Cryptic diversity in the Mexican highlands: Thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: *Crotalus*)

Christopher Blair\(^1,2\) | Robert W. Bryson Jr\(^3,4\) | Charles W. Linkem\(^3\) | David Lazcano\(^5\) | John Klicka\(^3\) | John E. McCormack\(^4\)

\(^1\)Department of Biological Sciences, New York City College of Technology, The City University of New York, Brooklyn, New York
\(^2\)Biology PhD Program, CUNY Graduate Center, New York, New York
\(^3\)Department of Biology and Burke Museum of Natural History and Culture, University of Washington, Seattle, Washington
\(^4\)Moore Laboratory of Zoology, Occidental College, Los Angeles, California
\(^5\)Laboratorio de Herpetología, Facultad de Ciencias Biológicas, Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Nuevo León, México

**Abstract**

With the continued adoption of genome-scale data in evolutionary biology comes the challenge of adequately harnessing the information to make accurate phylogenetic inferences. Coalescent-based methods of species tree inference have become common, and concatenation has been shown in simulation to perform well, particularly when levels of incomplete lineage sorting are low. However, simulation conditions are often overly simplistic, leaving empiricists with uncertainty regarding analytical tools. We use a large ultraconserved element data set (>3,000 loci) from rattlesnakes of the *Crotalus triseriatus* group to delimit lineages and estimate species trees using concatenation and several coalescent-based methods. Unpartitioned and partitioned maximum likelihood and Bayesian analysis of the concatenated matrix yield a topology identical to coalescent analysis of a subset of the data in BPP. ASTRAL analysis on a subset of the more variable loci also results in a tree consistent with concatenation and BPP, whereas the SVDQUARTETS phylogeny differs at additional nodes. The size of the concatenated matrix has a strong effect on species tree inference using SVDQUARTETS, warranting additional investigation on optimal data characteristics for this method. Species delimitation analyses suggest up to 16 unique lineages may be present within the *C. triseriatus* group, with divergences occurring during the Neogene and Quaternary. Network analyses suggest hybridization within the group is relatively rare. Altogether, our results reaffirm the Mexican highlands as a biodiversity hotspot and suggest that coalescent-based species tree inference on data subsets can provide a strongly supported species tree consistent with concatenation of all loci with a large amount of missing data.

**KEYWORDS**

Bayesian, coalescence, genomics, rattlesnakes, speciation, systematics
1 | INTRODUCTION

Phylogenetics and phylogeography continue to experience a progressive transformation with the evolving power of next-generation DNA sequencing (NGS) technologies and sophisticated new analytical models (Edwards, 2009, Edwards et al., 2016; Liu, Yu, Kubatko, Pearl, & Edwards, 2009; McCormack, Hird, Zeller, Carstens, & Brumfield, 2013; McCormack & Faircloth, 2013). Although phylogenetic studies have been conducted with whole genomes (Jarvis et al., 2014), reduced-representation approaches like restriction enzyme-based methods (e.g., RADseq, ddRADseq, GBS) and DNA sequence capture have emerged as particularly flexible methods for interrogating large eukaryotic genomes for hundreds to thousands of orthologous loci suitable for phylogenetics (Baird et al., 2008; Elshire et al., 2011; Faircloth et al., 2012; Lemmon, Emme, & Lemmon, 2012; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). The large number of loci provided by these methods can help resolve difficult nodes in species trees, particularly those involved in rapid evolutionary radiations (Giarla & Esselstyn, 2015; Leaché et al., 2015; Leaché, Banbury, Linkem, & Oca, 2016; Meiklejohn, Faircloth, Glenn, Kimball, & Braun, 2016; Wagner et al., 2013). A specific class of molecular marker used with sequence capture, ultraconserved elements (UCEs; Bejerano et al., 2004; Faircloth et al., 2012), has helped resolve longstanding evolutionary questions across taxonomic groups at multiple spatial and temporal scales (Crawford et al., 2012; Faircloth, Sorenson, Santini, & Alfaro, 2013; McCormack et al., 2012; McCormack, Harvey, Faircloth, Crawford, Glenn, & Brumfield, 2013; Leaché et al., 2016; Smith, Harvey, Faircloth, Glenn, & Brumfield, 2013; Streicher, Schulte, & Wiens, 2016). In UCEs, the conserved core regions facilitate the collection of orthologous loci among diverse organisms, whereas the flanking regions provide the sequence variability necessary for phylogenetics (Faircloth et al., 2012).

A major challenge of working with UCE and other types of large data sets is deciding on the best approaches to analyse these large, often heterogeneous data sets. Although traditional concatenation is relatively straightforward and computationally feasible, it can be statistically inconsistent in the presence of incomplete lineage sorting (Chou et al., 2015; Kubatko & Degnan, 2007; Mirarab, Bayzid, & Warnow, 2016; Roch & Steel, 2014). Meanwhile, fully probabilistic multispecies coalescent methods such as STABBEAST2 (Ogilvie, Bouckaert, & Drummond, 2017) are attractive as they model stochastic lineage sorting and simultaneously estimate gene trees, the species tree, divergence times and substitution rates. However, likelihood-based multispecies coalescent methods are not easily scalable to large NGS data sets. To address this, gene tree reconciliation methods estimate species trees relatively quickly and directly from gene trees, instead of sequence data (Liu et al., 2009; Liu, Yu, & Edwards, 2010; Vachaspati & Warnow, 2015; Zhang, Sayyari, & Mirarab, 2017). However, poorly resolved gene trees can have a large impact on the species tree (Meiklejohn et al., 2016; Xi, Liu, & Davis, 2015; Xu & Yang, 2016). A relatively new coalescent method, SVDQUARTETS, bypasses the estimation of gene trees and uses singular value decomposition scores to estimate relationships among quartets prior to assembling quartets into a full species tree containing all taxa (Chifman & Kubatko, 2014, 2015). The method appears statistically consistent under the multispecies coalescent, is scalable to NGS data sets and performs well under simulations (Chou et al., 2015), but it has not been well tested with empirical data sets.

Rattlesnakes (Crotalus and Sistrurus) represent a radiation of around 45 pitvipers distributed throughout the New World from Canada to Argentina (Blair & Sánchez-Ramírez, 2016; Campbell & Lamar, 2004; Gloyd, 1940; Klauber, 1956). There continues to be broad interest in elucidating the evolutionary history of these enigmatic snakes due, in part, to their medicinal importance and contributions to yearly envenomations (Campbell & Lamar, 2004). Despite extensive effort over many decades, however, relationships among key groups remain unresolved. Early studies using morphology and protein data conflict with those using mtDNA sequences, which in turn often conflict with phylogenies generated from nuclear genes (Blair & Sánchez-Ramírez, 2016; Brattstrom, 1964; Bryson, Murphy, Lathrop, & Lazzano-Villareal, 2011; Foote & MacMahon, 1977; Gloyd, 1940; Klauber, 1956; Murphy, Fu, Lathrop, Feltham, & Kovac, 2002; Reyes-Velasco, Melk, Smith, & Castoe, 2013). Part of the discrepancy seems to result from the fact that many molecular studies rely solely on mtDNA (Bryson et al., 2011; Castoe & Parkinson, 2006; Murphy et al., 2002; Parkinson, 1999; Parkinson, Campbell, Chippindale, & Schuett, 2002), which often conflict with phylogenies from other sources of data (Grummer, Bryson, & Reeder, 2014; Leaché & McGuire, 2006). An additional difficulty in resolving the rattlesnake tree is that many groups rapidly diversified during the late Miocene during periods of global climate change (Zachos, Pagani, Sloan, Thomas, & Billups, 2001) and mountain formation in Mexico (Blair & Sánchez-Ramírez, 2016; Bryson et al., 2014, 2011). Rapid radiations increase the likelihood of gene tree species tree conflicts, particularly in species with large effective population sizes (Maddison, 1997). Thus, genome-level data sets coupled with multispecies coalescent methods are often needed to provide both additional informative variation and the necessary analytical tools to resolve difficult nodes.

Rattlesnakes of the Crotalus triseriatus species group inhabit mountainous regions in Mexico and the south-western United States (Campbell & Lamar, 2004). Previous mtDNA work on the group indicates multiple instances of paraphyly and the presence of cryptic species (Bryson et al., 2011). A subsequent study based on multilocus and morphological data described two new species and elevated one subspecies to species status (Bryson et al., 2014). Nine species are now recognized in the group, although phylogenetic relationships among species remain uncertain, and it is likely that additional species-level diversity awaits discovery. Previous mtDNA evidence suggests that the C. triseriatus group rapidly diversified during formation of the Trans-Mexican Volcanic Belt (Bryson et al., 2011), and UCE data and multispecies coalescent models may provide additional insight to better resolve the divergence history of the clade.

The first objective of our study is to use a large UCE data set to estimate phylogenetic relationships and species limits within the C. triseriatus group. Our second goal is to estimate divergence times...
and compare results to previous rattlesnake studies based on fewer loci (Blair & Sánchez-Ramírez, 2016; Bryson et al., 2011). Finally, we compare traditional concatenation and coalescent methods, including the newer SVDQUARTETS, using a large empirical data set. We hope that these comparisons facilitate additional discussion on best practices for extracting meaningful phylogenetic information from genomic data sets.

2 | MATERIALS AND METHODS

2.1 | Sampling and library preparation

We sampled 54 rattlesnakes of the C. triseriatus group from throughout the Mexican highlands (Figure 1; Supplementary Table S1). We included multiple exemplars of all currently described taxa to assess both species limits and phylogenetic relationships. Taxa in the group include C. aequosus, C. armstrongi, C. campbelli, C. lepidus (C. I. lepidus, C. I. klauberi and C. I. maculosus), C. morulus, C. pusillus, C. ravis (C. r. ravis, C. r. bruneus and C. r. exiguus), C. talarici and C. triseriatus (Bryson et al., 2014, 2011). We included Agkistrodon piscivorus as an outgroup taxon to root tree topologies (Alencar et al., 2016). Handling of animals followed animal use protocols approved by the University of Nevada at Las Vegas Animal Care Committee (R701-1105-203) and the University of Washington Institutional Animal Care and Use Committee (4209-01).

Genomic DNA was extracted from tissue samples using the Qia-gen DNeasy Blood & Tissue Kit (Qiagen Inc., Valencia, CA, USA). Ali-quotsof extracts were shipped to RAPID Genomics (Gainsville, FL, USA) for UCE library preparation and sequencing following Bryson et al. (2017). Briefly, 5,472 custom-designed probes (MYbaits; MYcroarray, Inc., Ann Arbor, MI, USA) were used to enrich 5,060 UCE loci using protocols specified in Faircloth et al. (2012). Libraries were sequenced on an Illumina HiSeq 3,000 (100 paired-end) at the University of Florida ICBR Facility. Data quality filtering and assembly were performed using the phyLUCE pipeline (Faircloth, 2016) following Bryson et al. (2017). A locus was retained in the final assembly if it was represented by >75% of taxa. Alignments were deposited in Dryad (https://doi.org/10.5061/dryad.4467407).

2.2 | Concatenated phylogenetic analysis

All phylogenetic analyses were performed using the high-performance computer cluster at the Center for Theoretical Physics at New York City College of Technology (CUNY). We performed both maximum likelihood (ML) and Bayesian phylogenetic analyses on the concatenated (75% complete) matrix. Unpartitioned ML analyses were performed using RAxML v. 8.2.10 (Stamatakis, 2014) under a GTR GAMMA substitution model. We used the -f a option to calculate bootstrap support values and infer the ML tree in a single analysis. Nodal support was assessed through 100 rapid bootstrap replicates. All trees were rooted using A. piscivorus.

Unpartitioned Bayesian analyses were performed using ExaBayes v. 1.5 (Aberer, Kober, & Stamatakis, 2014). ExaBayes is a MCMC package specifically geared towards Bayesian inference of large phylogenetic data sets, providing multiple analytical solutions for high-level parallelism. All analyses utilized a general time reversible model with gamma distributed rate heterogeneity (GTR GAMMA). To make sure that chains were not getting stuck in local optima, we imple-mented two independent runs (-R 2) with Metropolis-Coupling (three heated and one cold chain per run [-C 4]) in parallel to better sample parameter space. Sampling proposal weights included the following: likeSpr = 4, parsimonySPR = 8, stNNI = 4, eSPR = 4, biDistGamma = 7, branchMulti = 2. All analyses were implemented on 64 cores for at least 1 million generations (sampling every 500) using default pri-ors for all parameters. Mixing and stationarity were monitored using TRACER v. 1.6.0 (Rambaut & Drummond, 2007) with a target effective sample size (ESS) value of 200 for all parameters. The average standard deviation of split frequencies (ASDSF) was also used as a measure of convergence using the default threshold value of 5%. Extended majority-rule (MRE) consensus trees were generated through the consense script. The default burn-in of 25% was used for all post-run analysis. Unrooted topologies were rooted using A. piscivorus.

The optimal partitioning strategy for all partitioned analyses was determined using PARTITIONFINDER v. 2.1.1 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2017). Because of the large size of the data set, we used the relaxed hierarchical clustering algorithm to assign sites to different partitions (Lanfear, Calcott, Kainer, Mayer, & Stamatakis, 2014). This clustering algorithm has been shown to be nearly as accurate as the greedy algorithm implemented in PARTITIONFINDER, but allows for the scaling of up to thousands of loci. Although the strict hierarchical clustering algorithm can further improve computational performance, the relatively high error rates preclude its utility in large phylogenomic data sets (Lanfear et al., 2014). We limited substitution models to GTR + G as this is the standard model implemented for nucleotide data in RAxML and currently the only nucleotide model in ExaBayes. Initial data blocks were assigned by locus from the output provided by the phyLUCE pipeline. Default weights (i.e., 1.0,0.0) were used to calculate similarity among partitions (subsets), which were based solely on differences in substitution rate among sets. We spec-ified an rcluster-max of 1,000 and rcluster-percent of 10. AICc was used for all partitioning scheme selection and the --raxml option was executed for analysis (Stamatakis, 2014). We used the ML tree from the unpartitioned RAxML analysis as the starting tree for PARTI-TIONFINDER. The optimal partitioning scheme was then utilized in partioned RAxML and ExaBayes analyses using the same parameter settings as the unpartitioned analyses.

2.3 | Coalescent-based species delimitation and species tree analysis

We first used SVDQUARTETS (Chifman & Kubatko, 2014, 2015) in PAUP* v. 4.0a152 (Swofford, 2001) for coalescent-based phylog-emonic analysis. SVDQUARTETS uses site pattern frequencies in SNP or multilocus sequence data to infer singular value decomposition (SVD) scores among alternative unrooted quartet trees (lower score
A tree amalgamation step is subsequently used to assemble a full tree containing all taxa. We first ran an analysis utilizing every sequence as a terminal (i.e., no species assignments) to estimate a lineage tree statistically consistent with the multispecies coalescent. We exhaustively sampled all 341,055 quartets. Following the calculation of SVD scores for alternative quartets, Quartet FM (QFM; Reaz, Bayzid, & Rahman, 2014) was used in PAUP* to assemble the full tree. Support for nodes was assessed using nonparametric bootstrapping with 100 replicates. Trees were rooted with *A. piscivorus*.

We used *BPP* v. 3.3a (Yang & Rannala, 2010) for joint species delimitation and species tree inference (i.e., unguided species delimitation; Yang & Rannala, 2014). *BPP* utilizes the multispecies coalescent model in a fully probabilistic (Bayesian) framework for parameter estimation, explicitly accounting for gene tree/species tree discordance due to incomplete lineage sorting. Due to the relatively high computational burden of Bayesian methods like *BPP*, we performed inference on reduced data sets consisting of a subset of 100 of the most variable loci. Geneious v. 9.1.7 (Kearse et al., 2012) was used to select loci based on the percentage of variable sites. Assignment of individuals to putative species followed previous studies (Bryson et al., 2014, 2011) and the results of the concatenated phylogenetic analyses. Two samples, Ct143CikJAL and Ct12CaaAGS, were placed in strongly supported yet conflicting relationships in concatenated and coalescent analyses (see below) in a pattern suggestive of historical introgression. Both were subsequently excluded from the analysis to meet methodological assumptions of the method (i.e., no hybridization among species; Yang & Rannala, 2010, 2014), leaving a total of 17 groupings and 53 individuals. *Crotalus lepidus* was divided into six populations consistent with geography and our phylogenetic analyses: *C. l. lepidus*, *C. l. maculosus*, *C. l. klauberi*-3 Northwest (samples from the north-eastern Sierra Madre Occidental and adjacent sky islands), *C. l. klauberi*-4 Northeast (samples from the dry slopes of the north-eastern Sierra Madre Occidental and adjacent uplands), *C. l. klauberi*-2 Durango (samples from the Sierra Madre Occidental of Durango) and *C. l. klauberi*-1 South (samples from the Sierra Madre Occidental south of the Rio Mezquital drainage). We specified gamma priors of $G(2, 1,000)$ for both population sizes ($\theta$) and the divergence time of the root ($\tau$). The locusrate parameter was used to account for rate variation among loci using $D(\lambda) = 10$. Uniform rooted trees were used as the species model prior. Following a burn-in of 10,000 generations, chains were sampled every five generations for a total of 20,000 samples. Multiple analyses were run using different species delimitation algorithms (Algorithms 0 and 1) to check for indications of convergence. We also tested the influence of the prior on $\theta_s$ and $\tau_0$ as previous research suggests that these priors may influence species delimitation results.

**FIGURE 1** Map of all taxa used in this study. Species are colour-coded based on previous taxonomy (Bryson et al., 2014, 2011); distributions adapted from Campbell and Lamar (2004). Two samples indicated with stars were identified as probable hybrids in our study.
We performed a coalescent catenated matrix of 3,384 UCE loci in SVD QUARTETS. Individuals were used to calculate support values.

We also used bpp for species tree inference (Analysis 01) under the multispecies coalescent model using the 500 loci data set. Individuals were partitioned into 17 “species” following the results of the species delimitation analyses (see Results). Three independent MCMC runs were implemented with a burn-in of 10,000, a sample frequency of 5, and 20,000 total samples. We specified gamma priors of G(2, 1,000) for both population sizes (θs) and the divergence time of the root (t0). The species model prior assumed uniform rooted trees.

Based on the species delimitation results obtained from bpp, we performed a coalescent-based species tree analysis on the full concatenated matrix of 3,384 UCE loci in SVDS QuarTets. Individuals (n = 53) were assigned to 17 species, and Ct143CikJAL and Ct2Ca-qAGS were excluded, matching our bpp species tree analyses. We performed a species tree search using all quartets (195,833) and used QFM to assemble the species tree. Node support was assessed using 100 nonparametric bootstrap replicates. We also performed additional species tree analyses in SVDS QuarTets using concatenated subsets of the total data (50 UCE loci, 100 UCE loci, 1,000 UCE loci, 2,000 UCE loci) to assess levels of congruence. All quartets were evaluated for each subset and 100 nonparametric bootstrap replicates were used to calculate support values.

Finally, we used the gene tree reconciliation method ASTRAL-III v. 5.5.6 (Mirarab et al., 2014; Mirarab & Warnow, 2015; Zhang et al., 2017) to infer a species tree that is statistically consistent with the multispecies coalescent. ASTRAL takes as input a set of unrooted gene trees and finds the species tree containing the maximum number of induced quartets that are consistent with the gene trees. Several studies have indicated that ASTRAL is one of the more accurate gene tree reconciliation methods currently available (e.g., Chou et al., 2015; Mirarab et al., 2014). However, caution is often warranted as poor resolution gene trees can substantially influence the accuracy of the inferred species tree (Mirarab et al., 2014; Xi et al., 2015). We began by manually inspecting the 500 loci used in bpp. We used GENEIOUS to remove individuals with all missing data (?'s), the two putative hybrid samples, individuals with highly fragmented data (Hosner, Faircloth, Glenn, Braun, & Kimball, 2016) and loci without the outgroup sequence. We also removed loci with extremely low levels of variability that were unlikely to yield useful gene trees. RAxML was used with the raxml_wrapper.pl script (https://sco.h-its.org/exelixis/web/software/raxml/index.html) to process the 351 alignments (-f d search under GTRGAMMA). Initial runs were used to prune identical sequences from each alignment and subsequent runs searched for ML trees using unique haplotypes. All 351 ML gene trees were used in ASTRAL, and we created a file to assign individuals to species. Support for relationships in the ASTRAL species tree was quantified using local posterior probabilities (Sayyari & Mirarab, 2016). We did not include ML bootstrap trees in analyses (i.e., multilocus bootstrapping) as recent research suggests that local posterior probabilities may be more a more accurate measure of support (Sayyari & Mirarab, 2016). The unrooted ASTRAL species tree was subsequently rooted with A. piscivorus.

### 2.4 Divergence time estimation

Because of the large size of the data set, we estimated divergence times on the fixed bpp species tree topology using MCMC Tree within PAML v. 4.9e (Rannala & Yang, 2007; Yang, 2007; Yang & Rannala, 2006). MCMC Tree is particularly suitable for Bayesian estimation of divergence times under alternative clock models in large, next-generation data sets where joint estimation of tree topology and divergence times is computationally problematic (Reis & Yang, 2011). Divergence times were estimated using a pruned alignment containing a single gene per lineage. PAUP* 4.0a161 (Swofford, 2001) was used to test for a strict molecular clock using likelihood ratio tests, AICc, and BIC under a HKY + I + G substitution model. As the results strongly rejected the molecular clock (see Results), we used the approximate likelihood method and independent rates clock model in MCMCTree (Reis & Yang, 2011). Analyses used a HKY + G4 substitution model, default priors for kappa (G[6, 2]) and alpha (G[1, 1]), G(2, 2,000) for substitution rate, and G(1, 10) for sigma. A single alignment partition of the full 3,384 loci was used to minimize the number of parameters and ensure adequate mixing and convergence. Sampling included a burn-in of 1 million generations followed by 10,000 posterior samples drawn every 5,000 generations. We calibrated the root node using information from a recent fossil-calibrated rattlesnake phylogeny (Blair & Sánchez-Ramírez, 2016). Specifically, soft bounds between 20 and 12 Ma were placed on the MRCA of Crotalus and Agkistrodon. Convergence was assessed by running all analyses twice and monitoring effective sample sizes (ESS) in Tracer v. 1.6 (Rambaut & Drummond, 2007). We also tested the impact of the sigma prior on divergence times by re-running analyses using a G(1, 100) prior. Finally, MCMC Tree was run without data (i.e., sampling the prior only; usedata = 0) to quantify the information content in the sequence data and make sure that the choice of priors for divergence times was reasonable.

### 2.5 Network estimation

Because strictly bifurcating trees may not adequately depict evolutionary history, we used the pseudolikelihood method SNaQ (Solís-Lemus & Ané, 2016) in the PhyloNetworks package (Solís-Lemus, Bastide, & Ané, 2017) to estimate a phylogenomic network. SNaQ is a recently developed algorithm that can explicitly account for both incomplete lineage sorting and introgression/gene flow. Previous studies have suggested that a large number of loci may be needed to accurately detect hybridization events (Solís-Lemus & Ané, 2016), so we used all 3,384 loci and all 55 individuals, each of which was
assigned to species following our species tree analyses. We began by running RAxML on each of the 3,384 alignments (ML + bootstrap search). We then used the best trees to determine concordance factors at the level of individuals. This data frame was then modified to create a concordance factor table at the level of species that was used as input for SNaQ. A more comprehensive discussion of running SNaQ using multiple alleles per species can be found in the online documentation of SNaQ/PhyloNetworks. We tested hmax values of 0–4 and selected the best value by plotting hmax vs. loglik. Ten independent runs were implemented for each value of hmax. For hmax = 0, we used the ASTRAL tree as the starting topology. For all remaining hmax values, we used the best network estimated from previous hmax values as the starting topology.

3 | RESULTS

3.1 | Data set characteristics

We sequenced 3,384 UCE loci (after quality filtering) for 54 specimens of montane rattlesnakes in the C. triseriatus species group and the outgroup. Out of a total of 2,057,530 characters, 1,978,984 were constant, 40,907 were variable and parsimony-uninformative, and 37,639 were variable and parsimony-informative. The proportion of gaps and undetermined characters in the concatenated data matrix was 18.4%, and pairwise identity was 99.4% with 94.7% of sites identical (including the Agkistrodon outgroup), illustrating the high degree of conservatism in UCE loci. Also typical of UCE data, the data showed an AT-bias, with a %GC-value of 38.1.

3.2 | Concatenated analysis

Unpartitioned ML analyses of the concatenated matrix under a GTRGAMMA substitution model yielded a topology with strong support for most nodes (Figure 2). The majority of presently defined species and subspecies were monophyletic, with the exception of C. aquilus and C. lepidus klauberi, the latter of which was placed throughout a large clade in the phylogeny with strong bootstrap support. The two samples of C. lepidus lepidus were deeply nested within a clade of C. lepidus klauberi. The majority of nodes towards the base of the tree received high bootstrap support, with the newly described C. tlapoci placed sister to C. pusillus, and C. campbellii sister to C. armstrongi. Crotalus raus was strongly supported as sister to the remaining sampled rattlesnakes. Crotalus morulus was also fully resolved and placed sister to a clade containing C. lepidus maculosus, C. lepidus klauberi and C. lepidus lepidus. One sample of C. aquilus (Ct2CaqAGS) was placed within a clade of C. l. klauberi.

All unpartitioned ExaBayes analyses indicated adequate convergence (ESS values >200, ASDSF < 0.05). In fact, both runs converged to the identical posterior distribution (ASDSF = 0) after 20,000 generations. However, chains were run for a full 1 million generations to adequately sample all parameters. The topology of the ExaBayes MRE tree was identical to the ML tree with all nodes fully resolved (posterior probability = 1.0).

The rcluster algorithm in PartitionFinder clustered the concatenated alignment into 2,005 subsets consisting of 20,156 parameters (INL = −3,796,380.79; AICc = 7,633,472.41). This partitioning scheme was then used to perform partitioned ML and Bayesian analyses in RAxML and ExaBayes, respectively, assigning independent GTRGAMMA models to each partition. The ML tree from the partitioned analysis was identical to the unpartitioned ML and Bayesian trees, with high bootstrap support for the majority of nodes. Similar to the unpartitioned analysis, the partitioned ExaBayes runs attained excellent convergence relatively rapidly (ASDSF = 0). However, ESS values for two parameters (LnPr and TL) were low due to large fluctuations in values throughout the runs. All other parameters had excellent ESS values (>200). This indicated potential issues with the estimation of branch lengths on large partitioned ExaBayes analyses. Nevertheless, we created a MRE consensus tree from the two partitioned ExaBayes runs and the topology was identical to the unpartitioned ExaBayes analysis and the partitioned and unpartitioned RAxML analyses. However, branch length estimates were off by an order of magnitude.

3.3 | Coalescent-based species delimitation and species tree analysis

The bootstrap consensus tree from the lineage-based (without a priori species assignments) SVDQUARTETS analysis conflicted with the inferred topology from the RAxML and ExaBayes analyses, although support for conflicting relationships was mixed (Supplementary Fig. S1). In the SVDQUARTETS tree, following the divergence of C. raus, a C. armstrongi/C. campbellii clade was strongly supported as sister to all remaining rattlesnakes. Similar to the concatenation (i.e., non-coalescent) analyses, C. tlapoci was strongly supported as sister to C. pusillus. Relationships within the C. aquilus, C. lepidus and C. morulus clade were ambiguous and contradictory to relationships inferred from the concatenated RAxML and ExaBayes analyses. However, many nodes in this clade suffered from low bootstrap support. One sample of C. l. klauberi (Ct143ClkJAL) was placed within a clade of C. aquilus.

We performed multiple coalescent-based species delimitation analyses in bpp to help determine if the genetic lineages inferred in our phylogenetic analyses may represent distinct species or reproductively isolated populations. Alternative runs, priors and search algorithms gave consistent results, indicating that chains were run for a sufficient length of time. There was overwhelming support for a species delimitation model where each of the 17 lineages (including the outgroup) represent distinct populations or species (Figure 3; Supplementary Table S2). Posterior probability values for each “species” were 1.0 for the majority of taxa in the tree in both the 100 and 500 loci data sets. Only four taxa received less than full support in our analyses: C. lepidus klauberi-2 (Durango), C. lepidus klauberi-3 (Northwest), C. lepidus klauberi-4 (Northeast) and C. lepidus lepidus. However, posterior probability values remained high, particularly for analyses based on 500 loci (Figure 3). Based on these results, we treated these taxa as distinct lineages for downstream species tree analysis.
A. piscivorus

C. ravus bruneus

C. ravus ravus

C. ravus exigus

C. tlaloci

P. pusillus

C. campbelli

C. armstrongi

C. triseriatus

C. lepidus klauberi + C. aquilus

C. lepidus klauberi-1 (South)

C. aquilus

C. morulus

C. lepidus maculosus

C. lepidus klauberi-2 (Durango)

C. lepidus klauberi-3 (Northwest)

C. lepidus lepidus

C. lepidus klauberi-4 (Northeast)

**FIGURE 2** Phylogenomic relationships of montane rattlesnakes of the *Crotalus triseriatus* group based on unpartitioned concatenated maximum likelihood (ML) analysis of 3,384 UCE loci (2,057,530 bp) under a GTR+GAMMA model of substitution. Values adjacent to nodes represent rapid bootstrap values from a joint bootstrap + ML search in RAxML (-f a option). Blue boxes indicate geographical lineages within *C. lepidus klauberi*; grey box indicates putative hybridization/introgression between *C. aquilus* and *C. lepidus klauberi*. MRE consensus tree from unpartitioned ExaBayes analyses was identical, with full posterior probability values (1.0) for all nodes.
We used the species delimitation model with the highest posterior probability for additional coalescent-based species tree analyses. The three independent BPP runs were largely consistent with one another, suggesting strong signal in the data and that chains were run for an acceptable duration. The only topological difference among runs occurred in the clade containing C. lepidus klauberi-3 (Northwest), C. lepidus klauberi-4 (Northeast) and C. lepidus lepidus. Two of the three runs placed C. lepidus lepidus as sister to C. lepidus klauberi-4 (Northeast), whereas the third run placed C. lepidus klauberi-3 (Northwest) sister to C. lepidus klauberi-4 (Northeast). The tree topology of the former two runs was identical to the concatenated ML and Bayesian trees (Figures 2 and 4a).

We also used the species delimitation results from BPP for subsequent species tree analysis in SVD QUARTETS. The resulting species tree differed from the RAxML and ExaBayes trees, the BPP species tree and the SVD QUARTETS lineage tree, particularly with regard to deeper relationships (Figure 4b). For example, BPP and the concatenated ML and Bayesian analyses placed C. campbelli and C. armstrongi as sister to a clade containing C. tlaocoi and C. pusillus. In contrast, the coalescent-based species tree analysis in SVD QUARTETS placed C. campbelli and C. armstrongi as sister to all remaining rattlesnakes except C. ravus. In addition, the concatenated analyses and BPP placed C. aquilus sister to one of the C. lepidus klauberi lineages (Figure 2 and 4a), whereas SVD QUARTETS placed C. aquilus as sister to a clade consisting of C. monulus, C. lepidus maculosus, C. lepidus lepidus and the three remaining C. lepidus klauberi lineages (Figure 4b). Bootstrap support was high for the majority of nodes in the SVD QUARTETS species tree (>70), except for the node uniting C. campbelli and C. armstrongi (63) and the node representing the MRCA of the majority of the species group (59).

The species tree inferred from ASTRAL analysis of 351 gene trees was largely consistent with the trees inferred from concatenated ML and Bayesian analyses and coalescent analyses in BPP (Figure 4c). The majority of differences involved nodes with low support values. Crotalus campbelli and C. armstrongi were placed sister to a clade containing C. tlaocoi and C. pusillus, albeit with low local

![Figure 3](image-url)  
**FIGURE 3** Bayesian posterior probability values for different species within the *Crotalus triseriatus* species group. Values represent averages over eight different BPP analyses using different search algorithms, priors and independent MCMC runs. Orange bars represent averages based on 100 loci and blue bars represent averages based on 500 loci.
posterior probability. In the ASTRAL tree, C. lepidus klauberi-1 (South) was more closely related to C. morulus and other C. lepidus lineages vs. C. aquilus, which differed from relationships inferred using the other species tree methods utilized.

We explored how the size of the concatenated data matrix might influence species tree inference in SVDQUARTETS by inferring trees based on 50, 100, 1,000 and 2000 UCE loci. Discordant results were obtained from each analysis, although most incongruence was restricted to poorly supported nodes (Figure 5). No tree was identified from SVDQUARTETS analysis. Tree is the SVDQUARTETS tree inferred using the concatenated matrix of 2,057,530 bp. Values indicate influence species tree inference in SVDQUARTETS by inferring trees

![Diagram](image)

**FIGURE 4** Coalescent-based species trees of rattlesnakes of the Crotalus triseriatus group based on alternative search algorithms. Individuals (53) were assigned to 17 species based on results from prior coalescent-based species delimitation analyses. (a) Majority-rule consensus tree from BPP analysis of 500 of the more variable loci. Values at nodes represent posterior probability values. (b) Species tree inferred from SVDQUARTETS analysis. Tree is the SVDQUARTETS tree inferred using the concatenated matrix of 2,057,530 bp. Values indicate nonparametric bootstrap support from 100 replicates. "X" indicates nodes that were not recovered in the bootstrap consensus tree. (c) Species tree inferred from ASTRAL analysis of 351 gene trees inferred under a GTRGAMMA model in RAxML. Values at nodes represent local posterior probability values.

**3.4 Divergence time estimation**

A strict clock model was strongly rejected using a likelihood ratio test (difference = 184.8485; \( p < 0.001 \), AICc (difference = 339.6926) and BIC (difference = 151.6418). Adequate convergence was obtained in multiple runs of MCMCTree (Figure 6), with all ESS values >200. Virtually identical results were obtained using smaller gamma priors for sigma \( \Gamma(1, 100) \), and thus, we present the results using \( \Gamma(1, 10) \) only. The MRCA of the C. triseriatus group dated back to around 10 Ma. Divergence within the three subspecies of C. ravus occurred in the Pliocene around 4–5 Ma. The newly described C. tlaploci diverged from its sister taxon, C. pusillus, close to 5 Ma. Crotalus campbelli diverged from C. armstrongi about 5.5 Ma. Independent lineages within C. lepidus klauberi originated within the past 5 million years. The estimated mean rate of substitution (\( \mu \)) was 3.57E-4 subs/site/My (95% HPD = 2.496E-4–4.696E-4). We also ran MCMCTree without data (i.e., sampling from the prior only) to determine the information content in the sequence data and make sure that the priors on divergence times were adequate. All divergence time priors were reasonable, yet different from posterior estimates that included the sequences (Supplementary Table S3). Thus, we concluded that the data were efficient to estimate divergence times and that our root calibration scheme was adequate.

**3.5 Network analysis**

There was a sharp increase in pseudolikelihood score from hmax = 0 to hmax = 1 (−766.252 vs. −541.526) followed by a more gradual improvement with additional reticulations (Supplementary Fig. S2). Thus, a single hybridization event best represented the data. The SNaQ major tree was virtually identical to the topologies from ASTRAL, BPP and concatenation (Figure 7). Topological differences were predominantly restricted to the clade consisting of C. lepidus lepidus and the different populations of C. lepidus klauberi. The single inferred reticulation event involved the mrca of the C. lepidus + C. morulus clade contributing genes (\( \gamma = 0.194 \)) to C. armstrongi.

**4 DISCUSSION**

**4.1 Phylogenomic analysis**

One of our main objectives was to examine levels of discordance in commonly used phylogenomic packages on a large empirical data
set. With the realization that gene tree topologies do not always agree with the species tree due to multiple processes including incomplete lineage sorting (Degnan & Rosenberg, 2006, 2009; Edwards, 2009; Maddison, 1997), substantial effort has been placed on the development of sophisticated algorithms to explicitly account for gene tree discordance when estimating species trees (e.g., Heled & Drummond, 2010; Liu et al., 2009; Liu et al., 2010; Mirarab et al., 2014; Vachaspati & Warnow, 2015). The species tree methods we select for comparison include concatenation, a fully probabilistic coalescent method (BPP), a gene tree reconciliation method (ASTRAL-III), and others.

**FIGURE 5** Species trees inferred by SVDQUARTETS based on different subsets of the total data set. Loci were selected randomly and concatenated for each analysis. Individual–species assignments followed analyses on the full data and were based off results from BPP. Trees shown are the SVDQUARTETS trees with bootstrap support values mapped on branches (100 replicates). Numbers in bold indicate bootstrap values below 70%. “X” indicates nodes that were not recovered in the bootstrap consensus tree.
and a site-specific technique based on algebraic statistics (SVDQUARTETS). These methods are selected based on both our data set characteristics and the relative performance of methods in simulation and empirical studies (Chifman & Kubatko, 2014; Chou et al., 2015; Mirarab et al., 2014; Mirarab & Warnow, 2015; Yang, 2015; Yang & Rannala, 2010). Both our concatenated ML and Bayesian analyses provide a well-supported phylogenomic hypothesis for rattlesnakes in the C. triseriatus group. Interestingly, the topology inferred from concatenation (Figure 2) is identical to the topology inferred by fully probabilistic coalescent analyses in BPP (Figure 4a). Due to the computational burden of methods like BPP, we restrict our analysis to 500 UCE loci. Our results coupled with those of previous studies suggest that BPP is likely to be an accurate estimator of species trees even with subsets of the total data (Caviedes-Solis, Bouzid, Banbury, & Leaché, 2015; Shi & Yang, 2018). We are a bit surprised that BPP is not used more often for species tree inference, although it is widely used for the purposes of species delimitation (e.g., Blair et al., 2015; Leaché & Fujita, 2010). We encourage empiricists to explore the method with additional phylogenomic data sets.

There are now myriad gene tree reconciliation techniques available to estimate species trees (Liu et al., 2010, 2009; Mirarab et al., 2014; Mirarab & Warnow, 2015; Zhang et al., 2017). These two-step procedures first use individual alignments to estimate gene trees, which are subsequently used as input data for species tree inference. We examined ASTRAL as simulation studies have shown it to be one of the more accurate techniques (Chou et al., 2015; Mirarab et al., 2014). It is also one of the few methods that can handle multiple individuals per species. To try to minimize potential errors in species tree inference, we followed the recommendations of recent studies (e.g., Hosner et al., 2016; Meiklejohn et al., 2016; Xi et al., 2015) and limited our ASTRAL analysis to 351 of the more variable UCE loci with little fragmentary data and use RAxML for gene tree

**FIGURE 6** Divergence times of rattlesnakes of the Crotalus triseriatus group based on 2,057,530 bp. Parameter inference was based on approximate likelihood calculations in MCMCTree using the BPP species tree as a fixed topology and one individual per lineage. Horizontal bars represent 95% HPD confidence intervals for divergence times. Inset plot shows estimated mean divergence times from two independent MCMCTree runs. High levels of concordance between runs indicate that chains were run for a sufficient length of time and convergence was reached.
reconstruction. The resulting species tree was highly congruent to the species trees inferred using concatenation and BPP. One difference that was strongly supported is the placement of C. aquilus and C. lepidus klauberi-1 (South), which were inferred as sister taxa in concatenation and BPP. Conversely, in the ASTRAL tree, C. lepidus klauberi-1 (South) was sister to a clade consisting of C. morulus and the remaining C. lepidus taxa. Other differences in the ASTRAL tree involved lineages within C. lepidus klauberi, but these relationships were weakly supported. Thus, although our ASTRAL results were consistent with other methods examined, additional studies are needed to help determine how large phylogenetic data sets should be pruned and subsampled prior to species tree inference.

Although these gene tree reconciliation methods are relatively quick, there are numerous additional considerations that researchers must be aware of when using these methods. For example, gene tree error has been shown to negatively influence the resulting species trees (Mirarb et al., 2016, 2014; Roch & Warnow, 2015; Xi et al., 2015). Unfortunately, gene tree error/uncertainty is often ubiquitous when working with loci with relatively few informative characters such as UCEs. Second, missing data can also negatively impact gene tree reconciliation techniques, most likely to a greater degree than concatenation and site-based methods (Burleigh, Kimball, & Braun, 2015; Hosner et al., 2016). Missing data can come as either an entire sequence missing for some species in an alignment or species having partial (fragmented) sequences that reduce information content. Other studies have also shown a measurable impact of the method used to reconstruct gene trees (Meiklejohn et al., 2016; Xi et al., 2015). Recently, Streicher et al. (2016) used a large UCE data set from iguanian lizards to quantify the effects of taxon and locus sampling on species tree inference. Their results based on concatenation and coalescent analyses (NJst) suggest that including a large proportion of missing data (including loci with up to 50% missing taxa) may be beneficial in phylogenomic studies. However, results and conclusions were based off branch support values in the unknown iguanian tree and differences were obtained between phylogenomic methods. In another recent study, Hosner et al. (2016) found that missing data (particularly partial/fragmented sequences) were a bigger issue for gene tree reconciliation approaches vs. concatenation and quartet-based methods like SVDQUARTETS. Their results also corroborated other recent studies (e.g., Meiklejohn et al., 2016) and suggest that gene tree reconciliation methods be restricted to a subset of the most variable loci to improve the quality of gene trees. In sum, there is still a great deal of uncertainty with how taxa and loci should be sampled in phylogenomics, particularly when using gene tree reconciliation approaches.

Because of the potential issues with gene tree reconciliation methods highlighted above, there has been substantial interest in species tree methods that work directly from the sequence data and are scalable to large phylogenomic data sets. One popular method that has emerged is the SVDQUARTETS algorithm currently implemented in PAUP*. The method is statistically consistent under the multispecies coalescent model, can work with both SNP data and concatenated UCE loci, is relatively quick, and can apparently work well with large quantities of missing data (Chifman & Kubatko, 2014; Chou et al., 2015; Leaché et al., 2016, 2015). These attributes make the methods particularly appealing to analyse large phylogenomic data sets generated through both sequence capture and RAD-seq-related approaches. The technique has been shown to perform well in simulation (Chou et al., 2015), but there are still little data available illustrating how SVDQUARTETS performs on empirical phylogenomic data (Hosner et al., 2016; Leaché et al., 2016, 2015). We performed multiple SVDQUARTETS analyses on different data sets to quantify discordant topological patterns. The lineage-based analysis (using individuals as terminals) differed from the concatenated ML and Bayesian analyses although bootstrap support values were relatively low. In the SVDQUARTETS lineage tree, C. morulus was not monophyletic and C. campbelli was nested with C. armstrongi. The SVDQUARTETS species tree based on all 3,384 loci also differed from species trees inferred from other methods. Of particular note is the early divergence of (C. campbelli + C. armstrongi) in the SVDQUARTETS tree, whereas other species tree analyses and concatenation placed these taxa as sister to (C. pusillus + C. tlaoci). However, bootstrap support for this placement was weak (59%), and it is likely that C. campbelli and C. armstrongi form a clade with C. pusillus and C. tlaoci.

To further probe how results from SVDQUARTETS may change based on data set size and characteristics, we performed additional species tree analyses on concatenated alignments consisting of 50, 100, 1,000, and 2000 UCE loci. Different topologies were recovered for each analysis, although bootstrap support at conflicting nodes was relatively weak. With only 50 UCE loci, the C. raus clade was
placed sister to *C. campbelli* and *C. armstrongi* with a bootstrap value of 65%. This result disappeared in all analyses with more data. Our results also suggested an unclear relationship between data set size and bootstrap support values, indicating that there are additional nuances besides number of loci that contribute to strongly supported species trees. In general, our results are consistent with other recent studies that found lower node support values for SVDQUARTETS trees vs. trees inferred through concatenation (Leaché & Linkem, 2015; Leaché et al., 2016, 2015; Hosner et al., 2016). In addition, our results are similar to a recent study that examined phylogenomic relationships among gibbons using real and simulated data and found that species trees inferred by BPP and ASTRAL were much more consistent than those inferred by SVDQUARTETS (Shi & Yang, 2018). Although additional simulation and empirical studies are needed to examine characteristics of the SVDQUARTETS method in greater detail, it appears that relatively large data sets are needed to reach a topology and support values consistent with other methods of species tree inference, and that full likelihood-based approaches that adequately model rate heterogeneity and coalescent stochasticity may be preferable.

Strictly bifurcating trees may not adequately depict evolutionary history when gene flow and hybridization are prevalent (Solís-Lemus & Ané, 2016; Wen, Yu, Zhu, Nakhlé, & Posada, 2018; Yu & Nakhlé, 2015). Using SNaQ, we found hybridization within the sampled taxa of the *C. triseriatus* group to be relatively rare. Results provided strong support for a model consisting of a single reticulation event from the mrca of *C. lepidus* + *C. morulus* into *C. armstrongi*. The major tree inferred by SNaQ was virtually identical to the ASTRAL tree and very similar to the trees inferred by BPP and concatenation. Interestingly, SNaQ detected no introgression between *C. aequorus* and *C. lepidus klauberi*-1 (South) even though this was suggested in our other phylogenetic analyses. It is possible that introgression between these two taxa was limited to two individuals in our study (Ct143ClnkJAL and Ct2CqcaGJS), and the inclusion of larger number of samples from each species masked the signal of hybridization. Additional sampling of both species from near where their ranges meet should further elucidate the propensity for introgression in these snakes.

Interest in phylogenomic networks continues to increase as more evidence accumulates suggesting that both incomplete lineage sorting and reticulation/hybridization may be a common component of the evolutionary history of several groups (Solís-Lemus & Ané, 2016; Solís-Lemus et al., 2017; Wen et al., 2018). Unfortunately, many of these explicit network methods are computationally demanding (Than, Ruths, & Nakhlé, 2008), necessitating the use of some method of approximation. New pseudolikelihood methods such as SNaQ and similar techniques in the PhyloNet package (Yu & Nakhlé, 2015; Wen et al., 2018) show promise for analysing large phylogenomic data sets. Combining tree-based and network-based phylogenomic methods, each with their own strengths and weaknesses, will likely provide the most robust conclusions regarding evolutionary history. Further, although our inferred divergence times are broadly concordant with previous studies (e.g., Blair & Sánchez-Ramírez, 2016; Bryson et al., 2011), more research is needed to determine how reticulated histories influence inferred divergence times.

### 4.2 Systematics and biogeography of the *C. triseriatus* group

The Mexican highlands are a known biodiversity hotspot (Mittermeier et al., 2017), and multiple phylogeographic studies have documented cryptic diversity distributed throughout high-elevation biomes within Mexico’s major mountain systems (e.g., Bryson, García-Vázquez, & Riddle, 2012; León-Paniagua, Navarro-Sigüenza, Hernández-Baños, & Morales, 2007; McCormack, Peterson, Bonaccorso, & Smith, 2008; Navarro-Sigüenza, Townsend Peterson, Nyari, García-Deras, & García-Moreno, 2008; Puebla-Olivares et al., 2008). Based on our analyses, it is clear that *C. ravus* is divergent from the remainder of the group. The three currently recognized subspecies of *C. ravus* are allopatric (Campbell & Lamar, 2004), morphologically differentiated (Campbell & Armstrong, 1979), form distinct lineages (Bryson et al., 2011), and are supported as distinct species by our species delimitation models using phylogenomic data. Based on these combined lines of evidence, we recommend that the three subspecies be elevated to full species status: *C. ravus*, *C. brunneus* and *C. exiguus*.

Nearly all species trees place *C. tlaloci* sister to *C. pusillus* and *C. campbelli* sister to *C. armstrongi*. These relationships are consistent with geography, with *C. pusillus* inhabiting the western regions of the Trans-Mexican Volcanic Belt and adjacent Sierra de Coalcomán, and *C. tlaloci* found along the central part of the Trans-Mexican Volcanic Belt. *Crotalus campbelli* is distributed along the western margin of the Trans-Mexican Volcanic Belt and is separated by low-elevation valleys from *C. armstrongi* inhabiting regions of the Trans-Mexican Volcanic Belt to the east (Bryson et al., 2014). None of these species are known to occur sympatrically, although *C. pusillus* and *C. armstrongi* are both found on Cerro Tancítaro in western Michoacán.

*Crotalus triseriatus* is the sister species to the large clade containing *C. aequorus*, *C. morulus* and *C. lepidus*. *Crotalus aequorus* and southern *C. l. klauberi* (*C. l. klauberi*-1 South) are strongly supported as sister taxa in several of our analyses. Previous results based on mtDNA suggested *C. aequorus* formed a closer relationship with *C. morulus* and placed southern *C. l. klauberi* as sister to the remaining *C. lepidus* (Bryson et al., 2011). Our species delimitation analyses indicate *C. lepidus* contains as many as six distinct lineages, all consistent with geography. These same lineages were inferred with mtDNA (Bryson et al., 2011). Although our sampling spanned the geographical distribution of most subspecies, we lack samples from critical regions where subspecies converge (Campbell & Lamar, 2004). We therefore hold off on formal taxonomic changes to *C. lepidus* until further range-wide sampling combined with a detailed examination of morphology can be made. Additional sampling and integrative approaches to species delimitation will also help alleviate concerns of overestimating the number of true species when using BPP and other multispecies coalescent programmes (Sukumaran & Knowles, 2017).
Our results based on UCE data and results based on mtDNA (Bryson et al., 2011) present evidence for limited hybridization of C. aquilus with southern C. lepidus and C. morulus across the northern region of the Central Mexican Plateau from San Luis Potosí to Jalisco. The patchy islands of suitable rocky habitat above 2,000 m across this region, similar to “sky islands” in northern Mexico and Arizona (McCormack, Huang, & Knowles, 2009), suggest populations in this area likely harbour genetic signatures of a former Pleistocene hybrid zone. Current gene flow at a low rate may also be possible in those few areas with contiguous suitable habitat. Divergence time estimates here and elsewhere (Bryson et al., 2011) indicate C. aquilus, southern C. l. klauberi and C. morulus likely diverged from each other prior to onset of the Pleistocene 2.6 Ma. After these divergences, montane woodlands covered much of the Central Mexican Plateau (McDonald, 1993; Metcalfe, 2006; Gugger, González-Rodríguez, Rodríguez-Correa, Sugita, & Cavender-Bares, 2011; Van Devender, 1990), and as a result, the distributions of these taxa probably extensively overlapped during most of the Pleistocene. However, within the core distributions of C. aquilus and C. morulus, no evidence of introgression has been uncovered and each species maintains phenotypic cohesion, suggesting these two species have maintained their distinctiveness throughout the past several million years. It is clear, however, that southern C. l. klauberi has a closer relationship to C. aquilus than to other C. lepidus, at least based on genomic and potentially phenotypic data. While mtDNA clearly supports the distinctiveness of southern C. l. klauberi, consistent with our species delimitation models, it unites this lineage with other C. lepidus (Bryson et al., 2011). This finding suggests that a history of introgression between southern C. l. klauberi and C. aquilus has been maintained in the nuclear genome, but not the mitochondrial genome. Finer-scale population-level sampling across the northern Central Mexican Plateau is needed to fully characterize potential hybrid-zone dynamics within the species group and help determine if southern C. l. klauberi deserves recognition as a distinct species.

ACKNOWLEDGEMENTS

We thank the following curators, institutions and individuals for providing tissue samples: R.W. Murphy and A. Lathrop (Royal Ontario Museum), O. Flores-Villela and A. Nieto-Montes (Universidad Nacional Autónoma de México), J.A. Campbell, C. Franklin and E.N. Smith (University of Texas at Arlington), J. Alvarado-Díaz (Universidad Michoacana de San Nicolás de Hidalgo, Michoacán), A. Kardon (San Antonio Zoo), U.O. García-Vázquez, J. Lemos-Espinal, L. Cansecoc-Marquez, G. Swinford, E. Mocio-Deloya, K. Setser and G. Quintero-Díaz. Funding was provided in part by NSF (DEB-1257785 and DEB-1258205). Collecting was conducted under permits granted by SEMARNAT to RWB, the late F. Mendoza-Quijano, and UNAM. For additional support and advice, we thank members of the McCormack and Klicka Labs, J. Jones, C. Grunwald, and the many volunteers that helped with field work. The authors are grateful to the Center for Theoretical Physics of New York College of Technology (CUNY) for providing computational resources. Finally, we thank A. Stamatakis and A. Aberer for assistance with ExaBayes and C. Ané for assistance with SNaQ.

AUTHOR CONTRIBUTIONS

C.B., R.W.B., J.K. and J.E.M. developed the conceptual framework for the project; R.W.B. and D.L. performed fieldwork and contributed samples; R.W.B. generated the data; C.B., R.W.B. and C.W.L. analysed the data; C.B. and R.W.B. led the writing, and all authors contributed to and approved the final manuscript.

DATA ACCESSIBILITY

Alignment and tree files are available from the Dryad Digital Repository (https://doi.org/10.5061/dryad.4467407).

ORCID

Christopher Blair ORCID iD https://orcid.org/0000-0002-4782-6788

REFERENCES


Crawford, N. G., Faircloth, B. C., McCormack, J. E., Brumfield, R. T., Winiker, K., & Glenn, T. C. (2012). More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Bioletters*, 8, 783–786.


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.