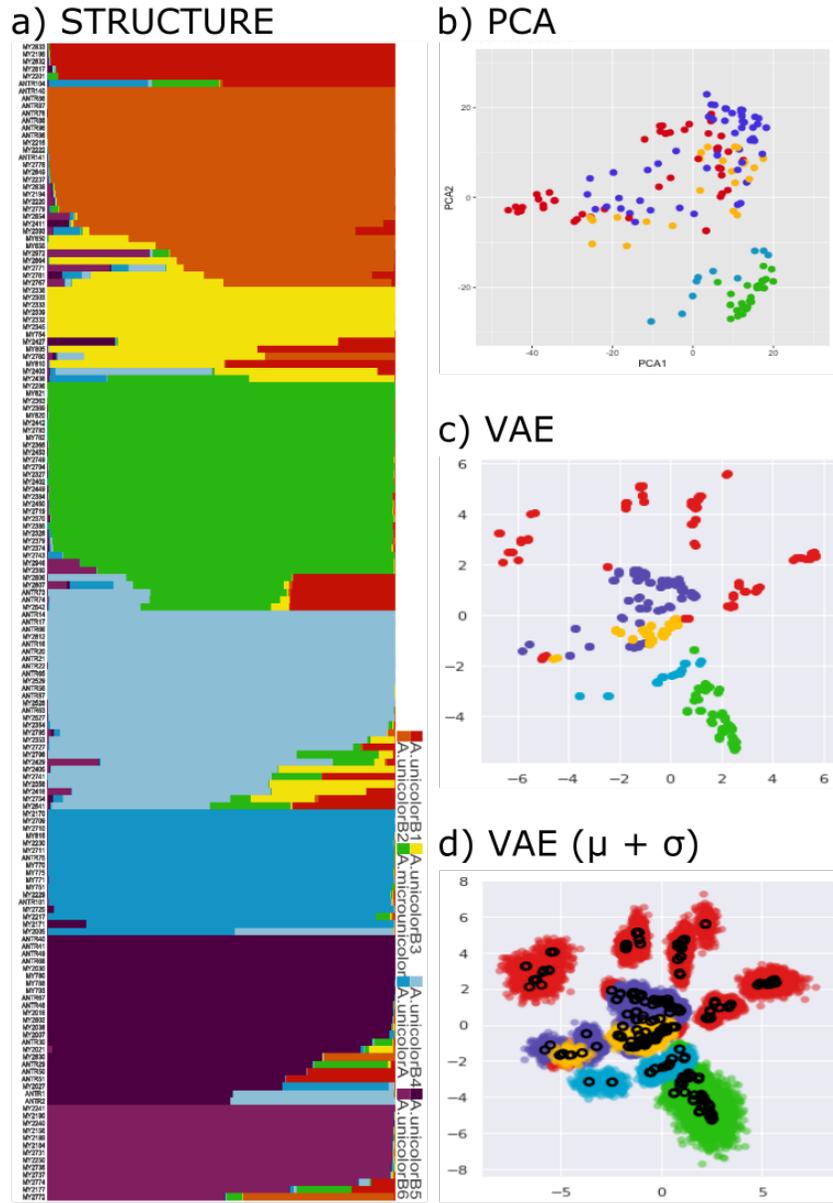


Figure 4. STRUCTURE and VAE analyses using the datasets that include only *A. unicolorB* individuals with varying locus occupancies (UniB_M30, UniB_M40, UniB_M50, UniB_M60, and UniB_M70; see Table 1 for dataset details). a) 30% locus occupancy. b) 40% locus occupancy. c) 50% locus occupancy. d) 60% locus occupancy. e) 70% locus occupancy.

