Phylogenomics of Vipers and the Role of Competition on Venom Evolution

---

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Biological Sciences

---

by
Rhett M. Rautsaw
August 2022

Accepted by:
Dr. Christopher L. Parkinson, Committee Chair
Dr. Margaret Ptacek
Dr. Samantha A. Price
Dr. Jason L. Brown

Abstract

Competition is a critical selective force for diversification; however, empirical studies on its role promoting differentiation of adaptive phenotypes have largely been limited to small spatial and taxonomic scales. Here, we test the effect of competition on the evolution of pitviper venoms – a cross-continental radiation of venomous snakes each with tens to hundreds of individual toxins varying in expression, composition, and overall complexity among species. By inferring a novel phylogenomic tree and reconstructing the biogeography of the Viperidae, we demonstrate a rapid radiation occurred upon invasion of the New World. Using >500 venom gland transcriptomes and phylogenetic comparative modeling, we reveal that venom phenotypes diverge when multiple species coexist in a given area through time via positive diversity dependence. Furthermore, we find that pitviper communities have evolved to maximize functional diversity despite comparatively low phylogenetic diversity, suggesting an evolutionary response of venom rather than communities accumulating phylogenetically diverse species. Together, these findings support competition as a likely selective pressure driving venom diversification in pitvipers.

Acknowledgments

Ever since I was a kid, my parents and family kindled my love and interest in nature; encouraging me to pursue my dreams and passions rather than simply finding a job. With this in mind – 10 years ago after I graduated high school – I began my journey. I waded through two years of undergraduate classes, three years of an M.S., and now five more years of a Ph.D. Thanks to them I have turned my original interest in "chasing snakes" into a legitimate scientific career – unraveling mysteries of ecology and evolution. I would not be where I am today without their endless love and support. I want to specifically thank my parents (Darci and Brett) for being there for me whenever I needed you and for raising me to be who I am today. I also want to thank my sisters (Leighana and Lorelei), grandparents, aunts, uncles, and cousins for lifting me up when I needed it, listening to me when I was down, and generally being there for me throughout this time and my life. I love you all so much and I cannot thank you enough for what you have done for me through the years.

To my partner and support system, Shelly Gaynor. You are literally my biggest fan. Every time we are around a group of people, you go out of your way to brag about me and my accomplishments. When I'm feeling worthless, you tell me how well I'm doing. When I'm feeling like I can't accomplish something, you motivate me and push me to be the best that I can be. When I'm working too hard, you remind me to take care of myself. The past five years have not been easy given that we have been 500 miles apart, but we've somehow made it work. I don't deserve you, but you're literally the best.

I could not have completed my dissertation without the support, reviews, edits, and most importantly general chats and camaraderie shared with my now best friends: Andrew Mason, Jason Strickland, Erich Hofmann, and Kathryn Mercier. I would not trade anything for the time I got to spend with you all in the lab. It was legitimately some of the best times of my life and the friendships we developed would probably last multiple lifetimes. I'm incredibly grateful to have met

you all and for all your help – including allowing me to treat as if you were my therapist and express my frustrations.

Many others have passed through the Parkinson Lab during my dissertation including: Erin Stiers, Mark DiMeo, Tristan Schramer, Jade Mellor, Ramses Rosales-Garcia, Matthew Holding, Mark Margres, Ed Myers, and the entire CUGBF team (especially Rooksie Noorai). You all made my time in the lab wonderful and I am similarly so happy I have gotten to meet you all and work with each of you. So thank you all for just being who you are and making my time at Clemson so much fun.

To my advisor Christopher Parkinson, I first want to thank you for taking a chance on me and bringing me into the lab. As I learned from graduate applications, not many people want to take a chance on students that don't have a lot of experience, but you took that chance on me – not once, but twice! First training me to be an ecologist during my M.S. and then an evolutionary biologist during my Ph.D. I cannot thank you enough for this opportunity. Second, I want to thank you for always looking out for me and my best interests. You and your family – Cynthia and Caroline – have always gone out of your way to ensure that I am doing alright, that I have food in my refrigerator, and even that I am paid adequately to live. You would always go to battle for me and any of your students to ensure that our needs our met. I very obviously could not have done this without you, but I want to thank you for going out of your way to facilitate a productive environment.

I also thank my committee – Margaret Ptacek, Sam Price, and Jason Brown – for the useful advice and helping improve this dissertation significantly. Finally, I have many collaborators to thank without whom this work would not be possible. In particular, the PIs of the NSF/FAPESP funded "Scales of Biodiversity" grant (H. Lisle Gibbs, Darin R. Rokyta, Inacio L. M. Junqueira-de-Azevedo, Felipe G. Grazziotin, Ana M. Moura-da-Silva, Erika Hingst-Zaher) as well as the numerous students and postdocs in each of their labs. I also want to thank R. Alexander Pyron for use of many anchored hybrid enrichment datasets for phylogenomics. Sample collection would not have been possible without numerous collaborations across Mexico, Costa Rica, Honduras, Ecuador, and Brazil – particularly the help from Miguel Borja, Gamaliel Castaneda-Gaytan, HerpMX, Mahmood Sasa, Josiah H. Townsend, Luis Herrera, Carlos R. Vasquez-Almazan, and David Salazar.

Table of Contents

List of Figures

Chapter 1

Introduction

Species interactions have long been recognized as strong and ubiquitous agents of selection influencing speciation, diversification, and trait evolution[1,2]. For example, closely related species are expected to diverge to avoid competition and hybridization with each other when they coexist (sympatry) relative to where they exist separately (allopatry) – a pattern now known as character displacement[3]. Character displacement can be generated by a variety of processes (see[4]), but ecological character displacement is the most commonly studied and is the result of competition over a shared, limited resource. Character displacement has been well studied in many iconic model systems such as Darwin's finches[5,6], *Anolis* lizards[7–9], and three-spined sticklebacks[10–12]. In these systems, ecological opportunity facilitates adaptive radiations and a "diversity-dependent" pattern of diversification where species rapidly evolve to fill niches and to avoid competition based on the diversity of competitors[13–16]. The impact of competition has been well studied in these model systems, in small groups of species, and in individual species pairs[17]; however, the influence of competition on larger macroevolutionary or cross-continental scales is only now being recognized thanks to advances in phylogenetic comparative methods that allow scientists to link ecological processes to macroevolutionary patterns of lineage diversification and trait evolution[18–25].

The evolution of phenotypes in Darwin's finches and three-spined sticklebacks is largely attributed to a few or even single large-effect loci[6,26–28]. Although these large-effect loci facilitate rapid evolution of a single trait, it may also reduce evolvability by constraining the direction and magnitude of phenotypic change that can occur[29–31]. By comparison, multidimensional traits facilitate adaptation and evolutionary lability through a combination of a large and many small-effect changes[29–31].

Multidimensional traits may, therefore, evolve across a multitude of dimensions, magnitudes, and complexity leading to niche specialization or expansion. Despite the high adaptive potential of multidimensional traits, few studies on character displacement have focused on these systems. Focusing on highly complex, multidimensional traits enable a more robust examination of competition's role in trait evolution by investigating several dimensions of niche space simultaneously.

Snake venom is a quantifiable complex, multidimensional trait derived from tens to hundreds of individual toxin genes which can vary drastically in composition, sequence, expression, and overall complexity between species and populations[32–34]. Venom has a near 1:1 relationship between genotype and phenotype making it easy to quantify changes in the genome as well as throughout transcription and translation[35,36] (but see[37]). Additionally, since venom is secreted, it is not subject to complex developmental or pleiotropic effects and has been shown to evolve rapidly between species[38–41]. In particular, the lack of pleiotropy facilitates evolutionary lability since evolutionary change may be constrained by other phenotypes under stabilizing selection. Finally, venom is closely linked to ecology through resource acquisition (i.e. prey capture)[30,42–44] making it a prime candidate for ecological character displacement as species may evolve divergent venom phenotypes when they occur in sympatry to avoid competition and focus on alternate prey resources.

Vipers (Viperidae) are a family of 367 species of charismatic venomous snakes divided into three subfamilies: true vipers (Viperinae), monogeneric Feae's vipers (Azemiopinae), and pitvipers (Crotalinae; Reptile Database, May 2021). Known for their long, hinged fangs which deliver venoms to immobilize prey, aid in digestion, and serve as a defense against threats, vipers (Viperidae) represent one of the most well-studied groups of snakes due to their medical significance across the globe[45]. In North and South America specifically (*i.e.*, New World), pitvipers are responsible for $> 98\%$ of snakebites[45] and their venoms have been shown to display extensive variation within and among species[41,46–52]. One notable axis of variation in pitviper venom is a well-documented dichotomy separating metalloprotease-rich hemorrhagic venoms (Type I) from phospholipase-rich neurotoxic venoms (Type II). For example, across species, the Black-Speckled Palm Pitviper (*Bothriechis nigroviridis*) maintains a Type II venom while its putatively sympatric congener, the Talamancan Palm Pitviper (*Bothriechis nubestris*), has a Type I venom[50]. Similarly, this dichotomy can be found within single species such as the Mojave Rattlesnake (*Crotalus scutulatus*) and Timber Rattlesnake (*Crotalus horridus*)[41,47,48]. Some of the variation in pitviper venom is the result of its close relationship with resource acquisition (i.e. prey capture) and coevolving prey venom resis-

tance[30,43] and recent studies have shown that venom complexity in rattlesnakes is correlated with dietary breadth[51,53]. However, no study has tested whether competition among pitvipers could be responsible for shifts in venom phenotypes.

Despite being one of the most well-studied groups of snakes, the phylogeny and biogeographic history of Viperidae remains uncertain for many genera and species[54–69]. Hypothesized to have originated in the late-Paleocene to mid-Eocene[61,65,68,69], vipers rapidly radiated across the globe with speciation bursts for pitvipers upon invasion of Southeast Asia and the New World[65]. The rapid radiation of vipers created significant challenges for phylogenetic and biogeographic inference which has been compounded by a lack of genetic resolution as only a handful of mitochondrial and nuclear loci have been used in previous studies, resulting in a high degree of uncertainty and unstable placement of many genera and species[61,64–69]. To adequately examine the role of competition on venom evolution in New World pitvipers, a robust phylogenetic and biogeographic reconstruction is necessary as phylogenetic uncertainty can greatly influence macroevolutionary inference[70–73].

Here, we aim to (1) infer a robust phylogeny for Viperidae, (2) reconstruct the biogeographic history, and (3) test for evidence of competition influencing the evolution of New World pitviper venom. To infer a new phylogenetic hypothesis for Viperidae, we use a multiomic approach; combining data from RNA-sequencing, phylogenomic sequence capture (SeqCap), and whole-genome sequencing (WGS) to collect over 2000 loci for gene tree estimation and species-tree inference. We then date the divergences of our phylogeny and thoroughly examine the biogeographic history of the group. We use this information and focus inward on the venom evolution of the New World pitviper clade specifically. Using RNA-Seq, we assemble and annotate the venom gland transcriptome for each species and perform cross-species transcriptome mapping to place expression into a joint manifold. We use these data to test whether venom diverged in response to competition via phylogenetic comparative modeling, community-based trait dispersion, and generalized dissimilarity modeling.

Chapter 2

Methods

2.1 Sample Collection & Sequencing

We first collected sequence data for phylogenomic inference by gathering datasets from NCBI's Sequence Read Archive totaling 100 samples (29 RNA, 3 WGS, and 68 SeqCap datasets) from 36 viper species. We supplemented these data by utilizing our previously published 216 venom gland transcriptomes[51] and sequenced an additional 290 venom gland and blood transcriptome libraries for a total of 506 libraries representing 115 viper species and 4 outgroup taxa. We also sequenced 4 whole-genome sequencing libraries from 2 viper species and 198 sequence capture/Anchored Hybrid Enrichment (AHE) libraries from 113 viper species and 13 outgroup species.

In total, our final dataset included 808 samples (535 RNA, 7 WGS, 266 SeqCap) for 194 viper species of 368 total species (53% species completeness) and 32 of 36 total viper genera (89% genus completeness; Fig. 2.1). Importantly, testing for evidence of competition on venom evolution focused on New World taxa, where we were able to sample 126 of 160 total New World pitviper species (79%; Fig. 2.1). The only genera not sampled were *Garthius*, *Macrovipera*, *Montatheris*, and *Proatheris*.

Outgroup sampling totaled 18 species from the families Dactyloidae (1 species), Aniliidae (1), Pythonidae (1) Boidae (2), Acrochordidae (1), Pareidae (1), Atractaspididae (1), Elapidae (2), Colubridae (2), Dipsadidae (3), and Natricidae (3). Taxonomy is based on the Reptile Database (May 2021; http://www.reptile-database.org/).

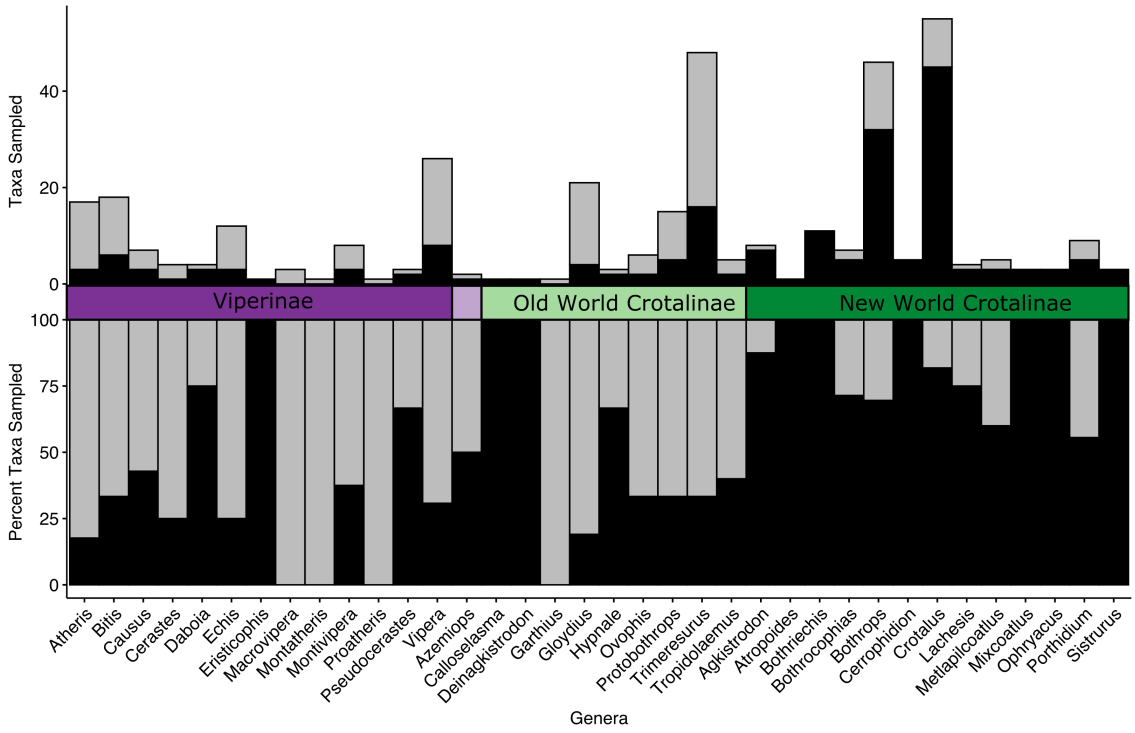

Figure 2.1: Sampling of species for each genus in Viperidae with black representing the number/percent of species sampled and gray representing species unsampled.

2.1.1 Transcriptomics (RNA-Seq)

We collected venom from each snake by allowing them to bite onto a parafilm-sealed sterile collection cup. Four days later when transcription is maximized[74], snakes were euthanized following AVMA[75] and ASIH guidelines[76]. We collected tissues and immediately transferred tissues to RNAlater for storage.

We isolated RNA via a standard TRIzol extraction following Hofmann & Rautsaw[36]. Briefly, tissues were finely diced and placed in TRIzol solution (Invitrogen, Waltham, MA). The mixture was homogenized and transferred to a phase lock heavy gel tube (5Prime). Once cells were lysed, total RNA was isolated using chloroform and purified via isopropyl alcohol and ethanol precipitation. RNA quantification was performed using a Qubit with the RNA BroadRange Kit (Thermo Fisher Scientific, Waltham, MA). RNA quality was checked using an Bioanalyzer 2100 with an RNA 6000 Pico Kit or TapeStation 2200/4200 with a RNA ScreenTape (Agilent Technologies, Santa Clara, CA) to ensure sufficient quantity and quality RNA for library preparation and sequencing.

We produced cDNA libraries from isolated mRNA using magnetic bead isolation of mRNA followed by cDNA synthesis and PCR amplification. First, we isolated mRNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490S; NEB, Ipswich, MA). Following bead isolation and cleanup, cDNA libraries were prepared from isolated mRNA using a NEB Next Ultra RNA Library Prep Kit for Illumina (NEB #E7530; NEB, Ipswich, MA) following manufacturer's recommendations. We used a fragmentation time of 13 minutes, 30 seconds to achieve a target mean fragment size of 400 bp, and 14 PCR cycles for amplification of double stranded cDNA libraries. We quantified library yield using a Qubit with the DNA BroadRange Kit (Thermo Fisher Scientific, Waltham, MA) and assessed library quality on a Bioanalyzer 2100 with a High Sensitivity DNA Kit or TapeStation 2200/4200 with a D1000 ScreenTape (Agilent Technologies, Santa Clara, CA). The total amplifiable concentration of cDNA in each library was then determined using KAPA qPCR (Roche, Basel, Switzerland) at the Florida State University Molecular Cloning Facility or in-house. Pooled libraries were sequenced with 150 base pair (bp) paired-end reads on an Illumina HiSeq 2500 or Illumina NovaSeq 6000 platform (Illumina, San Diego, CA) at the Florida State University College of Medicine Translational Science Laboratory (Tallahassee, FL, USA), Clemson University Genomics and Bioinformatics Facility (CUGBF, Clemson, SC), or Institute Butantan (São Paulo, São Paulo, Brazil).

2.1.2 Whole Genome Sequencing (WGS)

We extracted DNA from blood samples using the QIAGEN DNeasy Kit (QIAGEN, Valencia, CA) and quantified DNA using a Qubit DNA BroadRange Kit (Thermo Fisher Scientific, Waltham, MA). We checked DNA quality using a TapeStation 4200 with a D1000 ScreenTape (Agilent Technologies, Santa Clara, CA). We prepared libraries for sequencing using the NEB Ultra II FS DNA Library Prep Kit for Illumina (NEB #E6177; NEB, Ipswich, MA) following manufacturer recommendations. We quantified library yield using the Qubit DNA BroadRange Kit (Thermo Fisher Scientific, Waltham, MA) and assessed library quality on a TapeStation 4200 with a D1000 ScreenTape (Agilent Technologies, Santa Clara, CA). The total amplifiable concentration of cDNA in each library was then determined using KAPA qPCR (Roche, Basel, Switzerland) at the Florida State University Molecular Cloning Facility. Pooled libraries were sequenced with 150 base pair (bp) paired-end reads on an Illumina NovaSeq 6000 platform (Illumina, San Diego, CA) at the Florida State University College of Medicine Translational Science Laboratory (Tallahassee, FL, USA).

2.1.3 Sequence Capture (SeqCap)

We extracted DNA using QIAGEN DNeasy Kits (QIAGEN, Valencia, CA) and assessed DNA quality and concentration using gel electrophoresis and NanoDrop 2000c Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). Only samples showing bright bands in the high molecular-weight region were selected for sequencing and the amount of DNA in each sample was estimated to be over $2.6\mu g$.

We used anchored hybrid-enrichment generate DNA sequence data. First, probes were designed based on the conserved regions of vertebrate genomes identified by Lemmon et al.[77] (see online materials for details and probe sequences). Then, genomic DNA was sonicated to 300-700 bp fragments and amplified for preparing libraries. Next, indexed library adapters were added to the ends of the DNA fragments[77]. The prepared libraries were pooled at equal concentration in groups of 15 and enriched following Lemmon et al.[77]. After enrichment, library pools were pooled for sequencing at equal concentration. After quantification using qPCR, the libraries were sequenced on an Illumina Hiseq 2500 platform (Illumina, San Diego, CA), which produced 100bp paired-ends. All data collection and preliminary analysis was performed at the Center for Anchored Phylogenomics at Florida State University (www.anchoredphlogeny.com).

2.2 Loci Extraction

To extract loci for phylogenomics across all data types, we developed a custom pipeline (*PhyProbe*; Fig. 2.2) which utilizes the RNA-Seq data to identify loci followed by alignment-based sequence extraction from other data types using *AliBaSeq*[78]. Three types of loci were extracted: (1) benchmarking universal single-copy orthologs (BUSCOs), (2) squamate conserved loci (SqCL)[79] consisting of commonly used phylogenomic loci such as anchored hybrid enrichment loci (AHEs) and ultra-conserved elements (UCEs), and (3) novel transcript sequences.

2.2.1 Transcriptomics

We first trimmed adapters and low quality reads using *Trim Galore!* v0.4.4 (https://github.com/FelixKrueger/TrimGalore). Reads were removed if they had phred scores less than five and a length less than 75 bp. Next, we merged paired-end reads using *PEAR* v0.9.10[80] and performed *de novo* assembly using *Trinity* v2.11.0[81] or *rnaSPAdes* v3.13.1[82] with default settings.

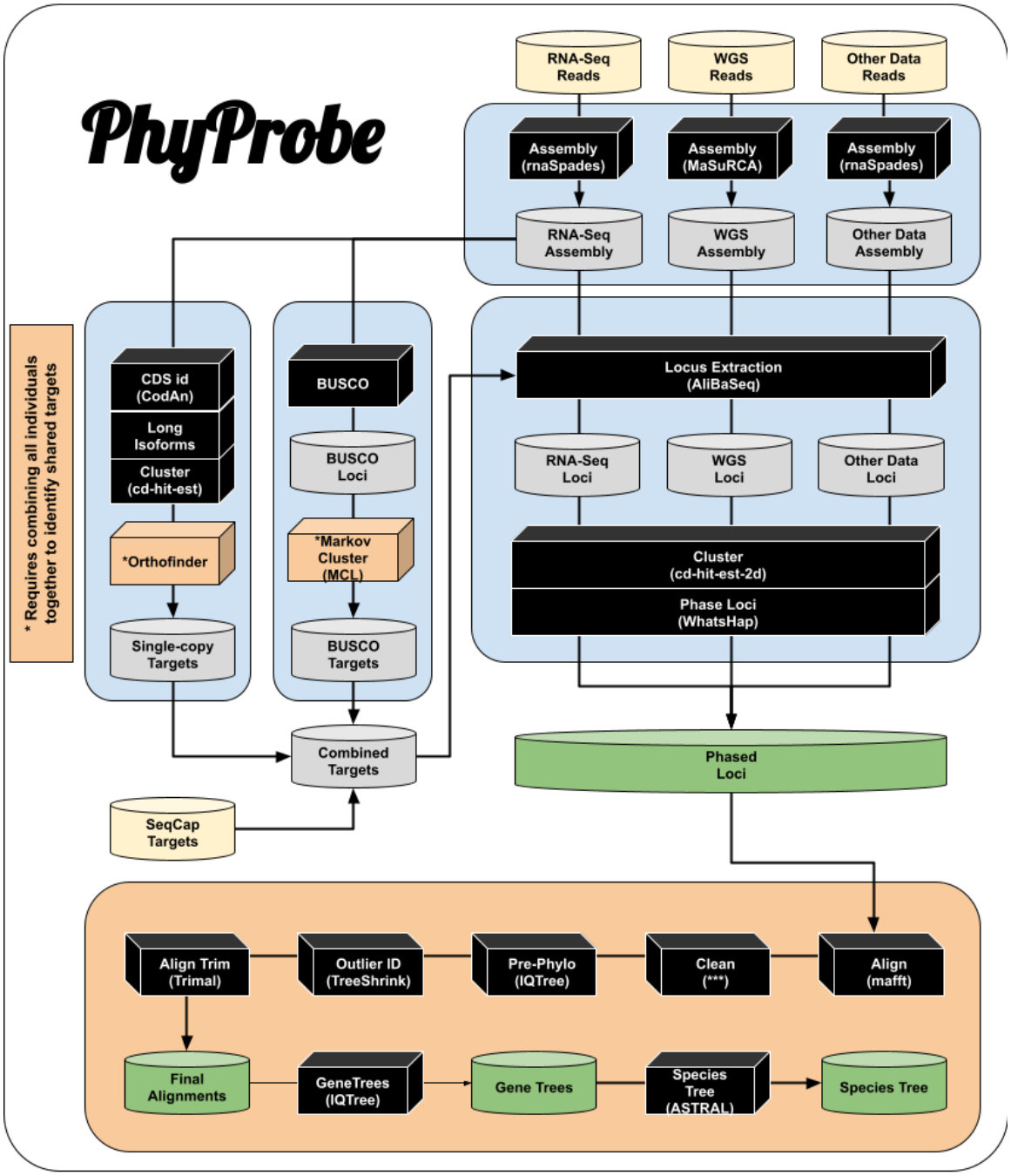

Figure 2.2: *PhyProbe* pipeline developed for multiomic loci extraction and phylogenomic inference. RNA-Seq assemblies were used to extract BUSCOs and novel transcripts. Individuals are clustered together to produce orthologous databases. Loci are then re-extracted along with other targets with *AliBaSeq*. Loci are clustered per individual to reduce redundancy across locus sets and then phased to produce haplotypes. Haplotypes are aligned for all individuals and then cleaned thoroughly prior to gene tree and species tree inference.

We used *BUSCO* v4.1.4 on the transcriptome assemblies to extract single-copy orthologs using the "tetrapod_odb10" set of 5,310 genes from OrthoDB[83,84]. We then used a custom script to search through the multi-copy loci for sequences with 100% identity (*i.e.*, false-negative single-copy loci) and added these sequences to the single-copy loci set. Given that *BUSCO* frequently assembles paralogous or incorrect loci, we then combined all individuals together and performed an all-vs-all *blastn* search (e-value = 1^{-10}) for each locus followed by Markov Clustering (*MCL*; -I 3.0). *MCL* separates dissimilar sequences thereby removing potential paralogs identified as an ortholog by *BUSCO*. We retained the largest cluster from each locus and used this as the reference database to re-extract BUSCO loci from each individual with *ALiBaSeq*[78]. Specifically, we performed *dc-megablast* to query the reference BUSCO database against our transcriptome assemblies (e-value = 1^{-10}). We then used *AliBaSeq* to amalgamate *BLAST* hits in our assembly and extract the full locus (`-x a -f S -e 1e-10 --is --amalgamate-hits`). We then cleaned *AliBaSeq* results by removing loci from each individual if they had less than 50% of the total reference length.

We also used *Python* and *Anolis* SqCL loci as reference sequences to extract commonly used phylogenomic loci (SqCL)[79] with *AliBaSeq* as above. Finally, we attempted to extract additional transcript sequences. First, we used *CodAn* to identify all coding regions in our transcriptome assembly with the "VERT_full" model[85]. We kept the longest assembled isoform and clustered all sequences at 100% sequence identity using *cd-hit-est*[86] to reduce redundancy. To create a finalized set of transcripts, we ran *Orthofinder* v2.5.4[87] on the clustered loci across all individuals. Single-copy orthologous loci identified by *Orthofinder* were then used as the final "orthogroup" set and were re-extracted from each transcriptome assembly using ALiBaSeq as above.

2.2.2 WGS & SeqCap

We trimmed adapters and low quality reads using *Trim Galore!* (v0.4.4; https://github.com/FelixKrueger/TrimGalore) with default settings (`-q 20 -l 20`). We *de novo* assembled the WGS data using *MaSuRCA* v4.0.5[88] with default settings. SeqCap data was *de novo* assembled using *Trinity* v2.11.0[81] or *rnaSPAdes* v3.13.1[82] with default settings. We then used *ALiBaSeq* as above to extract BUSCO, SqCL, and the novel transcripts from the assemblies using the RNA-established databases as reference.

2.2.3 Combining Datasets

With all three locus sets extracted (BUSCOs, SqCL, and orthogroups), we then combined the final extracted loci for each individual and used cd-hit-est-2d at 99% sequence identity to reduce redundancy, keeping the longest sequence. This was necessary because UCEs and other loci were often found nested within or were identical to longer BUSCO sequences. With the final loci, we then performed phasing for each individual by mapping reads to the final loci with *minimap2* (-x sr)[89] and identified SNPs with *bcftools mpileup/call* with default settings. We removed indels and used *WhatsHap*[90] to phase the loci and *bcftools consensus* with the original reference loci to generate the final phased haplotype sequences.

2.3 Phylogenomics

2.3.1 Sequence Filtering

With the final combined dataset, consisting of 11,327 loci, we performed several filtering steps to ensure orthology and clean alignments prior to phylogenetic inference. First, we aligned each gene using *MAFFT* (`--auto --adjustdirectionaccurately`)[91]. We then used *CIAlign*[92] to clean the alignment and remove divergent sequences ($> 80\%$ divergence) which may be the result of extracting a paralogous gene. We also removed short sequences, insertions, and cropped ends with default *CIAlign* settings. We then used *trimal* (`-gappyout`)[93] to remove gappy sites from the alignment and then removed sequences where $> 30\%$ of the total alignment length consisted of gaps with a custom script. Next, we used these preliminary alignments to infer gene trees using *IQTree2* with default settings and 1000 UltraFast bootstraps[94–96].

We used the preliminary gene trees as input to *TreeShrink*[97], which searches for abnormally long branches across trees and removes these taxa from the tree/alignment. *TreeShrink* uses the gene trees to estimate a species or taxon-specific expectation for it's influence on tree diameter. It then systematically removes taxa if its inclusion in the tree is significantly greater than its own expectation/threshold (*i.e.*, long branch outliers). *TreeShrink* is conservative in taxon removal, only removing taxa if they exceed their own expectation which is likely due to sequence extraction errors such as paralogous gene inclusion. *TreeShrink* also adequately controls for taxa with ubiquitous long branches such as outgroups – ensuring they are not wrongly removed.

We re-cleaned the alignments with *CIAlign* (default settings) and *trimal* (`-gappyout`). Finally, we filtered for missing data using a series of three filters. We removed (1) sequences where $> 30\%$ of the total alignment length consisted of gaps, (2) loci that had less than 25% of the samples represented, and (3) samples that had less than 5% of the loci represented.

2.3.2 Species Tree Inference

To infer our final species tree, we employed two different methodologies: (1) *Astral-III*[98–100] and (2) *SVDquartets*[101]. For all analyses, we used *Anolis carolinensis* as the outgroup. For Astral-III, we inferred gene trees using *IQTree2* as above. We collapsed low support branches (*i.e.*, bootstrap support $< 10\%$) and then used *Astral-III* to infer the final lineage and species trees. For *SVDquartets*, we concatenated loci together and used *SNP-Sites*[102] to extract the polymorphic sites. We sampled 1,000,000 quartets with 1,000 bootstrap replicates.

2.3.3 Informativeness & Discordance

To examine phylogenetic informativeness of the different types of loci, we estimated site-specific substitution rates using *IQTree2*. We then plotted the average informativeness profiles generated with *PhyInformR*[103] for each locus type (*i.e.*, BUSCO, AHE, UCE, orthogroups). To assess discordance in our tree we visualized the alternate topologies produced by *ASTRAL*.

2.4 Divergence Dating, Biogeography, & Diversification

To estimate divergence times, we used a penalized likelihood approach implemented in treePL[104]. Briefly, for species tree divergence dating, we first subset our data to the individual with the highest gene coverage per species. We then used *IQTree2* to constrain the topology of our tree and rescale our branch lengths from coalescent units to substitutions per site using a concatenated gene matrix under the GTR+G substitution model. We chose this model because it is very general; commonly used for phylogenetics and divergence dating[51] and leads to similar inferences as other models while avoiding time-consuming model selection steps[105]. We calibrated our tree using eight commonly used fossil calibration points in snakes following Zaher et al.[69] and Burbrink et al.[106] (see Appendix B). Calibration points included stem Alethinophidia (Haasiophis terrasanctus; 93 Mya), stem Boinae (Titanoboa cerrejonensis; ¿58 Mya), stem Colubroides (Procerophis sahnii; ¿54

Mya), stem Elapidae (Elapidae indet.; ¿24.9 Mya), crown Natricidae (Natrix longivertebrata; ¿13.8 Mya),stem Dipsadidae (Paleoheterodon tiheni; ¿12.5 Mya), stem Viperinae (Vipera cf. V. antiqua; ¿22.1 Mya), and stem Crotalinae (Sistrurus sp. indet.; ¿10.3 Mya). For full justification in fossil choice and nodal placement see Zaher et al.[69] and Appendix B. Calibration was constrained to minimum ages to account for uncertainty in the maximum estimates. We then used treePL prime to determine the best optimization parameters and used cross validation to identify the best smoothing parameter. Final divergence dating was performed on 1,000 bootstrap replicates with varying branch lengths from *IQTree2*. We summarized results with TreeAnnotator (0% burnin) to calculate the mean node ages and 95% confidence intervals[107].

To reconstruct the biogeographic history of the group and to test for evidence of competition on venom evolution, we must first identify where each species occurs and which species may be competing. To do this, we developed *VenomMaps* - a database of manually curated distribution maps and species distribution models (SDMs) for all 160 species of New World pitvipers[108]. Distribution maps provided broad scale information on the distribution of species while SDMs can be used to assess more fine scale information on species presence/absence. *VenomMaps* is available in an intuitive user-friendly, publicly-accessible GitHub repository and Shiny application[108] (Fig. 2.3). See Rautsaw et al.[108] for full details on distribution curation and SDM inference. Briefly, we cleaned occurrence records obtained from numerous sources including GBIF (downloaded 2021-08-19)[109], Bison (downloaded 2021-08-19) Bison.usgs.gov2021BisonViperidae., HerpMapper (only New World taxa; downloaded 2021-08-19)[110], Brazilian Snake Atlas[111], BioWeb (downloaded 2021-07-07)[112], unpublished data/databases from collaborators, and georeferenced literature records totaling 373,673 species-level records, 292,425 of which are New World pitvipers. Next, we manually curated distribution maps in QGIS using previous distributions/publications, the occurrence records, a digital relief map (maps-for-free.com), and The Nature Conservancy Terrestrial Ecoregions (TNC.org) to identify clear distribution boundaries. Old World viper distributions were taken from Roll et al.[113]. We then used the cleaned occurrence records to infer SDMs for 140 species with sufficient data. To do this, we developed *autokuenm* - a unix-executable R script designed to take occurrence records and environmental data and prepare these data for species distribution modeling with *kuenm*[114] (Fig. 2.4). Briefly, *autokuenm* filters occurrence records for each species and identifies their respective M-areas based on alpha-hull overlap with biogeographic regions. It then crops environmental areas and a kernel density bias layer to the M-area. Environmental layers are selected to minimize

variable inflation factors (VIF) and maximize MaxEnt variable importance and occurrence records are partitioned into training and testing datasets for final modeling with *kuenm*. We created candidate calibration models with multiple combinations of regularization multipliers (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 8, 10), feature classes (l, q, h, lq, lp, lt, lh, qp, qt, qh, pt, ph, th, lqp, lqt, lqh, lpt, lph, lth, qpt, qph, qth, pth, lqpt, lqph, lqth, lpth, qpth, lqpth), and sets of environmental predictors (bioclimatic, topographic, combination) totaling 2,958 candidate models per species. Models were compared and the best models selected using statistical significance (partial ROC), prediction ability (omission rates; OR), and model complexity (AICc).

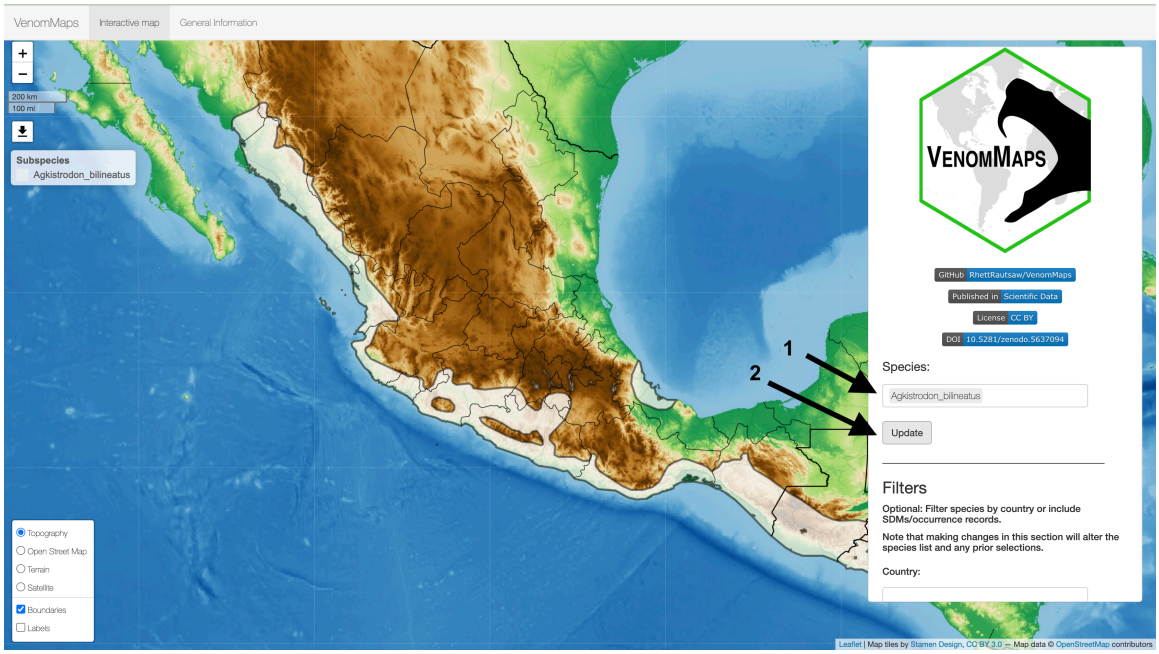

Figure 2.3: Screenshot from the *VenomMaps* server. See Rautsaw et al.[108] for more details.

We used *BioGeoBEARS*[115,116] to infer the biogeography of Viperidae. For biogeographic regions, we merged regions from several previous studies[117–120] and The Nature Conservancy Terrestrial Ecoregions. Species were assigned to regions if $> 15\%$ of their distribution (see *VenomMaps* above) occurred in that region or if $> 50\%$ of the region was overlapped by the species' distribution. All species biogeographic region assignments were then manually checked and updated. We calculated the distance between and adjacency of regions and used this as input in the *BioGeoBEARS* analysis. This reduced the number of biogeographic region combinations possible and ensured that species could not directly disperse across distant regions (*e.g.*, Africa to South America). We per-

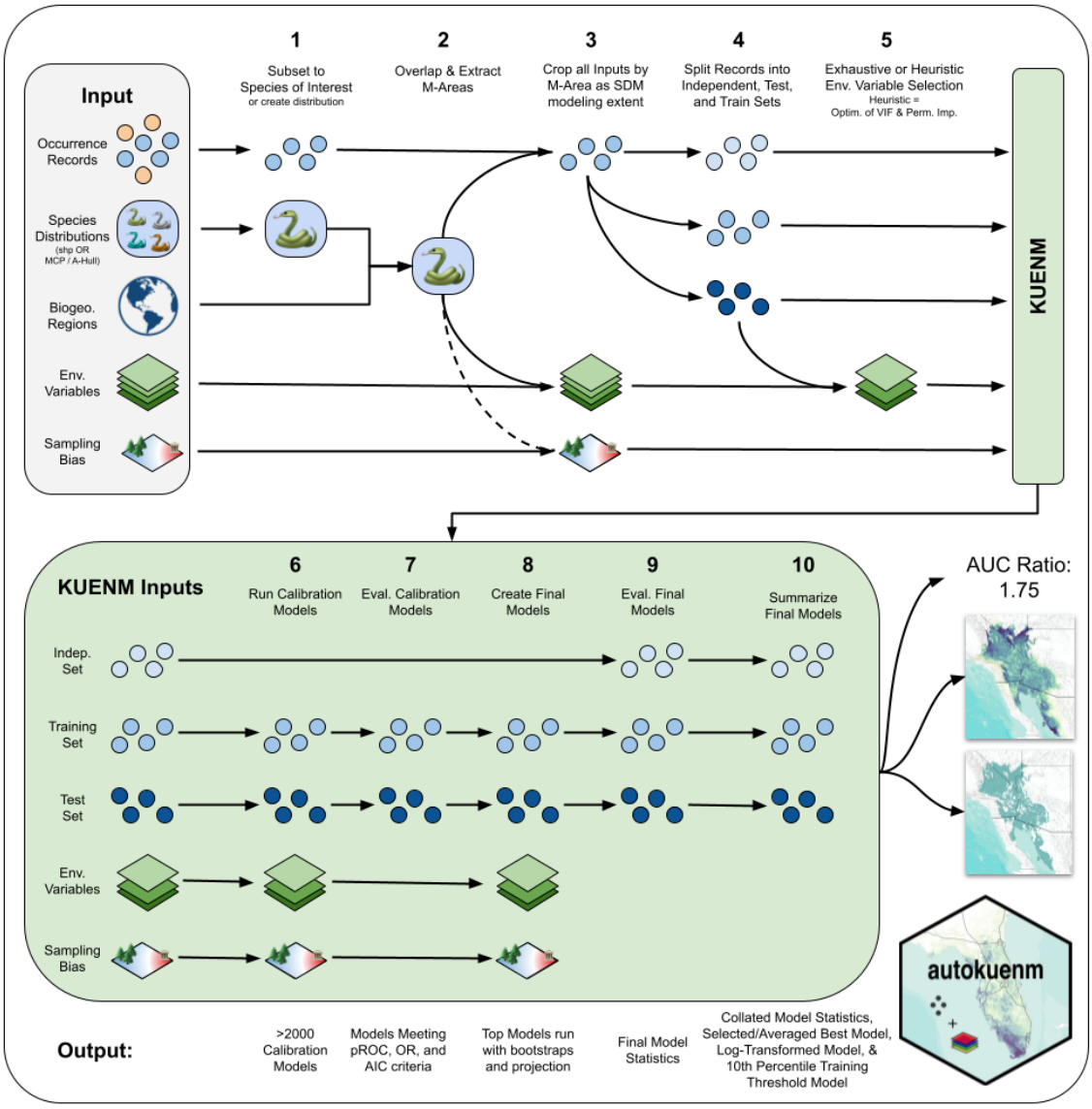

Figure 2.4: Flowchart of the *autokuenm* pipeline. See Rautsaw et al.[108] for more details.

formed three rounds of biogeographic inference with increasing levels of spatial resolution (regions, dominions, provinces). For each of these analyses, we restricted the maximum range size to two, three, and four regions, respectively. Additionally, given the greater number of areas and computational constraints in the province-level analysis, we analyzed subsets of the New World Crotalinae phylogeny separately. For all three spatial scales, we compared DEC, DIVALIKE, and BAYEAREALIKE models with AICc within *BioGeoBEARS*[115,116].

To test for shifts in diversification rates, we used `MiSSE` – a state-dependent speciation-

extinction model which uses hidden states to infer diversification rate shifts - from the R package hisse[121]. Due to computational constraints, we allowed for a maximum of 4 parameters for turnover and extinction processes (*e.g.*, 3 turnover processes:1 extinction).

2.5 Venom Gland Transcriptomics

To quantify venom expression, we used venom gland transcriptomics of New World pitvipers (*i.e.*, Old World taxa were excluded from all downstream analyses due to a lack of sampling). Briefly, we merged the cleaned paired-end reads (see above) using *PEAR*[80] and assembled the transcriptomes following Holding et al.[122] using three *de novo* assemblers: *Extender* (overlap parameter set to 120)[123], *SeqMan NGen* (default settings; Lasergene DNAStar, Madison, WI; https://www.dnastar.com/t-nextgen-seqman-ngen.aspx), and *Trinity* (default settings)[81]. We then combined assemblies and clustered at 100% identity using *cd-hit-est*[86] to remove redundancy. To identify and annotate venom toxins, we developed *ToxCodAn* which utilizes trained Hidden Markov Models to extract toxins from the transcriptome assemblies (Fig. 2.5; See Nachtigall & Rautsaw et al.[124] for more details). After annotation with *ToxCodAn*, we combined the redundancy-filtered toxins and signalp-filtered putative toxins as our preliminary toxin set. We further annotated the nontoxin sequences identified by *ToxCodAn* with *CodAn*[85].

Next, we applied a variety of filters to ensure quality and confident transcriptome annotation. First, we filtered chimeric sequences using a custom Python script (*i.e., ChimeraKiller*). In this script, reads are mapped to the annotated transcriptome and transcripts with zero coverage at any position are removed. Chimeric transcripts are then reported by searching for a difference $> 75\%$ (`-d 0.75`) in the average length of reads on either side of a given site based on the average read size. Second, we filtered for incomplete coding regions and putatively premature stop codons. Third, we filtered out sequences with unreliable read coverage defined as having $< 5x$ coverage for $> 10\%$ of transcript length. We then clustered transcriptomes for each individual at 99% identity using cd-hit-est[86] to cluster allelic variation[36,48]. For each species, the transcriptomes of multiple samples were then combined and similarly clustered at 98% sequence identity with *cd-hit-est*. We removed any nontoxin sequences from the transcriptome with a blast match (`-perc_identity 98 -qcov_hsp_perc 80`) to *ToxCodAn*-identified toxins in the transcriptome that were undetected by *ToxCodAn*. We then used *RSEM*[125] with default *Bowtie2*[126] settings to estimate expression and

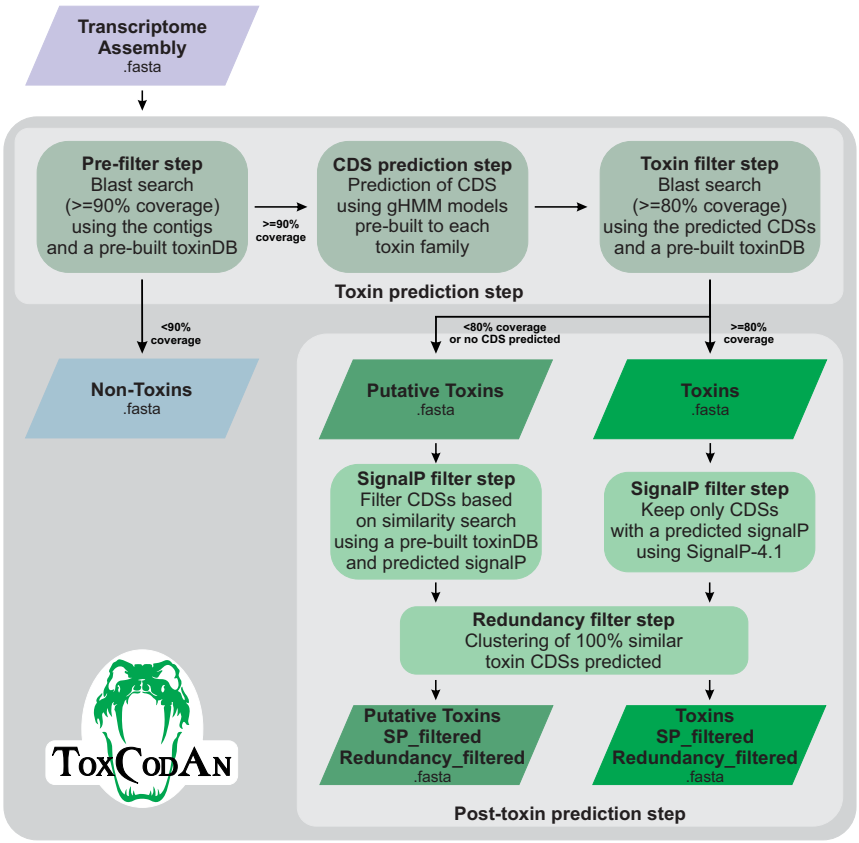

Figure 2.5: Flowchart of the *ToxCodAn* pipeline. See Nachtigall & Rautsaw et al.[124] for more details.

performed a final filtering of toxin transcripts by removing any sequences below the 95th percentile of nontoxin expression. We then recalculated expression with *RSEM* as above with the final transcriptome.

2.5.1 Homology-Weighted Transcriptome Expression Mapping

Comparing transcriptomes and gene expression across species has mostly been done using strict orthologous sequences (*i.e.*, gene lineages related exclusively by speciation events). However, this is often unrealistic in nature as paralogs are frequently produced via gene duplication and deletion events. This is particularly true for venom genes which are known to undergo frequent duplication and subsequent neofunctionalization[40,127,128].

To compare toxin expression across individuals and species, we first calculated TPM10K to normalize transcripts per million (TPM) expression estimates by each species' transcriptome com-

pleteness[129]. We then developed *TransMap* to perform pairwise homology-mapping of expression into a joint manifold – similar to the methodology of *SAMap*[130]. Specifically, we performed pairwise, reciprocal *blastn* searches of species' transcriptomes to build a homology graph (`-max_hsps 1 -qcov_hsp_perc 50 -evalue 1e-10`). We included a `-perc_identity` cutoff dependent on the phylogenetic divergence between two species using our divergence-dated tree above with closely-related taxa having a higher threshold requirement for defining homology than distantly related taxa. We used bitscores for the homology metric in the graph. We replaced any missing genes in the graph with zeros and inflated the graph by a power of 10 to ensure that closely-related genes are more heavily weighted. We then mapped the log2-transformed TPM10K expression of each individual/species to every other species' transcriptome by taking the homology-weighted sum of expression (Fig. 2.6).

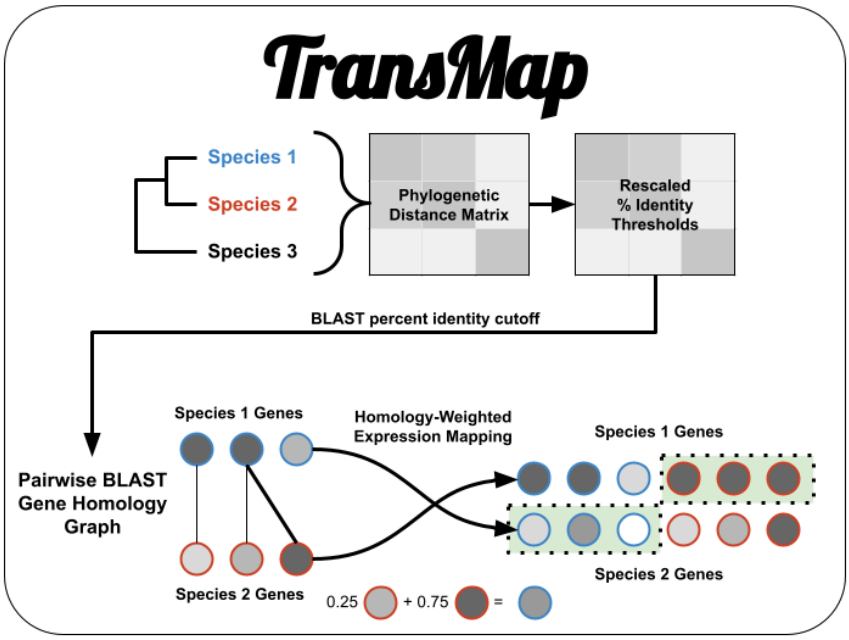

Figure 2.6: Flowchart of the *TransMap* pipeline. Species tree is used to construct a matrix of phylogenetic distance and is then rescaled as *BLAST* percent identity cutoffs. Pairwise, reciprocal *BLAST* searches are performed between species transcriptomes and expression of a species is mapped to all other species genes via homology-weighting.

There were over 8,000 toxin genes across all species; therefore, we then reduced dimensionality of the data for downstream analyses. To do this, we applied three different methodologies: (1) principal component analysis (PCA) and (2) phylogenetic PCA (pPCA) in R and (3) variational autoencoding (VAE). For the VAE, we developed *pyVAE* which is a modified version of *popVAE*[131] – a program for population structure analyses using VAEs from SNP data.

2.6 Phylogenetic Comparative Modeling

To test if competition plays a role in the diversification of venom in New world pitvipers, we used a phylogenetic comparative modeling approach. Specifically, we fit models of trait evolution to each of the latent axes (*i.e.*, 5 PCs, 5 pPCs, 2 VAEs) in *RPANDA*[132] and compared model fit with AICc. As null models, we fit Brownian Motion (BM)[133], Ornstein-Uhlenbeck (OU)[134], and Accelerating/Decelerating (ACDC)[135] models of trait evolution. Competition-based models of trait evolution occurred in two categories. The first set of models were Diversity-Dependent (DD) linear and exponential models[14] where the rate of trait evolution is dependent on the number of species that occur in a given area through time. The last set of models were Phenotype-Matching (PM) models[19,20,23,136] which specifically model the process of character displacement whereby trait evolution is dependent on the phenotypic values of other species. For the PM model set, we tested two variations with and without an internal OU parameter. Both the DD and PM models utilize biogeographic information to inform the model regarding what species may be interacting and thereby influence trait evolution; therefore, both represent models rooted in competition-driven evolution. For our analyses, we used the dominion-scale *BioGeoBEARs* DEC model in our DD and PM models as this scale of analysis allowed for more fine-scale resolution of biogeography (compared to the region-scale analysis) while retaining high confidence in ancestral states – compared to the province level analysis which could also only be performed on subsets of the tree at a time. To account for uncertainty in our DEC model, we used *BioGeoBEARS* to estimate 50 stochastic biogeographic maps and fit the latent axes to each of these 50 maps. We then looked for concordance of the best fit models across the 50 stochastic maps using AICc Weights as evidence for or against competition as a driver of venom evolution.

To further assess confidence in our results, we performed parametric bootstrapping[137] where the parameters for the top null model and top competition model for each stochastic map (n=50) and each trait (n=12) were pulled and used to simulate data 500 times. For each simulation, the best fit null model and competition model were refit with the simulated data sets (2 simulated datasets * 2 models). Overall, this totaled ~120,000 simulations. We then performed likelihood ratio tests to determine our statistical power in detecting a true difference between the null model and competition models in our venom data.

As there are known issues associated with using latent variables in phylogenetic comparative

modeling[138,139], we also fit PCMs for a subset of the toxin genes and their raw log-transformed TPM10K expression estimates. Specifically, we removed genes with correlations $r > 0.75$, leaving 819 of 8,292 total genes. For each of these 819 genes, we fit BM, OU, ACDC, DD, and PM models and looked for concordance across genes as additional evidence.

2.7 Spatial Community Dynamics

To further explore the influence of competition on trait evolution, we took a spatial community ecology approach to test for trait overdispersion – an expectation of species influenced by competition within a community. Specifically, we first created 0.5 degree grid across North and South America. In each of these cells, we used our *VenomMaps* distributions to determine the presence/absence of each species in each pseudo-community. We used this information to calculate species richness, phylogenetic diversity, and functional diversity in each of the pseudo-communities. Phylogenetic diversity was calculated using Faith's Phylogenetic Diversity - or the sum of the total phylogenetic branch length - in the R package *picante*[140]. Functional diversity was calculated with RaoQ using either the two VAE axes or 5 PC axes in the R package *FD*[141,142]. We chose to use RaoQ over other metrics as it is unweighted and therefore allows for use of presence-absence data and has been shown to be robust for community comparisons[23].

If species richness and functional diversity are related and communities have evolved to maximize functional diversity, this would support competition as a driver of venom evolution. To test this, we then randomized communities while maintaining species richness using the `randomizeMatrix` function in *picante*[140]. We then recalculated phylogenetic diversity and functional diversity in each of the randomized communities. We performed 100 permutations and calculated the mean and standard deviation of phylogenetic and functional diversity estimates across all permutations. We then compared these estimates to our observed phylogenetic and functional diversity by calculating the standard effect size (SES). Here, if values are less than 0, this indicates that the observed communities are less diverse than random chance while values greater than 0 indicate that the observed communities are more diverse.

2.8 Microevolutionary Dynamics

To explore the possibility of competition playing a role at the microevolutionary scale, we focused on a well-studied system in the American Southwest: the Mojave Rattlesnake (*Crotalus scutulatus*)[47,48,143–148]. This species is known to have significant intraspecific venom variation with many individuals maintaining neurotoxic (Type II) venoms and others having hemorrhagic (Type I) venoms. Therefore, to explore the role of competition in this system, we applied a Generalized Dissimilarity Model (GDM)[149] to test whether the presence of other species can explain trait shifts across the landscape. Specifically, GDMs are a non-linear permutation matrix regression that is used to determine what variables best explain dissimilarity among communities or individuals. Therefore, we calculated the euclidian distance in the first two PC axes of venom expression for this species as our metric of venom dissimilarity among individuals. We chose to use PCs here rather than the VAEs, because careful examination demonstrated that these axes better represented the known variation present in this system. For predictors of venom dissimilarity, we included geographic distance, phylogenetic distance, species richness, and several uncorrelated bioclimatic variables (https://www.worldclim.org/data/bioclim.html) – several of which have been previously identified as potentially significant in this system (*e.g.*, bio8, bio11). For species richness, we chose to use our estimated species distribution models as a proxy as these offer more fine-scale resolution of species presence/suitability. Specifically, we summed the suitability values of co-occurring species as our proxy for species richness.

Chapter 3

Results

3.1 Phylogenomics & Systematics

The final dataset consisted of 2,256 loci with an average coverage of 1,368 loci per sample. RNA and WGS datasets recovered significantly more loci ($\overline{x}$ = 1,977 & 583, respectively) while the SeqCap datasets were largely – but not entirely – limited to their targeted AHE loci ($\overline{x}$ = 173; Fig. 3.1). Analysis of phylogenetic informativeness suggested BUSCOs and AHE loci were the most informative loci and UCEs were the least informative at this timescale (Fig. 3.1). Additionally, we found that SVDquartets was largely concordant with the Astral topology with the exception of a few low-coverage samples/species which were placed in a seemingly haphazard fashion in the SVDquartet topology (Fig. 3.2). For this reason, we used the Astral topology for all downstream analyses.

Several generic relationships differed from previous phylogenetic estimates which primarily utilized mitochondrial DNA (Fig. 3.3). First, within Viperinae, we found that *Causus* – a previously unstable genus – was strongly supported as sister to a clade containing *Atheris* + *Bitis* (Fig. 3.4 & 3.5). Together, *Causus*, *Atheris*, and *Bitis* were sister to all other Viperinae (Fig. 3.4 & 3.5). Next, we found *Echis* and *Cerastes* formed two monophyletic lineages; however, we recovered low support for the placement of *Cerastes* (39%; Fig. 3.4 & 3.5). The alternate topologies recovered by Astral for this group consisted of a reversed branching pattern (*i.e., [Cerastes,[Echis,[Other Viperinae]]]*; 35%) or with *Echis* and *Cerastes* as sister genera (27%; Appendix A). We recovered *Pseudocerastes* and *Eristicophis* as sister to each other which together formed the sister clade to

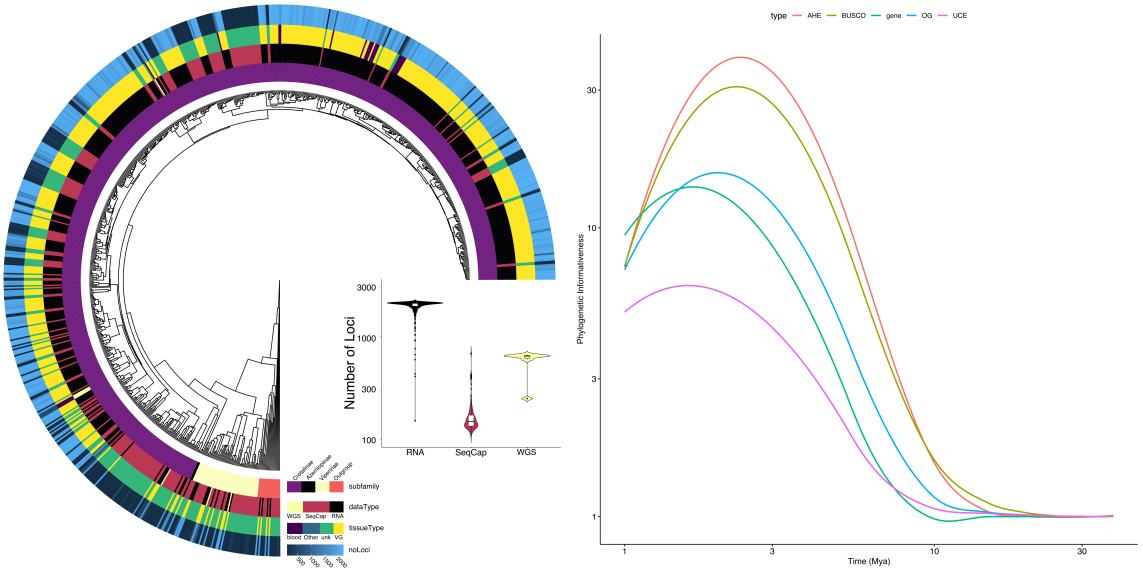

Figure 3.1: Left: Number of loci recovered by each sample categorized by family, data type, and tissue type. SeqCap datasets recover far fewer loci than RNA and WGS datasets. Right: Average phylogenetic informativeness through time reveals that AHE and BUSCOs as the most informative locus types. Abbreviations: AHE (Anchored Hybrid Enrichment); BUSCO (Benchmarking Universal Single Copy Orthologs); gene (Misc. squamate conserved loci); OG (Orthogroups derived from RNA-Seq); UCE (Ultra-Conserved Elements)

(*Vipera* + (*Daboia* + *Montivipera*)) (Fig. 3.4 & 3.5). Notably, we identified *Daboia* as paraphyletic with respect to *Montivipera* due to the basal placement of *Daboia palestinae*; however, this was not strongly supported and additional analyses are needed to confirm this finding (Fig. 3.4 & 3.5).

Azemiopinae was found sister to Crotalinae as expected; however, many of the relationships within Crotalinae were novel (Fig. 3.3, 3.4, & 3.5). Within Old World Crotalinae, we first recovered a strongly supported clade containing two pairs of sister genera: *Calloselasma* + *Hypnale* and *Deinagkistrodon* + *Tropidolaemus* (Fig. 3.4 & 3.5). Our phylogeny also strongly supports *Trimeresurus* and *Protobothrops* as monophyletic sister genera (Fig. 3.4 & 3.5). Finally, *Gloydius* and *Ovophis* also form monophyletic sister genera which together form the sister group to New World Crotalinae (Fig. 3.4 & 3.5). Importantly, we found *Trimeresurus gracilis* and *Ovophis okinavensis* nested within *Ovophis* (see discussion for more details; Fig. 3.4 & 3.5).

Herein, we define several clades for ease of discussion regarding New World Crotalinae. We refer to the collective group of *Agkistrodon*, *Crotalus*, and *Sistrurus* as the "Temperate" clade (Node A in Fig. 3.4), in reference to their biogeographic distributions primarily in temperate North America. The "Tropical" clade includes all remaining New World genera (*Atropoides*, *Bothriechis*,

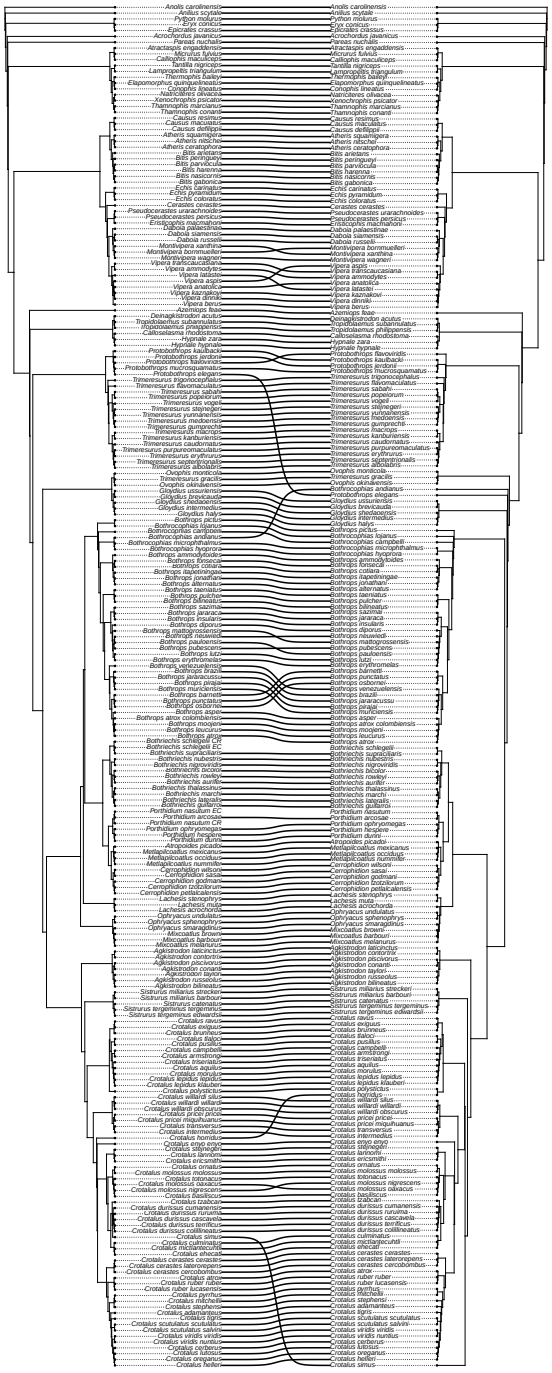

Figure 3.2: Concordance between Astral-III (left) and SVDquartets (right). Generally, the two methods agree with the exception of SVDquartets struggling to place many taxa with low locus coverage or sampling coverage.

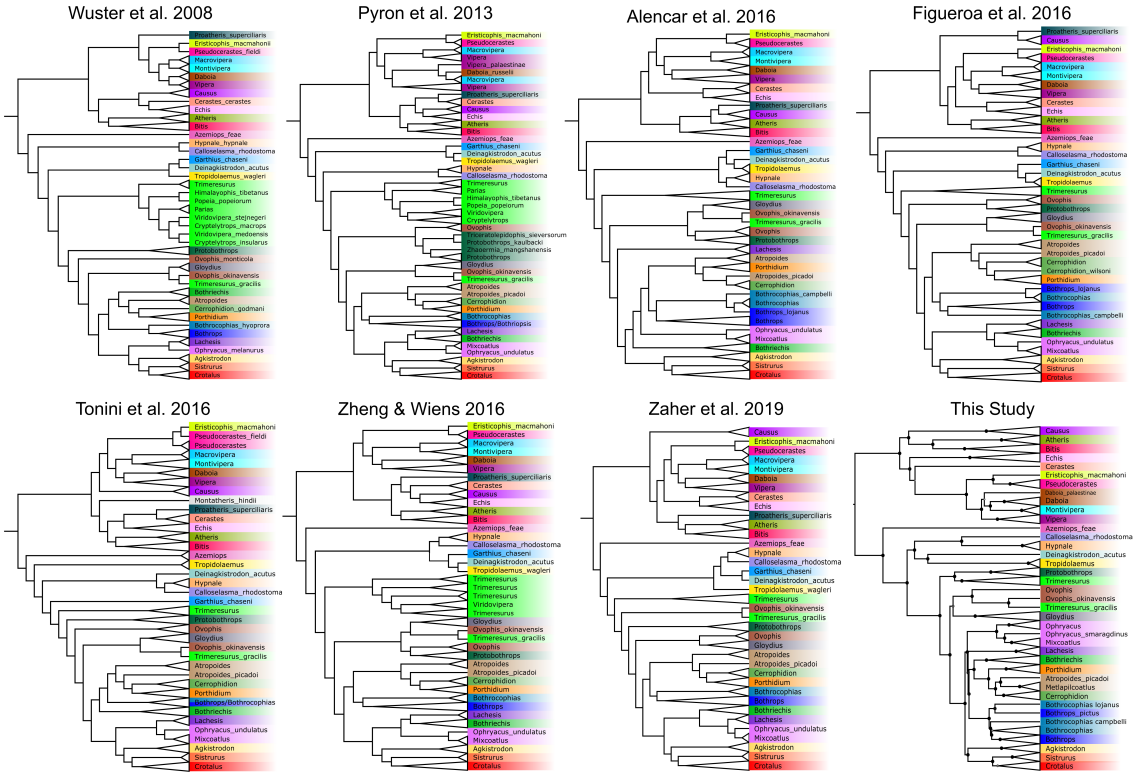

Figure 3.3: Previous phylogenies produced by[61,64–69] as well as our Astral-III topology for comparison. Using phylogenomics, our phylogeny revealed novel relationships across many parts of the tree while supporting several previously recovered relationships.

Bothrocophias, *Bothrops*, *Bothriopsis*, *Cerrophidion*, *Lachesis*, *Metlapilcoatlus*,*Mixcoatlus*, *Ophryacus*, and *Porthidium*; Node B in Fig. 3.4). Within the Tropical clade, the "Latin American" clade is defined as the monophyletic group including *Atropoides*, *Bothriechis*, *Bothrocophias*, *Bothrops*, *Cerrophidion*, *Metlapilcoatlus*, and *Porthidium* (Node C in Fig. 3.4). The "Central American" clade is defined as the monophyletic group including *Atropoides*, *Bothriechis*, *Cerrophidion*, *Metlapilcoatlus*, *Porthidium* (Node D in Fig. 3.4). Lastly, the "Bothropoid" clade includes *Bothrocophias* and *Bothrops* (Node E in Fig. 3.4).

Within New World Crotalinae, we found strong support for initial divergence between the Temperate and Tropical clades with the latter containing previously unstable genera such as *Ophryacus*, *Mixcoatlus*, *Lachesis*, and *Bothriechis* (Fig. 3.4 & 3.5). Within the Tropical clade, we found that *Ophryacus* and *Mixcoatlus* form a monophyletic group which diverged from *Lachesis* and the Latin American clade; however, we recovered *Ophryacus* as paraphyletic with respect to *Mixcoatlus* (see discussion for more details; Fig. 3.4 & 3.5). Next, we recovered *Lachesis* as a monophyletic

lineage sister to the Latin American clade (Fig. 3.4 & 3.5). Within the Latin American clade, initial divergence occurred between the Central American and Bothropoid clades (Fig. 3.4 & 3.5). Interestingly, within Bothropoids, we recovered *Bothrocophias* as paraphyletic (see discussion for more details; Fig. 3.4 & 3.5). Within the Central American clade, *Bothriechis* formed the basal clade sister to the remaining genera (*Porthidium, Cerrophidion, Atropoides, Metlalpilcoatlus*; Fig. 3.4 & 3.5). *Porthidium* formed the next diverging lineage sister to (*Cerrophidion*,(*Atropoides, Metlalpilcoatlus*)) (Fig. 3.4 & 3.5). Interestingly, we recover *Atropoides* and *Metlapilcoatlus* as monophyletic (see discussion for more details; Fig. 3.4 & 3.5).

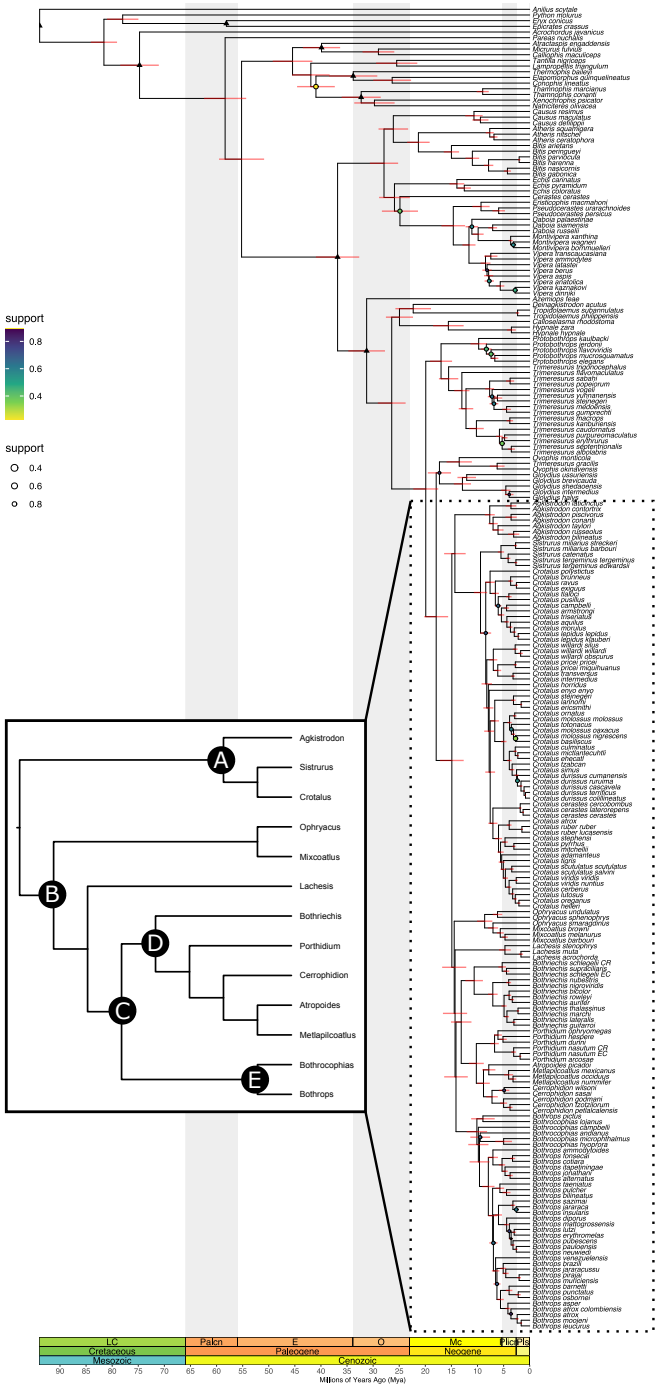

Figure 3.4: Dated species tree of Viperidae with outgroups produced with Astral-III and dated with TreePL. Fossil calibration points are denoted with triangles. Nodes with $< 80\%$ support are indicated with colored circles. Bars indicate the 95% confidence interval for divergence estimates. Inset phylogeny defines clades: (A) Temperate (B) Tropical, (C) Latin American, (D) Central American, (E) Bothropoids

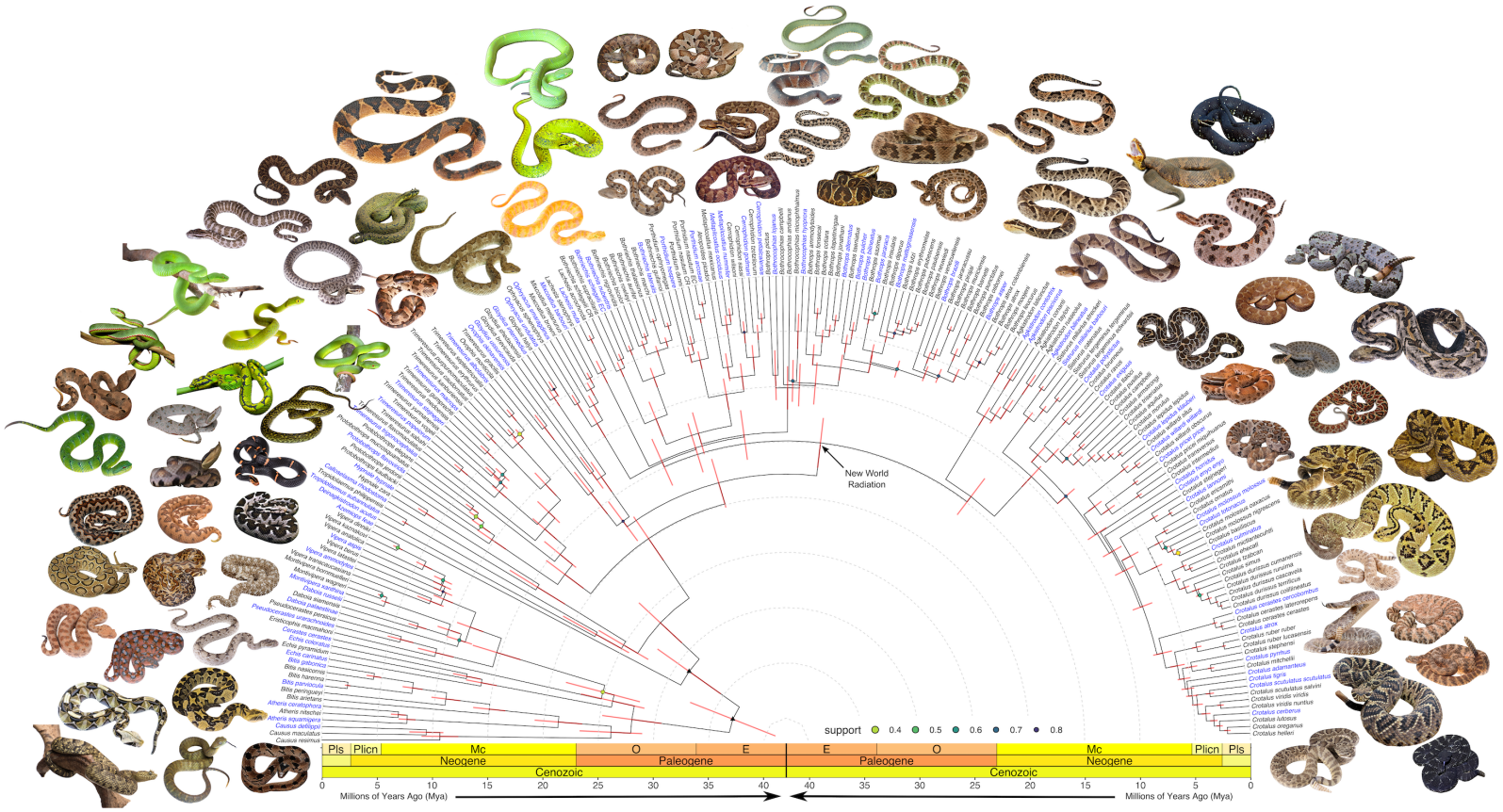

Figure 3.5: Dated species tree for Viperidae produced with Astral-III and dated with TreePL. Fossil calibration points are denoted with triangles. Nodes with $< 80\%$ support are indicated with colored circles. Bars indicate the 95% confidence interval for divergence estimates. Taxa colored in blue represent the taxa pictured next to the phylogeny. Image Credits: Appendix C

3.2 Divergence Dating, Biogeography, & Diversification

Overall, our divergence time estimates are more recent than previous estimates for this clade which used only a few mitochondrial and nuclear loci (Appendix B)[61,65]. Divergence dating of our phylogeny suggests that the ancestor of vipers diverged from its sister group in the late-Paleocene or early-Eocene (59.8-51.2 Mya); however, Viperidae did not begin to diversify until the late-Eocene (41.5-33.2 Mya; Fig. 3.5). Biogeographic reconstruction suggests that Viperidae originated in the Middle East ranging from the Arabian Peninsula to India (Fig. 3.6, & 3.7).

Initial divergence in the Viperinae occurred in the mid-Oligocene (31.5-26.1 Mya) between the predominantly African and Euro-Arabian subclades with origins in present day Ethiopia and Arabia, respectively (Fig. 3.5 & 3.7). An additional deep divergence occurred in the Euro-Arabian clade 28.3-21.3 Mya separating the predominantly European taxa from the Middle Eastern taxa (Fig. 3.7). Azemiopinae and Crotalinae diverged in the early to mid-Oligocene (35.3-28.3 Mya) with a suggested origin in present-day India (Fig. 3.5 & 3.7). Lineages within Crotalinae began to diverge in the mid to late-Oligocene (29.9-24.6 Mya) with the majority of diversification occurring in the early-Miocene (23.0-18.2 Mya) throughout southeast Asia (Fig. 3.5 & 3.7).

Invasion of the New World by Crotalinae occurred in the mid-Miocene (21.1-16.5 Mya) (Fig. 3.5 & 3.7). Following this, there was a rapid radiation of New World pitvipers as shown with high net diversification rates by MiSSE (Fig. 3.8). This was followed by a comparative slowdown in diversification (Fig. 3.8). The most recent common ancestor of New World pitvipers is suggested to have occurred in the mid-Miocene (18.1-13.6 Mya) with a widespread distribution occurring across present-day central North America and south along the eastern coast of Middle America to Costa Rica (Fig. 3.5 & 3.7). *Agkistrodon*, *Sistrurus*, and *Crotalus* originated in the northern extent of this range, largely diversifying within central North America and Tamaulipas while the remaining taxa originated in the southern extent of this range in Middle America (Fig. 3.7).

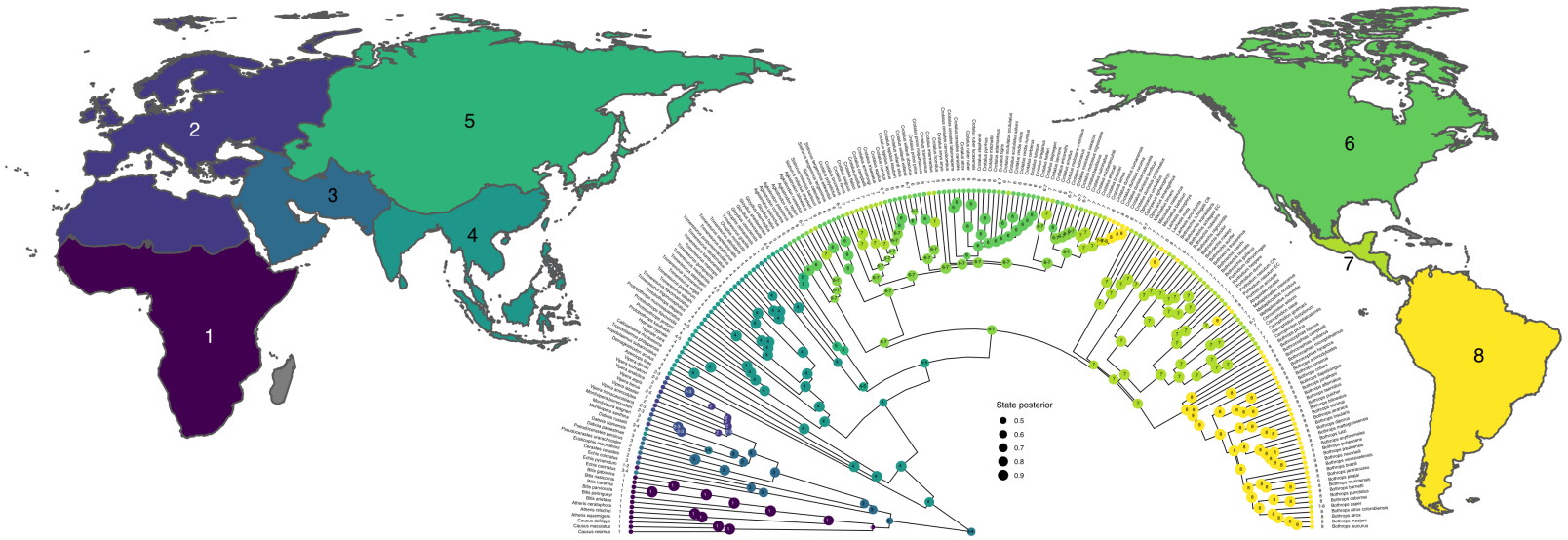

Figure 3.6: BioGeoBEARS analysis of region supports the BAYAREALIKE model.

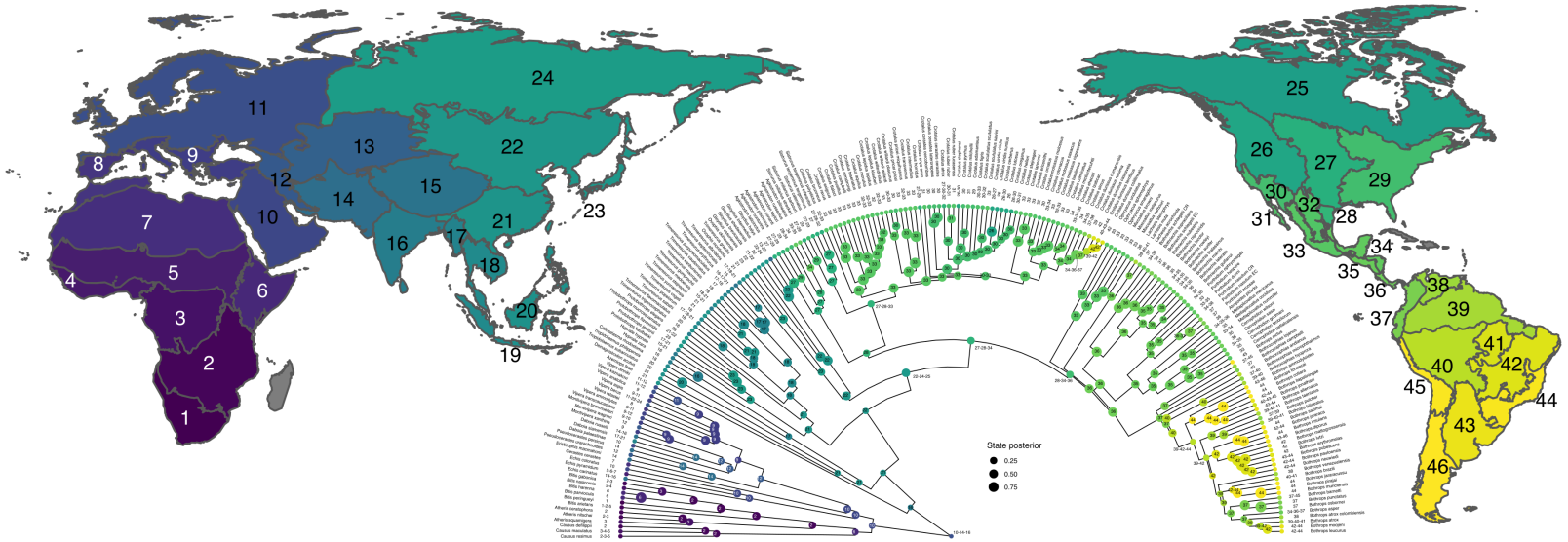

Figure 3.7: BioGeoBEARS analysis of dominions supports the DEC model.

Figure 3.8: Net diversification rates estimated by MiSSE demonstrate rapid radiation of New World Crotalinae followed by subsequent slow downs. Rapid radiations also occurred in the most speciose genera *Crotalus* and *Bothrops*.

Within the Temperate clade, our biogeographic reconstruction suggests a Chihuahuan desert origin for *Crotalus* with the first major divergence occurring 9.8-7.7 Mya between montane, small-bodied rattlesnakes and lowland, large-bodied rattlesnakes with hypothesized origins in the Trans-Volcanic Mexican Belt (TVMB) and Sonoran desert, respectively (Fig. 3.5 & 3.9). The large-bodied rattlesnakes began to diversify ∼9.6-7.6 Mya with an eastern dispersal by *C. horridus* to Central and Eastern North America, a western dispersal into Baja California by *C. enyo*, and a split leading to two larger lineages of rattlesnakes (Fig. 3.9). One lineage rapidly dispersed south along the Mexican Pacific lowlands into the Balsa's Basin and further into Central and South America (Fig. 3.9). The

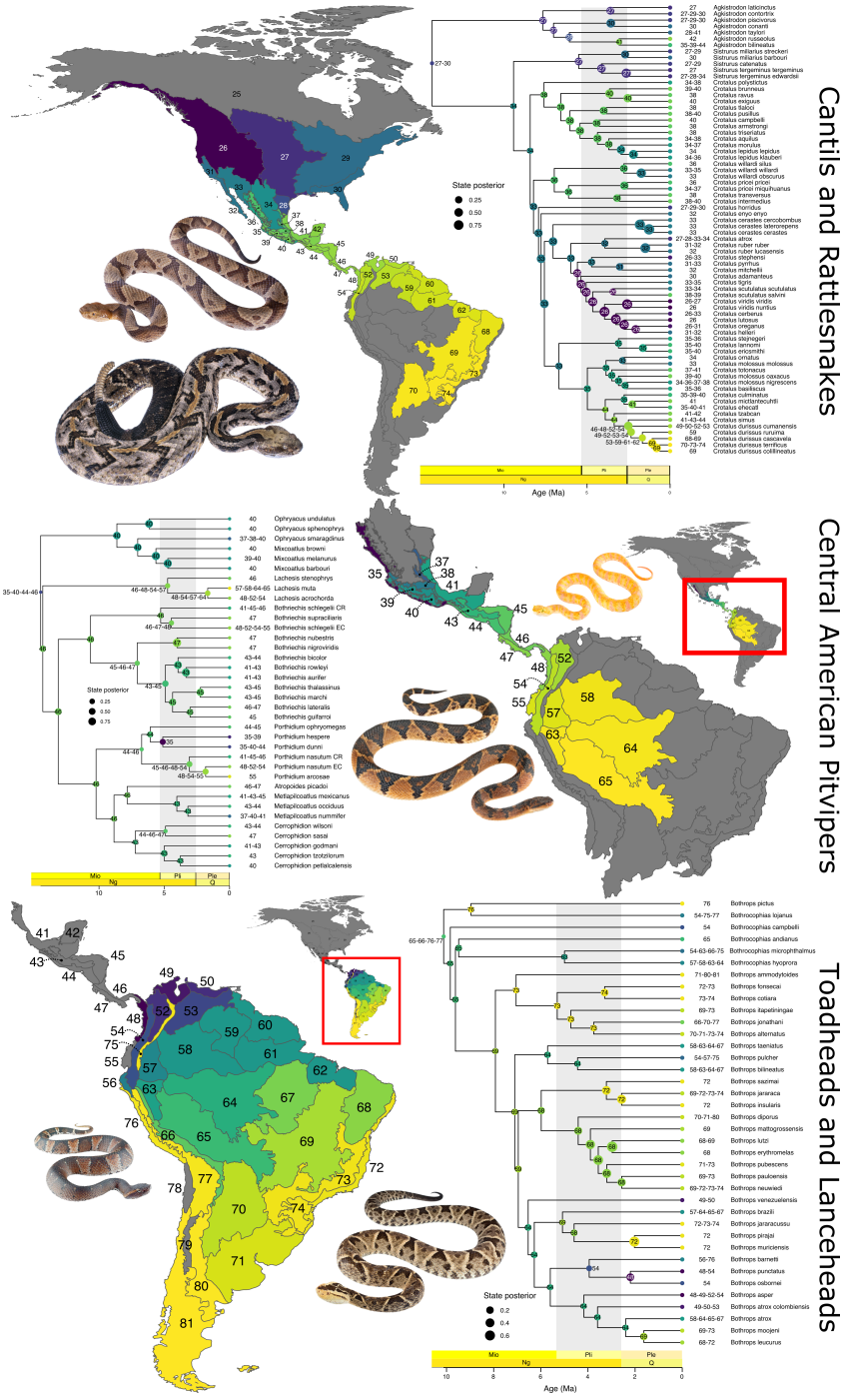

Figure 3.9: BioGeoBEARS analysis of provinces supports DIVALIKE models. Three subset analyses were performed on (1) Cantils and Rattlesnakes, (2) Central American Pitvipers, and (3) Toadheads and Lanceheads.

other lineage diversified across the Chihuahuan, Sonoran, and Mojave deserts (Fig. 3.9). Within the desert lineage, a secondary eastern dispersal occurred by *C. adamanteus* 6.1-5.0 Mya (Fig. 3.5 & 3.9).

The Tropical clade began to diversify 17.7-13.1 Mya, shortly after invasion of the New World with a hypothesized origin from present day southern Mexico to Costa Rica (Fig. 3.5, 3.7, & 3.9). *Ophryacus* and *Mixcoatlus* first diverged from (*Lachesis* + Latin American clade) with suggested origins in the Sierra Madre del Sur of Mexico and Costa Rica, respectively (Fig. 3.5 & 3.9). Within the Latin American clade, Bothropoids diverged from the Central American clade 17.3-12.8 Mya with hypothesized origins in northern South America and Costa Rica, respectively (Fig. 3.5, 3.6, & 3.9). The Central American clade then diversified via multiple dispersal and vicariance events across Central America (Fig. 3.5 & 3.9). For example, *Lachesis* and *Bothriechis schlegelii* dispersed south across the Isthmus of Panama into South America 6.0-3.8 Mya while *Porthidium* crossed more recently 2.1-1.6 Mya (Fig. 3.5 & 3.9). Additionally, *Bothriechis*, *Atropoides/Metlapilcoatlus*, *Cerrophidion*, and *Porthidium* dispersed northward back into Mexico (Fig. 3.5 & 3.9).

Our results indicate that the *Bothropoids* dispersed into South America across the Isthmus of Panama during the mid Miocene (17.3-12.8 Mya; Fig. 3.5 & 3.9). *Bothrocophias* formed several paraphyletic lineages with origins throughout the northern Andes (Fig. 3.5 & 3.9). *Bothrops* form a monophyletic lineage which dispersed into the Brazilian Shield and present-day Cerrado and Chaco regions (12.3-8.5 Mya; Fig. 3.5 & 3.9). Following this, rapid diversification of *Bothrops* began 9.5-7.1 Mya with dispersals into the Patagonian Steppe, Atlantic Forest, and back northward into the Amazon basin and Guiana Shield (Fig. 3.5 & 3.9).

3.3 Phylogenetic Comparative Modeling

3.3.1 Venom Space

We observe significant variation in venom expression (Fig. 3.10 & 3.11). We used the first five axes of our PCA and pPCA in downstream axes which accounted for 35% and 29% of the variation in venom expression, respectively (Fig. 3.12 & 3.13). We used both axes of our VAE; however, as VAE is a machine-learning technique, we are not able to examine contributions. Crotoxin-like Phospholipase A2 (PLA2) homologs explain the most variance in both PC1 and pPC1 (Fig. 3.12 & 3.13). Additionally, C-type lectins (CTLs) contribute highly to PC1 while Snake Venom

Metalloproteases (SVMPs) contribute highly to pPC1 (Fig. 3.12 & 3.13). PC2 is largely separated by SVMPs and Bradykinin-potentiating peptides (BPP) while pPC2 is largely separated by CTLs and Snake Venom Serine Proteases (SVSPs; Fig. 3.12 & 3.13).

These patterns align with several well-known patterns in snake venom variation within New World pitvipers. For example, species which are known to have simple, neurotoxic venoms dominated by crotoxin-like homologs are found clustered closely while *Bothrops* – which have significantly different venoms dominated by CTLs – are found oppositely clustered in venom space (Fig. 3.10 & 3.11).

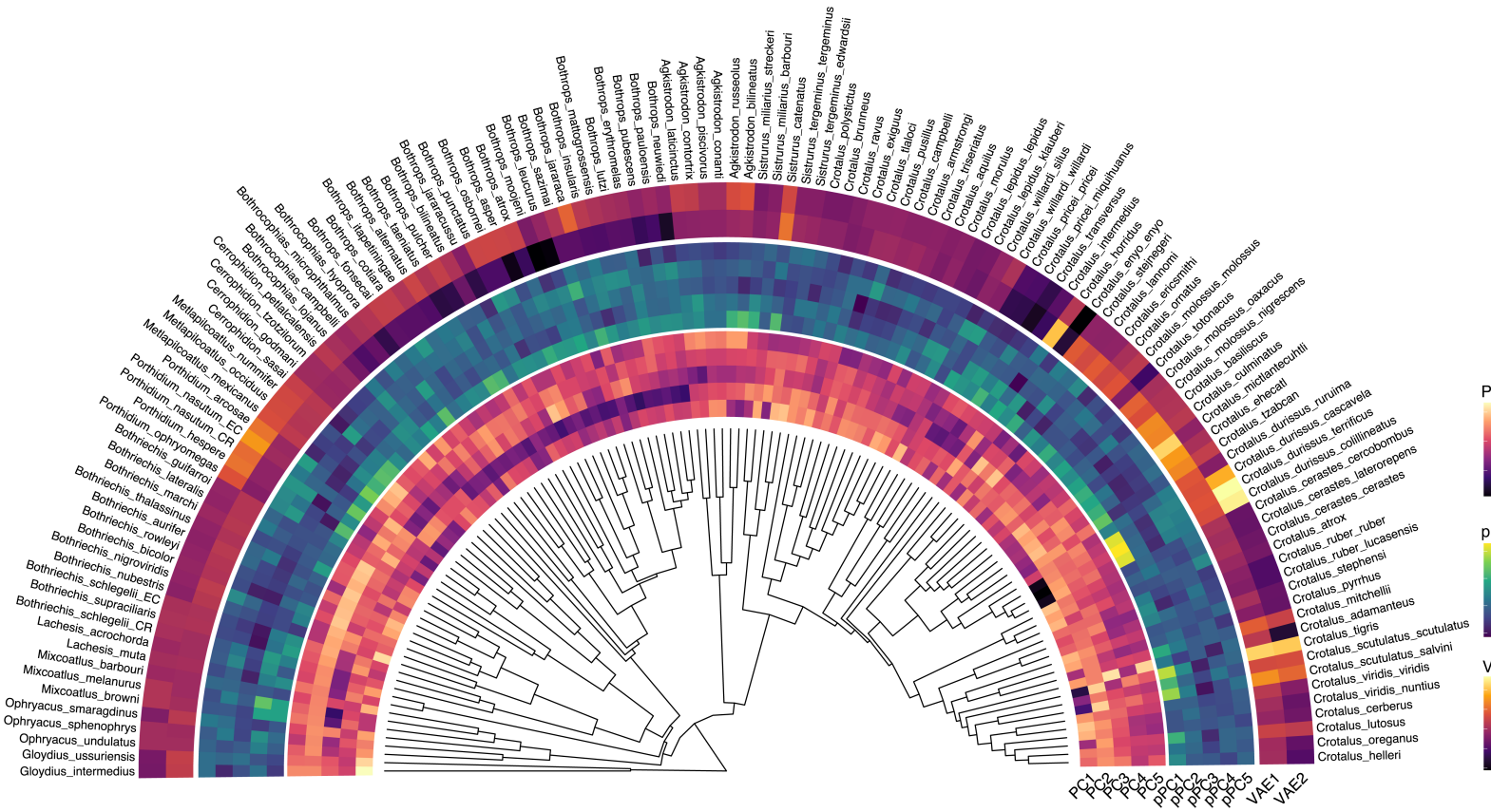

Figure 3.10: Species tree with heatmap of the mean value for the first five PCA and pPCA axes as well as the two VAE axes. Axes are created from the homology-weighted transcriptome expression map. Crotoxin-dominated venoms found in taxa such as *Crotalus tigris*, *C. durissus*, *C. scutulatus*, *C. horridus*, and *Sistrurus catenatus* make up the majority of variation in the first two axes of each technique.

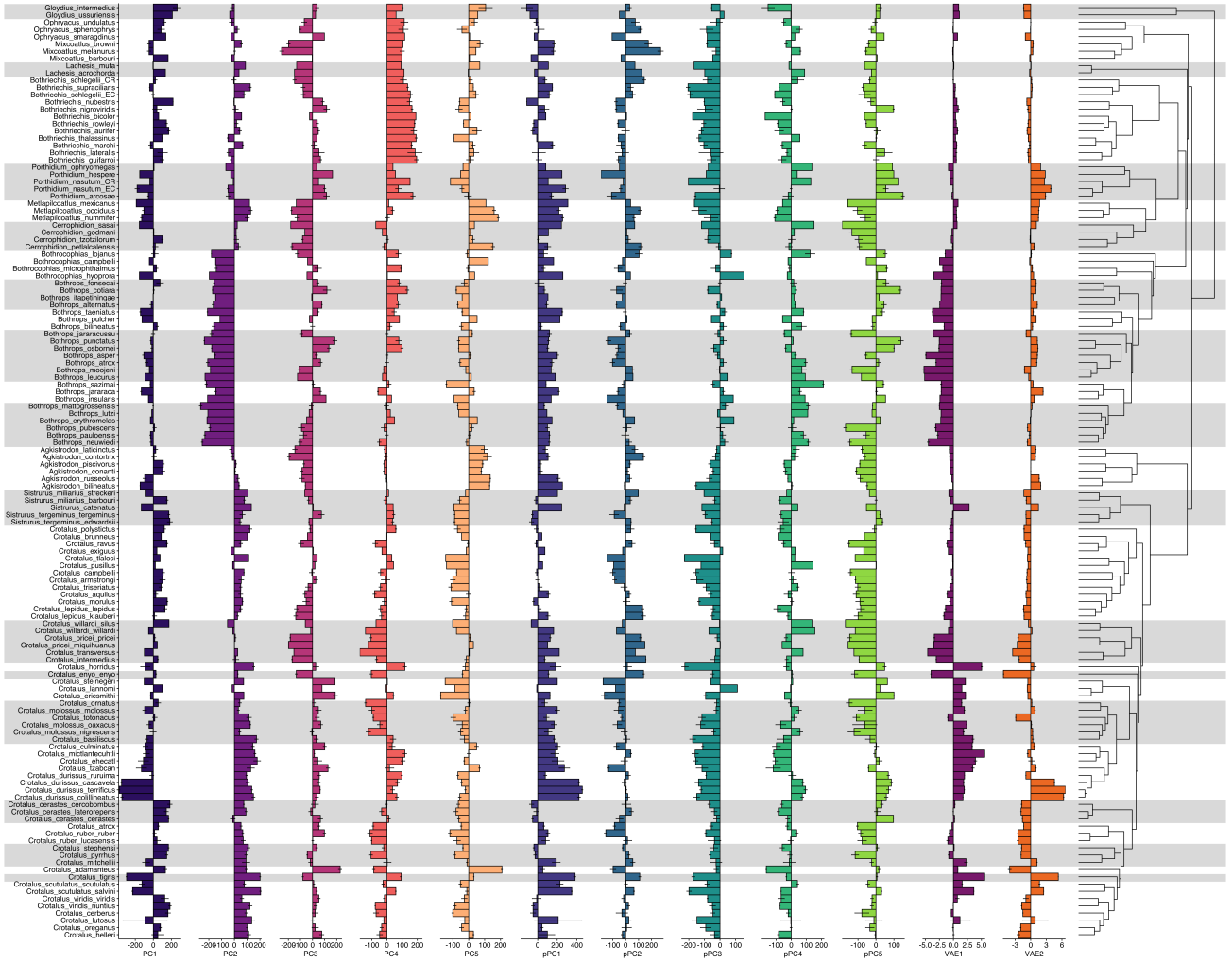

Figure 3.11: Species tree with barplots of the mean and standard errors for the first five PCA and pPCA axes as well as the two VAE axes. Axes are created from the homology-weighted transcriptome expression map. Crotoxin-dominated venoms found in taxa such as *Crotalus tigris*, *C. durissus*, *C. scutulatus*, *C. horridus*, and *Sistrurus catenatus* make up the majority of variation in the first two axes of each technique.

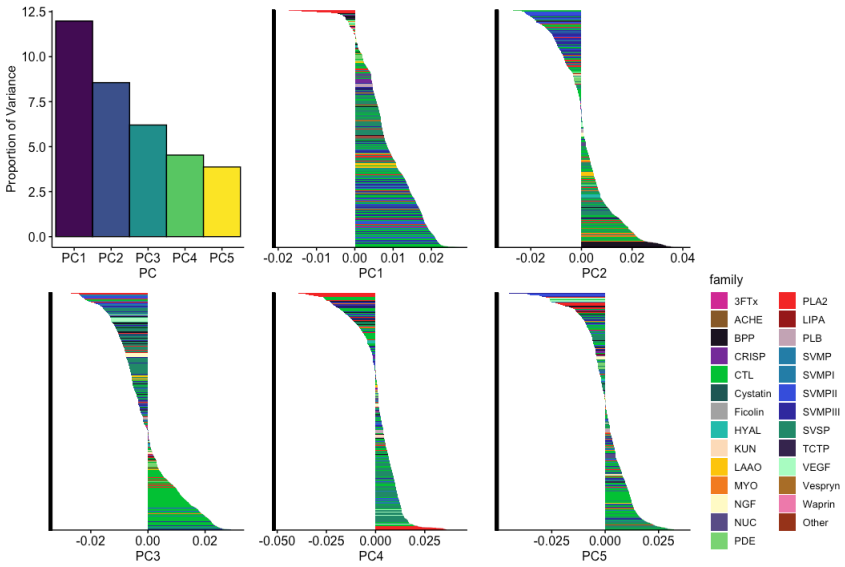

Figure 3.12: The proportion of total variance explained by the first five PCs as well as the variable contributions in each axis colored by toxin family. Crotoxin-like PLA2 and CTL expression largely separate PC1, while BPP and SVMP expression separate PC2. Toxin Abbreviations: 3-finger toxin (3FTx), Acetylcholinesterase (ACHE), Bradykinin-potentiating peptides (BPP), Cysteine-rich secretory proteins (CRISP), C-type lectins (CTL), Hyaluronidase (HYAL), Kunitz-type proteinase inhibitor (KUN), L-amino acid oxidase (LAAO), Myotoxin (MYO), Nerve growth factor (NGF), Ecto 5' nucleotidase (NUC), Phosphodiesterase (PDE), Phospholipase A2 (PLA2), Lipase (LIPA), Phospholipase B (PLB), Snake venom metalloproteinase (SVMP), Snake venom serine protease (SVSP), Translationally Controlled Tumor Protein (TCTP), and Vascular endothelial growth factor (VEGF).

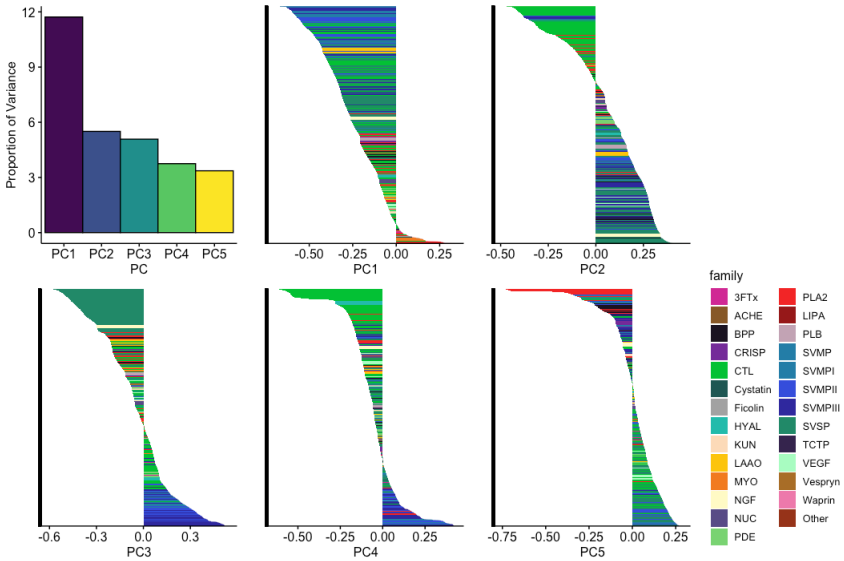

Figure 3.13: The proportion of total variance explained by the first five pPCs as well as the variable contributions in each axis colored by toxin family. Crotoxin-like PLA2 and SVMP expression largely separate pPC1, while CTL and SVSP expression separate pPC2. Toxin Abbreviations: See Fig. 3.12.

3.3.2 Modeling Results

We compared several traditional models of trait evolution (*i.e.*, Brownian Motion [BM], Ornstein Uhlenbeck [OU], Accelerating/Decelerating [ACDC]) to competition-based models which are dependent on species density and traits present across geographic space (Diversity-Dependent [DD] and Phenotype Matching [PM]). Across most biogeographic stochastic maps for each latent PC, pPC, and VAE axis, competition-based models were unanimously favored (Fig. 3.14, 3.15, 3.16). In particular, the DD models were largely favored with strongly positive diversity dependence ($\bar{r}_{DDexp}$ = 3.98; $\bar{b}_{DDlin}$ = 10.24) suggesting that venom diversity increases with lineage diversity. Similarly, although the PM models were only supported for PC2 and pPC3, on average the PM models resulted in strongly negative interaction parameters ($\bar{S}_{PM}$ = -0.93; $\bar{S}_{PMOU}$ = -0.28) further suggesting an influence of competition and potentially character displacement.

Parametric bootstrapping showed low model identifiability and statistical power for the PC axes and first several pPC axes. However, we have good model support and power for the VAE axes and later pPCs (Fig. 3.17). Close examination of these results demonstrated that low-support and statistical power was largely attributable to bimodal support across stochastic biogeographic maps with 25-50% of the stochastic maps favoring null models over competition-based models depending on the trait/axis being examined. To provide additional support for our results, we fit models using the original TPM10K expression of 819 uncorrelated ($r < 0.75$) genes and looked for concordance across these genes. As expected, we see more variability in model support across genes, with some strongly favoring the "null" models (*i.e.*, BM, OU, & ACDC); however, DD models remained the most highly supported models (Fig. 3.18) with support across multiple toxin families (Fig. 3.19).

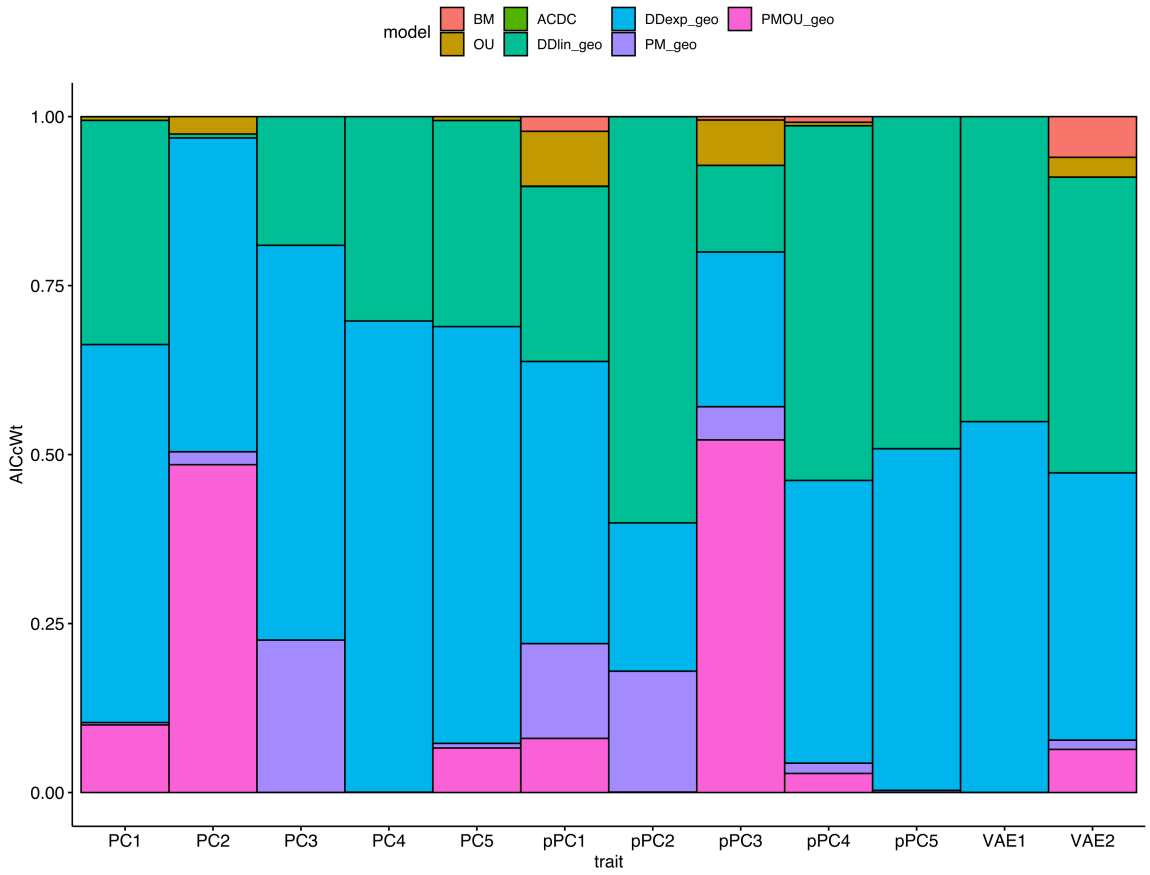

Figure 3.14: Stacked barplot of phylogenetic comparative model support for traditional models (BM, OU, ACDC) and models incorporating interspecific interactions and competition (DDlin, DDexp, PM, PMOU) on latent venom expression PCs, pPCs, and VAE axes. Unanimous support is found for competition based models, particularly DD models.

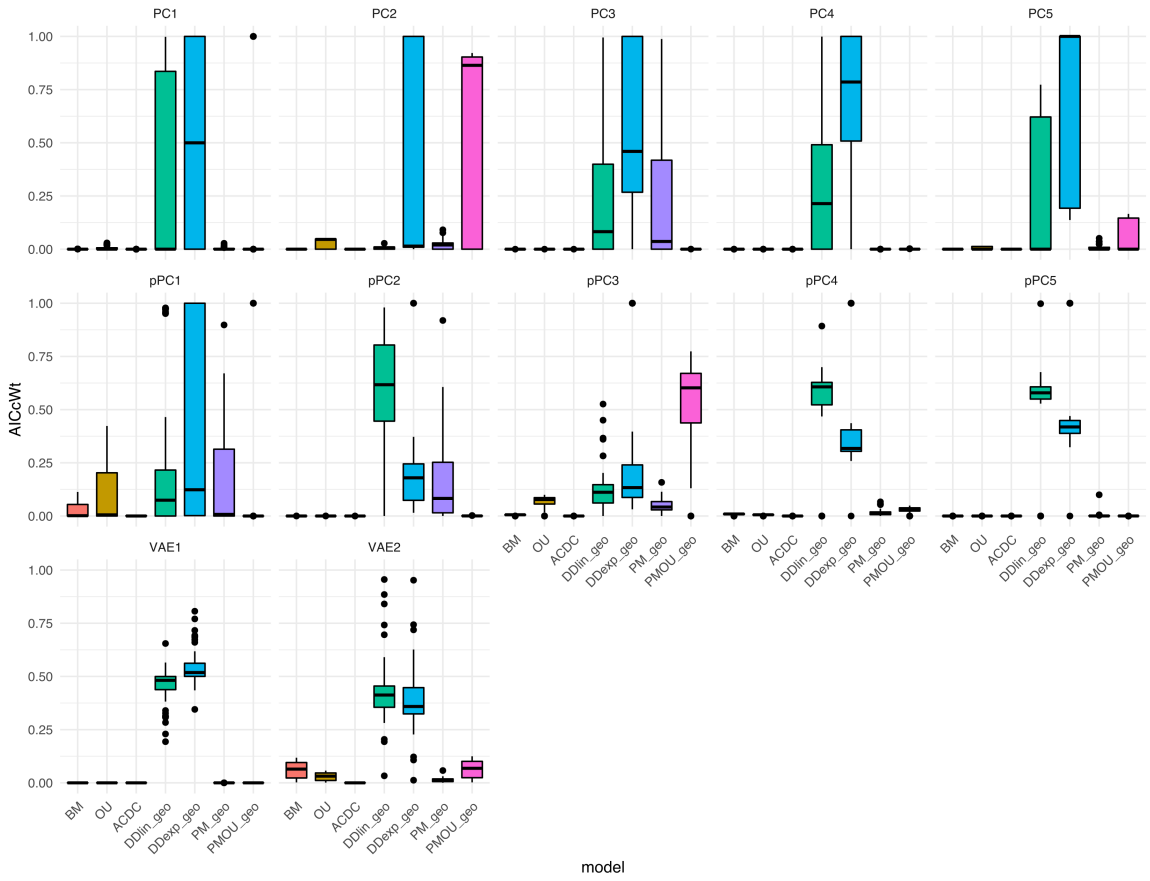

Figure 3.15: Boxplots of phylogenetic comparative model support for traditional models (BM, OU, ACDC) and models incorporating interspecific interactions and competition (DDlin, DDexp, PM, PMOU) on latent venom expression PCs, pPCs, and VAE axes. Unanimous support is found for competition based models, particularly DD models.

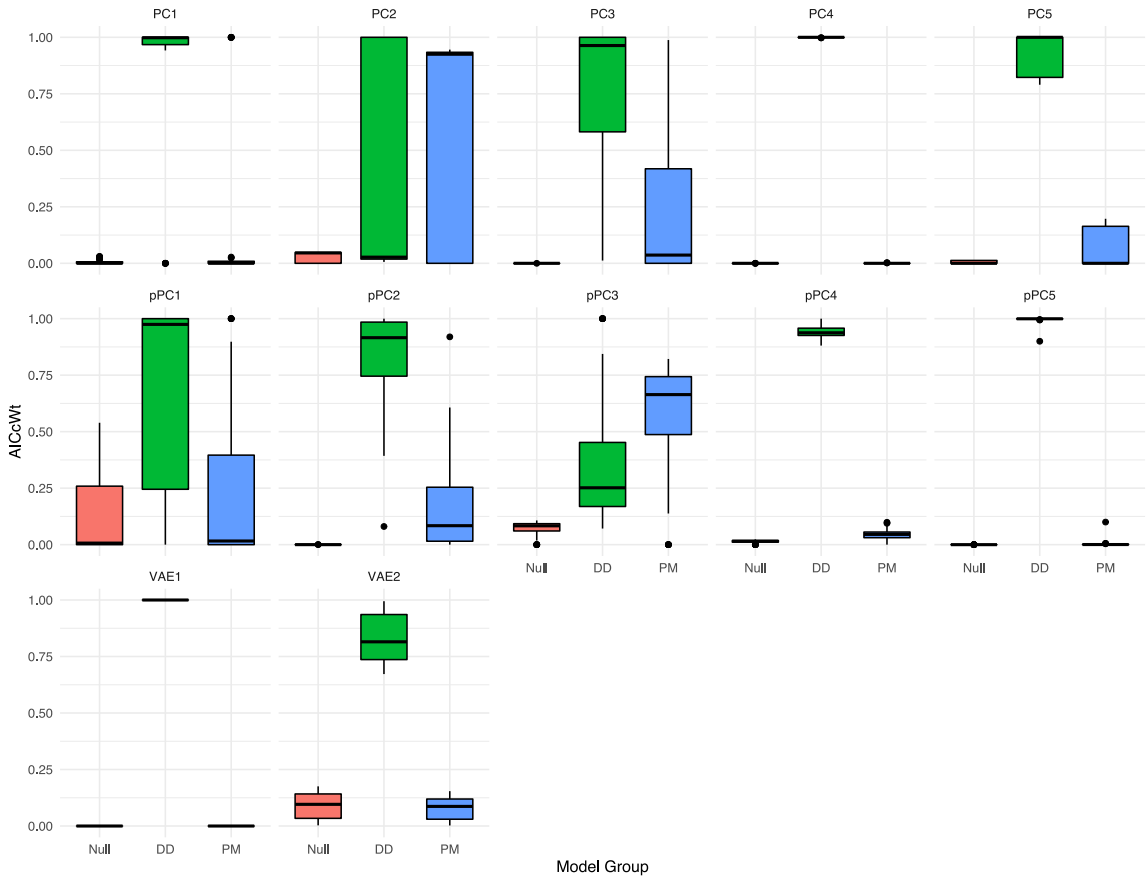

Figure 3.16: Boxplots of phylogenetic comparative model support summed for all traditional models (BM, OU, ACDC), all diversity-dependent models (DDlin, DDexp), and all phenotype-matching models (PM, PMOU) on latent venom expression PCs, pPCs, and VAE axes. Unanimous support is found for competition based models, particularly DD models.

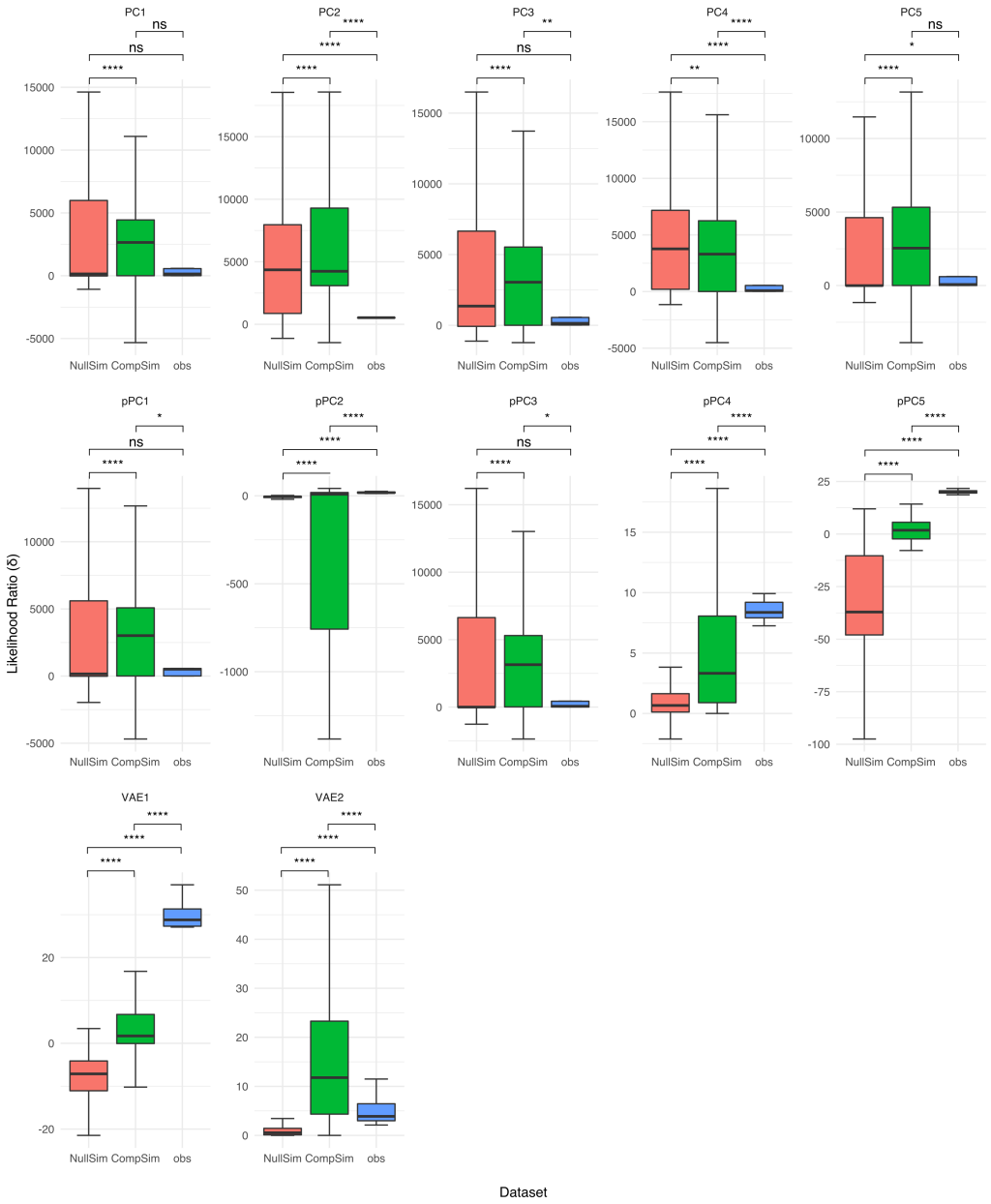

Figure 3.17: Likelihood ratios of the best competition model compared to the best null model (*i.e.*, positive values represent stronger support for the competition model). Three datasets are shown: NullSim for data simulated under the best null model and refit to both models; CompSim for data simulated under the best competition model and refit to both models; obs for comparison of the observed maximum likelihood estimates. Strong support is seen for pPC4, pPC5, VAE1, and VAE2; however, the remaining traits show low support largely due to bimodal patterns across stochastic biogeographic maps.

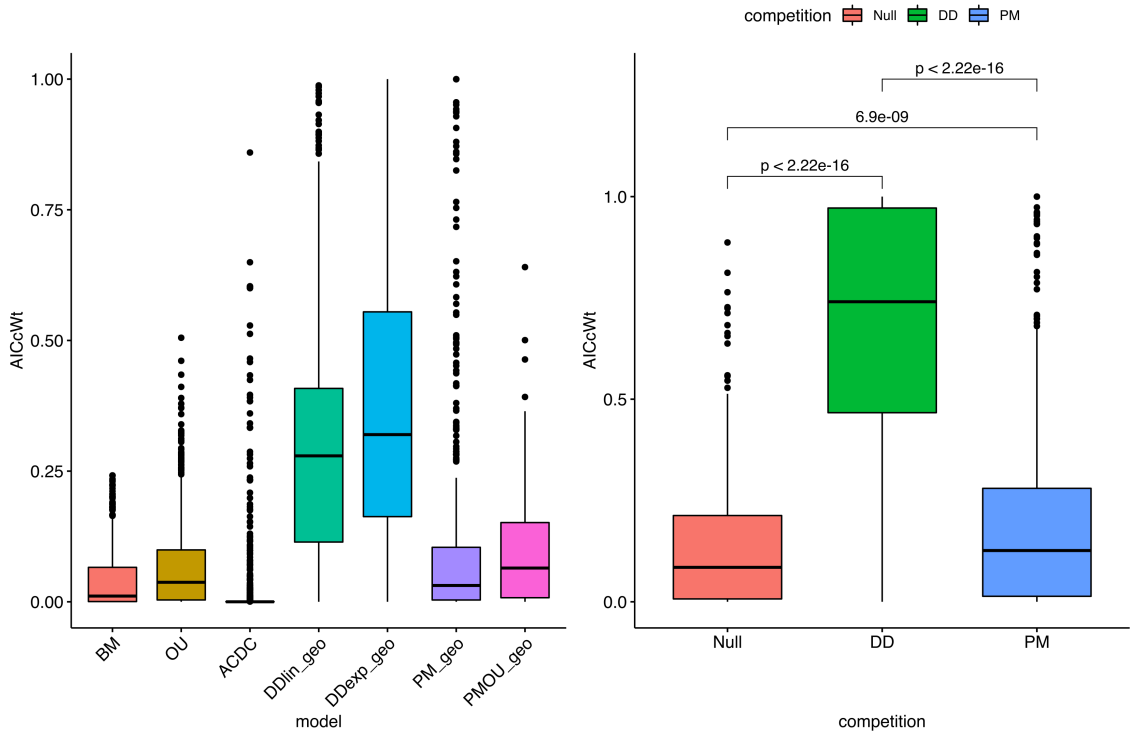

Figure 3.18: Left: Boxplots of phylogenetic comparative model support for traditional models (BM, OU, ACDC) and models incorporating interspecific interactions and competition (DDlin, DDexp, PM, PMOU) on the raw TPM10K expression of 819 genes. Right: Boxplots of phylogenetic comparative model support summed for all traditional models (BM, OU, ACDC), all diversity-dependent models (DDlin, DDexp), and all phenotype-matching models (PM, PMOU) on the raw TPM10K expression of 819 genes. Near unanimous support is found for competition based models, particularly DD models.

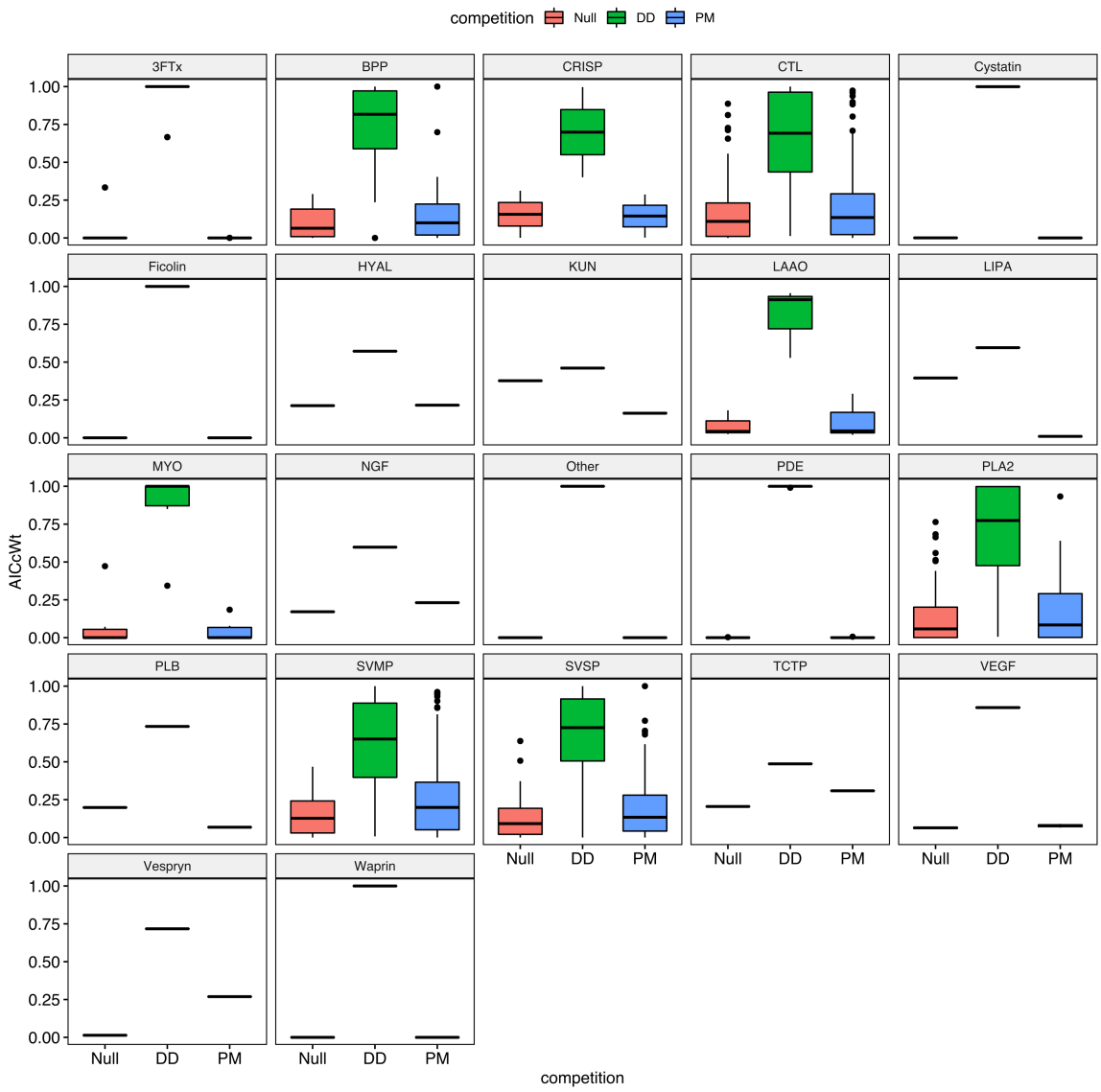

Figure 3.19: Boxplots of phylogenetic comparative model support summed for all traditional models (BM, OU, ACDC), all diversity-dependent models (DDlin, DDexp), and all phenotype-matching models (PM, PMOU) on the raw TPM10K expression of 819 genes, separated by toxin family. Near unanimous support is found for competition based models, particularly DD models.

3.4 Spatial Community Dynamics

To further explore the patterns of diversity dependence, we looked for trait overdispersion in communities across the New World. As expected, we found positive correlations among species richness, phylogenetic diversity, and functional diversity (Fig. 3.20 & 3.21). The communities with the highest species richness and functional diversity were found in the American Southwest, Vercruzan province in Mexico, and southern Brazil (Fig. 3.20 & 3.21). After randomizing communities to maintain species richness and recalculating phylogenetic and functional diversity, we found support for phylogenetic underdispersion with observed communities having significantly less phylogenetic diversity than expected by chance (SES < 0; Fig. 3.20 & 3.21). However, we found functional/trait overdispersion with observed communities having significantly higher functional diversity than expected by chance (SES > 0; Fig. 3.20 & 3.21).

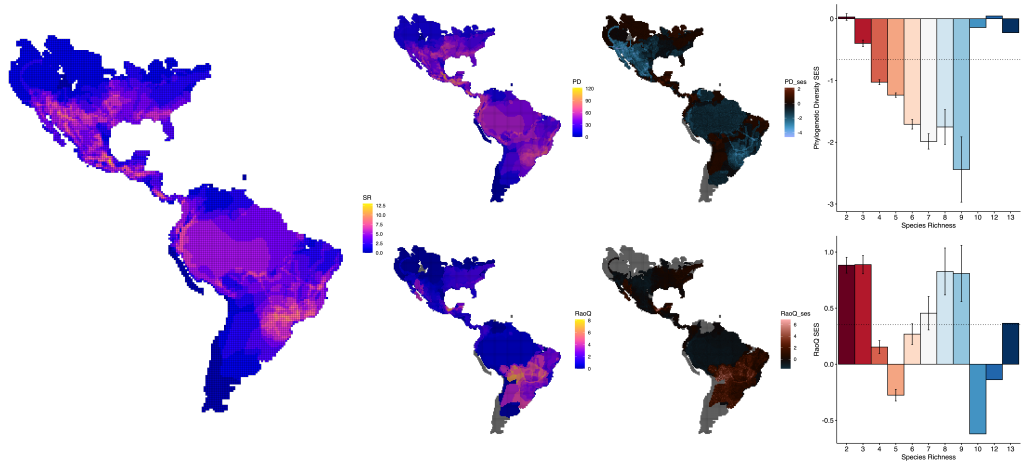

Figure 3.20: Half-degree communities across North and South America and the communities corresponding species richness (left), phylogenetic diversity(top), and functional diversity as calculated using the two VAE axes (bottom). Maps and barplots on the right show standard effect size difference of randomized communities to observed communities with values < 0 indicative of underdispersion and values ?0 indicative of over dispersion. Dotted line represents the mean across all communities. Overall, we see support for phylogenetic underdispersion, but functional overdispersion suggesting a role of competition in driving divergence within communities.

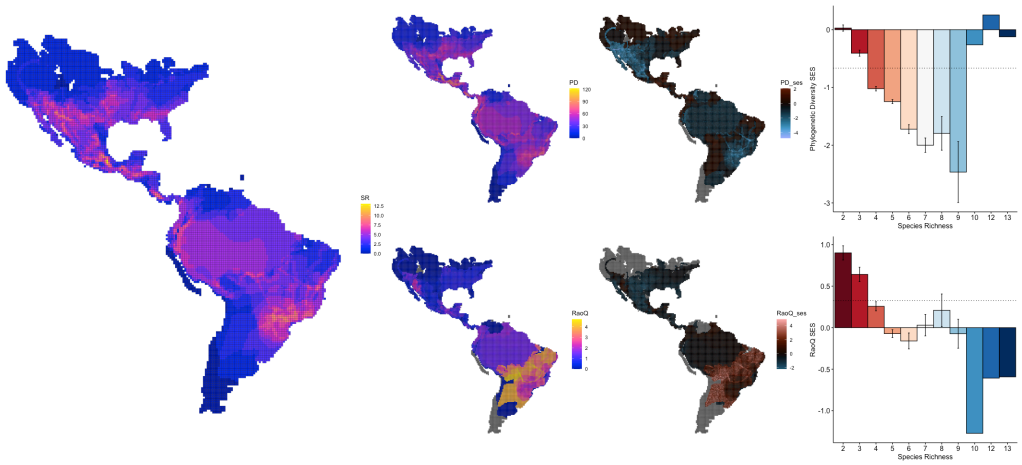

Figure 3.21: Half-degree communities across North and South America and the communities corresponding species richness (left), phylogenetic diversity(top), and functional diversity as calculated using the first five PC axes (bottom). Maps and barplots on the right show standard effect size difference of randomized communities to observed communities with values < 0 indicative of underdispersion and values ?0 indicative of over dispersion. Dotted line represents the mean across all communities. Overall, we see support for phylogenetic underdispersion, but functional overdispersion suggesting a role of competition in driving divergence within communities.

3.5 Microevolutionary Dynamics

Finally, to explore the possibility of competition playing a role at the microevolutionary scale, we focused on the Mojave Rattlesnake (*Crotalus scutulatus*) utilizing Generalized Dissimilarity Modeling (GDM). Overall, the GDM found no significant explanatory variables for venom dissimilarity across the landscape out of geographic distance, phylogenetic distance, species richness, and several uncorrelated bioclimatic variables. Nonetheless, the variables which explained the highest proportion of the variation in venom phenotype included geographic distance, bio11 (Mean Temperature of Coldest Quarter), and species richness.

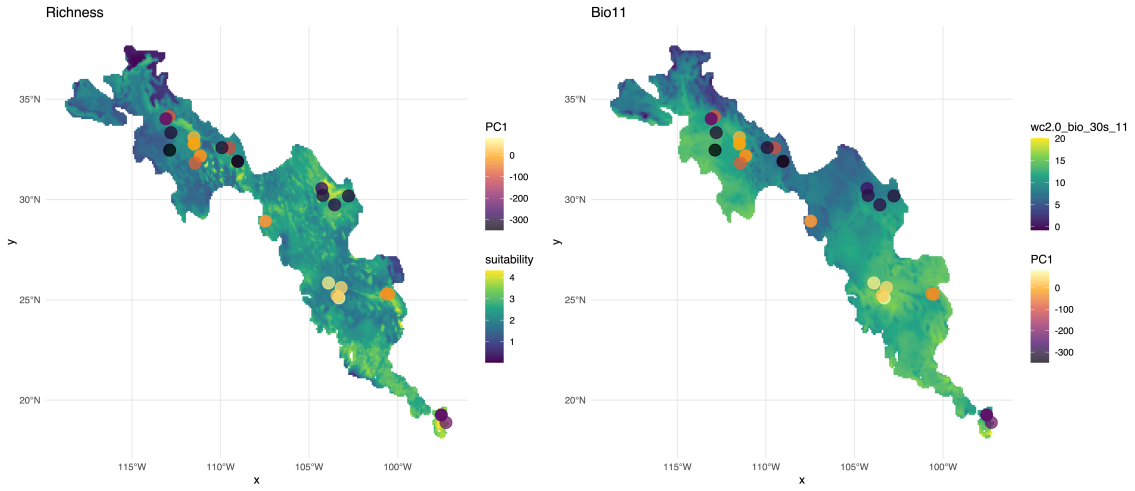

Figure 3.22: Map of *Crotalus scutulatus* distribution with sampling localities plotted and colored by venom phenotype (PC1). Darker points represent more neurotoxic Type A venoms while lighter colors represent hemorrhagic Type B venoms. Background for the two maps are summed SDM suitability across coexisting species which was used as a metric for species richness (left) and the top contributing bioclimatic variable (bio11; right).

Chapter 4

Discussion

Since Darwin, competition and species interactions have been recognized as strong and ubiquitous agents of divergent selection influencing speciation, diversification, and trait evolution; however, studies have largely been limited to relatively simple traits across small geographic and taxonomic scales. In this study, we provide an abundance of evidence using novel phylogenomic and biogeographic inference from hundreds of venom gland transcriptomes for a role of competition, specifically positive diversity dependence, on the evolution of venom in the New World pitviper radiation.

4.1 Systematics & Biogeography

Relationships of species – and even genera – within vipers has been heavily debated, with multiple phylogenetic hypotheses put forth over the past two decades alone (Fig. 3.3)[61,64–69]. Overall, our novel phylogenomic analysis is largely consistent with the most recent hypotheses using only a few mitochondrial and nuclear loci[57,61,64–69]; however, we also uncover several novel relationships across the viper tree of life (Fig. 3.3, 3.4, & 3.5). Our results also suggest a more recent diversification of vipers compared to previous studies (Fig. 3.4 & 3.5;[61,65,68,150]). Overall, more recent divergence date estimates are expected as the use of limited data during divergence dating, such as a few loci derived from mitochondria, often results in significantly older divergence estimates compared to phylogenomic datasets[151]. Importantly, our estimates of divergence closely align with known geological events throughout the Viperidae radiation which we detail below.

4.1.1 Origins of Vipers

First, we estimate that the ancestor of vipers diverged in the late-Paleocene or early-Eocene (59.8-51.2 Mya) and diversification of Viperidae began in the late-Eocene (Viperidae crown age: 41.5-33.2 Mya; Fig. 3.4 & 3.5). In comparison, Wuster et al.[61] and Alencar et al.[65] estimate slightly older dates for Stem-Viperidae (61.5 & 64.5 Mya, respectively) and Crown-Viperidae (47.4 & 52.4 Mya, respectively) while Fenwick et al.[63] and Zaher et al.[69] estimate much younger dates (Stem: ∼42.5 Mya; Crown: 25.5 & 30.7 Mya, respectively). Our estimates closely coincide with the Paleocene-Eocene Thermal Maximum, upheaval of the Himalayan mountains, and with the disappearance of the Tethys Sea as the Indian subcontinent began colliding with Asia and Africa+Asia became more directly connected through present day Arabia[152,153]. Importantly, this area (*i.e.*, India and the Middle East) represents the hypothesized area of origin for Viperidae (Fig. 3.6 & 3.7). This hypothesized origin for Viperidae intuitively makes sense as all early-diverging lineages of caenophidian snakes (*i.e.*, Acrochordus, Xenodermidae, Pareidae)[69] are restricted to Australasia; therefore, divergence likely occurred between the Middle East and southeast Asia during the initial collision of the Indian subcontinent with Eurasia[152,153]. Nonetheless, we recovered low support for this node biogeographic origin and as we did not include any outgroups, our results should be interpreted with caution and additional studies are needed.

Wuster et al.[61] similarly suggested an Asiatic origin with subsequent dispersal to Africa. While Smid & Tolley[150] present three equally plausible scenarios for origins in Eurasia and Afro-Arabia. However, both studies assume much older divergence dates for Viperidae and complete separation of Afro-Arabia from Eurasia by the Tethys Sea[61,150]. Our divergence dates are much younger and coincide with complete collision of the Indian subcontinent with Asia and recession of the Tethys Sea[152,153]. By the time of initial divergence between Viperinae and Azemiopinae + Crotalinae (41.5-33.2 Mya), Arabia and Asia had largely become connected; facilitating dispersal into Arabia in partial agreement with Wuster et al.[61] and "Scenario 2" by Smid & Tolley[150]. Concurrently, geologic upheaval by the Himalayan mountains likely facilitated divergence either through direct vicariance or indirect disruption of ecosystem stability[154] in the Middle East leading to a western radiation in Europe, Arabia, and Africa for Viperinae and a eastern radiation into India and southeast Asia for Azemiopinae and Crotalinae (Fig. 3.6 & 3.7).

4.1.2 Viperinae

Within Viperinae, Wuster et al.[61] and Smid & Tolley[150] suggested an African origin with dispersal back into Eurasia by *Eristicophis, Pseudocerastes, Montivipera, Macrovipera, Daboia*, and *Vipera*. Importantly, the placement of *Causus* as either basal to all Viperinae[150] or as sister to (*Echis* + *Cerastes*)[61] strongly influences biogeographic inference favoring an African origin. Additionally, monophyly of Saharan clades (*Cerastes* and *Echis*) with sub-Saharan clades (*Atheris* + *Bitis*) also favors inference of a singular African origin[61,150]. In contrast, we infer an origin for Viperinae in Arabia and separate dispersals into Saharan and sub-Saharan Africa (Fig. 3.7).

Initial divergence within Viperinae occurred between predominantly Euro-Arabian taxa and sub-Saharan African taxa within present day Arabia and Ethiopia, respectively (Fig. 3.7). Our phylogeny strongly supports the placement of *Causus* as sister to a clade containing (*Atheris* + *Bitis*) (Fig. 3.3, 3.4, & 3.5). No previous phylogenetic hypothesis has recovered this relationship for *Causus* (Fig. 3.3)[61,64–69]; however, Alencar et al.[65] did similarly recover *Causus* as monophyletic with *Atheris* and *Bitis*. This monophyletic relationship is intuitive given that all three genera are generally found in sub-Saharan Africa. As sea levels were low at the time of divergence and these areas were likely connected, we hypothesize that divergence from Arabia likely corresponds with volcanic activity from the Ethiopia-Yemen Continental Flood Basalts which began $\sim$ 29 Mya[152,153,155]. This hypothesis is consistent with previous studies on other taxa such as in mammals which diverged between Africa and Arabia $\sim$27 Mya due to this volcanic activity[156]; however, to our knowledge, this biogeographic break has not been recovered in other genera of snakes. The sub-Saharan lineage then likely diversified into unique ecological niches (*Atheris*: arboreal; *Bitis*: terrestrial, heavy-bodied; *Causus*: terrestrial, small-bodied) and our hypothesized origin for all three genera in the Zambezian region agrees with recent work by Barlow et al.[157]. Barlow et al.[157] suggested that habitats shifts do not play a role in the diversification within *Bitis*; however, their analysis was missing *Atheris* and *Causus* and additional studies should be done to test whether habitat shifts are responsible for divergence of these three genera from each other.

In the Euro-Arabian clade, a deep split 28.3-21.3 Mya separated *Echis* and *Cerastes* – genera predominantly found in Arabia and Saharan Africa – from European genera; suggesting secondary dispersal from Arabia into Saharan Africa as well as dispersal back into Eurasia (Fig. 3.5, 3.6 & 3.7). We hypothesize that complete recession of the Tethys Sea likely facilitated dispersal into Eurasia

at this time at this time, but fluctuations in sea level and initial separation of Arabia from Africa resulted in divergence of these lineages throughout this period[152,153]. However, our phylogenetic placement of *Cerastes* was uncertain and we have poor taxon sampling of both *Echis* and *Cerastes*. Previous studies have often identified *Cerastes* and *Echis* as sister genera which together are sister to the sub-Saharan clade[61,64,66,67]. Therefore, a single African dispersal may be possible and additional studies are needed to test these hypotheses more thoroughly and determine if the observed topology is a result of taxon/gene sampling or is indicative of true discordance or polytomy.

The European clade repeatedly dispersed northwest into Anatolia, the Balkans, and southern Europe from Arabia and the Middle East throughout the Pliocene and Pleistocene likely in response to the concurrently rising Alps as the Paratethys and Mediterranean seas became connected and separated with fluctuating sea levels[152,153]. The Alps likely also facilitated several vicariant speciation events as species continued to disperse north. In agreement, Smid and Tolley[150] suggest vicariant speciation via expansions of the Paratethys and fragmentation of Europe into a series of islands. Additionally, biogeographic breaks and multiple dispersals between the Balkans/Europe and Antolia caused to Paratethys fluctuations are common throughout the Pliocene and Pleistocene as seen in beetles and lizards[158,159].

4.1.3 Old World Crotalinae

Our phylogeny resolves the placement of several previously contentious taxa within Crotalinae. We recover a well-supported clade with two pairs of sister genera: *Calloselasma* + *Hypnale* and *Deinagkistrodon* + *Tropidolaemus* (Fig. 3.3, 3.4, & 3.5). This topology has been previously supported by Alencar et al.[65] and Zheng & Wiens[68]; however, many previous studies have recovered these as independent lineages (Fig. 3.3)[55,57,60,61,64,66]. Early phylogenetic studies also frequently recovered *Trimereserus* as paraphyletic, leading to the recognition of *Protobothrops*[54–57,59,160]. Interestingly, in contrast, we recover strong support for *Trimeresurus* and *Protobothrops* as monophyletic sister genera. We also recover strong support for *Gloydius* and *Ovophis* as sister genera (Fig. 3.3, 3.4, & 3.5). Together, *Gloydius* + *Ovophis* are sister to New World Crotalinae (Fig. 3.3, 3.4, & 3.5). Until recently, the topology of *Trimeresurus* + *Protobothrops* and *Gloydius* + *Ovophis* had not been recovered with each genus often recovered as independent lineages and *Ovophis* frequently recovered as paraphyletic with some species sister to *Gloydius* in agreement with our hypothesis and other species sister to *Protobothrops* (Fig. 3.3)[55–57,59–61,63–69]. However, a more recent multilocus phylo-

genetic analysis of Chinese snakes similarly supports the topology we recovered[161]. Importantly, we recovered orphan taxa *Trimeresurus gracilis* and *Ovophis okinavensis* as nested within *Ovophis* with strong support. These taxa have been difficult to place previously (Fig. 3.3)[56,59–61,63–69]. Although additional taxon sampling is needed to confirm these findings, we reassign *Trimeresurus gracilis* to the genus *Ovophis*; this taxa should now be defined as *Ovophis gracilis*.

We hypothesize that Old World Crotalinae diversified over the course of the late Oligocene and early Miocene within Southeast Asia (Fig. 3.5 & 3.7). Interestingly, although India is the hypothesized origin for Azemiopinae and Crotalinae, there is little viper diversity present in India compared to Southeast Asia[108,113]. The Oligocene is considered a transition period from the warm-tropical Eocene to the cooler, more temperate Miocene[162]. At this time, the Himilayan mountains were also forming driving a combination of environmental changes and vicariance[152,153]. As such, during the Oligocene-Miocene transition, India experienced significant aridification resulting in regression of tropical forests and expansion of grasslands[162]. Therefore, we hypothesize that Azemiopinae and Crotalinae likely occurred across India, but retracted to more tropical Southeast Asian habitats where they diversified (Fig. 3.5 & 3.7). Previous studies of Viperidae grouped all Asian taxa into one large region[61]; the lack of resolution prohibits thorough comparison of Old World Crotalinae; however, several studies show significant faunal exchange and close phylogenetic relationships between taxa in southern India and Southeast Asia following initial collision followed by decreases in exchange as climates shifted (summarized in[163,164]).

Additionally, at this time, Southeast Asia was much more connected across the present-day Indonesian archipelago which likely facilitated rapid dispersal across this area[152,153]. Bursts of speciation in Crotalinae[65] likely then occurred as the Himilayan mountains and sea levels rose; allopatrically separating many populations[152,153]. The Southeast Asian islands are a region of special interest due to the high lineage diversity of the region likely generated through allopatry and dispersal. However, our limited sampling precludes us from making robust insights on diversification of the regions and detailed future studies would be valuable contribution to our understanding of Old World Crotalinae diversification.

4.1.4 New World Crotalinae

Our phylogeny suggests invasion of the New World by Crotalinae during the mid-Miocene (21.1-16.5 Mya) coinciding with the mid-Miocene Climatic Optimum (MMCO; $\sim$ 18-14 Mya) – a

period of time representing the maximum temperatures for the Miocene Epoch[165,166] (Fig. 3.5 & 3.7). During the MMCO, warmer temperatures facilitated rapid, widespread dispersal for many ectothermic species[165]. In particular, this period represents a peak in dispersal across Beringia[167] due to a northward shift in temperature isotherms[168], expansion of tropical habitats in western Asia[169], and drops in sea level (< -15 m) forming land-bridges or insular "stepping-stones" across Beringia[152,153]. Indeed, our estimate for dispersal across Beringia aligns with peaks of biotic exchange between Eurasia and North America across many other taxonomic groups, such as plants, mammals, and amphibians[167]. By comparison, Wuster et al.[61] and Alencar et al.[65] suggests an older invasion of the New World 34-22 Mya; however, at this time temperatures were much lower. Although fluctuations in temperature and glaciation during the hypothesized periods of invasion by Wuster[61] and Alencar[65] may have facilitated dispersal during this period, it seems more likely that dispersal occurred during the MMCO when temperatures were much higher and more stable[166,170].

Following invasion of the New World, we hypothesize that there was rapid southern dispersal and diversification into new habitats/niches to keep pace with the dropping temperatures of the mid-Miocene disruption which began $\sim$ 14 Mya and would make the northern-most latitudes uninhabitable to ectotherms[165,166]. In support of this, we recover high net diversification rates following invasion of the New World and a hypothesized origin for diversification widespread across Central North America, Tamaulipas, and south along the eastern coast of Middle America to Costa Rica (Fig. 3.7 & 3.8). After initial rapid diversification, we hypothesize that slowdowns in diversification occurred as niches were filled across North and South America (Fig. 3.8). These results are consistent with the bursts of speciation recovered by Alencar et al.[65] and inference by Wuster et al.[61]. Our phylogeny supports initial divergence occurred between temperate North American genera (*Agkistrodon*, *Sistrurus*, and *Crotalus*) and tropical genera in Middle and South America (Fig. 3.5 & 3.7). Importantly, we provide strong support for the placement of several topologically unstable Middle American genera including *Ophryacus*, *Mixcoatlus*, *Lachesis*, and *Bothriechis* within the monophyletic Tropical clade (Fig. 3.5 & 3.7) unlike previous studies which have often recovered these genera as independent lineages or as sister to the Temperate clade[60,61,63–69] (Fig. 3.3). However, our results closely match previous topologies by Parkinson et al.[55,57] with the exception of placement of *Bothriechis* within the Tropical clade and sister relationship between *Ophryacus* and *Lachesis*. Our phylogeny supports that Bothropoids dispersed into South America and diverged from the Central American clade soon after divergence from the North American clade (Fig. 3.5

& 3.7). Therefore, we hypothesize that three distinct biogeographic radiations occurred in North, Middle, and South America.

4.1.4.1 North America

In the North American, Temperate clade, *Agkistrodon* diverged early upon invasion of the New World in Central North America (17.1-13.0 Mya) followed by a later diversification of rattlesnakes (11.1-8.6 Mya; *Sistrurus* and *Crotalus*). Within *Crotalus*, the first major divergence occurred 9.8-7.7 Mya between the montane, small-bodied clade in the Trans-Volcanic Mexican Belt (TVMB) and the lowland, large-bodied clade in the Chihuahuan desert (Fig. 3.5 & 3.9). These divergence date estimates are largely consistent with Wuster et al.[61] and Holding et al.[51] and closely coincides with the completion of pulses of maphic volcanism moving from west to east reaching the Gulf coast and completing the creation of the Trans-Volcanic Mexican Belt (TVMB) ∼7 Mya[171]. With a hypothesized origin in the TVMB for the montane clade (Fig. 3.9), the TVMB likely acted as a biogeographic barrier between these two clades while subsequent pulses of silicic volcanism which occurred until ∼3 Mya[171] expanded this region and likely contributed to frequent allopatric speciation and local extinction of populations across the TVMB – as evidenced by the island-like distribution of species within the montane clade. The formation of the TVMB is a well-known biogeographic break for many species including *Phrynosoma*[172], *Sceloporus*[173], *Barisia*[174] which all diverged across this area ∼10-7 Mya coinciding with both our estimate and previous estimates for the divergence of montane rattlesnakes[175,176].

Crotalus willardi and *C. pricei* are also small-bodied, montane specialists; however, we did not recover this clade with the remaining montane species (Fig. 3.5); instead, this clade branched separately. As *C. willardi* and *C. pricei* are predominantly found in the northern Sierra Madre Occidental, we hypothesize that this group was similarly allopatrically separated by the TVMB (Fig. 3.9). The low topological support for this clade may suggest continued gene flow with the ancestral large-bodied rattlesnake following divergence from the montane clade (Fig. 3.5). Blair et al.[120] similarly suggest an origin for this clade and the montane clade throughout southern and western Mexico; however, unstable placement of this clade alongside a lack of geographic resolution – grouping together the Sierra Madre Occidental and TMVB – limit biogeographic inference.

The large-bodied rattlesnakes dispersed into the Sonoran desert ∼9.6-7.6 Mya (Fig. 3.5 & 3.9). At this time, the Sonoran and Chihuahuan deserts were occupied by more forests, but grass-

lands were rapidly expanding due to the Miocene disruption leading to cooler, more arid climates[177]. This shift in ecosystems likely facilitated eastern dispersal and subsequent divergence of *C. horridus* (Fig. 3.9). Additionally, at this time, Baja California had only recently begun to separate from mainland Mexico, facilitating dispersal to Baja California by *C. enyo* (Fig. 3.9)[152,153]. Soon after their dispersal into the Sonoran desert, the large-bodied rattlesnakes split, leading to two large, independent radiations of rattlesnakes (Fig. 3.9). One lineage rapidly dispersed south along the Mexican Pacific lowlands into the Balsa's Basin, Central America, and across the Isthmus of Panama into South America (Fig. 3.9; additional discussion in next section). The other lineage radiated across the Chihuahuan, Sonoran, and Mojave deserts (Fig. 3.9). Within this desert radiation, a secondary eastern dispersal occurred by *C. adamanteus* 6.1-5.0 Mya (Fig. 3.5 & 3.9). Blair et al.[120] also lacks geographic resolution to infer dispersal/vicariance between the Chihuahuan and Sonoran deserts as well as between the Mexican Pacific Lowlands, Sierra Madre Occidental, and TVMB. Therefore, they hypothesize the origin of *Crotalus* was widespread across all of these regions with the large bodied clade originating specifically in the TVMB/Mexican Pacific Lowland region. Their results were also likely highly influenced by the lowly supported basal placement of *C. horridus*, *C. enyo*, and *C. cerastes*, and the *C. willardi* clade[120].

4.1.4.2 Middle America

Phylogenetic placement of Middle American pitviper genera has been challenging[55,57,60,61,63–69]; however, we recovered strong support for all previously contentious genera in a single monophyletic "Tropical" group (Fig. 3.3, 3.4, & 3.5). First, we found that *Ophryacus* and *Mixcoatlus* are sister genera; however, we identify *Ophryacus* as paraphyletic and synonymize *Mixcoatlus* with *Ophryacus* (Fig. 3.3, 3.4, & 3.5); reassigning *Mixcoatlus barbouri*, *M. browni*, and *M. melanurus* as *Ophryacus barbouri*, *O. browni*, and *O. melanurus*, respectively. This result is also consistent with concurrent work using mitochondrial DNA (Reyes-Velasco et al. in prep).

Lachesis has most commonly been recovered as sister to *Ophryacus* and sometimes as basal to all other New World Crotalinae[55,57,60,61,63–69]; however, our results strongly support *Lachesis* as an independent lineage within the Middle American clade sister to the remaining Middle American taxa (Fig. 3.3, 3.4, & 3.5). Our results within *Lachesis* suggest a Costa Rican origin and southern dispersal into South America, consistent with Zamudio and Greene[178]. *Bothriechis* has most commonly been recovered as sister to *Lachesis*/*Ophryacus* or as an monophyletic lineage[50,55,57,60,63–69];

however, our results strongly support *Bothriechis* as sister to *Porthidium* + *Cerrophidion* + *Atropoides* consistent with the hypothesis from Wuster et al.[61] (Fig. 3.3, 3.4, & 3.5). Contrary to previous studies which have most commonly recovered *Atropoides* as sister to *Cerrophidion* + *Porthidium*[55,57,60,61,64–69], we identified *Porthidium* as basal in this group consistent with Fenwick et al.[63].

Finally, a recent paper by Campbell et al.[179] resulted in the formal creation of a new genus – *Metlapilcoatlus* – to recognize paraphyly frequently identified within *Atropoides*[60,61,63–69]. Although our phylogeny finds deep divergence between the now monotypic *Atropoides picadoi* and *Metlapilcoatlus sp.*, we strongly support monophyly of this group and therefore recommend synonomy of *Metlapilcoatlus* with *Atropoides* and resurrect *Atropoides nummifer*, *A. olmec*, *A. indomitus*, *A. occiduus*, and *A. mexicanus* (Fig. 3.4 & 3.5).

Rapid cladogenesis in Middle America likely occurred as a result of niche partitioning of available habitats (Fig. 3.8) with *Ophryacus* and *Bothriechis* occupying semi-arboreal and arboreal niches in Mexico and Central America, respectively, *Lachesis* occupying the large-bodied, terrestrial niche, and the remaining taxa (*i.e., Porthidium, Atropoides*, and *Cerrophidion*) filling a small to mid-bodied, terrestrial niche across Middle America with different elevation preferences (*e.g., Cerrophidion* in montane habitats and *Porthidium* in lowland habitats). Diversification across Middle American clades was then aided by a series of geological events that occurred in this region during the late-Miocene and throughout the Pliocene[180,181]. The Nicaraguan Depression - which is a result of volcanic back-arc formation from the Central American volcanic arc - is an area of low elevation between the Chortis Block highlands and Lower Central America highlands which expanded following invasion of the Middle American pitviper genera. This low elevation area is sensitive to changes in sea levels and the formation of marine gaps - as partially evidenced by the presence of marine taxa in Lake Nicaragua - partially or possibly completely separating the two highland regions. In support of previous studies[180,181], we find concordance in our divergence time estimates between taxa from the two highland areas (Fig. 3.5 & 3.9). Specifically, the MRCA of: (*Atropoides picadoi* + *Atropoides sp.*): 9.2-6.7 Mya; (*Cerrophidion godmani* + *C. sasai*): 8.4-6.2 Mya; and (*Bothriechis nigroviridis* + *B. bicolor*): 8.2-6.4 Mya (Fig. 3.5 & 3.9). We also have partial support from the MRCA of (*Porthidium nasutum* + *P. ophryomegas*): 7.8-6.1 Mya; however, as these species occur in lowland habitats they have likely re-dispersed across this gap (Fig. 3.5 & 3.9).

We additionally find support for the Motagua-Polochic Fault as a biogeographic break across

Honduras and Guatamala. This region separates the Chortis and Mayan highlands and mountain formation in these regions have led to diversification of montane lineages[180,181]. We see concordant divergence time estimates for the MRCA of (*Bothriechis bicolor* + *B. guifarroi*): 5.6-4.6 Mya and (*Porthidium ophryomegas* + *P. dunni*): 7.0-5.5 Mya (Fig. 3.5 & 3.9). However, contrary to Castoe et al.[180], our topology does not support similar divergence time between *Cerrophidion godmani* and *C. wilsoni* and we have insufficient sampling for *Atropoides* to fully assess this biogeographic break. The Isthmus of Tehuantepec is a low elevation area separating the Sierra Madres in Mexico from the Mayan highlands. We find concordant support for this area resulting in allopatric speciation for (*Cerrophidion petlalcalensis* + *C. tzotzilorum*): 4.3-3.4 Mya, (*Atropoides nummifer* + *A. occiduus*): 3.9-2.7 Mya, and (*Bothriechis aurifer* + *B. rowleyi*): 3.8-3.1 Mya (Fig. 3.5 & 3.9).

Finally, we see concordant divergence dates for a dispersal across the Isthmus of Panama into South America for *Lachesis* (6.0-3.8 Mya) and *Bothriechis schlegelli* (5.1-4.1 Mya) while other groups including *Porthidium* and *Crotalus durissus* dispersed into South America much later at 2.1-1.6 Mya and 2.9-2.1 Mya, respectively (Fig. 3.5 & 3.9). These estimates closely match estimates of other taxa following the closing of the Isthmus of Panama ∼4 Mya during the Great American Biotic Interchange[182]. Additionally, our estimates for dispersal across the Isthmus of Panama by *Crotalus durissus* closely aligns with previous estimates[183,184]. Interestingly, while most lineages followed a southwestern dispersal through the Pacific forests, *Crotalus durissus* dispersed southeast through the Guiana Shield (Fig. 3.5 & 3.9) – a result similarly supported by previous studies[183,184]. From here, dispersal to the Brazilian shield is hypothesized to have occurred via a continuous dispersal corridor[183,184] (but see[185]).

4.1.4.3 South America

Bothropoids diverged from the Central American clade during the mid Miocene via dispersal into South America (17.3-12.8 Mya; Fig. 3.5 & 3.9). Dispersal across the Isthmus of Panama by Bothropoids occurred much earlier than in the Middle American clade (see discussion above). At this time, the Isthmus of Panama had not yet fully formed and instead volcanic activity from the Panama Arc resulted in a semi-emergent island chain across which taxa could disperse or raft[182]. Our estimates align with the first land mammals to disperse across this region ∼20 Mya[182] and previous estimates for Bothropoid dispersal[61,65]. Upon invasion into South America, diversification was largely influenced by the concurrently rising Andes mountains and hydrological changes. In

particular, dispersal of Bothropoids would have been initially restricted to a southern disperal along the Andean mountains along the western coast of South America as eastern dispersal was restricted by marine incursion from the Pebas Formation (*alias*, the Amazonian Sea) and Paranan Sea at that time[186–191]. The upheaval of the Andes mountains resulted in a drainage of the Paranan Sea and the Pebas Formation with a shift in the Pebas drainage from the northward Orinoco river to the eastward Amazon River[186,187]. Interestingly, we recover *Bothrocophias* as paraphyletic; supporting either the synonymy of *Bothrocophias* into *Bothrops* or reclassification of several lineages as genera (Fig. 3.5). We hypothesize that the multiple paraphyletic clades of *Bothrocophias* result from southern dispersal across the Andes followed by vicariance as the Andes continued to rise (Fig. 3.9).

We infer a southeastern dispersal of ancestral *Bothrops* into the Brazilian Shield and present-day Cerrado and Chaco regions (12.3-8.5 Mya; Fig. 3.5 & 3.9). We hypothesize that this occurred through the Chapare Buttress which was one of the first regions to connect the western and eastern parts of South America as the Pebas Formation and Paranan Sea continued draining[189]. Dispersal across the Chapare Buttress by *Bothrops* is similarly supported by divergence estimates from several lineages of fishes between the Amazonian and Paranan Basins following drainage of these regions ∼15-8 Mya[192,193], dispersal/divergence of *Anolis* between the Andes and Brazil ∼11.5 Mya[191], and by plants (*i.e., Croton*) between the Amazonian and Atlantic forests ∼14-6 Mya[194]. Coinciding with western dispersal into Brazil in the late Miocene was a period of global cooling which increased aridity forming the present-day environments of the Cerrado/Chaco regions[195]. These regions act as known biogeographic breaks for many taxa (summarized in[195]), prohibiting dispersal and facilitating vicariance and disjunction; however, several lines of evidence suggest corridors occurred intermittently between the Amazonian and Atlantic tropical forests for the past 50 My with the southern-most corridor near the Chapare Buttress being the most stable[194,196].

In our analysis, we recover the Cerrado as the origin of *Bothrops* diversification with low support; however, we did not perform time-stratified biogeographic analysis and this region was likely chosen by BioGeoBEARS because it is geographically centered between the Amazonian and Atlantic forests (Fig. 3.9). It is more likely that *Bothrops* dispersed across the Cerrado to the Atlantic forest via known corridors. In support of this conclusion, recent work by Hamdan et al.[197] inferred similar estimates for diversification of Bothropoids (∼12 Mya) in northern South America and dispersing from the Central Andes into the Brazilian shield; however, this analysis lacked geographic resolution; combining Central America + Andes + Amazonian basin + Guiana shield into one region and the

Cerrado + Chaco + Atlantic Forest region into another. In contrast, Pontes-Nogueira et al.[196] performed a more detailed time-stratified biogeographic reconstruction and similarly inferred jump-dispersal occurring between the Andes and Atlantic Forest. Additional studies on specific taxa such as *B. bilineatus* and *B. brazili/jararacussu* similarly infer disjunction between the Amazonian and Atlantic forests and dispersal between these regions facilitated by forest corridors[198,199]. Following this, we infer dispersals into the Patagonian Steppe, back northward into the Amazon basin and Guiana Shield, and finally into the Cerrado/Chaco regions closely matching the results of[196] (Fig. 3.9).

4.1.4.4 Remaining Phylogenetic Uncertainty

Several phylogenetic and taxonomic questions remain due to gaps in sampling both taxonomically and spatially which are known to have significant impacts on phylogenetic inference[200,201]. Future studies should focus on the further exploration of paraphyletic relationships within *Daboia* and *Bothrocophias* as well as the lack of topological support for *Cerastes* (Fig. 3.4 & 3.5). Additionally, our phylogeny recovered lowly supported specific relationships within several genera such as *Vipera* and *Protobothrops* as well as specific clades such as the *Bothrops neuwiedi* and *Crotalus molossus* complexes (Fig. 3.4 & 3.5). It is likely that high interspecific gene flow both historically and contemporarily within these groups preclude confident phylogenetic placement. Therefore, future studies should focus on these groups with both high-resolution sampling and genetic data to fully understand the evolutionary relationships within these groups. Additionally, as many groups remain difficult to resolve in spite of high-resolution genomic data, recent work using gene genealogy interrogation (GGI) offers a promising approach to further address the phylogenetic relationships of these taxa[202].

4.2 Competition & Venom Evolution

Snake venom is a complex, multidimensional trait which evolves rapidly in response to coevolutionary pressures from predators and prey[30,43,203]. Although the general composition of snake venom is constrained to a few toxin families[204], relative expression and protein sequence diversity of those toxins results in a diverse functional array across species[51,53,205] and unique coevolutionary responses[30,43]. In particular, recent studies have demonstrated that toxin sequence and expression

diversity in rattlesnakes and their allies is directly associated with phylogenetic diversity of prey with more complex venoms having similarly more complex diets[51,53]. Given the close link of venom diversity and diet, there is an intuitive connection for a role of competition in venom diversification similar to beak sizes in birds[5,6,22].

In this study, we recover significant venom diversity across New World pitvipers with our PCA, pPCA, and VAE displaying several well-known patterns (Fig. 3.10 & 3.11). Specifically, we found that species with simple, neurotoxin-dominated venoms dominated by crotoxin-like homologs – such as those found in *Crotalus tigris*[206], *C. durissus*[45], *C. scutulatus*[47,48,143,144], and *C. horridus*[41,207] – are found clustered close to one another as extreme axis values across all three methods.

Our phylogenetic comparative modeling results suggest that the evolution of snake venom composition and expression is associated with competition, particularly, positive diversity dependence, where venom diversification increases when multiple species occur in sympatry thus, minimizing competition by occupying different prey niche space (Fig. 3.14, 3.15, 3.16, 3.18, & 3.19). Many studies have shown support for negative diversity dependence[14,208] and early bursts[209,210] where diversification rates decay with time or as species diversity increases and niches are filled ("niche-filling" hypothesis); however, few studies have shown positive diversity dependence with the exception of songbird beak shape and size[22,211]. We hypothesize that the multidimensional, polygenic nature of venom facilitates evolutionary lability or a less constrained niche space for venom to evolve in a multitude of directions and magnitudes.

In further support for a role of competition on venom evolution, we find support for PM models in some of our latent axes and infer negative values for the interaction parameter ($S < -0.25$); suggesting potential character displacement. Furthermore, by examining community structure and the functional venom diversity present in each community, we demonstrate higher than expected functional diversity indicative of overdispersion (Fig. 3.20 & 3.21). By comparison, we see phylogenetic underdispersion – an expected pattern since community assembly is limited by biogeographic constraints (Fig. 3.20 & 3.21). Therefore, despite communities being phylogenetically depauparate, we see high functional diversity lending support to a role of competition facilitating functional venom divergence within each community. Therefore, we are confident in our results and support a role of competition driving venom diversification across New World pitvipers.

Nonetheless, our parametric bootstrapping recovered low statistical power and model iden-

tifiability (Fig. 3.17), we found that this was largely due to bimodality in model support across stochastic biogeographic maps. Unfortunately, there does not to appear to be clear differences in biogeographic reconstruction between the two groups of stochastic maps supporting and unable to distinguish competition models from null models. Therefore, additional exploration is necessary and we interpret our results cautiously as previous studies have found that these models can be confidently identified from each other[20,23]. Additionally, we suspect the low power found for all PCs and the first three pPCs may be associated with known issues regarding the use of PCAs and other dimension-reduction methodologies in phylogenetic comparative modeling[138,139]. The use of reduced-dimension datasets typically results in support for negative diversity dependence or early burst models even when data are generated under a BM or OU model[138,139]. However, despite our use of dimensional reductions on our venom data, we do not find strong support for negative diversity dependence or early-burst across any of our trait axes. Additionally, the raw TPM10K expression similarly supports positive diversity dependent models.

Measurement error can also bias PCM results and lead to a spurious support for models where evolution is faster in the recent past, as is the case in positive DD models[22]. Snake venom is known to vary significantly within species leading to high measurement error in our estimates of mean venom phenotype per species/lineage. Given this, we cannot exclude the possibility of measurement error leading to spurious support for positive DD models. Nonetheless, measurement error should not significantly impact support for PM models where trait evolution is solely impacted by the trait values of coexisting species.

Most studies of character displacement and other influences of competition on trait evolution such as diversity dependence have focused on shallow scales examining pairs of sympatric and allopatric species[2,3,10,17,18]. While these studies have driven the field forward, recent process-generating models such as the diversity dependent models[14], phenotype-matching models[19,20], generalist matching mutualism models[136], and expansions of these models[23] are able to detect signals of competition in deep-time datasets where there are not clear patterns from comparisons of present-day allopatric and sympatric species[21]. To better model species interactions through time, many of these models have been adapted to utilize biogeographic reconstructions[20,23,136] and more recent models have attempted to jointly infer biogeography and trait evolution since colonization and extirpation rates can be directly impacted by biotic interactions[212]. In line with our study, recent work on Australian monitor lizards (*Varanus*) found similar support for character displacement and

functional overdispersion in communities with regard to niche partitioning of body size[23]. Songbirds similarly show support for a role of positive diversity dependence and character displacement on ecological traits such as beak size and shape, but not in other characteristics such as song[22,213].

Although we accounted for measurement error in our estimates when performing analyses, a tremendous amount of intraspecific venom variation is seen within New World pitviper species[41,47–49,207,214,215]. No study has examined the influence of competition and character displacement on venom evolution at a microevolutionary level. To address this, we focused on the Mojave Rattlesnake and used generalized dissimilarity modeling to identify key predictor variables. Importantly, previous studies identified bio11 as an important variable predicting the presence/absence of Type A and Type B venoms within Mojave Rattlesnakes[47]. In our analysis, we found that species richness – a proxy for the presence of competition inferred from fine scale species distribution modeling[108] – explained a similar amount of the venom variance as bio11 (Fig. 3.22). Although our results were not significant, this analysis may provide preliminary support for character displacement or diversity dependence occurring on smaller scales. We recommend that future studies test for influence of competition on a microevolutionary scale more thoroughly by focusing on systems with disjunct distributions and various levels of sympatry and allopatry such as insular and montane species.

Overall, our results highlight the utility in process-generative phylogenetic comparative models in exploring species interactions and the influence of competition over deep timescales. Additionally, we stress the importance of incorporating ecology and biotic interactions when attempting to understand the evolution of traits and diversification processes. However, our study represents only the beginning for further exploration on the influence of interspecific interactions on the evolution of pitvipers and their venom. For example, our study focused solely on the influence of biogeography and species interactions on trait evolution; however, new models which jointly infer trait evolution and biogeography and allow traits to similarly influence colonization and extirpation rates have been developed to provide a more holistic view on the evolution of species and their traits[23,212]. Similarly, more generalized models of inferring the effect of species interactions on trait evolution have been developed to account for population-level dynamics such predation using Approximate Bayesian Computation[25]. These models could potentially be used to jointly model the effect of competition and predator-prey coevolution to better understand the interaction among these drivers of venom evolution[25]. Evolution of key-innovations, such as venom, are also expected to facilitate

increased diversification rates. Although we used MiSSE to infer diversification rates across New World pitvipers, we did not search for diversification rate shifts. State-dependent Speciation and Extinction (SSE) models such as HiSSE[216], QuaSSE[217], and non-parametric models such as ES-sim[218,219] will allow us to explore the role of venom evolution on diversification rate shifts. Finally, new models which integrate phenotypic and species diversification can be used to jointly infer the influence of competition on venom evolution and species diversification[220].

Chapter 5

Conclusions

Vipers are a family of charismatic venomous snakes occupying a variety of ecological niches from large, terrestrial species to small, arboreal species; however, the evolutionary relationships of species and genera within this family has been largely clouded by a lack of genetic resolution and extremely rapid radiations, particularly upon invasion of the New World. Here, we infer the most robust phylogenomic and biogeographic hypotheses to date by amassing hundreds of next-generation sequencing datasets and confidently resolving many of the previously unstable clades within Viperidae. We infer younger divergence dates than previous estimates; however, these dates largely coincide with known biogeographic breaks in other taxa as well as geological events. In agreement with previous research, we hypothesize that vipers likely originated in Asia and Arabian peninsula in the late-Paleocene or early-Eocene with diversification beginning in the late Eocene as Viperinae radiated west into Africa and Europe and Crotalinae radiated east into Asia and the Americas. Upon invasion of the New World, rapid cladogenesis occurred as species began to fill the multitude of available niches and venom concurrently diversified. Using geographically-aware phylogenetic comparative models to infer species interactions through time, we propose that venom evolves in a diversity-dependent manner with species attempting to reduce competition over prey resources by maximizing venom disparity based on the diversity of species present in an area. We fortify this result by showing that the current Viperid communities across the New World have higher than expected venom functional diversity despite lower than expected phylogenetic diversity; supporting a competition-driven diversification of venom across the New World. Finally, analysis of intraspecific variation suggests that species richness – a proxy for competition – may help explain

venom dissimilarity on a microevolutionary scale.

References

1. Darwin, C. *On the Origin of the Species* (1859).
2. Pfennig, D. W. & Pfennig, K. S. Character displacement and the origins of diversity. *The American Naturalist* **176,** S26–S44 (2010).
3. Brown Jr., W. L. & Wilson, E. O. Character Displacement. *Systematic Zoology* **5,** 49–64 (1956).
4. Stuart, Y. E., Inkpen, S. A., Hopkins, R. & Bolnick, D. I. Character displacement is a pattern: So, what causes it? *Biological Journal of the Linnean Society* **121,** 711–715 (2017).
5. Grant, P. R. & Grant, B. R. Evolution of Character Displacement in Darwin's Finches. *Science* **313,** 224–226 (2006).
6. Lamichhaney, S. *et al.* A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* **352,** 470–474 (Apr. 2016).
7. Losos, J. B. A Phylogenetic Analysis of Character Displacement in Caribbean ¡i¿Anolis¡/i¿ Lizards. *Evolution* **44,** 558–569 (May 1990).
8. Losos, J. B. *Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles* 528 (University of California Press, Berkeley, CA, Aug. 2009).
9. Kamath, A. *et al.* Character displacement in the midst of background evolution in island populations of Anolis lizards: A spatiotemporal perspective. *Evolution* **74,** 2250–2264 (Oct. 2020).
10. Schluter, D. & McPhail, J. D. Ecological character displacement and speciation in sticklebacks. *The American Naturalist* **140,** 85–108 (1992).
11. Schluter, D. Frequency dependent natural selection during character displacement in sticklebacks. *Evolution* **57,** 1142–1150 (2003).
12. Rundle, H. D., Vamosi, S. M. & Schluter, D. Experimental test of predation's effect on divergent selection during character displacement in sticklebacks. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 14943–14948 (2003).
13. Schluter, D. Ecological Character Displacement in Adaptive Radiation. *The American Naturalist* **156,** S4 (2000).
14. Weir, J. T. & Mursleen, S. Diversity-dependent cladogenesis and trait evolution in the adaptive radiation of the auks (Aves: Alcidae). *Evolution* **67,** 403–416 (2013).
15. Stroud, J. T. & Losos, J. B. Ecological Opportunity and Adaptive Radiation. *Annual Review of Ecology, Evolution, and Systematics* **47,** 507–532 (Nov. 2016).
16. Rabosky, D. L. Diversity-Dependence, Ecological Speciation, and the Role of Competition in Macroevolution. *Annual Review of Ecology, Evolution, and Systematics* **44,** 481–502 (2013).
17. Lemmon, E. M. Diversification of conspecific signals in sympatry: Geographic overlap drives multidimensional reproductive character displacement in frogs. *Evolution* **63,** 1155–1170 (2009).

18. Tobias, J. A. *et al.* Species coexistence and the dynamics of phenotypic evolution in adaptive radiation. *Nature* **506,** 359–363 (2014).

19. Nuismer, S. L. & Harmon, L. J. Predicting rates of interspecific interaction from phylogenetic trees. *Ecology Letters* **18,** 17–27 (2015).

20. Drury, J. P., Clavel, J., Manceau, M. & Morlon, H. Estimating the effect of competition on trait evolution using maximum likelihood inference. *Systematic Biology* **65,** 700–710 (2016).

21. Drury, J. P., Grether, G. F., Garland, T. & Morlon, H. An Assessment of Phylogenetic Tools for Analyzing the Interplay Between Interspecific Interactions and Phenotypic Evolution. *Systematic Biology* **67** (ed Harmon, L.) 413–427 (May 2018).

22. Drury, J. P. *et al.* Contrasting impacts of competition on ecological and social trait evolution in songbirds. *PLoS Biology* **16,** 1–23 (2018).

23. Brennan, I. G. *et al.* Phylogenomics of monitor lizards and the role of competition in dictating body size disparity. *Systematic Biology* **70,** 120–132 (2021).

24. Anderson, S. A. S. & Weir, J. T. Character displacement drives trait divergence in a continental fauna. *Proceedings of the National Academy of Sciences* **118,** e2021209118 (May 2021).

25. Xu, L., Van Doorn, S., Hildenbrandt, H. & Etienne, R. S. Inferring the Effect of Species Interactions on Trait Evolution. *Systematic Biology* **70,** 463–479 (2021).

26. Jones, F. C. *et al.* A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology* **22,** 83–90 (2012).

27. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484,** 55–61 (2012).

28. Schluter, D. *et al.* Fitness maps to a large-effect locus in introduced stickleback populations. *Proceedings of the National Academy of Sciences of the United States of America* **118,** 1–10 (2021).

29. Daub, J. T. *et al.* Evidence for polygenic adaptation to pathogens in the human genome. *Molecular Biology and Evolution* **30,** 1544–1558 (2013).

30. Margres, M. J. *et al.* Quantity, not quality: Rapid adaptation in a polygenic trait proceeded exclusively through expression differentiation. *Molecular Biology and Evolution* **34,** 3099–3110 (Dec. 2017).

31. Barghi, N., Hermisson, J. & Schlötterer, C. Polygenic adaptation: a unifying framework to understand positive selection. *Nature Reviews Genetics* **21,** 769–781 (2020).

32. Chippaux, J.-P., Williams, V. & White, J. Snake venom variability: methods of study, results and interpretation. *Toxicon* **29,** 1279–1303 (Jan. 1991).

33. Casewell, N. R., Wüster, W., Vonk, F. J., Harrison, R. A. & Fry, B. G. Complex cocktails: the evolutionary novelty of venoms. *Trends in Ecology & Evolution* **28,** 219–229 (Apr. 2013).

34. Casewell, N. R., Jackson, T. N., Laustsen, A. H. & Sunagar, K. Causes and Consequences of Snake Venom Variation. *Trends in Pharmacological Sciences* **41,** 570–581 (2020).

35. Rokyta, D. R., Margres, M. J. & Calvin, K. Post-transcriptional mechanisms contribute little to phenotypic variation in snake venoms. *G3: Genes, Genomes, Genetics* **5,** 2375–2382 (Nov. 2015).

36. Hofmann, E. P. *et al.* Comparative venom-gland transcriptomics and venom proteomics of four Sidewinder Rattlesnake (Crotalus cerastes) lineages reveal little differential expression despite individual variation. *Scientific Reports* **8,** 15534 (2018).

37. Casewell, N. R. *et al.* Medically important differences in snake venom composition are dictated by distinct postgenomic mechanisms. *Proceedings of the National Academy of Sciences* **111,** 9205–9210 (June 2014).

38. Nakashima, K.-I. *et al.* Accelerated evolution in the protein-coding regions is universal in crotalinae snake venom gland phospholipase A2 isozyme genes. *Proceedings of the National Academy of Sciences* **92,** 5605–5609 (1995).

39. Gibbs, H. L. & Rossiter, W. Rapid evolution by positive selection and gene gain and loss: PLA2 venom genes in closely related ¡i¿Sistrurus¡/i¿ rattlesnakes with divergent diets. *Journal of Molecular Evolution* **66,** 151–166 (Feb. 2008).

40. Casewell, N. R., Wagstaff, S. C., Harrison, R. A., Renjifo, C. & Wüster, W. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Molecular Biology and Evolution* **28,** 2637–2649 (Sept. 2011).

41. Rokyta, D. R., Wray, K. P. & Margres, M. J. The genesis of an exceptionally lethal venom in the timber rattlesnake (Crotalus horridus) revealed through comparative venom-gland transcriptomics. *BMC Genomics* **14,** 394 (2013).

42. Daltry, J. C., Wüster, W. & Thorpe, R. S. Diet and snake venom evolution. *Nature* **379,** 537–540 (Feb. 1996).

43. Holding, M. L., Biardi, J. E. & Gibbs, H. L. Coevolution of venom function and venom resistance in a rattlesnake predator and its squirrel prey. *Proceedings of the Royal Society B: Biological Sciences* **283,** 20152841 (Apr. 2016).

44. Holding, M. L., Drabeck, D. H., Jansa, S. A. & Gibbs, H. L. Venom Resistance as a Model for Understanding the Molecular Basis of Complex Coevolutionary Adaptations. *Integrative and Comparative Biology* **56,** 1032–1043 (Nov. 2016).

45. Mackessy, S. P. in *The Biology of Rattlesnakes* (eds Hayes, W., Beaman, K., Cardwell, M. & Bush, S.) 495–510 (Loma Linda University Press, Loma Linda, CA USA, 2008).

46. Queiroz, G. P., Pessoa, L. A., Portaro, F. C., Furtado, M. d. F. D. & Tambourgi, D. V. Interspecific variation in venom composition and toxicity of Brazilian snakes from Bothrops genus. *Toxicon* **52,** 842–851 (2008).

47. Strickland, J. L. *et al.* Evidence for divergent patterns of local selection driving venom variation in Mojave Rattlesnakes (Crotalus scutulatus). *Scientific Reports* **8,** 17622 (Dec. 2018).

48. Strickland, J. L., Mason, A. J., Rokyta, D. R. & Parkinson, C. L. Phenotypic variation in mojave rattlesnake (Crotalus scutulatus) venom is driven by four toxin families. *Toxins* **10,** 1–23 (Mar. 2018).

49. Amazonas, D. R. *et al.* Molecular mechanisms underlying intraspecific variation in snake venom. *Journal of Proteomics* **181,** 60–72 (Apr. 2018).

50. Mason, A. J. *et al.* Trait differentiation and modular toxin expression in palm-pitvipers. *BMC Genomics* **21,** 147 (Dec. 2020).

51. Holding, M. L. *et al.* Phylogenetically diverse diets favor more complex venoms in North American pitvipers. *Proceedings of the National Academy of Sciences of the United States of America* **118,** e2015579118 (Apr. 2021).

52. Nachtigall, P. G. *et al.* Differences in PLA2 Constitution Distinguish the Venom of Two Endemic Brazilian Mountain Lanceheads, Bothrops cotiara and Bothrops fonsecai. *Toxins* **14,** 237 (Mar. 2022).

53. Mason, A. J. *et al.* Venom gene sequence diversity and expression jointly shape diet adaptation in pitvipers. *Molecular Biology and Evolution* **39** (ed Yoder, A.) 1–14 (Apr. 2022).

54. Kraus, F., Mink, D. G. & Brown, W. M. Crotaline Intergeneric Relationships Based on Mitochondrial DNA Sequence Data. *Copeia* **1996,** 763 (Dec. 1996).

55. Parkinson, C. L. Molecular Systematics and Biogeographical History of Pitvipers as Determined by Mitochondrial Ribosomal DNA Sequences. *Copeia* **1999,** 576 (Aug. 1999).

56. Malhotra, A. & Thorpe, R. S. A phylogeny of the Trimeresurus group of pit vipers: New evidence from a mitochondrial gene tree. *Molecular Phylogenetics and Evolution* **16,** 199–211 (2000).

57. Parkinson, C. L., Campbell, J. A. & Chippindale, P. T. in *Biology of the Vipers* (eds Schuett, G. W., Hoggren, M., Douglas, M. E. & Greene, H. W.) 93–110 (Eagle Mountain Publishing, Salt Lake City, Utah, USA, 2002).

58. Wüster, W. *et al.* Origin and evolution of the South American pitviper fauna: evidence from mitochondrial DNA sequence analysis. *Biology of the Vipers,* 111–128 (2002).

59. Malhotra, A. & Thorpe, R. S. A phylogeny of four mitochondrial gene regions suggests a revised taxonomy for Asian pitvipers (Trimeresurus and Ovophis). *Molecular Phylogenetics and Evolution* **32,** 83–100 (2004).

60. Castoe, T. A. & Parkinson, C. L. Bayesian mixed models and the phylogeny of pitvipers (Viperidae: Serpentes). *Molecular Phylogenetics and Evolution* **39,** 91–110 (2006).

61. Wüster, W., Peppin, L., Pook, C. E. & Walker, D. E. A nesting of vipers: Phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). *Molecular Phylogenetics and Evolution* **49,** 445–459 (2008).

62. Fenwick, A. M., Gutberlet Jr, R. L., Evans, J. A. & Parkinson, C. L. Morphological and molecular evidence for phylogeny and classification of South American pitvipers, genera. *Zoological Journal of the Linnean Society* **156,** 617–640 (2009).

63. Fenwick, A. M., Greene, H. W. & Parkinson, C. L. The serpent and the egg: Unidirectional evolution of reproductive mode in vipers? *Journal of Zoological Systematics and Evolutionary Research* **50,** 59–66 (2012).

64. Pyron, R., Burbrink, F. T. & Wiens, J. J. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evolutionary Biology* **13,** 93 (2013).

65. Alencar, L. R. V. *et al.* Diversification in vipers: Phylogenetic relationships, time of divergence and shifts in speciation rates. *Molecular Phylogenetics and Evolution* **105,** 50–62 (2016).

66. Figueroa, A., McKelvy, A. D., Grismer, L. L., Bell, C. D. & Lailvaux, S. P. A species-level phylogeny of extant snakes with description of a new colubrid subfamily and genus. *PLOS ONE* **11** (ed Parmakelis, A.) e0161070 (Sept. 2016).

67. Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W. & Pyron, R. A. Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biological Conservation* **204,** 23–31 (2016).

68. Zheng, Y. & Wiens, J. J. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution* **94,** 537–547 (2016).

69. Zaher, H. *et al.* Large-scale molecular phylogeny, morphology, divergence-time estimation, and the fossil record of advanced caenophidian snakes (Squamata: Serpentes). *PLOS ONE* **14** (ed Joger, U.) e0216148 (May 2019).

70. Donoghue, M. J. & Ackerly, D. D. Phylogenetic uncertainties and sensitivity analyses in comparative biology. *Philosophical Transactions of the Royal Society B: Biological Sciences* **351,** 1241–1249 (1996).

71. Hernández, C. E. *et al.* Using phylogenetic information and the comparative method to evaluate hypotheses in macroecology. *Methods in Ecology and Evolution* **4,** 401–415 (2013).

72. Rangel, T. F. *et al.* Phylogenetic uncertainty revisited: Implications for ecological analyses. *Evolution* **69,** 1301–1312 (2015).

73. Paterno, G. B., Penone, C. & Werner, G. D. sensiPhy: An r-package for sensitivity analysis in phylogenetic comparative methods. *Methods in Ecology and Evolution* **9,** 1461–1467 (2018).

74. Rotenberg, D., Bamberger, E. S. & Kochva, E. Studies on ribonucleic acid synthesis in the venom glands of ¡i¿Vipera palaestinae¡/i¿ (Ophidia, Reptilia). *The Biochemical Journal* **121,** 609–612 (Feb. 1971).

75. Leary, S. *et al. ¡i¿AVMA Guidelines for the Euthanasia of Animals: 2013 Edition¡/i¿* tech. rep. (American Veterinary Medical Association, 2013), 1–102.

76. Beaupre, S. J., Jacobson, E. R., Lillywhite, H. B. & Zamudio, K. R. *Guidelines for Use of Live Amphibians and Reptiles in Field and Laboratory Research* 2nd ed. (American Society of Ichthyologists and Herpetologists, 2004).

77. Lemmon, A. R., Emme, S. A. & Lemmon, E. M. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* **61,** 727–744 (2012).

78. Knyshov, A., Gordon, E. R. & Weirauch, C. New alignment-based sequence extraction software (ALiBaSeq) and its utility for deep level phylogenetics. *PeerJ* **9,** e11019 (Mar. 2021).

79. Singhal, S., Grundler, M., Colli, G. & Rabosky, D. L. Squamate Conserved Loci (SqCL): A unified set of conserved loci for phylogenomics and population genetics of squamate reptiles. *Molecular Ecology Resources* **17,** e12–e24 (2017).

80. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30,** 614–620 (Mar. 2014).

81. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8,** 1494–1512 (2013).

82. Bushmanova, E., Antipov, D., Lapidus, A. & Prjibelski, A. D. rnaSPAdes: A ¡i¿de novo¡/i¿ transcriptome assembler and its application to RNA-Seq data. *GigaScience* **8,** 1–13 (2019).

83. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31,** 3210–3212 (2015).

84. Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution* **35,** 543–548 (2018).

85. Nachtigall, P. G., Kashiwabara, A. Y. & Durham, A. M. CodAn: predictive models for precise identification of coding regions in eukaryotic transcripts. *Briefings in Bioinformatics* **22,** 1–11 (May 2021).

86. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28,** 3150–3152 (Dec. 2012).

87. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *bioRxiv* (2019).

88. Zimin, A. V. *et al.* The MaSuRCA genome assembler. *Bioinformatics* **29,** 2669–2677 (Nov. 2013).

89. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34,** 3094–3100 (Sept. 2018).

90. Martin, M. *et al.* WhatsHap: fast and accurate read-based phasing. *bioRxiv* (2016).

91. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30,** 772–780 (Apr. 2013).

92. Tumescheit, C., Firth, A. E. & Brown, K. CIAlign: A highly customisable command line tool to clean, interpret and visualise multiple sequence alignments. *PeerJ* **10,** e12983 (2022).

93. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25,** 1972–1973 (2009).

94. Minh, B. Q., Nguyen, M. A. T. & Von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* **30,** 1188–1195 (2013).

95. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* **14,** 587–589 (2017).

96. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37,** 1530–1534 (2020).

97. Mai, U. & Mirarab, S. TreeShrink: Fast and accurate detection of outlier long branches in collections of phylogenetic trees. *BMC Genomics* **19** (2018).

98. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19,** 15–30 (2018).

99. Yin, J., Zhang, C. & Mirarab, S. ASTRAL-MP: scaling ASTRAL to very large datasets using randomization and parallelization. *Bioinformatics* **35,** 3961–3969 (Oct. 2019).

100. Rabiee, M., Sayyari, E. & Mirarab, S. Multi-allele species reconstruction using ASTRAL. *Molecular Phylogenetics and Evolution* **130,** 286–296 (Jan. 2019).

101. Chifman, J. & Kubatko, L. Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30,** 3317–3324 (2014).

102. Page, A. J. *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial genomics* **2,** e000056 (Apr. 2016).

103. Dornburg, A., Fisk, J. N., Tamagnan, J. & Townsend, J. P. PhyInformR: phylogenetic experimental design and phylogenomic data exploration in R. *BMC Evolutionary Biology* **16,** 1–7 (2016).

104. Smith, S. A. & O'Meara, B. C. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28,** 2689–2690 (Oct. 2012).

105. Abadi, S., Azouri, D., Pupko, T. & Mayrose, I. Model selection may not be a mandatory step for phylogeny reconstruction. *Nature Communications* **10** (2019).

106. Burbrink, F. T. *et al.* Interrogating Genomic-Scale Data for Squamata (Lizards, Snakes, and Amphisbaenians) Shows no Support for Key Traditional Morphological Relationships. *Systematic Biology* **69** (ed Thomson, R.) 502–520 (May 2020).

107. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* **15,** e1006650 (2019).

108. Rautsaw, R. M. *et al.* VenomMaps: Updated species distribution maps and models for New World pitvipers (Viperidae: Crotalinae). *Scientific Data* **9,** 232 (Dec. 2022).

109. GBIF.org. *GBIF Occurrence Download. Viperidae.* 2021.

110. HerpMapper. *HerpMapper - A Global Herp Atlas and Data Hub (New World Crotalinae). Iowa, U.S.A.* 2021.

111. Nogueira, C. C. *et al.* Atlas of Brazilian Snakes: Verified Point-Locality Maps to Mitigate the Wallacean Shortfall in a Megadiverse Snake Fauna. *South American Journal of Herpetology* **14,** 1–274 (Dec. 2019).

112. BioWeb.bio. *BioWeb Occurrence Download. Viperidae.* 2021.

113. Roll, U. *et al.* The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology and Evolution* **1,** 1677–1682 (2017).

114. Cobos, M. E., Townsend Peterson, A., Barve, N. & Osorio-Olvera, L. Kuenm: An R package for detailed development of ecological niche models using Maxent. *PeerJ* **2019,** 1–15 (2019).

115. Matzke, N. J. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* **5** (2013).

116. Matzke, N. J. *BioGeoBEARS: BioGeography with Bayesian (and likelihood) Evolutionary Analysis with R Scripts* 2018.

117. Morrone, J. J. *Biogeographical regionalisation of the neotropical region* **1,** 1–110 (2014).

118. Löwenberg-Neto, P. *Neotropical region: A shapefile of Morrone's (2014) biogeographical regionalisation* May 2014.

119. Morrone, J. J., Escalante, T. & Rodríguez-Tapia, G. Mexican biogeographic provinces: Map and shapefiles. *Zootaxa* **4277,** 277–279 (2017).

120. Blair, C. & Sánchez-Ramírez, S. Diversity-dependent cladogenesis throughout western Mexico: Evolutionary biogeography of rattlesnakes (Viperidae: Crotalinae: Crotalus and Sistrurus). *Molecular Phylogenetics and Evolution* **97,** 145–154 (Apr. 2016).

121. Beaulieu, J. M. & O'Meara, B. C. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology* **65,** 583–601 (2016).

122. Holding, M. L., Margres, M. J., Mason, A. J., Parkinson, C. L. & Rokyta, D. R. Evaluating the performance of de novo assembly methods for venom-gland transcriptomics. *Toxins* **10,** 249 (June 2018).

123. Rokyta, D. R., Lemmon, A. R., Margres, M. J. & Aronow, K. The venom-gland transcriptome of the eastern diamondback rattlesnake (¡i¿Crotalus adamanteus¡/i¿). *BMC Genomics* **13,** 312 (2012).

124. Nachtigall, P. G. *et al.* ToxCodAn: A new toxin annotator and guide to venom gland transcriptomics. *Briefings in Bioinformatics* **22,** 1–16 (Apr. 2021).

125. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).

126. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (Apr. 2012).

127. Hargreaves, A. D., Swain, M. T., Hegarty, M. J., Logan, D. W. & Mulley, J. F. Restriction and recruitment—gene duplication and the origin and evolution of snake venom toxins. *Genome Biology and Evolution* **6,** 2088–2095 (Aug. 2014).

128. Bayona-Serrano, J. D. *et al.* Replacement and parallel simplification of nonhomologous proteinases maintain venom phenotypes in rear-fanged snakes. *Molecular Biology and Evolution* **37** (ed True, J.) 3563–3575 (Dec. 2020).

129. Munro, C., Zapata, F., Howison, M., Siebert, S. & Dunn, C. W. Evolution of Gene Expression across Species and Specialized Zooids in Siphonophora. *Molecular Biology and Evolution* **39** (ed Rogers, R.) 1–16 (Feb. 2022).

130. Tarashansky, A. J. *et al.* Mapping single-cell atlases throughout metazoa unravels cell type evolution. *eLife* **10** (May 2021).

131. Battey, C. J., Coffing, G. C. & Kern, A. D. Visualizing population structure with variational autoencoders. *G3: Genes, Genomes, Genetics* **11** (2021).

132. Morlon, H. *et al.* RPANDA: An R package for macroevolutionary analyses on phylogenetic trees. *Methods in Ecology and Evolution* **7,** 589–597 (2016).

133. Felsenstein, J. Phylogenies and the Comparative Method. *The American Naturalist* **125,** 1–15 (Jan. 1985).

134. Hansen, T. F. Stabilizing Selection and the Comparative Analysis of Adaptation. *Evolution* **51,** 1341 (1997).

135. Blomberg, S. P., Garland, T. & Ives, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57,** 717–745 (Apr. 2003).

136. Manceau, M., Lambert, A. & Morlon, H. A Unifying Comparative Phylogenetic Framework Including Traits Coevolving Across Interacting Lineages. *Systematic Biology* **66,** 551–568 (2017).

137. Boettiger, C., Coop, G. & Ralph, P. Is your phylogeny informative? Measuring the power of comparative methods. *Evolution* **66,** 2240–2251 (2012).

138. Uyeda, J. C., Caetano, D. S. & Pennell, M. W. Comparative Analysis of Principal Components Can be Misleading. *Systematic Biology* **64,** 677–689 (July 2015).

139. Adams, D. C. & Collyer, M. L. Multivariate Phylogenetic Comparative Methods: Evaluations, Comparisons, and Recommendations. *Systematic Biology* **67,** 14–31 (2018).

140. Kembel, S. W. *et al.* Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26,** 1463–1464 (June 2010).

141. Laliberte, E. & Legendre, P. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **91,** 299–305 (2010).

142. Laliberté, E., Legendre, P. & Shipley, B. *FD: measuring functional diversity from multiple traits, and other tools for functional ecology* 2014.

143. Glenn, J. & Straight, R. Mojave rattlesnake ¡i¿Crotalus scutulatus scutulatus¡/i¿ venom: variation in toxicity with geographical origin. *Toxicon* **16,** 81–84 (Jan. 1978).

144. Glenn, J. L., Straight, R. C., Wolfe, M. C. & Hardy, D. L. Geographical variation in ¡i¿Crotalus scutulatus scutulatus¡/i¿ (Mojave rattlesnake) venom properties. *Toxicon* **21,** 119–130 (Jan. 1983).

145. Dowell, N. L. *et al.* The Deep Origin and Recent Loss of Venom Toxin Genes in Rattlesnakes. *Current Biology* **26,** 2434–2445 (Sept. 2016).

146. Borja, M. *et al.* Biological and proteolytic variation in the venom of crotalus scutulatus scutulatus from Mexico. *Toxins* **10,** 35 (Jan. 2018).

147. Dowell, N. L. *et al.* Extremely Divergent Haplotypes in Two Toxin Gene Complexes Encode Alternative Venom Types within Rattlesnake Species. *Current Biology* **28,** 1016–1026 (Apr. 2018).

148. Zancolli, G. *et al.* When one phenotype is not enough: Divergent evolutionary trajectories govern venom variation in a widespread rattlesnake species. *Proceedings of the Royal Society B: Biological Sciences* **286** (2019).

149. Fitzpatrick, M., Mokany, K., Manion, G., Nieto-Lugilde, D. & Ferrier, S. *gdm: Generalized Dissimilarity Modeling* 2022.

150. Šmíd, J. & Tolley, K. A. Calibrating the tree of vipers under the fossilized birth-death model. *Scientific Reports* **9,** 5510 (2019).

151. Mulcahy, D. G. *et al.* Estimating divergence dates and evaluating dating methods using phylogenomic and mitochondrial data in squamate reptiles. *Molecular Phylogenetics and Evolution* **65,** 974–991 (2012).

152. Scotese, C. R. & Wright, N. M. *PALEOMAP Paleodigital Elevation Models (PaleoDEMS) for the Phanerozoic* tech. rep. (2018), 1–26.

153. Miller, K. G. *et al.* Cenozoic sea-level and cryospheric evolution from deep-sea geochemical and continental margin records. *Science Advances* **6** (2020).

154. Manish, K. & Pandit, M. K. Geophysical upheavals and evolutionary diversification of plant species in the Himalaya. *PeerJ* **2018,** 1–25 (2018).

155. Baker, J., Snee, L. & Menzies, M. A brief Oligocene period of flood volcanism in Yemen: Implications for the duration and rate of continental flood volcanism at the Afro-Arabian triple junction. *Earth and Planetary Science Letters* **138,** 39–55 (1996).

156. Kappelman, J. *et al.* Oligocene mammals from Ethiopia and faunal exchange between Afro-Arabia and Eurasia. *Nature* **426,** 549–552 (2003).

157. Barlow, A. *et al.* Ancient habitat shifts and organismal diversification are decoupled in the African viper genus Bitis (Serpentes: Viperidae). *Journal of Biogeography* **46,** 1234–1248 (2019).

158. Fattorini, S. Biogeography of the tenebrionid beetles (Coleoptera, Tenebrionidae) on the Aegean Islands (Greece). *Journal of Biogeography* **29,** 49–67 (2002).

159. Ahmadzadeh, F. *et al.* Multiple dispersal out of Anatolia: Biogeography and evolution of oriental green lizards. *Biological Journal of the Linnean Society* **110,** 398–408 (2013).

160. Hoge, A. & Romano-Hoge, S. Notes on micro and ultrastructure of "Oberhautschen" in Viperoidea. *Mem. Inst. Butantan* **44/45,** 81–118 (1983).

161. Li, J. N. *et al.* A large-scale systematic framework of Chinese snakes based on a unified multilocus marker system. *Molecular Phylogenetics and Evolution* **148,** 106807 (2020).

162. Lajmi, A. & Karanth, P. K. Eocene–Oligocene cooling and the diversification of Hemidactylus geckos in Peninsular India. *Molecular Phylogenetics and Evolution* **142,** 106637 (2020).

163. Turner, H., Hovenkamp, P. & Van Welzen, P. C. Biogeography of Southeast Asia and the West Pacific. *Journal of Biogeography* **28,** 217–230 (2001).

164. Klaus, S., Morley, R. J., Plath, M., Zhang, Y. P. & Li, J. T. Biotic interchange between the Indian subcontinent and mainland Asia through time. *Nature Communications* **7,** 1–6 (2016).

165. Böhme, M. The Miocene Climatic Optimum: Evidence from ectothermic vertebrates of Central Europe. *Palaeogeography, Palaeoclimatology, Palaeoecology* **195,** 389–401 (2003).

166. Methner, K. *et al.* Middle Miocene long-term continental temperature change in and out of pace with marine climate records. *Scientific Reports* **10,** 1–10 (2020).

167. Jiang, D., Klaus, S., Zhang, Y. P., Hillis, D. M. & Li, J. T. Asymmetric biotic interchange across the Bering land bridge between Eurasia and North America. *National Science Review* **6,** 739–745 (2019).

168. Herold, N., Huber, M. & Müller, R. D. Modeling the miocene climatic optimum. Part I: Land and atmosphere. *Journal of Climate* **24,** 6353–6373 (2011).

169. Wang, B. *et al.* The mid-Miocene Zhangpu biota reveals an outstandingly rich rainforest biome in East Asia. *Science Advances* **7,** 1–8 (2021).

170. Zhang, Y. G. *et al.* A 40-million-year history of atmospheric CO 2 A -million-year history of atmospheric CO. *Phil. Trans. R. Soc. B* **271** (2013).

171. Ferrari, L., Orozco-Esquivel, T., Manea, V. & Manea, M. The dynamic history of the Trans-Mexican Volcanic Belt and the Mexico subduction zone. *Tectonophysics* **522-523,** 122–149 (2012).

172. Bryson, R. W., García-Vázquez, U. O. & Riddle, B. R. Diversification in the Mexican horned lizard Phrynosoma orbiculare across a dynamic landscape. *Molecular Phylogenetics and Evolution* **62,** 87–96 (2012).

173. Bryson, R. W., García-Vázquez, U. O. & Riddle, B. R. Relative roles of Neogene vicariance and Quaternary climate change on the historical diversification of bunchgrass lizards (Sceloporus scalaris group) in Mexico. *Molecular Phylogenetics and Evolution* **62,** 447–457 (2012).

174. Bryson, R. W. & Riddle, B. R. Tracing the origins of widespread highland species: A case of Neogene diversification across the Mexican sierras in an endemic lizard. *Biological Journal of the Linnean Society* **105,** 382–394 (2012).

175. Bryson Jr., R. W., Murphy, R. W., Lathrop, A. & Lazcano-Villareal, D. Evolutionary drivers of phylogeographical diversity in the highlands of Mexico: A case study of the Crotalus triseriatus species group of montane rattlesnakes. *Journal of Biogeography* **38,** 697–710 (2011).

176. Bryson, R. W. *et al.* Multilocus species delimitation in the Crotalus triseriatus species group (serpentes: Viperidae: Crotalinae), with the description of two new species. *Zootaxa* **3826,** 475–496 (2014).

177. Osborne, C. P. Atmosphere, ecology and evolution: What drove the Miocene expansion of C4 grasslands? *Journal of Ecology* **96,** 35–45 (2008).

178. Zamudio, K. R. & Greene, H. W. Phylogeography of the bushmaster (¡i¿Lachesis muta¡/i¿: Viperidae): implications for neotropical biogeography, systematics, and conservation. *Biological Journal of the Linnean Society* **62,** 421–442 (1997).

179. Campbell, J. A., Frost, D. R. & Castoe, T. A. New Generic name for Jumping Pitvipers (Serpentes: Viperidae). *Revista Latinoamericana de Herpetologia* **02,** 52–53 (2019).

180. Castoe, T. A. *et al.* Comparative phylogeography of pitvipers suggests a consensus of ancient Middle American highland biogeography. *Journal of Biogeography* **36,** 88–103 (2009).

181. Daza, J. M., Castoe, T. A. & Parkinson, C. L. Using regional comparative phylogeographic data from snake lineages to infer historical processes in Middle America. *Ecography* **33,** 343–354 (2010).

182. O'Dea, A. *et al.* Formation of the Isthmus of Panama. *Science Advances* **2,** 1–12 (2016).

183. Wüster, W. *et al.* Tracing an invasion: Landbridges, refugia, and the phylogeography of the Neotropical rattlesnake (Serpentes: Viperidae: Crotalus durissus). *Molecular Ecology* **14,** 1095–1108 (2005).

184. Adrian Quijada-Mascareñas, J. *et al.* Phylogeographic patterns of trans-Amazonian vicariants and Amazonian biogeography: The Neotropical rattlesnake (Crotalus durissus complex) as an example. *Journal of Biogeography* **34,** 1296–1312 (2007).

185. Gosling, W. D. & Bush, M. B. A biogeographic comment on: Wüster et al. (2005) Tracing an invasion: Landbridges, refugia, and the phylogeography of the Neotropical rattlesnake (Serpentes: Viperidae: Crotalus durissus). *Molecular Ecology* **14,** 3615–3617 (2005).

186. Hoorn, C. An environmental reconstruction of the palaeo-Amazon River system (Middle-Late Miocene, NW Amazonia). *Palaeogeography, Palaeoclimatology, Palaeoecology* **112,** 187–238 (1994).

187. Hoorn, C., Guerrero, J., Sarmiento, G. A. & Lorente, M. A. Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. *Geology* **23,** 237–240 (1995).

188. Díaz De Gamero, M. L. The changing course of the Orinoco River during the Neogene: A review. *Palaeogeography, Palaeoclimatology, Palaeoecology* **123,** 385–402 (1996).

189. Lundberg, J. G. *et al.* The stage for neotropical fish diversification a history of tropical south american rivers. *Phylogeny and Classification of Neotropical Fishes,* 603 (1998).

190. Santos, J. C. *et al.* Amazonian amphibian diversity is primarily derived from late Miocene Andean lineages. *PLoS Biology* **7,** 0448–0461 (2009).

191. Prates, I. *et al.* Biogeographic links between southern Atlantic Forest and western South America: Rediscovery, re-description, and phylogenetic relationships of two rare montane anole lizards from Brazil. *Molecular Phylogenetics and Evolution* **113,** 49–58 (2017).

192. Cooke, G. M., Chao, N. L. & Beheregaray, L. B. Marine incursions, cryptic species and ecological diversification in Amazonia: The biogeographic history of the croaker genus Plagioscion (Sciaenidae). *Journal of Biogeography* **39,** 724–738 (2012).

193. Cardoso, Y. P., Jardim de Queiroz, L., Bahechar, I. A., Posadas, P. E. & Montoya-Burgos, J. I. Multilocus phylogeny and historical biogeography of Hypostomus shed light on the processes of fish diversification in La Plata Basin. *Scientific Reports* **11,** 1–14 (2021).

194. Masa-Iranzo, I., Sanmartín, I., Caruzo, M. B. R. & Riina, R. Skipping the dry diagonal: Spatiotemporal evolution of Croton section Cleodora (Euphorbiaceae) in the Neotropics. *Botanical Journal of the Linnean Society* **197,** 61–84 (2021).

195. Luebert, F. The two South American dry diagonals. *Frontiers of Biogeography* **13,** 1–14 (2021).

196. Pontes-Nogueira, M., Martins, M., Alencar, L. R. V. & Sawaya, R. J. The role of vicariance and dispersal on the temporal range dynamics of forest vipers in the Neotropical region. *PLOS ONE* **16** (ed Chiang, T.-Y.) e0257519 (Sept. 2021).

197. Hamdan, B., Guedes, T. B., Carrasco, P. A. & Melville, J. A complex biogeographic history of diversification in Neotropical lancehead pitvipers (Serpentes, Viperidae). *Zoologica Scripta,* zsc.12398 (Dec. 2019).

198. Dal Vechio, F., Prates, I., Grazziotin, F. G., Zaher, H. & Rodrigues, M. T. Phylogeography and historical demography of the arboreal pit viper Bothrops bilineatus (Serpentes, Crotalinae) reveal multiple connections between Amazonian and Atlantic rain forests. *Journal of Biogeography* **45,** 2415–2426 (2018).

199. Dal Vechio, F. *et al.* Rain forest shifts through time and riverine barriers shaped the diversification of South American terrestrial pit vipers (Bothrops jararacussu species group). *Journal of Biogeography* **47,** 516–526 (2020).

200. Betancur-R., R. *et al.* Phylogenomic incongruence, hypothesis testing, and taxonomic sampling: The monophyly of characiform fishes*. *Evolution* **73,** 329–345 (2019).

201. Hancock, Z. B., Lehmberg, E. S. & Blackmon, H. Phylogenetics in Space: How Continuous Spatial Structure Impacts Tree Inference. *Molecular Phylogenetics and Evolution,* 107505 (May 2022).

202. Arcila, D. *et al.* Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. *Nature Ecology and Evolution* **1,** 1–10 (2017).

203. Jansa, S. A. & Voss, R. S. Adaptive evolution of the venom-targeted vWF protein in opossums that eat pitvipers. *PLoS ONE* **6** (ed Ho, P. L.) e20997 (June 2011).

204. Barua, A., Mikheyev, A. S. & Russo, C. Many Options, Few Solutions: Over 60 My Snakes Converged on a Few Optimal Venom Formulations. *Molecular Biology and Evolution* **36,** 1964–1974 (2019).

205. Barua, A. & Mikheyev, A. S. Toxin expression in snake venom evolves rapidly with constant shifts in evolutionary rates. *Proceedings of the Royal Society B: Biological Sciences* **287,** 20200613 (May 2020).

206. Margres, M. J. *et al.* The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *Proceedings of the National Academy of Sciences of the United States of America* **118,** e2014634118 (Jan. 2021).

207. Rokyta, D. R., Wray, K. P., McGivern, J. J. & Margres, M. J. The transcriptomic and proteomic basis for the evolution of a novel venom phenotype within the Timber Rattlesnake (Crotalus horridus). *Toxicon* **98,** 34–48 (May 2015).

208. Mahler, D. L., Revell, L. J., Glor, R. E. & Losos, J. B. Ecological opportunity and the rate of morphological evolution in the diversification of greater Antillean anoles. *Evolution* **64,** 2731–2745 (2010).

209. Harmon, L. J. *et al.* Early bursts of body size and shape evolution are rare in comparative data. *Evolution* **64,** 2385–2396 (2010).

210. Clavel, J., Aristide, L. & Morlon, H. A Penalized Likelihood Framework for High-Dimensional Phylogenetic Comparative Methods and an Application to New-World Monkeys Brain Evolution. *Systematic Biology* **68,** 93–116 (Jan. 2019).

211. Chira, A. M. *et al.* The signature of competition in ecomorphological traits across the avian radiation. *Proceedings of the Royal Society B: Biological Sciences* **287** (2020).

212. Quintero, I. & Landis, M. J. Interdependent Phenotypic and Biogeographic Evolution Driven by Biotic Interactions. *Systematic Biology* **69** (ed Hahn, M.) 739–755 (July 2020).

213. Drury, J. P. *et al.* Tempo and mode of morphological evolution are decoupled from latitude in birds. *PLOS Biology* **19** (ed Quental, T. B.) e3001270 (Aug. 2021).

214. Margres, M. J., Bigelow, A. T., Lemmon, E. M., Lemmon, A. R. & Rokyta, D. R. Selection to increase expression, not sequence diversity, precedes gene family origin and expansion in rattlesnake venom. *Genetics* **206,** 1569–1580 (July 2017).

215. Sunagar, K. *et al.* Intraspecific venom variation in the medically significant Southern Pacific Rattlesnake (¡i¿Crotalus oreganus helleri¡/i¿): biodiscovery, clinical and evolutionary implications. *Journal of Proteomics* **99,** 68–83 (Mar. 2014).

216. Caetano, D. S., O'Meara, B. C. & Beaulieu, J. M. Hidden state models improve state-dependent diversification approaches, including biogeographical models. *Evolution* **72,** 2308–2324 (2018).

217. Fitzjohn, R. G. Quantitative traits and diversification. *Systematic Biology* **59,** 619–633 (2010).

218. Harvey, M. G. & Rabosky, D. L. Continuous traits and speciation rates: Alternatives to state-dependent diversification models. *Methods in Ecology and Evolution* **9,** 984–993 (2018).

219. Moreira, M. O., Fonseca, C. & Rojas, D. ES-sim-GLM, a Multiple Regression Trait-Dependent Diversification Approach. *Evolutionary Biology* **49,** 92–101 (2022).

220. Aristide, L. & Morlon, H. Understanding the effect of competition during evolutionary radiations: an integrated model of phenotypic and species diversification. *Ecology Letters* (ed Mooers, A.) ele.13385 (Sept. 2019).